Nick T. Thomopoulos

Elements of Manufacturing, Distribution and Logistics

Quantitative Methods for Planning and Control

Springer

Elements of Manufacturing, Distribution and Logistics

Nick T. Thomopoulos

Elements of Manufacturing, Distribution and Logistics

Quantitative Methods for Planning and Control

Springer

Nick T. Thomopoulos
Stuart School of Business
Illinois Institute of Technology
Chicago, IL, USA

ISBN 978-3-319-26861-3 ISBN 978-3-319-26862-0 (eBook)
DOI 10.1007/978-3-319-26862-0

Library of Congress Control Number: 2015957219

Springer Cham Heidelberg New York Dordrecht London

Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

For my wife, my children, and
my grandchildren

Preface

I was fortunate to have a rich and diverse career in industry and academia as it unfolded. This included working at International Harvester as supervisor of operations research in the corporate headquarters, at IIT Research Institute (IITRI) as a senior scientist with worldwide applications, and as a professor in the Industrial Engineering Department and in the Stuart School of Business at the Illinois Institute of Technology (Illinois Tech) and the many years of consulting assignments with industry and government throughout the world. At Illinois Tech, I was lucky to be assigned a broad array of courses, gaining a wide breadth of analytical knowledge with the variety of topics, and with the added knowledge I acquired with every repeat of the course. I also was privileged to serve as the advisor to 36 bright Ph.D. students as they carried on their dissertation research. Bits of knowledge from the various courses I taught and continual research helped me in the classroom and also in my consulting assignments. I used my industry knowledge in classroom lectures so the students could see how some of the textbook methodologies are actually applied in industry. At the same time, the knowledge I gained from the classroom helped me to formulate and develop solutions to industry applications as they unfolded. This variety of experience allowed me to view the quantitative tools that are doable and useful in industry, and this book is based on this total experience.

Thanks especially to my wife, Elaine Thomopoulos, who encouraged me to write this book and who gave consultation whenever needed. Thanks also to the many people who have helped and inspired me over the years. I can name only a few here. These are:

Brian Ahearn, Komatsu-Dresser; Wayne Bancroft, Walgreens; Owen Baxter, Wolverine World Wide; Sven Berg, L.M. Ericsson; Harry Bock, Florsheim Shoe Co.; Randy Braun, Komatsu-Dresser; Toemchai Bunnag, PTT Public; Dan Cahill, International Truck and Engine; Michel Chaussumien, Citroen; Charles Clerval, Citroen; John DeMotts, International Harvester; Frank Donahue, Navistar; Ralph Edgington, NCR; Nick Faklis, Patterson Office Supplies; Peter Fikaris, CMA-Flodyne-Hydradyne; Willard Huson, Navistar; Jim Gleason, International

Harvester; Mark Graham, Inspirage; Scott Haligas, Florsheim Shoe Co.; Jim Harrington, International Harvester; Joe Hoffman, Springfield Remanufacturing Corporation; Linkert Horn, IBM; Dick Ilseman, International Truck and Engine; Keith Mahal, Navistar; Nick Malham, FIC; Nico Mantelli, Olivetti; Craig Marecek, Komatsu-Dresser; Chad Myers, Springfield Remanufacturing Corporation; Paul O'Donnell, Westinghouse; Dino Olivetti, Olivetti; Nikos Papaioannou, Core Centrics USA; Pricha Pantimsuchai, M-Focus; Ted Prenting, Marist College; Harry Rindeskar, ASEA; Doloris Santucci, International Truck and Engine; Mark Spieglan, FIC; Ryan Stack, Springfield Remanufacturing Corporation; Ola Tannous, Electro Motive Diesel; Badal Wadia, Core Centrics USA; Tor Wadell, ASEA; and Charles Wright, Just Right Enterprises.

Chicago, IL Nick T. Thomopoulos

Contents

Chapter Summaries

Chapter 1. Forecasting: The forecast is perhaps the most important function in controlling the inventory. In a typical inventory entity, forecasts are needed for each of the future months up to the planning horizon, typically 12 months. Data from the past demands is needed to generate the forecasts.

Assuming monthly time buckets, the demand for a fixed number of history months (usually 12, 24, or 36) is saved in the database. The monthly demands represent the number of pieces ordered on an item for each month. Many entities also save the lines per month. These are the number of customer orders that arrive during a month. The history months are mostly saved in calendar months, but some entities use fiscal months of the 4, 4, 5 type. It behooves the entity to process the demand history prior to forecasting, by running a routine to seek if any outlier demands are found, and, if so, to adjust them prior to forecasting. Another useful routine would seek to adjust the demand history if any demands are returned due to a ship-in-error by quantity or by part.

To control the inventory, the total requirement of the part is needed for each future month. The total requirement is the sum of the forecast plus any other requirements that may occur. The other requirements are any demands outside of the normal customer base, like foreign country or military. Also, the forecast of a future month may need to be adjusted when a customer brings in an order for a quantity of an item, but not to be delivered until a future month. In inventory systems, four methods of generating the forecasts are generally in use: moving average, regression, discounting, and smoothing. The three basic forecast models are horizontal, trend, and smoothing. The smoothing model is of two types: seasonal multiplicative and seasonal additive. In most situations, the forecasts are converted from fractions to integers for ease of use by the management. Finally, a way to convert the monthly forecasts to weekly forecasts is developed for entities that require weekly forecasts.

Chapter 2. Forecast Error: The standard deviation of the 1-month-ahead forecast error is used to determine how much safety stock to have available to satisfy the level of service to customers. An exact measure of the standard deviation is not easy to generate; however, estimates are sometimes used in its place. The way to estimate the standard deviation varies depending on the forecasting method in use (moving average, regression, discounting, smoothing). This chapter describes how to estimate the standard deviation for each of the forecast models presented in Chap. 1.

Another important statistic for each forecast model is the coefficient of variation (cov). The closer the cov is to zero, the better. When cov is near 0.30, the distribution of the forecast errors is similar to normal distribution. When the cov is closer to 1.00, the distribution of the forecast errors is not like a normal distribution. In Chap. 4, the concept of the truncated normal is presented. The cov is used to estimate the parameter to use for the truncated normal. When the distribution of the forecast errors is not normally distributed, the truncated normal distribution should be applied to determine how much safety stock is needed. In determining the safety stock, the standard deviation and cov of the lead-time demand are used in the computations. The methods to generate these statistics are included in the chapter.

When the forecast and its standard deviation are generated for a total collection of SKUs, estimates are needed for each SKU. A way to measure the standard deviation for each SKU is presented. As an example, this need occurs for a style shoe where the forecast is based on all sizes combined, and estimates for inventory control are needed for each size of the style.

Chapter 3. Order Quantity: The order quantity, Q, is a planned amount of stock to replenish the inventory from the supplier when new stock is needed. In the ideal situation, Q is calculated to give the minimum cost of holding and ordering the stock, called the economic order quantity, EOQ. The EOQ is obtained from several factors: A is the annual forecast of the item, Co is the cost per order, c is the cost per unit, and h is the annual holding rate. Note, A comes from the forecast, c from the supplier, and Co and h are parameters from the stocking facility. In some facilities, the order quantity is restricted to fall in monthly buckets, but in most situations it is not limited in that way. A related cost measure is the stocking rate denoted as s, which measures the rate of added cost on the item due to stocking the item via hold plus order. The stocking rate yields another measure labeled the stocking cost per unit denoted as $(c \times s)$ which is a hidden added cost due to hold plus order. The effective cost per unit becomes $c' = c(1 + s)$. It is sometimes useful for the inventory management to realize the sensitivity between the four ingredients and the order quantity. When A changes, how does this affect Q? The same queries can be made with changes in Co, c, and h. Another measure of interest is the change in the hold plus order cost when using an order quantity different from the EOQ. Note also, as the order quantity changes, the on-hand inventory also changes.

Oftentimes, the simple economic order quantity is restricted due to a series of constraints. The more typical type are described below:

- Price breaks are when the supplier offers discounts on the cost per unit or cost per order as the buy quantity increases. The more you buy, the less you pay.
- Min and max constraints are when the supplier specifies that the buy quantity cannot be higher than a max constraint and cannot be lower than a min constraint.

- Multiple quantity constraints occur when the supplier specifies a multiple number for which the buy amount must conform.
- Bin max constraint is when the stock holding entity has limited space to stock the item. The on-hand plus on-order quantity cannot exceed this bin max quantity for the item.
- Job quantity constraint is when the stock holding entity typically sells the units in multiples of a fixed number, like four tires on an auto vehicle. The order quantity to buy is restricted to conform to the job quantity and the current on-hand for the item.

In service parts, two important order quantities are the initial buy quantity and the final buy quantity.

- The initial buy quantity occurs when a new part is introduced and there is no demand history to base a forecast. The service facility is obliged to carry stock on this item and must generate a first time buy.
- The final buy quantity occurs when the supplier notifies the stocking facility they will stop supplying the item, and the facility is obliged to have the item available for a prolonged obligation period. A last time buy is needed.

Chapter 4. Safety Stock: Safety stock is needed to attain a desired level of service when the demands exceed the forecasts over the lead-time duration. Four popular methods to generate the safety stock are in use: month's supply, service level, percent fill, and Lagrange. For each method, a safety stock parameter is assigned by the management to specify how to compute the safety stock. The month's supply method is based on a specified number of future forecasts. The service level method is set so the probability of not having an out-of-stock condition during the lead-time is satisfied. The percent fill method seeks to accommodate the ratio of (demand fill)/(total demand) over the lead-time. The Lagrange method is a cost-benefit way to set the safety stock. The latter three methods are based on the distribution of forecast for the lead-time demand. The distribution for the lead-time demand can be normally distributed or not normally distributed. When normally distributed, the standard normal distribution is applied to set the safety stock. When not normally distributed, the truncated normal distribution is used to set the safety stock. The coefficient of variation for the lead-time demand is used to identify which of the two distributions to use.

The Lagrange method works well when the safety stock for a group of items is controlled simultaneously. The goal is to have a desired service level for the entire group at the total minimum cost. The safety stock depends on the lead-time provided by the supplier; but for suppliers who are late in the delivery of the replenish stock, adjustments to the safety stock are needed. Each supplier is measured as to its delivery score, and the safety stock adjustment for the supplier is based on his score. Sometimes loss sales occur when stock is not available to complete a customer demand. A way to measure and control the amount of loss sales is developed.

Chapter 5. Replenishments: Replenishing the stock is one of the most important functions in the control of the inventory. The goal is to replenish the stock that properly covers the customer demands at minimum cost in inventory. To accomplish, two measures are regularly computed: the order point and order level. Each day, these are compared to the current on-hand plus on-order inventory to determine if a new replenishment is needed now and, if so, how much. Each replenishment must conform with any constraints provided by the supplier, as minimum buy quantity and multiple buy quantity. For analysis sake, the inventory is partitioned into cycle stock and safety stock, where the sum is the total stock. The ingredients that affect the stock levels are the percent fill, lead-time, coefficient of variation, and month-in-buy. The sensitivity of each with the stock levels and with the turnover is described. In many retail entities, the demand for each item is low, and thereby the Poisson distribution is used to determine the order point and order level. For entities with low demand items, table entries are provided to guide the management on how to set the order point and order level.

Chapter 6. Distribution Control: For convenience here, a network (NW) is defined when two or more stock holding facilities are connected with one entity. This is a distribution system with two or more locations, could be a system with distribution centers (DC) or two or more retail outlets. The individual stocking locations are here referred as locations. The goal is to control the inventory for the NW and for each location. Typically, once a month at the NW, forecasts, standard deviation of the forecast error, and the planned order quantity are generated. Also, at each location, the forecast, standard deviation, and order quantity are also needed to compute the order point and order level. At each location, the on-hand and on-order are observed and compared to the order point and order level to determine if a location needs a replenishment buy, and if so, how much. Sometimes the location buys directly with the supplier, and other times, the NW buys for the total system

When the NW buys the stock and the replenish quantity arrives from the supplier, a way to distribute the stock to the locations is needed. This process is called allocation, with a goal to allot the stock to the locations in a fair-share manner. Another important process at the NW is called transfer. Transfer occurs periodically, perhaps weekly, where the stock status of each location is measured, and if a location is low and another is high, a transfer quantity is computed. Share is another process that takes place on occasion. One location is selected as the source, and each of the parts at the location is analyzed in a share manner as follows: if the inventory status of a part is high at the share location, and it is low at another location, a share quantity is computed, whereby the quantity is transferred accordingly. Limit buy is another process that sometimes occurs in the NW. The total buy for a supplier cannot exceed a given limit, could be in dollars, pieces, and so forth. Discount buy is when the total buy, all parts, must reach a goal to qualify for a discount, could be for a supplier discount, for a truckload min limit, and so forth.

Chapter 7. Manufacturing Control: The manufacturing plant consists of an assortment of raw materials, components, machines of various types, and a variety of skilled workers, with the goal to produce goods to a higher level and some to finished-good-items. The management's task of coordinating all this activity is difficult indeed. A first concern of the management is to periodically generate a production plan for the coming planning horizon. This plan pertains to the aggregate of all items in the plant and yields the volume of production for all. The plan depends on the type of items to produce: make-to-stock, make-to-order, or a combination of make-to-stock and make-to-order. For each finished-good-item, a master production plan is generated to coordinate the schedule that satisfies the inventory status and customer demands. This schedule also yields the available-to-promise quantity that is a vital tool to the sales force. The concept of raw load and level load by future time periods for various production centers in the plant is described. To ensure the schedule of all items is doable, a rough-cut capacity planning analysis is calculated for each of the production centers in the plant. When the capacity does not meet the load, adjustments are required. For every item to be produced, a bill-of-material is used to identify each of the parts and components that are required in the build. From here, a material requirement planning set of computations determines the build schedule for every item.

Chapter 8. Just-in-Time: Just-in-time (JIT) is a philosophy of production based on the concept of adding value and eliminating waste. JIT and lean manufacturing have very similar goals. Value is added only by work performed on the product, and waste is anything other than a minimal amount of necessary resources—material, manpower, and the capital equipment—that is required for production and does not add value to the product. The process called Kanban is a system where cards are used between send and receive stations in a way so that the stations produce only the necessary quantity of goods at the necessary time. JIT examples are presented where the components to a product are received from a supplier shortly after the customer order arrives. The relation between lean manufacturing and JIT is described. Smaller batch sizes are preferred in production, and this is accomplished as the setup time at a production process is reduced. The smaller the setup time, the lower the economic batch size. The safety stock to achieve a service level to the customer depends largely on the lead-time of replenish time from the supplier. As the lead-time becomes smaller, the amount of safety stock needed is lowered accordingly. The management should also seek to level the week-to-week aggregate production loads to avoid excess cost of overtime, backorders, and outsourcing. When the finished-good-items are on a make-to-order basis, the strategy of postponement reduces the lead-time to the customers and also eliminates much of the complication in the assembly.

Chapter 9. Assembly: A variety of assembly lines are applied in industry. For the smaller type of products, the assembly lines are often of the single model type, where one model of the product is produced on the line. The work elements to assemble one unit are gathered and set in a precedence diagram showing the feasible work relations between the elements. The number of operators needed is computed, and the work elements are assigned to each operator in a fair-share way. This latter function is called line balancing.

For larger products that are stocked in a warehouse or distribution center waiting for subsequent customer demands (washing machines, refrigerators, and so forth), the assembly is called a mixed model make-to-stock line. Two or more models of the product are assembled on the line at the same time. The plan schedule to produce the units is stated by the management, and the total time to complete all the work elements is tallied. The number of operators needed to conform to the schedule is computed, and the work elements are then assigned to the operators, via line balancing. Because two or more models are mixed on the line, another task of the management is to determine the sequencing of the models down the line.

On the higher price items (trucks, tractors, and so on), the units produced on the line each have a variety of features and options assigned by the customers. This is a make-to-order line. Because of this, each unit going down the line is unique. Each day, a different schedule of units, called jobs, is registered to be produced. The work elements are assigned (via line balancing) to the operators in a way to satisfy an average day of production. The sequencing of the units down the line is arranged so the operator times are efficient as the units roll down the line from one unit in the sequence to the next unit.

In the latter two assembly lines described, robots are often scattered on the line and perform major tasks of the units, like in the paint booth or in lifting of heavy components as what takes place in the body shop of the assembly line. The line balancing and sequencing tasks are still in need to accommodate the robots.

Chapter 10. Statistical Process Control: A process can be defined as a combination of components, tools, people, and machines that together produce an item, like a part that subsequently is inserted into a finished-good-item. Although, the output of the process may vary, the final customer requires the product to satisfy the standards specified by the engineers. Management's goal is to control the output units of a process so that it conforms to the standards set by the engineers. The output measures are of two types: attribute or variable. Attribute is when the output measure is defective or not defective. Variable is when some measure is taken from each sampled unit.

Four attribute type control methods are described and labeled as the following: p-chart, np-chart, c-chart, and u-chart. A pre-study is set up for each of the methods. N subgroups are studied at intermittent time intervals with a number of samples taken from each subgroup and the number of defectives tallied, if any. Statistics on the capability of the process is measured after the pre-study, and if acceptable, lower and upper control limits are computed. The process is monitored in the same way for the future subgroups to ensure the process is performing with high standards.

Four variable type control methods are described and labeled as the following: x-bar and R-chart, median and R-chart, individual-chart, and x-bar and s-chart. A pre-study is arranged with N subgroups at systematic time intervals, where sample data is taken. Specification limits are provided from the engineers, and these are compared with the sample output data from the samples. The goal is for the output to be well within the specification limits and ideally within six-sigma limits. Measures of process potential and process capability are computed using the results gathered in the pre-study. If these are acceptable, a set of statistical coefficients is used to compute control limits for the particular testing method. Thereupon, the process is monitored over future subgroups to ensure the output measures remain within the control limits.

Chapter 11. Distribution Network: In a typical original equipment manufacturing (OEM) system, the suppliers ship stock to one or more distribution centers that serve as the source to a series of retailers (dealers, stores), and the retailers sell the products to the customers. This could be a network that holds service parts for an automotive corporation. The dealers sell the autos to the customers, and when repair or maintenance is needed, the customer seeks service from the dealer. The dealer carries a limited supply of parts for this purpose and relies on the DC to have a full set of parts as needed. The OEM is faced with holding the minimal amount of stock needed to properly service the demands from the dealers. In the ideal situation, the DC system consists of a network of locations that are strategically near the dealers and to some extent near to the suppliers. One location is the master stock location that serves as the headquarters for the entire system. The other locations are branch locations. The master location is often a much larger location and holds more variety of stock. Sometimes a small location is provided and is called a 2-level-service location. This location holds minimal stock and is supplied by the master stock location. The parts that have a high amount of demand are stocked in most locations and are called distributed parts. The low demand parts are labeled as non-distributed and are mostly stocked only from the master stock location.

Periodically, the management runs an analysis to seek what is the best arrangement of the DC network. The analysis seeks answers to: how many locations, where should the locations be sited, and how to allot the stock by locations. A series of parameters are provided in a “what if” arrangement. The management can run a variety of options with various settings of the parameters, and the system provides summary results for each option.

Sometimes the parts supplier does not package the parts as needed by the DC system and thereby sends the units produced to a packager who places the parts in containers with appropriate labeling. The packager then ships the packaged parts to the DC network. Cross-docking is also a strategy where the suppliers ship in large vehicles to a cross-dock facility and the goods are taken off the larger vehicles, sorted, and placed in smaller vehicles for transport directly to the stocking facility (DC or dealer). Back at the DC network, the integrity of the inventory count is vital to the efficiency of the service to customers. It behooves the DC network to have a continual cycle counting system in place to maintain the count of the amount of units in stock for each part. Three cycle counting methods are popular: random, ABC, and Pareto. Queuing theory is introduced showing how it has a wide variety of applications in the operation of a network.

Chapter 12. Supply Chain Management: The supply chain management team in a manufacturing firm is concerned with the flow of raw materials, basic goods, components, work-in-process, and finished goods from suppliers to warehouses, plants, distribution centers, retailers, and finally to the customers. The goal is to minimize throughput times and expenditure costs while achieving a high level of service to the customers with a clean environment. To accomplish, the team remains vigilant on the use of all technology available to them. Data sharing with the customers and throughout the operation is needed so all can work efficiently. The team applies the technology of scan-based tracking, electronic data interchange, bar codes, data matrix bar codes, quick response codes, and radio frequency identification where and when appropriate. In the inventory operations, vendor managed inventory and continuous replenishments are applied when suitable. In transportation, incoterms are applied along with tracing and tracking of the shipments; and whenever feasible, outsourcing and third-party providers are employed. Five examples are sited: a heavy-duty make-to-order manufacturer, a network of men's shoe stores, an offshore automotive dealership, a replenishment of inventory at retail stores for seasonal style good, and a delivery system to stores for a popular bakery item.

Chapter 13. Transportation: Transportation pertains to the movement of items from one location to another, could be from a plant to a distribution center to a retailer. A way to classify the type of goods is by shipping categories: household, express, parcel, or freight. Transport modes are the access ways to transport: rail, road, air, water, or pipes. Transport vehicles are the trucks, trailers, barges, aircraft, and cargo ships. Cargo handling is by ports, container terminals, shipping containers, forklift trucks, cranes, pallets, and dunnage. The shipment of goods also concern dispatchers, bill of lading, manifest, truckload shipping, less than truckload shipping, parcel carriers, bulk cargo, and break bulk cargo. Some industrial trading terms are free on board, carriage and freight, carriage, insurance and freight, and best way.

Chapter 14. Reverse Logistics: Reverse logistics is the process of managing the operations concerned with any returned goods to the manufacturer. On average, near 5 % of the goods sold are returned to the original manufacturer for a variety of reasons as: worn out goods, damaged goods, unsold goods, recall goods, and so forth. The manufacturer is obliged to receive and process the returned goods seeking any revenue that can be gained or arranging for proper disposal. The role of processing returned goods has expanded ever more as the environmental mandate of industrial growing green has gained strength. In the typical forward logistic way, the goods' final destination is with the customer. As the goods become old, they may be replaced and returned to the manufacturer to begin the reverse logistics cycle. The returned goods are mostly one-model-at-a-time occurring in a disjointed manner. The return process is costly, perhaps 10 % of the total cost of the original sales price, requiring the manufacturer to seek as much value from the item as possible. In the typical situation, the returned goods are cleaned, tested for worthiness, and, if accepted, are repaired and refurbished for resale. When not accepted, they are disposed in the proper manner.

The returned goods are often from various reasons: worn out and no longer usable goods, damaged goods from shipment or use, new goods that are unsold at the retailer, goods ordered on the internet with cash-on-delivery terms and payment refusal when delivered, recalled goods, discontinued goods, and so forth. The returned items are also from warrantee claims, some from fraudulent claims, and some are stolen goods. Other returned goods are of the reusable type whereby they are recycled after cleaning and repair. These include pallets, containers, packages, bottles, cans, cylinders, and the like.

Two examples are described. The first concerns appliances where a customer purchases an appliance, and later in time, the unit does not operate properly, and the defective product is sent back to a return facility that refurbishes the product as new for future use. The second pertains to the automotive industry from the truck industry where worn out and discarded engines, brakes, or transmissions are removed from the vehicle and are remanufactured to like-new status.

Chapter 1
Forecasting

1.1 Introduction

The forecast is perhaps the most important function in controlling the inventory. In a typical inventory entity, forecasts are needed for each of the future months up to the planning horizon, typically 12 months. Data from the past demands is needed to generate the forecasts. Assuming monthly time buckets, the demand for a fixed number of history months, (usually 12, 24 or 36), is saved in the database. The monthly demands represent the number of pieces ordered on an item for each month. Many entities also save the lines per month. These are the number of customer orders that arrive during a month. The history months are mostly saved in calendar months, but some entities use fiscal months of the 4, 4, 5 type. It behooves the entity to process the demand history prior to forecasting, by running a routine to seek if any outlier demands are found, and if so, to adjust them prior to forecasting. Another useful routine would seek to adjust the demand history if any demands are returned due to a ship-in-error by quantity or by part.

To control the inventory, the total requirement of the part is needed for each future month. The total requirement is the sum of the forecast plus any other requirements that may occur. The other requirements are any demands outside of the normal customer base, like foreign country or military. Also, the forecast of a future month may need to be adjusted when a customer brings in an order for a quantity of an item, but not to be delivered until a future month. In inventory systems, four methods of generating the forecasts are generally in use: moving average, regression, discounting and smoothing; and three basic forecast models are: horizontal, trend and smoothing. The smoothing model is of two type: seasonal multiplicative and seasonal additive. In most situations, the forecasts are converted from fractions to integers for ease of use by the management. Finally, a way to convert the monthly forecasts to weekly forecasts is developed for entities that require weekly forecasts.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_1

1.2 Forecast Pioneers

Three individuals, Robert G. Brown, Charles C. Holt and Peter Winters, were the pioneers in demand forecasting. They developed forecast models that are still in common use today. Brown developed the horizontal smoothing model; Holt, the trend smoothing model; and Winters the seasonal smoothing model, all in the early 1950s. All three used a method called exponential smoothing that requires only the demand from the most current time period. In the 1950s, it was not common to store and save multiple time periods of demand, since computer data processing was just getting started.

1.3 Data Collection

Demand history data is saved on a database for each item in a stocking location. The data includes the pieces ordered and sometimes the number of orders; and could be collected in buckets of calendar months or in fiscal months.

1.3.1 Stock Keeping Units

A stock-keeping unit is any item (part, product, style), that is stocked in a location to fill a forthcoming demand need. In the ideal situation, the database of the entity holds demand and stock information on each SKU. For example, when a part number is stocked in three distribution centers, DC, of a company, the part represents three SKUs, since inventory decisions are needed on the part at each DC.

Example 1.1 If Part × 123 is stocked at locations A and B, the part is an SKU at location A, and is an SKU at location B. Altogether, the part is set as two SKUs.

1.3.2 Monthly Buckets

In most stocking locations, the time periods for the planning horizon are monthly, and thereby the demand history is saved in monthly durations, and the future forecasts are generated in monthly durations. For convenience, the descriptions here will assume monthly durations. A common term in industry for monthly durations is monthly buckets. A way to convert the monthly forecasts to weekly forecasts is described later in this chapter.

Example 1.2 If the current month is April, 2015 and the system saves the most current 24 months of demands. The demand history (May 2013 to April 2015) in the database would appear as below:

year	J	F	M	A	M	J	J	A	S	O	N	D
2013					x	x	x	x	x	x	x	x
2014	x	x	x	x	x	x	x	x	x	x	x	x
2015	x	x	x	x								

1.3.3 Database

The forecasts are based on history demands, and as such, a collection of the history demands needs to be saved and stored in a database. Assuming monthly buckets, the management specifies how many months of history, Nd, to save; and typically, Nd = 12, 24 or 36 months. For an individual SKU in the inventory, NMH denotes the number months of history since the SKU was introduced. But because only Nd months of history is saved on the database, the actual number of months on the database becomes: N = min(Nd, NMH).

Example 1.3 Let t = 1 to N identify the history months on the database for a part, where t = 1 is the oldest history month and t = N is the most current. Assume Nd = 24 is the number of months to be saved on the database. Below shows how NMH and N are related for four skus, A, B, C, D.

If NMH = 7 for SKU A: N = 7
If NMH = 12 for SKU B: N = 12
If NMH = 24 for SKU C: N = 24
If NMH = 38 for SKU D: N = 24

1.3.4 Demands and Lines

The type of demand history in the database for each item depends on the need of the stocking location. The most common data is the demands by month, and also the lines by month. The monthly demands represent the number of units (pieces) requested by the customers for the item during a month's duration. The lines are the number of orders that come in from the customer for the month, where each customer order calls for a quantity of units. Every customer order is a line. The number of demands at month t is denoted as x_t, and the corresponding number of lines is n_t.

Example 1.4 Below shows a part with N = 5 months of history and the lines and demand history are as listed. Note, t = 1 is the oldest month and t = 5 is the most current month.

history months:	$t = 1\ \ 2\ \ 3\ \ 4\ \ 5$
lines:	$n_t = 2\ \ 0\ \ 1\ \ 4\ \ 2$
demands:	$x_t = 3\ \ 0\ \ 1\ \ 6\ \ 5$

1.3.5 Calendar and Fiscal Months

For planning purposes, the monthly buckets that hold the history demands are either in calendar months or in fiscal months. Most inventory holding operations use calendar months that are the same as the 12 regular months of the year. For scheduling convenience, manufacturing operations often use 12 fiscal months where each month is exactly 4 or 5 weeks in length. Each week starts on a specified day of the week, usually, Sunday or Monday. This way, every fiscal month is 28 or 35 days in length.

The fiscal months are one of the following: 445, 454 or 544 as described below.

Fiscal months :	445 has 4 weeks in January, 4 in February, and 5 in March.
	454 has 4 weeks in January, 5 in February, and 4 in March
	544 has 5 weeks in January, 4 in February, and 4 in March

The pattern repeats for the next 9 fiscal months of the year.

1.4 Past Demands

Prior to forecasting, care is needed to safeguard the integrity of the demand history. An algorithm to seek out any outlier demands in the history and adjust such demands is important. When return demands occur, another algorithm should seek to find and adjust the source in the history month, accordingly.

1.4.1 Demand History

The data held for each SKU is the most current N months of demand history. The number of months of history, NMH, could be one or larger depending on when the SKU became active. If NMH is of size Nd (the database size) or less, then all the demand history is saved; but should NMH exceed Nd, only Nd of the history demands are saved. The history demands are labeled by x_t, $t = 1$ to N, where N denotes the months of history on the database. Figure 1.1 depicts a plot of the demand history when N months of history.

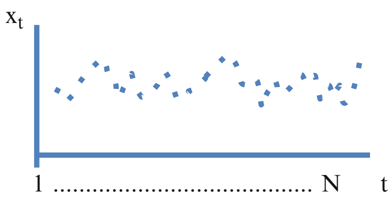


Fig. 1.1 Plot of demand history

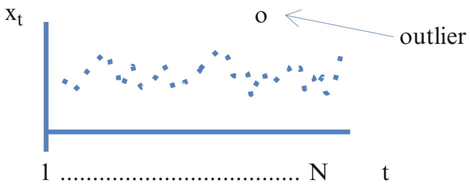


Fig. 1.2 An outlier demand in the demand history

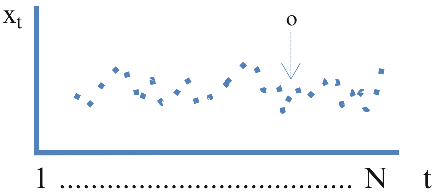


Fig. 1.3 Adjust the outlier demand

1.4.2 Outlier Demands

An outlier is a monthly demand in the history that is not typical of the normal flow of monthly demands. See Fig. 1.2 for an example scenario. The outlier can occur from many situations: the part number is mistyped in data entry; the demand amount on the order is not correct; a unusual weather condition calls for a spike in demand for an sku, so forth. No matter the cause, outliers are damaging and cause large errors in the forecasts, and the subsequent measures of forecast errors. The larger the increase in forecast error, the more safety stock is needed in the inventory. See Thomopoulos, 2015, pp. 119–136 on outliers.

As much as possible, prior to forecasting, an algorithm to scan the demand history and search for any outlier demands is needed. In the ideal situation, the algorithm will find the outlier and adjust the monthly demand prior to forecasting. The adjusted demand is set close to the neighbor monthly demands for the sku, as depicted in Fig. 1.3. To keep integrity in the database, the outlier demand in the database is not altered, however, the adjusted demand is used to generate the forecast for the SKU. In some systems, prior to forecasting, the difference in the outlier demand is spread to the closest neighbor monthly demands.

1.4.3 Return Demands

A return demand occurs when a customer sends back to the stock location a quantity of units of an sku that was previously shipped to them. The return quantity is entered in the database as a return demand. The return demand could occur from any of several events: the wrong part is shipped to the customer, the wrong quantity is shipped to the customer, one or more of the units are damaged, and so forth. The demand history should be adjusted prior to forecasting to account for the returned units.

In the ideal situation, an algorithm is developed to search the prior months of the demand history to identify when the demand was originally recorded. This typically is the closest prior month where the original demand could have occurred. When the original demand is identified, the demand history is adjusted accordingly, prior to forecasting. The entry in the database is not changed, only the demand history used subsequently in generating the forecasts. The return adjustment algorithm is described below.

For notational ease, let the return quantity be denoted as follows:

$$xr_t = \text{demand returned in month t.}$$

Also, let

$$t_o = \text{first prior month where } x_{to} \geq xr_{t.}.$$

So now, the adjustment to the history demands reduces the demand quantity in month t_o as:

$$x'_{to} = (x_{to} - xr_t)$$

1.4.4 Demand-to-Date (D2D)

The demand to date is the cumulative demands coming in for the current month on each SKU. A corresponding measure, w, is the portion of the current month that has so far elapsed. The D2D could change each day and even several times each day. The measure, w, will change every day. These demands are not generally used in generating a forecast on the SKU for the future months. Figure 1.4 shows where the typical database houses the most current N months of history demands and also the demand-to-data for the current month N + 1. At the start of the next month, the total D2D from the prior month becomes the history demand for the most current month.

Fig. 1.4 Example of demand-to-date

	History months	D2D
t	1 .... N	N+1
Demand history	x_1...................... x_N	x_N+1

Example 1.5 Assume a situation with 30 % of the current month elapsed, the demand-to-data for an item is four pieces. Hence, the data recorded for the item on the database is: $w = 0.3$ and $x_{N+1} = 4$.

1.5 Future Requirements

The stocking facility is obliged to carry enough stock to satisfy the forecasts and any other requirements. As such, the database houses the forecast for the future months and the other requirements. The sum is called the total requirements.

1.5.1 Forecasts

The future months are denoted as τ=1, 2, . . ., and the demand history is used to generate the forecasts for each of the future months. A fit of the flow from the history demands is generated and is projected forward to become the forecasts for the future months. The flow pattern is usually from one of the following patterns: horizontal, trend or seasonal. Should forecasts for the 12 future months be needed, the forecasts are then denoted as: $f\tau$ $\tau = 1–12$. For notation sake, $t = 1$ to N identifies the history months, and $\tau = 1$ to N\` identifies the future months. Figure 1.5 lists the demands and forecasts over the history and future months.

1.5.2 Advance Demands

Sometimes a customer will place an order of quantity size, x_o, for an sku that is not wanted now, but in a future month τ. This type of demand is called an advance demand. The demand is not like the normal type that forms the demand history. Instead this is a demand that is not needed until some future month. The advance demand is not used to generate the forecasts for the future months, but instead can be used to adjust a previously generated forecast for the τth future month.

Each advance demand, x_o, for future month τ is analyzed as shown below in four steps to adjust the forecast, f, for month τ. This is as follows:

Fig. 1.5 N history demands and N\` future forecasts

history	future
demands (x_t)	forecasts ($f\tau$)
x_1........ x_N	f_1...... $f_{N`}$

1. Collect, x_o, f and σ.

$$x_o = \text{advance demand for future month } \tau$$
$$f = \text{forecast for future month } \tau$$
$$\sigma = \text{standard deviation for future month } \tau$$

2. $k = (x_o - f)/\sigma$ = a value from the standard normal distribution
3. Using k, get $E(z > k)$ from the standard normal table (Table 4.2)

$$E(z > k) = \text{partial expectation}$$

4. $f^{`} = x_o + E(z > k)\sigma$ = adjusted forecast for future month τ

The adjusted forecasts is demoted as f`. For a full discussion, see Thomopoulos, 1980, pp. 249–251 on advance demands.

Example 1.6 Find the adjusted demand for a part with the following advance demand data: $x_o = 50$, $\tau = 4$ $f = 60$ and $\sigma = 10$. Note the following;

$$k = [50 - 60]/10 = -1.00$$

Using k in Table 4.2, $E(z > k) = 1.083$.
Thereby,

$$f^{`} = 50 + 1.083 \times 10 = 60.83$$

is the adjusted forecast for the month.

1.5.3 Other Requirements

The demand history is based on the cumulative monthly orders from a regular set of customers. In the typical way, a DC will receive orders from a collection of dealers who directly sell the units to the customers. The dealer orders arrive to the DC one-by-one and the accumulation of orders form the monthly demand history. The forecast is a projection of the future demands from this same set of customers. Occasionally, an order will arrive from a source that is not a typical customer. Could be from a foreign company, the military or another OEM. Usually, this type of order is for a large quantity of units and is for a future month. This type of demand is not the same as an advance demand, and is called other requirement. For notation sake, the other requirement order for a future month τ, is denoted as or_τ. In the database, the other requirements are saved for each of the future 12 months.

Example 1.7 An order comes in from a foreign customer for 100 pieces of part × 123 for the third future month. Thereby, $or_3 = 100$. The orders for each of the 12 future months are saved on the database as below.

$$\text{other requirements:} \quad or_1, \ldots\ldots\ldots, or_{12}$$

1.5.4 Total Requirements

The stocking location is obliged to carry enough stock to fill the sum of the forecast of customer demands, and also the other requirements from customers. This sum is called the total requirements and is tallied for each future month in the planning period. The forecasts are estimates of the customer demands, and the other requirements are exact quantities on customer orders. The total requirement for future month τ is the sum of the forecast, f_τ, plus other requirements, or_τ, for future month τ. The notation of total requirement for future month τ is tr_τ and is computed by $tr_\tau = (f_\tau + or_\tau)$. The database houses the following information for the N` future months.

future months:	1	N`
forecast:	f_1	$f_{N`}$
other requirements:	or_1	$or_{N`}$
total requirements:	tr_1	$tr_{N`}$

1.6 Horizontal, Trend and Seasonal Models

The basic forecast demand patterns are the following: horizontal, trend and seasonal. The forecast models for these patterns are described below.

1.6.1 Horizontal Forecast Model

The most common flow of history demands is flat where the monthly demands are randomly scattered above and below an average value. This is denoted as a horizontal demand pattern and calls for a horizontal forecast model to project the monthly demands for the future months. The horizontal forecast has one coefficient referred as the level, denoted as, a. The level is a measure of the average monthly demand. The forecast for future month τ, *denoted as* f_τ, is simply $f_\tau = a$, as depicted in Fig. 1.6.

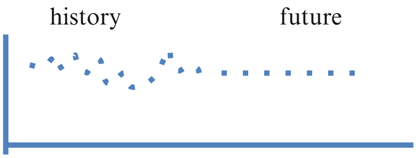


Fig. 1.6 Horizontal demand pattern and horizontal forecast

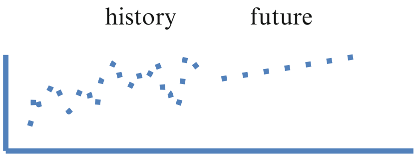


Fig. 1.7 Trend demand pattern and trend forecast

Example 1.8 Assume a part with the horizontal demand pattern where a horizontal forecast model is used and the level is estimated as a` = 10. The forecast for the first 3 future months become: $f_1 = 10$, $f_2 = 10$ and $f_3 = 10$.

1.6.2 Trend Forecast Model

Sometimes the demands along the history months are fluctuating around a trend line that is steadily moving either in an up or a down direction as depicted in Fig. 1.7. This is denoted as a trend demand pattern, and thereby calls for a trend forecast model. The trend forecast model has two coefficients labeled as the level, a, and the slope, b. The level represents the average demand as of the most current history month, $t = N$. The forecast for future month τ becomes, $f_\tau = a + b\tau$, where $\tau = 1, 2, \ldots$.

Example 1.9 Assume a part with a trend demand pattern where a trend forecast model is applied and the coefficients are estimated as a` = 10 and b` = 1. The forecast for the first 3 future months become: $f_1 = 11$, $f_2 = 12$ and $f_3 = 13$.

1.6.3 Seasonal Models

In some situations the flow of monthly demands fluctuate around a cyclical pattern that repeats in a consistent manner from year to year. This type of demand flow is called a seasonal demand pattern. Two seasonal forecast models are typically used to forecast the demands for the future months: the seasonal multiplicative forecast model, and the seasonal additive forecast model. These are described below.

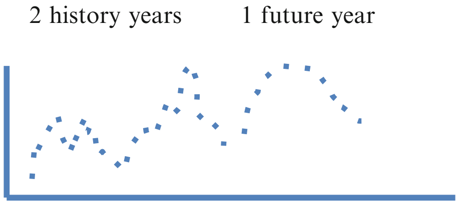


Fig. 1.8 Seasonal demand pattern with seasonal multiplicative forecast

1.6.4 Seasonal Multiplicative Forecast Model

The seasonal multiplicative forecast model has three coefficients, a = level, b = slope and r_τ = seasonal ratio for future month τ. The model has twelve unique seasonal ratios, one for each month of the year. The level represents the seasonally adjusted average value for the most current history month. The trend is $(a + b\tau)$ and represents the seasonally adjusted average for the future month τ. When $r_\tau = 1.00$ for future month τ, the forecast is the same as the trend. When $r_\tau > 1.00$, the forecast for future month τ is projected higher than the trend; and when $r_\tau < 1.00$, the forecast is lower than the trend. The forecast for future month τ becomes: $f_\tau = (a + b\tau)\ r_\tau$. See Fig. 1.8.

Example 1.10 Consider a part with a seasonal demand pattern and where the seasonal multiplicative forecast model is applied. Assume the coefficients are estimated as the following; a` = 10, b` = 1, r_τ = (0.7, 0.8, 0.9, 1.0, 1.1, 1.2, 1.3, 1.2, 1.1, 1.0, 0.9, 0.8) for τ = 1–12. The forecast model is: $f_\tau = (10 + 1\tau)r_\tau$ for $\tau = 1,2,\ldots$, whereby the forecasts for the first 3 future months become:

$$f_1 = (10 + 1)0.7 = 7.7$$
$$f_2 = (10 + 2)0.8 = 9.6$$
$$f_3 = (10 + 3)0.9 = 11.7$$

1.6.5 Seasonal Additive Forecast Model

The seasonal additive forecast model has three coefficients, a = level, b = slope and $d\tau$ = seasonal increment for future month τ. The model has twelve unique seasonal increments, one for each month of the year. The level represents the average value as of the most current history month. The trend is $(a + b\tau)$ and represents the seasonally adjusted average for the future month τ. When $d\tau = 0.00$ for future month τ, the forecast is the same as the trend. When $d_\tau > 0.00$, the forecast for future month τ is projected higher than the trend; and when $d_\tau < 0.00$, the forecast is lower than the trend. The forecast for future month τ becomes: $f_\tau = (a + b\tau) + d_\tau$. See Fig. 1.9.

Fig. 1.9 Seasonal demand pattern with seasonal additive forecast

Fig. 1.10 Weights for the moving average forecast method

Example 1.11 Consider a part with a seasonal demand pattern and where the seasonal additive forecast mode is applied. Assume the coefficients are estimated as the following; a` = 10, b` = 1, $d_\tau = (-3, -2, -1, 0, 1, 2, 3, 2, 1, 0, -1, -2)$ for $\tau = 1$ to 12. The forecast model is: $f_\tau = [(10 + 1\tau) + d_\tau]$ for $\tau = 1,2,\ldots$, whereby the forecasts for the first 3 future months become:

$$f_1 = (10 + 1) - 3 = 8$$
$$f_2 = (10 + 2) - 2 = 10$$
$$f_3 = (10 + 3) - 1 = 12$$

1.7 Forecasting Methods

In demand forecasting, four methods are in primary use. These are the following: moving average, regression, discounting and smoothing methods. A description on each is below.

1.7.1 Moving Average

The oldest forecasting method is by way of the moving average, where each month, a new average is computed. This method yields good results when the demand pattern is rather flat and of the horizontal type. This moving average method requires a parameter, N, that specifies the number of most current demands to use in generating the average. Each of the history demands $(x_1,\ldots,x_N)$ is given an equal weight in computing the average as shown in Fig. 1.10. The average becomes the forecast for the future months.

1.7.2 Regression

The regression method serves when the month-to-month demands are fluctuating about a trend line that is either moving up or moving down in a steady manner. The method calls for one parameter, N, that identifies the number of history demands to use in the forecast. The data needed are the N most current history demands ($x_1,\ldots,x_N$). Each of the demands is assigned equal weight. See Fig. 1.11.

1.7.3 Discounting

Discounting uses the most recent N history demands, $x_1,\ldots,x_N$, to generate the forecasts, and applies when the demand pattern is of the horizontal or of the trend type. This method uses a discount parameter, β, ($0 < \beta < 1$) that specifies how much weight to assign to each older monthly demand entry in the forecast calculations. The typical setting of the parameter is $\beta \cong 0.90$. Upon applying the discount parameter for N months, a set of monthly weights, ($w_1,\ldots, w_N$) evolve, where $w_N = 1.00$ and

$$w_{t-1} = \beta w_t \quad \text{for } t = N \text{ to } 2$$

Note, w_N has the largest weight and w_1 the smallest as depicted in Fig. 1.12.

1.7.4 Smoothing

The smoothing method is used when the demands follow a horizontal, trend or seasonal demand pattern. This method is also called Exponential Smoothing.

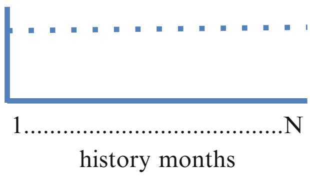


Fig. 1.11 Weights for the regression forecast method

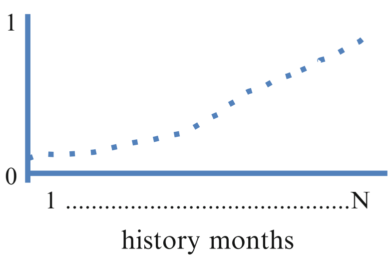


Fig. 1.12 Weights for the discount forecast method

Parameters: α = smooth the level β = smooth the slope γ = smooth seasonal indices	Forecast models: Horizontal (H) Trend (T) Seasonal multiplicative (SM) Seasonal additive (SA)
Data: x_N= most current demand entry	
Coefficients from prior month by forecast model: H: (a_{N-1}) T: (a_{N-1}, b_{N-1}) SM: $(a_{N-1}, b_{N-1}, r_{N-1})$ SA: $(a_{N-1}, b_{N-1}, d_{N-1})$	Coefficients: a = level b = slope r = seasonal ratio d = seasonal additive

Fig. 1.13 Parameters and data for the smoothing models

Smoothing parameter(s) are used to revise the forecast coefficients as each new demand entry becomes available. The horizontal smoothing model has one coefficient, a, and requires only one smooth parameter, α. The trend smoothing model has two coefficients, a, b, and uses two smooth parameters, α, β. The seasonal models have three coefficients and require three smooth parameters, α, β, γ.

Assuming the current month is N, the demand is denoted as x_N, and the coefficient estimates from the prior month are a_{N-1} for the horizontal model, a_{N-1}, b_{N-1} for the trend model, a_{N-1}, b_{N-1}, r_N for the seasonal multiplicative model, and a_{N-1}, b_{N-1}, d_N for the seasonal additive model. Smoothing parameters are needed to compute the revised estimates. The smoothing coefficients are typically set close to 0.10. Figure 1.13 lists the parameters, data, forecast models, and coefficients for the smoothing forecasts.

1.8 Horizontal Moving Average Model

The horizontal moving average model is called when the month-to-month demand flow is rather flat. The model is based on a parameter N that specifies the number of history months to use to compute the average demand per month. The N demands are $x_1, \ldots, x_N$ where x_N is the most current monthly demand. The average of the demands, called the level and denoted by, a, is calculated as below:

$$a = [x_1 + \ldots + x_N]/N$$

The forecast for each future month τ is simply

$$f_\tau = a \quad \tau = 1, 2, \ldots$$

Note, the moving average model gives equal weight to each of the N history demands in computing the forecast.

Example 1.12 Suppose forecasts are needed for a part with the N = 12 months of demand history, as listed below:

$$[6, 10, 6, 2, 7, 10, 3, 8, 3, 8, 7, 5]$$

where $x_1 = 6$ is the oldest demand entry, and $x_{12} = 5$ is the most current.

Applying the moving average method, the average (called the level) becomes a = 6.25, and thereby the forecast for future months $\tau = 1, 2, 3$, are the following: $f_1 = 6.25$, $f_2 = 6.25$, $f_3 = 6.25$, and so forth.

1.9 Trend Regression Model

The trend regression model applies for items when the history demands are trending in either an up or a down manner from month-to-month. The model is based on a parameter N that specifies the number of history demands, (x_1 to x_N) to use in the calculations; and the model gives equal weight to each history demand in the computations. The forecast for future month τ is of the form:

$$f_\tau = (a + b\tau)$$

where a is the level at the most current month, and b is the slope. The least squares method is used to estimate the coefficients from the history demands. For simplicity, the shortened version of the summations needed in the computations is listed below.

$$\Sigma x = \sum_{t=1}^{N} x_t$$
$$\Sigma t = \sum_{t=1}^{N} t$$
$$\Sigma tx = \sum_{t=1}^{N} tx_t$$
$$\Sigma t^2 = \sum_{t=1}^{N} t^2$$

Via the least squares method, the estimates of the slope, b, and the intercept, a_o, are the following:

$$\grave{b} = [\Sigma x \Sigma t - N \Sigma tx] / \left[(\Sigma t)^2 - N \Sigma t^2\right] = \text{slope}$$
$$\grave{a_0} = \left[\Sigma x - \grave{b} \Sigma t\right] / N = \text{intercept at } t = 0$$

Note, a_o` is the intercept and the level at t = 0. The level at t = N is computed in the following way:

$$a` = a_0` + Nb` = \text{level at } t = N$$

The estimates of the model coefficients are labeled as a` for the level and b` for the slope. Thereby, the forecast for future month t becomes, $f\tau = (a` + b`\tau)$.

Example 1.13 Assume an item with parameter N = 12 designating the number of history months to use in the calculations, and the most recent 12 history demands are the following:

$$[4, 7, 8, 10, 6, 8, 6, 8, 8, 8, 9, 11]$$

where $x_1 = 4$ is the oldest demand entry, and $x_{12} = 11$ is the most current. The summations are listed below:

$$\Sigma x = 93$$
$$\Sigma tx = 649$$
$$\Sigma t = 78$$
$$\Sigma t^2 = 650$$

So now, the estimate for the slope and the intercept are:

$$b` = 0.31$$
$$a_0` = 5.735$$

Finally, the level becomes:

$$a` = (5.735 + 12 \times 0.31) = 9.46$$

The forecast equation for future month τ is:

$$f_\tau = (9.46 + 0.31\tau).$$

Applying the equation for future months, 1, 2, 3, yield: $f_1 = 9.77$, $f_2 = 10.08$, and $f_3 = 10.39$, respectively.

1.10 Horizontal Discount Model

The horizontal discount model is used for items whose flow of demands is relatively flat and the forecaster wants to assign relatively less weight to each older demand. The model has two parameters, N and β. N is the number of months of demand history to use in the calculations, and β is the discount weight that specifies how much less weight to give to each older demand entry. The computations need the N most current history demands, x_t, t = 1 to N, and the corresponding set of monthly discount weights, w_t. The monthly discount weight for month t, w_t, is obtained using β and N as shown below:

$$w_t = \beta^{N-t} \quad \text{for } t = 1 \text{ to } N$$

So now, the data used to estimate the level is the following:

$$\begin{array}{lll} t & 1\ldots\ldots\ldots N & \text{(history months)} \\ x_t & x_1\ldots\ldots\ldots x_N & \text{(history demands)} \\ w_t & w_1\ldots\ldots\ldots w_N & \text{(monthly weights)} \end{array}$$

Applying the weighted least squares method, the following summations are needed:

$$\Sigma wx = \sum_{t=1}^{N} w_t x_t = \text{weighted sum of history demands}$$
$$\Sigma w = \sum_{t=1}^{N} w_t = \text{sum of weights}$$

The weighted estimate of the level now is obtained as follows:

$$a^{`} = \Sigma wx / \Sigma w$$

Thus the forecast for future month τ is set as:

$$f\tau = a^{`} \quad \tau = 1, 2, \ldots..$$

Example 1.14 Assume a part where the discounted horizontal forecast model is called using N = 12 for the number of history months, and β = 0.9 for the discount weight. The demand entries and the corresponding weights are listed below where t = 1 is the oldest month and t = N = 12 is the most current month. Note how the weights get smaller for each older month.

t	x_t	w_t
1	8	0.314
2	10	0.349
3	6	0.387
4	4	0.430
5	7	0.478
6	10	0.531
7	3	0.590
8	8	0.656
9	3	0.729
10	8	0.810
11	7	0.900
12	5	1.000

The tabulations for the weighted summations needed in the analysis are listed below:

$$\Sigma wx = 45.685$$
$$\Sigma w = 7.176$$

Using the above two summations, the weighted average demand, a`, is the following:

$$a` = 45.685/7.176 = 6.37$$

Finally, the monthly forecasts for the future months are the following:

$$f_\tau = 6.37 \quad t = 1, 2, \ldots..$$

1.11 Trend Discount Model

The trend discount model is used when the month-to-month history demands are either rising or are falling in a steady manner, and the forecaster wants to assign a higher weight to each more current demand in generating the forecasts. The model has two parameters, N and β. N is the number of months of demand history to use in the calculations, and β is the discount weight that specifies how much less weight to give to each older demand entry. The computations are based on the history demands, and a corresponding set of monthly weights. The monthly weights are computed using the discount parameter as $w_t = \beta^{N-t}$ for $t = 1$ to N. The data and corresponding weights are below.

t	1 . . . . N	(history months)
x_t	$x_1 \ldots\ldots\ldots\, x_N$	(history demands)
w_t	$w_1 \ldots\ldots\ldots\, w_N$	(weights)

The weighted summations are the following:

$$\Sigma w = \sum_{t=1}^{N} w_t$$
$$\Sigma wx = \sum_{t=1}^{N} w_t x_t$$
$$\Sigma wt = \sum_{t=1}^{N} w_t t$$
$$\Sigma wtx = \sum_{t=1}^{N} w_t t x_t$$
$$\Sigma wt^2 = \sum_{t=1}^{N} w_t t^2$$

and the computations for the slope, b, intercept, a_o, and level, a, are as follows:

$$b` = [\Sigma wx\Sigma wt - \Sigma w\Sigma wtx]/\left[(\Sigma wt)^2 - \Sigma w\Sigma wt^2\right] \quad = \text{slope}$$
$$a_0` = \left[\Sigma wx - b`\Sigma wt\right]/\Sigma w \quad = \text{intercept at } t = 0$$
$$a` = a_0` + Nb` \quad = \text{level at } t = N$$

Finally, the forecast of future month τ is set as:

$$f_\tau = a` + b`\tau \quad \text{for } \tau = 1, 2, \ldots.$$

Example 1.15 An example is provided with N = 12 history months, and β = 0.9 as the discount parameter. The N = 12 history demands are listed below along with the calculated monthly discount weights.

t	x_t	w_t
1	4	0.314
2	7	0.340
3	8	0.387
4	10	0.430
5	6	0.478
6	8	0.531
7	6	0.590
8	8	0.656
9	8	0.729
10	8	0.810
11	9	0.900
12	11	1.000

The tabulations yield the weighted sums below:

$$\Sigma w = 7.176$$
$$\Sigma wx = 58.425$$
$$\Sigma wtx = 477.691$$
$$\Sigma wt = 55.419$$
$$\Sigma wt^2 = 507.050$$

The slope, intercept, and level are computed as shown above, and the results are below:

$$b` = 0.335$$
$$a_0` = 5.556$$
$$a` = 9.574$$

The forecast equation for future month τ is the following:

$$f_{\tau} = 9.574 + 0.335\tau \quad \tau = 1, 2, \ldots$$

The first 3 monthly forecasts become: $f_1 = 9.909$, $f_2 = 10.244$, $f_3 = 10.579$.

1.12 Horizontal Smoothing Model

The horizontal smoothing model applies when only the most current monthly demand is applied in computing the forecasts, and the flow of monthly demands is flat. With this method, at the end of each month, the current month's demand, x_N, is gathered and used to revise the level from the prior month, $a`_{N-1}$. The two quantities are blended in a smoothing way to compute the revised estimate of the level, $a`_N$. A smoothing parameter, α, typically near $\alpha = 0.10$, is applied in this process. The smoothing method is described below:

$$\grave{a}_N = \alpha x_N + (1 - \alpha)\grave{a}_{N\text{-}1}$$

So now, with the current estimate of the level, the new forecasts become,

$$f\tau = \grave{a}_N \quad \tau = 1, 2, \ldots$$

This forecasting method is often called exponential smoothing.

Example 1.16 Assume the forecaster is using the horizontal smoothing model to forecast a part's demand where the smooth parameter is $\alpha = 0.10$, the prior estimate of the level is $a`_{N-1} = 10$, and the current demand is $x_N = 13$. The blending process to estimate the current level is shown below:

$$\grave{a}_N = 0.10 \times 13 + (1 - 0.10) \times 10 = 10.30$$

Finally the forecast equation for each future month is as follows:

$$f_{\tau} = 10.30 \quad \tau = 1, 2, \ldots.$$

1.13 Trend Smoothing Model

The trend smoothing model, also called exponential smoothing, applies when the flow of the month-to-month demands is either trending up or down in a steady manner. With this method, only the most current history demand is used in adjusting the forecast coefficients. The forecast for future month τ is,

$$f_{\tau} = a + b\tau$$

where a is the level, and b the slope. Denoting the current month as $t = N$, estimates of the forecast coefficients from the prior month, a_{N-1}, b_{N-1}, are saved. At the end of a month, the current month demand, x_N, is gathered and is used to revise the forecast coefficients.

Two smoothing parameters are applied in this model, α is the smooth parameter for the level, and β is for the slope. The prior estimate of the level and slope, along with the current demand, are blended as shown below in a smoothing way to revise the estimates.

$$\grave{a}_N = \alpha x_N + (1 - \alpha)\left(\grave{a}_{N-1} + \grave{b}_{N-1}\right)$$
$$\grave{b}_N = \beta\left(\grave{a}_N - \grave{a}_{N-1}\right) + (1 - \beta)\grave{b}_{N-1}$$

The forecast for the τth future month becomes:

$$f\tau = \grave{a}_N + \grave{b}_N\tau$$

Example 1.17 Suppose a part with a trend demand pattern, and where the trend smoothing model is applied with parameters: $\alpha = 0.20$ and $\beta = 0.10$. Assume the prior month's coefficient estimates are: $a_{N-1} = 10$, $b_{N-1} = 1$, and the current demand is $x_N = 14$.

The computations for the new level and slope are below:

$$\grave{a}_N = 0.20 \times 14 + (1 - 0.20) \times (10 + 1) = 11.60$$
$$\grave{b}_N = 0.10(11.60 - 10.00) + 0.90(1.00) = 1.06$$

So now, the forecast for each future month τ is the following:

$$f\tau = (11.60 + 1.06\tau) \quad \tau = 1, 2, \ldots.$$

The forecasts for first 3 future months are: $f_1 = 12.66$, $f_2 = 13.72$ and $f_3 = 14.78$, respectively.

1.14 Seasonal Multiplicative Model

The seasonal multiplicative model is used when the month-to-month flow of demands form a cyclical pattern that repeats each year. The forecast for future month τ is defined by

$$f_\tau = (a + b\tau)r_\tau$$

where a is the level, b the slope, and r_τ is the seasonal ratio for month τ. Note $(a + b\tau)$ is the trend at future month τ. When $\rho_\tau = 1$, there is no seasonal influence at month τ. At $\rho_\tau > 1$, the expected demand is higher than the trend, and at $\rho_\tau < 1$, the expected demand is lower than the trend.

1.14.1 Initialize

At least 12 months of demand history is needed to initially estimate the coefficients for the model. Thereby, the initial use of the model can begin when the item has $N = 12$ months of demand history, hence, $t = N = 12$. Below shows how the initial estimates of the 14 coefficients are estimated:

$$
\begin{aligned}
a_N &= [x_1 + \ldots + x_N]/12 &&= \text{estimate of the level} \\
b_N &= 0 &&= \text{estimate of the slope} \\
r_{N+t} &= x_t/a_N(t = 1 \text{ to } 12) &&= \text{estimate of the seasonal ratio for month } N + t
\end{aligned}
$$

At $t = 12$, the coefficient estimates are: a_{12}, b_{12}, r_{13}, …, r_{24}, and the forecast are generated by:

$$f_\tau = (a_{12} + b_{12} \times \tau) \times r_{12+\tau} \quad \text{for} \;\; \tau = 1, 2, \ldots$$

Note the forecasts will be the same as the first year demands.

1.14.2 Monthly Revise

For each month beyond $t = 12$, three of the coefficients are revised using the smoothing method along with the current month's demand. Applying the notation of $t = N$ for the current month, the new demand is denoted as x_N, and the estimates of the forecast coefficients from the prior month that have been saved are denoted as: a_{N-1}, b_{N-1}, and (r_N to r_{N+11}). Note, the seasonal ratios are for the 12 future months.

Three smoothing parameters are applied in this model, α is the smooth parameter for the level, β is for the slope, and γ is for the seasonal ratios. The revised estimates of three of the coefficients are obtained as below:

$$
\begin{aligned}
a_N &= \alpha(x_N/r_N) + (1 - \alpha)(a_{N-1} + b_{N-1}) \\
b_N &= \beta(a_N - a_{N-1}) + (1 - \beta)b_{N-1} \\
r_{N+12} &= \gamma(x_N/a_N) + (1 - \gamma)r_N
\end{aligned}
$$

1.14.3 Forecast

The forecast for the τth future month becomes:

$$f_\tau = (a_N + b_N\tau)r_{N+\tau} \quad \tau = 1, 2, \ldots$$

Example 1.18 Consider a part where the forecaster wants to apply the seasonal multiplicative model and chooses to use parameters: $\alpha = 0.10$, $\beta = 0.20$ and $\gamma = 0.30$. Note, of the 12 seasonal ratios from the past month that are saved (r_N, ..., r_{N+11}), only r_N is revised each month. Suppose, the prior coefficient estimates are: $a_{N-1} = 10$, $b_{N-1} = 2$, $r_N = 1.4$ and the current demand is $x_N = 20$. Applying the revision calculations yields the following:

$$a_N = 0.1(20/1.4) + 0.9(10 + 2) = 12.22$$
$$b_N = 0.2(12.22 - 10.00) + 0.8(2.00) = 2.044$$
$$r_{N+12} = 0.3(20/12.22) + 0.7(1.4) = 1.47$$

Using the revised coefficient estimates, the forecast for each future month τ is computed as below:

$$f_\tau = (12.22 + 2.044\tau)r_{N+\tau} \quad \tau = 1, 2, \ldots$$

1.15 Seasonal Additive Model

The seasonal additive model applies when the flow of monthly demands form a pattern that cycles about a trend line and the cycle repeats from year to year. The forecast for future month τ is:

$$f_\tau = (a + b\tau) + d_\tau,$$

where a is the level, b the slope, and d_τ is the seasonal increment for month τ. Note $(a + b\tau)$ is the trend at future month τ, and when $d_\tau = 0$, there is no seasonal influence at month τ. At $d_\tau > 0$, the expected demand is higher than the trend, and at $d_\tau < 0$, the expected demand is lower than the trend.

1.15.1 Initialize

At least 12 months of demand history is needed to initially estimate the coefficients for the model. Because of this, the initial use of the model can begin when the item has $N = 12$ months of demand history, whereby, $t = N = 12$. Below shows how the initial estimates of the 14 coefficients are estimated:

$$a_N = [x_1 + \ldots + x_N]/12 = \text{estimate of the level}$$
$$b_N = 0 = \text{estimate of the slope}$$
$$d_{N+t} = [x_t - a_N](t = 1 \text{ to } 12) = \text{estimate of the seasonal increment for month } N + t$$

At $t = 12$, the initial estimates are: a_{12}, b_{12}, d_{13}, …, d_{24}; and the forecast are obtained from:

$$f_\tau = (a_{12} + b_{12} \times t) + d_{12+\tau} \quad \text{for } \tau = 1, 2, \ldots$$

Note the forecasts will be the same as the first year demands.

1.15.2 Monthly Revise

Denoting the current month as $t = N$, estimates of the forecast coefficients from the prior month are the following: a_{N-1}, b_{N-1}, and 12 seasonal increments (d_N to d_{N+11}). At the end of each month, the current month demand, x_N, is gathered and used to revise three of the forecast coefficients, a_N, b_N and d_{N+12}.

Three smoothing parameters are applied in this model, α is the smooth parameter for the level, β is for the slope, and γ is for the seasonal increments. Note, the 12 seasonal increments saved from the past month (d_N, …, d_{N+11}), only d_N is revised each month. The prior estimates of the coefficients, along with the current demand, x_N, are blended in a smoothing way to revise three of the estimates. At $t = N$, the revised estimates of the three coefficients are as follows:

$$a_N = \alpha(x_N - d_N) + (1 - \alpha)(a_{N-1} + b_{N-1})$$
$$b_N = \beta(a_N - a_{N-1}) + (1 - \beta)b_{N-1}$$
$$d_{N+12} = \gamma(x_N - a_N) + (1 - \gamma)d_N$$

1.15.3 Forecast

The forecast for the τth future month becomes:

$$f_\tau = (a_N + b_N\tau) + d_{N+\tau}.$$

Example 1.19 Consider a part where the forecaster wants to employ the seasonal additive model to forecast the future demands. Assume the parameters are: $\alpha = 0.10$, $\beta = 0.20$ and $\gamma = 0.30$. The prior coefficient estimates are: $a_{N-1} = 10$, $b_{N-1} = 2$, $d_N = 4$ and the current demand is $x_N = 20$. The revised coefficients are below:

$$a_N = 0.1(20 - 4) + 0.9(10 + 2) = 12.40$$
$$b_N = 0.2(12.40 - 10.00) + 0.8(2.00) = 2.08$$
$$d_{N+12} = 0.3(20 - 12.40) + 0.7(4) = 5.08$$

The example shows where the revised coefficients are: $a_N = 12.4$, $b_N = 2.08$, $d_{N+12} = 5.08$, and thereby the forecast for future month τ is:

$$f_\tau = (12.4 + 2.08\tau) + d_{N+\tau}.$$

1.16 Cumulative, Integer and Weekly Forecasts

The forecasts described so far are generated for the future months and are in fractional form. In practice, however, the management often wants to use the forecasts in other ways, such as the following: for cumulative future durations; in integer formats only; and in weekly future time periods. The discussion below shows how to convert the monthly forecast to cumulative durations, integer forms, and weekly time periods.

1.16.1 Cumulative Forecasts

Forecasts are often needed in time lengths that are not monthly, but are for durations of T months in the future, could be integers or fractional intervals. In inventory planning, the lead-time, L, could be less or more than 1 month. The cumulative forecast for time duration T is denoted as F_T. The cumulative forecasts are computed as illustrated below where $f\tau$ is the future forecast for month τ.

$$\begin{aligned}
&\text{If} \quad T = 2, \quad F_2 = f_1 + f_2 \\
&\text{if} \quad T = 0.2, \ F_{0.2} = 0.2f_1 \\
&\text{if} \quad T = 1.2, \ F_{1.2} = f_1 + 0.2f_2
\end{aligned}$$

and so forth.

Example 1.20 Assume a part with monthly forecast $f_1 = 10$, $f_2 = 11$ and $f_3 = 12$. The associated cumulative forecasts when $T = 0.6$, 1.5 and 2.3 for the part are the following:

$$\begin{aligned}
&\text{if} \quad T = 0.6, F_{0.6} = 0.6 \times 10 = 6.0 \\
&\text{if} \quad T = 1.5, F_{1.5} = 10 + 0.5 \times 11 = 15.5 \\
&\text{if} \quad T = 2.3, F_{2.3} = 10 + 11 + 0.3 \times 12 = 24.6
\end{aligned}$$

1.16.2 Integer Forecasts

Most forecasting systems will display the forecasts in integers rather than in fractional form. For discussion ease, the fractional forecasts are called raw forecasts. In most inventory systems, it is common to convert the raw forecasts to integer forecasts. An algorithm to convert raw forecasts, f_τ, to integer forecasts, $f`_\tau$, is shown below:

$$\text{Set } d = 0$$
$$f`_\tau = \text{integer}[f_\tau + d + 0.5] \quad \text{and} \quad d = d + \left[f_\tau - f`_\tau\right] \quad \text{for } \tau = 1, 2, \ldots$$

Example 1.21 Suppose a part where the forecast for months 1–6 are the following: 1.3, 1.4, 1.5, 1.4, 1.3, 1.2. Converting the raw forecasts to integer forecasts yield the following: 1, 2, 1, 2, 1, 1

1.16.3 Weekly Forecasts

Some entities require forecasts on a weekly basis, rather than monthly. A way to convert the monthly forecast into weekly forecast is described below. The notation for the forecasts is the following:

$$F_\tau = \text{monthly forecast for future month } \tau, \quad \tau = 1 \text{ to N}$$
$$f_w = \text{weekly forecast for future week w}, w = 1, 2, \ldots.$$

The method is described using an example.

The example assumes the first future month of the monthly forecast is October; and the forecasts for the first three future months are 100, 120 and 150, respectively. Note, October has 31 days, November has 30 and December has 31. Table 1.1 is a

Table 1.1 Worksheet to convert monthly forecasts to weekly forecasts

(1) τ	(2) Month	(3) Days	(4) F_τ	(5) Week	(6) Cum	(7) p	(8) f_w
1	Oct	31	100	1	7	7/31	22.6
				2	14	7/31	22.6
				3	21	7/31	22.6
				4	28	7/31	22.6
				5	31	3/31	
2	Nov	30	120	5	4	4/30	25.6
				6	11	7/30	28.0
				7	18	7/30	28.0
				8	25	7/30	28.0
				9	30	5/30	
3	Dec	31	150	9	2	2/31	29.7
				10	9	7/31	33.9
				11	16	7/31	33.9
				12	23	7/31	33.9
				13	30	7/31	33.9

worksheet to convert the monthly forecast to weekly forecast for the first 13 future weeks.

Column (1) lists the notation for the future months, τ. Column (2) identifies the month, Oct, Nov, Dec. Column (3) gives the number of days in each calendar month, days. Column (4) is the forecast for the month, F_{τ}. Column (5) is the week index, w. Column (6) has the cumulative days in the month, cum. Column (7) shows the portion of the month to use for the weekly forecast, p. Column (8) records the weekly forecasts, f_w.

Note, each week, w, has 7 days. Sometimes the week is totally in 1 month, and sometimes it is in two adjacent months. In the latter case, the weekly forecast is composed of two computations, one for each month. The weekly forecasts are listed in column (8). Note, the forecasts are in fractions. In production, the forecasts are often converted to integers.

1.17 Summary

The chapter assumes the demand history for each item is saved in monthly buckets, and the forecasts are to be generated in monthly units. The history is of two type: demands and lines, where the demands is the number of pieces, and the lines is the number of orders. The monthly time buckets are either in calendar months or fiscal months. The inventory system is encouraged to develop an outlier seeking routine to find and adjust any outlier demands. Also, a return routine is recommended to adjust the demand history if any demand is returned due to ship-in-error. Four forecasting methods are described: moving averages, regression, discounting and smoothing, and three basic forecast models are developed: horizontal, trend and seasonal. In some entities, weekly forecasts are needed, and these can be converted from the monthly forecasts. The forecast routines generate forecasts in fractional form, and these forecasts are often converted to integers for ease in use.

Chapter 2
Forecast Error

2.1 Introduction

The standard deviation of the 1-month-ahead forecast error is used to determine how much safety stock is needed to satisfy the level of service to customers. An exact measure of the standard deviation is not easy to generate, however, estimates are used in its place. The way to estimate the standard deviation varies depending on the forecasting method in use (moving average, regression, discounting, smoothing). This chapter describes how to estimate the standard deviation for each of the forecast models presented in Chap. 1.

Another important statistic for each forecast model is the coefficient of variation (cov). The closer the cov is to zero, the better. When cov is near 0.30, the distribution of the forecast errors is similar to normal distribution. When the cov is closer to 1.00, the distribution of the forecast errors is not like a normal distribution. In Chap. 4, the concept of the truncated normal is presented. The cov is used to estimate the parameter to use for the truncated normal. When the distribution of the forecast errors is not normally distributed, the truncated normal distribution should be applied to determine how much safety stock is needed. In determining the safety stock, the standard deviation and cov of the lead-time demand are used in the computations. The methods to generate these statistics are included in the chapter.

When the forecast and its standard deviation is generated for a total collection of SKUs, estimates are needed for each SKU. A way to measure the standard deviation for each SKU is presented. As an example, this need occurs for a style shoe where the forecast is based on all sizes combined, and estimates for inventory control are needed for each size of the style. Estimates of the standard deviation for each forecast method is provided.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_2

2.2 Standard Deviation for the 1-Month-Ahead Forecast Error

The forecast is the main guide on how much stock to have available for each item over the future planning months. But since the forecast is an estimate and is not definitive, an extra layer of stock is needed to fill the customer demands that may occur above the forecast. This added stock is called the safety stock. A common way to measure the accuracy of the forecast is by way of the standard deviation, σ, of the 1-month forecast error. The standard deviation is used in subsequent computations that determine how much safety stock is needed to satisfy a desired level of service set by the management.

This chapter shows how to estimate the standard deviation for a 1-month forecast error, on the various forecasting methods described in Chap. 1. In addition, the chapter describes another measure, called the coefficient of variation, denoted as cov, that is a relative way to measure the accuracy of the forecasts. The cov is the ratio of the standard deviation over the most current level. The closer cov is to zero, the more accurate the forecast. Also when the cov is 0.50 or lower, the forecast errors are distributed similar to a normal distribution, and when cov is 1.00 or higher, the forecast errors are more like an exponential distribution.

2.3 Standard Deviation for the Horizontal Moving Average Model

The data available for the horizontal moving average model is the following:

$$\mathrm{x_t} \ \text{ for } \ \mathrm{t} = 1 \ \text{ to } \ \mathrm{N}$$

where x_t is the demand in history month t, N is the number of history months, and $t = N$ is the most current month. Also known from the forecast computation is the most current measure of the level, denoted as a`. The level is used to measure the residual error for each history month as below:

$$\mathrm{e_t} = \left(\mathrm{x_t} - \mathrm{a}^{`}\right) \quad \text{for } \ \mathrm{t} = 1 \ \text{ to } \ \mathrm{N}$$

So now, an estimate of the 1-month forecast error is computed as follows:

$$\sigma^{`} = \sqrt{\sum\nolimits_{t=1}^{N} e_t^2/(N-1)}$$

The coefficient of variation becomes:

$$\mathrm{cov} = \sigma^{`}/\mathrm{a}^{`}$$

2.4 Standard Deviation for the Trend Regression Model

The data available for the trend regression model is also the following:

$$x_t \text{ for } t = 1 \text{ to } N$$

where x_t identifies the history demand in month t, N is the number of history months, and $t = N$ is the most current month. Upon completing the forecast computations, the estimated coefficients are the following:

$$a_o = \text{intercept at } t = 0$$
$$b = \text{slope}$$
$$\grave{a} = \text{level at } t = N$$

The fit measures the forecast going backwards over the history months. The fit for history month t is obtained as follows:

$$f_t = (a_o + bt) \quad \text{for } t = 1 \text{ to } N$$

The residual error for each history month is measured as below:

$$e_t = (x_t - f_t) \quad \text{for } t = 1 \text{ to } N$$

and the estimate of the 1-month forecast error is obtained as follows:

$$\grave{\sigma} = \sqrt{\sum_{t=1}^{N} e_t^2/(N-2)}$$

In this situation, the estimate of the most current level is the fit at $t = N$, whereby,

$$\grave{a} = f_N = (a_o + bN)$$

and so, the measure of the coefficient of variations is as below:

$$cov = \grave{\sigma}/\grave{a}$$

2.5 Standard Deviation for the Horizontal Discount Model

The data available for the horizontal discount model is the following:

$$x_t \text{ for } t = 1 \text{ to } N$$

and

$$w_t \text{ for } t = 1 \text{ to } N$$

where x_t is the demand in history month t, w_t is the discount weight for month t, N is the number of history months, and t = N is the most current month.

Also known from the forecast computations is the most current measure of the level, denoted as a`. The level is used to measure the residual errors for each history month as below:

$$e_t = \left(x_t - a^`\right) \quad \text{for } t = 1 \text{ to } N$$

So now, an estimate of the standard deviation of the 1-month forecast error is as follows:

$$\sigma^` = \sqrt{\sum_{t=1}^{N} w_t e_t^2 / \sum_{t=1}^{N} w_t}$$

The coefficient of variation becomes:

$$cov = \sigma^` / a^`$$

2.6 Standard Deviation for the Trend Discount Model

The data available for the trend discount model is the following:

$$x_t \text{ for } t = 1 \text{ to } N$$

and

$$w_t \text{ for } t = 1 \text{ to } N$$

where x_t is the demand in history month t, w_t is the discount weight for month t, N is the number of history months, and t = N is the most current month.

Also known from the forecast computations are the following coefficient estimates: a_o = intercept at t = 0, b = slope, and a` = the current measure of the level. With these coefficients, the fit of the past history months are obtained as follows:

$$f_t = (a_o + bt) \quad \text{for } t = 1 \text{ to } N$$

The fit is used to measure the residual errors for each history month as below:

$$e_t = (x_t - f_t) \quad \text{for } t = 1 \text{ to } N$$

With the N set of residual errors, the estimate of the standard deviation of the 1-month forecast error is estimated as follows:

$$\grave{\sigma} = \sqrt{\sum_{t=1}^{N} w_t e_t^2 / \sum_{t=1}^{N} w_t}$$

The corresponding coefficient of variation becomes:

$$\text{cov} = \grave{\sigma} / \grave{a}$$

2.7 Standard Deviation for the Horizontal Smoothing Model

When horizontal smoothing is in use to forecast the monthly demands at month $t = N$, the data available are the following:

α = smoothing parameter
σ_{N-1} = estimate of the standard deviation from the prior month, $t = N-1$
a_{N-1} = estimate of the level from month $t = N-1$
x_N = demand at month $t = N$

The current 1-month-ahead forecast error is obtained as below:

$$e = (x_N - a_{N-1})$$

So now in keeping with the spirit of smoothing, the current forecast error, e, is blended with the prior estimate of the standard deviation, σ_{N-1}, to yield the current estimate of the standard deviation as shown below:

$$\sigma_N = \sqrt{\alpha e^2 + (1 - \alpha)(\sigma_{N-1})^2}$$

The estimate of the level at $t = N$ is denoted as a_N, and thereby the coefficient of variation is obtained as follows:

$$\text{cov} = \sigma_N / a_N$$

2.8 Standard Deviation for the Trend Smoothing Model

When the trend smoothing model is in use to forecast the monthly demands for month $t = N$, the data available are the following:

σ_{N-1} = estimate of the standard deviation from the prior month, $t = N-1$
a_{N-1} = estimate of the level from month $t = N-1$

b_{N-1} = estimate of the slope from month t = N−1
x_N = demand at month t = N

The forecast from month t = N−1 for month t = N is denoted as $x`_N$ and is obtained by the following:

$$x`_N = (a_{N-1} + b_{N-1})$$

Thereby, the current 1-month-ahead forecast error is derived as below:

$$e = (x_N - x`_N)$$

So now, the current forecast error, e, is blended with the prior estimate of the standard deviation, σ_{N-1}, to yield the current estimate of the standard deviation as shown below:

$$\sigma_N = \sqrt{\alpha e^2 + (1-\alpha)(\sigma_{N-1})^2}$$

At t = N, the level and slope are revised and are denoted as:

a_N = estimate of the level at month t = N
b_N = estimate of the slope at month t = N

The corresponding coefficient of variation is obtained as follows:

$$cov = \sigma_N / a_N$$

2.9 Standard Deviation for the Seasonal Multiplicative Model

When the seasonal-multiplicative model is in use to forecast the monthly demands for month t = N, the data available are the following:

σ_{N-1} = estimate of the standard deviation from the prior month, t = N−1
a_{N-1} = estimate of the level from month t = N−1
b_{N-1} = estimate of the slope from month t = N−1
$r_N, \ldots, r_{N+11}$ = estimate of 12 months of seasonal ratios as of t = N−1
x_N = demand at month t = N

The forecast from month t = N−1 for month t = N is denoted as $x`_N$ and is obtained by the following:

$$x`_N = (a_{N-1} + b_{N-1}) r_N$$

Thereby, the current 1-month-ahead forecast error is derived as below:

$$e = \left(x_N - x_N^{`}\right)$$

The current forecast error, e, is blended with the prior estimate of the standard deviation, σ_{N-1}, to yield the current estimate of the standard deviation as shown below:

$$\sigma_N = \sqrt{\alpha e^2 + (1-\alpha)\left(\sigma_{N-1}\right)^2}$$

At t = N, the level, slope and seasonal ratios are revised and denoted as:

a_N = estimate of the level at month t = N
b_N = estimate of the slope at month t = N
$r_{N+1}, \ldots, r_{N+12}$ = estimate of 12 months seasonal ratios as of t = N

The corresponding coefficient of variation is obtained as follows:

$$cov = \sigma_N / a_N$$

2.10 Standard Deviation for the Seasonal Additive Model

When the seasonal-additive model is in use to forecast the monthly demands for month t = N, the data available are the following:

σ_{N-1} = estimate of the standard deviation from the prior month, t = N−1
a_{N-1} = estimate of the level from month t = N−1
b_{N-1} = estimate of the slope from month t = N−1
$d_N, \ldots, d_{N+11}$ = estimate of 12 months of seasonal increments as of t = N−1
x_N = demand at month t = N

The forecast from month t = N−1 for month t = N is denoted as $x_N^{`}$ and is obtained by the following:

$$x_N^{`} = \left(a_{N-1} + b_{N-1}\right) + d_N$$

Thereby, the current 1-month-ahead forecast error is derived as below:

$$e = \left(x_N - x_N^{`}\right)$$

The current forecast error, e, is blended with the prior estimate of the standard deviation, σ_{N-1}, to yield the current estimate of the standard deviation as shown below:

$$\sigma_N = \sqrt{\alpha e^2 + (1-\alpha)\,(\sigma_{N-1})^2}$$

At t = N, the level, slope and seasonal additives are revised as noted below:

a_N = estimate of the level at month t = N
b_N = estimate of the slope at month t = N
$d_{N+1}, \ldots, d_{N+12}$ = estimate of 12 months of seasonal increments as of t = N

The corresponding coefficient of variation is obtained as follows:

$$\text{cov} = \sigma_N / a_N$$

2.11 Standard Deviation for the Cumulative Forecast

Recall, the cumulative forecast for time duration T is denoted as F_T, where T > 0. Whether T is an integer or a fraction, the estimate of the standard deviation of the cumulative forecast, denoted as σ_T, is the following:

$$\sigma_T = \sqrt{T}\sigma$$

The above relation assumes that the monthly future forecast errors are independent and each future month has the same standard deviation, σ.

Example 2.1 Suppose a situation where the monthly standard deviation of the forecast error is σ = 10. Below gives an example of the corresponding standard deviations for cumulative forecasts of length, T = 0.25, 1.7 and 4.0.

$$\sigma_{0.25} = \sqrt{0.25} \times 10 = 5.00$$
$$\sigma_{1.70} = \sqrt{1.70} \times 10 = 13.00$$
$$\sigma_{4.00} = \sqrt{4.00} \times 10 = 20.00$$

2.12 Standard Deviation for Weekly Forecasts

When the monthly forecasts are converted to weekly forecasts, the data available for the monthly forecasts are the following:

F = monthly forecast
σ = standard deviation for a monthly forecast error

The related notation for weekly forecasts is below:

f_w = weekly forecast
σ_w = standard deviation for the weekly forecast

In general, $f_w = F/(52/12) = F/4.33 = 0.23F$

Thereby, the weekly standard deviation becomes:

$$\sigma_w = \sqrt{0.23}\sigma = 0.48\sigma$$

2.13 Standard Deviation for an SKU

In many inventory situations, the units are stocked in a multitude of variations, like, models, sizes and locations. Baseball gloves are stocked in various models, shoes are in a variety of sizes, and cellular phones come in different colors and features. Each of the variations is called a stock-keeping-unit (SKU). A forecast is typically generated for the total but not directly for every SKU. For inventory control applications, however, a forecast and standard deviation is needed with each SKU.

Below shows how to generate the standard deviation by SKU as needed. The typical data known is the following:

F = one-month forecast for the total of all SKUs
σ = standard deviation for the one-month total
p_i = portion of demand for SKU i
n = number of SKUs
$\Sigma p_i = 1$

For a particular month, the following notation is used in the derivation:

Y = total one-month forecast
x_i = demand for SKU i
$\Sigma x_i = Y$

Let x = demand for SKU i. Also, for notational ease, let $p = p_i$ represent the portion of demand for SKU i.

The steps to find the forecast and standard deviation for an SKU is listed below:

$$E[x|Y] = pY$$
$$E[x] = pE[Y] = pF$$
$$V[x|Y] = E[x^2|Y] - E[x|Y]^2 = Yp(1-p)$$
$$E[x^2|Y] = V[x|Y] + E[x|Y]^2 = Yp(1-p) + p^2Y^2$$
$$\begin{aligned} E[x^2] &= E[Y]p(1-p) + p^2E[Y^2] \\ &= Fp(1-p) + p^2[F^2 + \sigma^2] \end{aligned}$$
$$\begin{aligned} V[x] &= E[x^2] - E[x]^2 \\ &= Fp(1-p) + p^2[F^2 + \sigma^2] - p^2F^2 \\ &= Fp(1-p) + p^2\sigma^2 \end{aligned}$$

Finally, the 1-month forecast for SKU i, f_i, and the corresponding standard deviation, σ_i, for SKU i are below:

$$f_i = p_i F$$
$$\sigma_i = \left[F p_i (1 - p_i) + p_i^2 \sigma^2\right]^{0.5}$$

Example 2.2 Suppose a shoe store where the forecast for a style shoe is F = 10 per month, and the corresponding standard deviation is $\sigma = 3$. A particular size (SKU) has an average of ten percent (p = 0.10) of the sales. The forecast, f_i, and standard deviation, σ_i, for the size shoe is computed as below:

$$f_i = 0.10 \times 10 = 1.0$$
$$V(x) = 10 \times 0.10(1 - 0.10) + 0.10^2 3^2 = 0.99$$
$$\sigma_i = \sqrt{0.99} = 0.99$$

Note, the coefficient of variation for the style is:

$$cov = \sigma/F = 0.30$$

and the cov for the size (SKU i) is:

$$cov_i = \sigma_i/f_i = 0.99$$

2.14 Summary

Estimates of the 1-month-ahead standard deviation are developed for each of the forecast models of Chap. 1. The estimates vary depending on the method of forecast, as per moving average, regression, discounting and smoothing. Special estimates are also presented for the seasonal multiplicative model and for the seasonal additive model. The coefficient of variation (cov) is likewise an important statistic generated for each model. The way to estimate these statistics also varies by forecast model in use. The standard deviation and the cov are also generated when the forecasts are for time durations not of 1 month, but instead are for less than 1 month, or for more than a month. The chapter also describes how to estimate the standard deviation for an individual SKU, for a situation when the forecasting system measures the standard deviation for the aggregate collection of all the SKUs, and also for weekly forecasts.

Chapter 3
Order Quantity

3.1 Introduction

The order quantity, Q, is a planned amount of stock to replenish the inventory from the supplier when new stock is needed. In the ideal situation, Q is calculated to give the minimum cost of holding and ordering the stock, called the economic order quantity, EOQ. The EOQ is obtained from several factors: A is the annual forecast of the item, Co is the cost per order, c is the cost per unit, and h is the annual holding rate. Note, A comes from the forecast, c from the supplier, and Co and h are parameters from the stocking facility. In some facilities, the order quantity is restricted to fall in monthly buckets, but in most situations it is not limited in that way. A related cost measure is the stocking rate denoted as, s, which measures the rate of added cost on the item due to stocking the item via hold plus order. The stocking rate yields another measure labeled the stocking cost per unit denoted as $(c \times s)$ which is a hidden added cost due to hold plus order. The effective cost per unit becomes: $c` = c(1 + s)$. It is sometimes useful for the inventory management to realize the sensitivity between the four ingredients and the order quantity. When A changes, how does this affect Q? The same queries can be made with changes in Co, c and h. Another measure of interest is the change in the hold plus order cost when using an order quantity different from the EOQ. Note also, as the order quantity changes, the on-hand inventory also changes.

Oftentimes, the simple economic order quantity is restricted due to a series of constraints. The more typical type are described below.

- Price breaks are when the supplier offers discounts on the cost per unit or cost per order as the buy quantity increases. The more you buy, the less you pay.
- Min and Max constraints are when the supplier specifies that the buy quantity cannot be higher than a Max constraint, and cannot be lower than a Min constraint.
- Multiple quantity constraints occur when the supplier specifies a multiple number for which the buy amount must conform.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_3

- Bin Max constraint is when the stock holding entity has limited space to stock the item. The on-hand plus on-order quantity cannot exceed this Bin Max quantity for the item.
- Job Quantity constraint is when the stock holding entity typically sells the units in multiples of a fixed number, like four tires on an auto vehicle. The order quantity to buy is restricted to conform to the Job Quantity and the current on-hand for the item.

In service parts, two important order quantities are the initial buy quantity and the final buy quantity.

- The initial buy quantity occurs when a new part is introduced and there is no demand history to base a forecast. The service facility is obliged to carry stock on this item and must generate a first time buy.
- The final buy quantity occurs when the supplier notifies the stocking facility they will stop supplying the item, and the facility is obliged to have the item available for a prolong obligation period. A last time buy is needed.

3.2 Order Quantity in Monthly Buckets

An important task at the stock locations is to determine the plan quantity to buy for each item in the inventory. Typically, this decision is needed every month as each new forecast becomes available. This section shows how the quantity is chosen to conform with the minimum order plus hold cost. In the computations, assume the order quantity is limited in monthly buckets, whereby the candidate quantities, Q, are for 1-month of forecast, 2-months of forecasts, and so forth. The method assumes the forecasts are the same each month. The optimal quantity is denoted as Q*. The data needed for each item to carry on this analysis is the following:

A is the annual forecast
c is the cost per unit
Co is the cost for each order
h is a parameter on the cost to hold in stock one-dollar of inventory for a whole year

The annual demand, A, is an estimate that comes from a forecast for the future 12 months. The cost per unit, c, is provided from the supplier of the item. The cost per order, Co, is an estimate of the expense associated with generating each new order quantity. This cost may include the expense associated with processing the order, plant setup, receiving, shipping, binning and transportation. It is essentially a parameter that the management sets as best it can. The annual holding rate, h, is a parameter that is attributed to holding a unit of the item in stock for 1 year. It includes the costs associated with the bin space, warehouse overhead, theft, deterioration, obsolescence, insurance, taxes and potential revenue lost from other investment opportunities.

With this information, the annual order cost, Ko, and the annual holding cost, Kh, are computed separately. Their sum, K, yields the annual order plus hold cost. The monthly quantity with the minimum annual cost is the economic quantity, Q*, to buy.

The computations are run for quantities of m = 1–12 months of supply. The computations for each of the 12 months are summarized below:

$$\begin{aligned}
&\text{m} = \text{months supply } (1, 2, \ldots\ldots, 12) \\
&\text{Q} = \text{mA}/12 = \text{m-months supply} \\
&\text{Ko} = \text{Co} \times \text{A}/\text{Q} = \text{annual order cost with Q} \\
&\text{Kh} = \text{c} \times \text{h} \times \text{Q}/2 = \text{annual holding cost with Q} \\
&\text{K} = (\text{Ko} + \text{Kh}) = \text{annual (order + hold) cost with Q}
\end{aligned}$$

The month m with the minimum K is noted as m* and the corresponding optimal order quantity becomes:

$$\text{Q}^* = \text{m}^*\text{A}/12.$$

Example 3.1 Consider an item where the annual forecast is 60 units, the cost per unit is $50, and the cost per order is $20. Further, the annual holding rate is 0.24. Since the annual forecast is 60 pieces, the average monthly forecast is five pieces. Table 3.1 shows the candidate order quantities are: 5, 10, 15, and so forth. For each these quantities, the annual order cost, Ko, and annual holding cost, Kh, are computed, and the associated sum, K, is listed. Since the minimum order plus hold cost, K = $170, occurs when Q = 15, the economic order quantity for the item is Q* = 15 pieces.

Table 3.1 Computations from Example 3.1

m	Q	Ko	Kh	K
1	5	240.00	30.00	270.00
2	10	120.00	60.00	180.00
3	15	80.00	90.00	170.00
4	20	60.00	120.00	180.00
5	25	48.00	150.00	198.00
6	30	40.00	180.00	220.00
7	35	34.29	210.00	244.29
8	40	30.00	240.00	270.00
9	45	26.67	270.00	296.67
10	50	24.00	300.00	324.00
11	55	21.82	330.00	351.82
12	60	20.00	360.00	380.00

m months supply, *Q* quantity, *Ko* annual order cost, *Kh* annual hold cost, *K* annual order plus hold cost

3.3 Stocking Rate

A related measure to the order quantity is called the stocking rate. This is a measure of the added expense to the item due to the need to order plus hold the item in the inventory. The optimal stocking rate is denoted as s, and represents an added portion of cost to each unit. The stocking rate is obtained from the following:

$$s = K^*/[A \times c]$$

where K* is the optimal annual cost of order plus hold, and $A \times c$ is the annual out-of-pocket cost.

3.4 Effective Cost-Per-Unit

Once the stocking rate, s, is computed, the effective cost per unit becomes:

$$c\grave{} = c(1 + s).$$

This effective cost represents the sum of the out-of-pocket cost plus order and hold cost for each unit.

Example 3.2 From the prior example, $K^* = \$170$ and $A \times c = [60 \times 50] = \$3000$. Thereby, $s = 170/3000 = 0.0566$ is the optimal stocking rate representing the ratio of order plus hold cost over the cost per unit. The associative effective cost per unit becomes:

$$c\grave{} = c \times [1 + s] = \$50 \times [1 + 0.0566] = \$52.83.$$

3.5 Economic Order Quantity

The order quantity, Q, that is not limited to monthly buckets, and which yields the minimum annual order-plus-hold cost for an item is called the economic order quantity (EOQ). The data needed in the computations is the same as in the prior section, and is relisted below:

A is the annual forecast
c is the cost per unit
Co is the cost per order
h is the annual holding rate

The annual order cost is: ACo/Q, and the annual hold cost is: chQ/2. Thereby, the annual order-plus-hold cost when Q is the order quantity is computed in the following way:

$$\mathrm{K(Q)} = \frac{\mathrm{ACo}}{\mathrm{Q}} + \frac{\mathrm{chQ}}{2}$$

With a bit of calculus, the quantity, Q, that yields the minimum annual order-plus-hold cost is denoted as Q* and is obtained as below:

$$\mathrm{Q}^* = \sqrt{2ACo/(ch)}$$

Plugging the optimal order quantity, Q*, into K(Q) gives the relation below:

$$\mathrm{K(Q^*)} = \frac{\mathrm{ACo}}{\mathrm{Q^*}} + \frac{\mathrm{chQ^*}}{2}$$

With some algebra, the following relation emerges:

$$\mathrm{K(Q^*)} = \sqrt{2AchCo}$$

3.6 Stocking Rate

Recall, the stocking rate is a measure of the ratio of order-plus-hold cost over the cost-per-unit. This can be calculated for any quantity Q in the following way:

$$\mathrm{s} = \mathrm{K(Q)/(Ac)}$$

3.7 Effective Cost-Per-Unit

The effective cost-per-unit gives the sum of the following costs: (out-of-pocket + order + hold), and is calculated as below:

$$\mathrm{c}^{`} = \mathrm{c}(1 + \mathrm{s})$$

Example 3.3 Suppose an item where the annual forecast is A = 60 pieces, the cost-per-unit is \$50, the cost per order is \$20, and the annual holding rate is 0.24. Applying the relation, the economic order quantity yields:

$$\mathrm{Q}^* = \sqrt{2 \times 60 \times 20/50 \times 0.24} = 14.14$$

and the annual order plus hold cost is the following:

$$\mathrm{K(Q^*)} = \sqrt{2 \times 60 \times 50 \times (0.24 \times 20)} = \$169.70$$

Continuing, the stocking rate and effective cost-per-unit are computed below.

$$s = 169.70/(60 \times 50) = 0.056$$
$$c^{`} = 50(1 + 0.056) = \$52.8$$

Example 3.4 Suppose the item of Example 3.3 again, and assume the stocking location orders Q = 5 pieces instead of the EOQ of Q* = 14. The computations below shows where the annual order plus hold cost becomes K(5) = $270, much higher than the optimal, K(14) = $169.70. Further computations show where the associated stocking rate with Q = 5 is s = 0.090, and the associated effective cost per unit is c` = $54.50.

$$K(Q) = 60 \times 20/5 + 50 \times 0.24 \times 5/2 = \$270$$
$$s = 270/(60 \times 50) = 0.090$$
$$c^{`} = 50(1 + 0.09) = \$54.50$$

3.8 Sensitivity

It is helpful to the management of the inventory to realize the relation between the order plus holding cost and the ingredients that are used to compute the order quantity. Below is a description on how the hold plus order cost, and also the on-hand inventory is sensitive to any changes in the input ingredients: annual demand, cost per order, cost per unit, and annual holding rate.

3.8.1 Sensitivity of the Annual Demand on Q

When the annual demand changes from A to A`, and the other ingredients (c, Co, h) remain the same, the economic order quantity changes from Q* to Q` by the relation listed below:

$$Q^{`} = Q^{*}\sqrt{A^{`}/A}$$

The average on-hand (OH) inventory, not including safety stock, when Q is the order quantity is Q/2. So a change in the annual demand also affects the on-hand inventory as described below. For notation, the relation between Q* and Q` on the on-hand is the following:

Q* yields OH
Q` yields OH`

Thereby, the change in on-hand inventory when the annual demand changes from A to A` is measured by the relation below:

$$OH^{`} = OH\sqrt{A^{`}/A}$$

Example 3.5 Suppose a part with a 10 % increase in the annual forecast (A`/A = 1.1). The computation below shows where the change from A to A` yields a 4.9 % increase in the order quantity.

$$Q^{`} = Q^{*}\sqrt{1.1} = 1.049$$

The 4.9 % increase in the order quantity causes the on-hand inventory to also rise by 4.9 %.

In the event the annual forecast on the part decreases by 10 % (A`/A = 0.9), the order quantity drops by 5.1 % since,

$$Q^{`} = Q\sqrt{0.9} = 0.949$$

Note, the on-hand inventory will also decrease by 5.1 %.

Example 3.6 Suppose the on-hand inventory on a group of N parts is $1,000,000 and the forecast for the group is projected to rise by 10 %. In this situation, the on-hand inventory is projected to increase to $1,048,000.

3.8.2 Sensitivity of the Cost Per Order on Q

When the cost per order changes from Co to Co`, and all the other components that affect the order quantity (A, c, h) remain the same, the economic order quantity changes as below:

$$Q^{`} = Q^{*}\sqrt{Co^{`}/Co}$$

Recall, the average on-hand (OH) inventory, not including safety stock, becomes Q/2. So a change in the cost per order also affects the on-hand inventory as described below. For notation, the relation between Q* and Q` on the on-hand is the following:

Q* yields OH
Q` yields OH`

Thereby, the change in on-hand inventory when the cost per order changes from Co to Co` is measured by the relation below:

$$OH^{`} = OH\sqrt{Co^{`}/Co}$$

Example 3.7 Suppose a situation where the cost per order is Co = $10 for each of N parts, and the sum of the on-hand inventory is $1,000,000. Management wishes to reduce the inventory to $900,000 and is seeking how much of a change in the cost per order is needed. To accomplish, the relation is stated:

$$\sqrt{Co^{`}/Co} = 0.9$$

Thereby, the adjusted cost per order becomes the following:

$$\text{Co}^{`} = 0.9^2\text{Co} = 0.81\text{Co} = \$8.10$$

Hence, the cost per order should be reduced to $8.10 to accomplish.

Example 3.8 In the event the management from Example 3.7, further inquires on the change in the cost per order needed to reduce the inventory of the N parts from $1,000,000 to $800,000, the relation needed is as below:

$$\sqrt{Co^{`}/Co} = 0.8$$

Thereby, the cost per order would need to be lowered to $6.40 as shown in the computations below.

$$\text{Co}^{`} = 0.8^2\text{Co} = 0.64\text{Co} = \$6.40.$$

3.8.3 Sensitivity of the Cost Per Unit on Q

In the event the cost per unit changes from c to c`, and all the other components that affect the economic order quantity (A, Co, h) remain the same, the economic order quantity, Q*, revises as below:

$$\text{Q}^{`} = \text{Q}^*\sqrt{\frac{c}{c^{`}}}$$

Note, a 10 % increase in the cost-per-unit causes the economic order quantity to decrease to $\text{Q}^{`} = \text{Q}^*\sqrt{1/1.1} = 0.953\text{Q}^*$. The corresponding change in on-hand inventory (from OH to OH`) is shown below:

$$\text{OH}^{`} = \text{OH}\sqrt{\frac{c}{c^{`}}}$$

Example 3.9 A distribution center (DC) has $1,000,000 of on-hand inventory on all the parts from a vendor who will increase the price on each part by 10 %. The price change will cause the on-hand inventory to change as below.

$$1{,}000{,}000\sqrt{1/1.1} = 1{,}000{,}000 \times 0.953 = \$953{,}000$$

at the current price. The value of the inventory at the new price becomes:

$$OH^{`} = 953{,}000 \times 1.1 = \$1{,}048{,}300$$

3.8.4 Sensitivity of the Annual Holding Rate on Q

The annual holding rate, h, is a strategic parameter set by the management to steer the size of the inventory. In the event the annual holding rate changes from h to h`, and all the other components that affect the economic order quantity (A, Co, c) remain the same, the economic order quantity, Q*, revises as below:

$$Q^{`} = Q^{*}\sqrt{h/h^{`}}$$

Note, a 10 % increase in the annual holding rate causes the economic order quantity to decrease to $Q^{`} = Q^{*}\sqrt{1/1.1} = 0.953Q^{*}$. The corresponding change in on-hand inventory (from OH to OH`) is shown below:

$$OH^{`} = OH\sqrt{h/h^{`}}$$

Example 3.10 Assume a distribution center has $1,000,000 of inventory for a group of parts and currently uses an annual holding rate of h = 0.24. The management wants to reduce the investment to $900,000 and is seeking how to do this with a change in the annual holding rate.

To accomplish, $\sqrt{h/h^{`}} = 0.9$ is noted, whereby, h/h` = 0.81, and thereby, h` = h/0.81 = 0.24/0.81 = 0.296 is the holding rate to apply.

Example 3.11 Suppose, the management from the example above further asks how to change the annual holding rate to reduce the inventory investment to $800,000.

In this situation, $\sqrt{h/h^{`}} = 0.8$ yields h/h` = 0.64, and h` = h/0.64 = 0.24/0.64 = 0.375.

3.8.5 Sensitivity When Q is Used Instead of the EOQ

Recall, the economic order quantity (EOQ) is denoted as Q*, and the associated annual hold plus order cost is labeled as K*. In some occasions the quantity Q is used in place of Q*, and thereby the annual order plus hold cost becomes K instead of K*. This difference in the order quantity could come about, say, when management has a policy to always order 1 week of stock for all items regardless of the EOQ. In other situations, the management may decide to order 6 months of stock, without regard to the EOQ. Below gives the relation that measures how K diverges from K* when Q differs from Q*:

$$K = K^* \times 0.5[Q/Q^* + Q^*/Q]$$

Example 3.12 Assume a part in a distribution center where the economic order quantity is Q* = 10, and the associated annual order plus hold cost is K* = $1000. If the management orders Q = 5, instead of Q* = 10, the annual order plus hold cost becomes:

$$K = 1000 \times 0.5[5/10 + 10/5] = \$1250$$

Example 3.13 Suppose the same part as Example 3.12 where Q* = 10 and K* = $1000. If the order quantity selected is Q = 50, instead of Q* = 10, the annual order plus hold cost becomes the following:

$$K = 1000 \times 0.5[50/10 + 10/50] = \$2600$$

3.9 Constraints

Some supplier contracts include constraints on the buy quantity of a part. Below describes the more common constraints and shows how to adjust the buy quantity to accommodate. The constrains are the following: price discounts, set-up cost discounts, min quantity, max quantity, multiple quantity, bin-max quantity, and job quantity.

3.9.1 Price Discounts

Some vendor contracts have discount terms that encourage their customers to reach for higher buy quantities to gain lower unit costs. These contracts have N tiers where each tier, i, has a minimum quantity, q(i), and a cost per unit, c(i). When the customer buy quantity is q(i) or larger, the cost per unit is c(i). The higher the tier, the lower the cost per unit. The corresponding part data is the following: A is the annual forecast, h is the annual holding rate, and Co is the cost per order.

For each tier, the economic order quantity is computed and is denoted as Q`. In the event Q` is not in the admissible range of the tier, the closest quantity, Q, in the range is selected. This adjustment is needed to ensure the quantity is appropriate to gain the cost per unit that the tier offers. So the data now associated with each tier i becomes the following: A, h, Co, q(i) and Q. Applying the admissible Q, the annual cost for the tier is computed. The annual cost, denoted as K, includes the out-of-pocket plus order and hold cost as shown below:

$$K = A \times c(i) + Co \times A/Q + c(i)h \times Q/2$$

The annual cost for each of the N tiers is compared and the order quantity, Q, that is associated with the minimum of the N values of K is selected as the economic order quantity, Q*.

Example 3.14 Consider an item with the annual forecast of A = 120, the annual holding rate is h = 0.24 and the cost per order is Co = $20. The vendor contract calls for price breaks with N = 4 tiers as follows:

Contract terms:

i	q(i)	c(i)
1	1	10.00
2	50	9.75
3	100	9.50
4	150	9.25

Applying the computations to find Q`, Q and K, one tier at a time, the results are listed below:

i	Q`	Q	K
1	44.7	45	1307
2	45.2	50	1276
3	45.8	100	1278
4	46.4	150	1292

At tier i = 1,

$$Q` = \sqrt{(2 \times 120 \times 20)/(10 \times 0.24)} = 44.7$$
$$Q = 45$$
$$K = 120 \times 10 + 20 \times 120/45 + 10 \times 0.24 \times 45/2 = 1307$$

At tier i = 2,

$$Q` = \sqrt{(2 \times 120 \times 20)/(9.75 \times 0.24)} = 45.2$$
$$Q = 50$$
$$K = 120 \times 9.75 + 20 \times 120/50 + 9.75 \times 0.24 \times 50/2 = 1276$$

Note, Q` = 45.2 is not in the range of 50 or larger, and thereby, the closest admissible quantity is Q = 50. Because tier i = 2 has the minimum annual cost, K = 1276, and the associated order quantity is Q = 50, the economic order quantity for this part becomes Q* = 50.

3.9.2 Setup Cost Discounts

Some supplier contracts offer discounts associated with the setup cost of an item. A set of target quantities are specified that reduces the setup cost charged to the customer. These contracts have N tiers where each tier, i, has a minimum quantity, q(i), and a setup cost SUC(i). When the customer buy quantity is q(i) or larger, the setup cost is SUC(i). The higher the tier, the lower the setup cost. The corresponding part data is the following: A is the annual forecast, c = cost-per-unit, h is the annual holding rate, and Co is the cost per order.

For each tier, the effective cost per order becomes: Co` = Co + SUC(i). So the data associated with each tier i is the following: A, c, h, q(i) and Co`. For each of the N tiers, this data is used to compute an economic order quantity, Q`, for the tier. In the event, Q`, is not in the admissible range of the tier, the closest quantity, Q, in the range is selected. This adjustment is needed to ensure the quantity is appropriate to gain the setup cost that the tier offers. Now using the admissible Q, the annual cost for the tier is computed. The annual cost, denoted as K, includes the order plus hold cost as shown below:

$$K = [Co + SUC(i)] \times A/Q + ch \times Q/2$$

The annual cost for each of the N tiers are compared and the order quantity, Q, that is associated with the minimum of the N values of K is selected as the economic order quantity, Q*.

Example 3.15 Consider an item where the annual forecast is A = 60, the cost per unit is $20, the annual holding rate is h = 0.24 and the cost per order is $10. Suppose also, the supplier contract has N = 2 tiers with minimum quantities and set-up-costs as listed below:

i	q(i)	SUC(i)
1	1	40
2	50	0

Note, the effective cost per order is Co` = 10 + 40 = 50 at tier i = 1, and is Co` = 10 at tier i = 2. At each higher tier the min quantity increases, and the setup cost decreases.

For each of the two tiers, the computations yield, Q`, Q and K as listed below:

i	Q`	Q	K
1	35.3	35	169.7
2	15.8	50	132.0

Because tier i = 2 has the minimum annual cost, K = 132, and the associated admissible quantity is Q = 50, the economic order quantity for this part becomes Q* = 50.

3.9.3 Min and Max Constraints

Many supplier contracts specify constraints on the size of the order quantity bestowed on them, whereby adjustments may need to be made to accommodate. Two of the common constraints are the minimum quantity, Q_{min}, and the maximum quantity, $Q_{max.}$

The minimum quantity states the smallest order size that is acceptable. In the event, the economic quantity is below the min quantity, the final quantity is adjusted upwards to meet the min quantity. In the description below, a Q-min algorithm is listed where Q` represents the economic quantity prior to a min quantity adjustment, and Q is the adjusted quantity to accommodate Q_{min}.

$$\begin{aligned}&\text{Q-min algorithm :}\\&\text{If } Q^{`} = 0, Q = 0\\&\text{else}, Q = \max\left[Q^{`}, Q_{min}\right]\end{aligned}$$

Example 3.17 Suppose a part where the economic quantity is Q` = 3 and the min quantity is $Q_{min} = 5$, whereby the adjusted quantity becomes: Q = max[3,5] = 5.

Sometimes, the vendor specifies a maximum quantity on a part, whereby the order size cannot exceed this max quantity. In the description below, a Q-max algorithm is listed where Q` represents the economic quantity prior to a max quantity adjustment, and Q is the adjusted quantity to accommodate Q_{max}.

$$\begin{aligned}&\text{Q-max algorithm :}\\&\quad Q = \min\left[Q^{`},\ Q_{max}\right]\end{aligned}$$

Example 3.18 Consider a component where the economic quantity is Q` = 17 and the max quantity is $Q_{max} = 10$. Since, Q` exceeds Q_{max}, the final quantity is: Q = min[17, 10] = 10

In some instances, the min and max constraints may cause a hardship to the stock holding location, whereby the buyer assigned of the item needs to contact the supplier to request a violation to the quantity constraint, usually at a surcharge cost.

3.9.4 Multiple Constraint

Perhaps the most common constraint on the order size is the multiple quantity for an item. Often, this constraint is called to be compatible with the number of units that are packaged together in a shipment. A Q-Mult Algorithm is provided below that converts the initial quantity, Q`, with multiple, M, to a final quantity, Q.

$$\begin{aligned}&\text{Q-Mult Algorithm :}\\&Q = M \times \text{integer}\left[\left(Q^{`} + M/2\right)/M\right]\\&Q = \max[Q, M]\\&\text{If } Q^{`} = 0, Q = 0\end{aligned}$$

Example 3.19 Suppose a situation where the multiple constraint is M pieces. Four examples are presented.

If $Q^{`} = 11$ and $M = 10,\ Q = 10 \times \text{integer}[(11 + 10/2)/10] = 10$
If $Q^{`} = 17$ and $M = 10,\ Q = 10 \times \text{integer}\left[(17 + 10/2)/10\right\} = 20$
If $Q^{`} = 3$ and $M = 10, Q = 10 \times \text{integer}[(3 + 10/2)/10] = 0 \,\&\, Q = \max[0, 10] = 10$
If $Q^{`} = 0$ and $M = 10,\ Q = 0.$

3.9.5 Bin-Max Constraint

Some stock holding locations with minimal space will place limits on how many units they can have in the bin space for specified items. Could be for items that need storage in special bins like barrels, refrigerated bins, heat-controlled bin, dark-room storage, or just plain limited space. These types of constraints are common in automotive dealerships and in pharmacies. For particular components, they may stipulate a limit on the number of units that they can stock, called a bin-max constraint and here denoted as Bm. A bin-max algorithm to accommodate this constraint is listed below, and the data required to run the algorithm is the following:

$$\begin{aligned}&Q^{`} = \text{raw order quantity}\\&Bm = \text{bin-max constraint}\\&OH = \text{on-hand inventory}\\&OO = \text{on-order}\end{aligned}$$

The final order quantity that satisfies the constraint, denoted as Q, is obtained as shown below by applying the bin-max algorithm. Note, with the algorithm, $OH + OO \leq Bm$.

Bin-Max algorithm:

$$Q = \min\left[Q^{`}, Bm - (OH + OO)\right]$$

Example 3.20 Suppose a truck dealer places a bin-max limit of $Bm = 10$ for a particular component. The raw order quantity is $Q = 9$ and the on-hand plus on-order is $\{OH + OO\} = 3$. The final order quantity becomes the following:

$$Q = \min[9, 10 - 3] = 7$$

3.9.6 Job Quantity Constraint

Another constraint that occurs in some stock holding locations pertains when a specified number of units are generally required to fill a customer order. Could be for tires on a car where four are needed for a customer, or the number of pills needed to fill a prescription at a pharmacy. Again, these types of constraints are common in automotive dealerships and in pharmacies. For particular components, the stock holding facility may tag certain items with this constraint by stipulating a quantity called a job-quantity, denoted as Qj. A job-quantity algorithm to accommodate this constraint is listed below. The algorithm is applied only when $Q` > 0$. The data required to run the algorithm is the following:

$$
\begin{aligned}
Q` &= \text{raw order quantity} \\
Qj &= \text{job-quantity constraint} \\
OH &= \text{on-hand inventory} \\
OO &= \text{on-order}
\end{aligned}
$$

The final order quantity that satisfies the constraint, denoted as Q, is obtained as shown below by applying the job-quantity algorithm.

Job-quantity algorithm:

$$
\begin{aligned}
&OHOO` = Qj \times \text{integer}\left[\left(Q` + OH + OO + Qj/2\right)/Qj\right] \\
&OHOO` = \max\left[OHOO`, Qj\right] \\
&Q = OHOO` - (OH + OO)
\end{aligned}
$$

Example 3.21 Assume a particular brand of tires at an automotive dealership when the job-quantity constraint is $Qj = 4$, and the current on-hand and on-order are $OH = 2$ and $OO = 0$, respectively. Also, the raw order quantity is $Q` = 7$ pieces. The job-quantity algorithm for this situation is followed below; whereby six items are ordered to conform with the job-quantity constraint.

$$
\begin{aligned}
&OHOO` = 4 \times \text{integer}[(7 + 2 + 0 + 4/2)/4] = 8 \\
&Q = 8 - 2 = 6
\end{aligned}
$$

3.10 Least Unit Cost

In production facilities, build quantities are needed for each of the items to plan the future production schedule. A common way to find the quantities is by the least unit cost method. The time periods could be in weekly or monthly buckets depending on the production scheduling system in use. The example of this section assumes monthly buckets. The goal is to find the order quantity per item that yields the least cost per unit. The data needed for each item are the following:

c = cost per unit
Co = cost per order
h = holding rate per time period
OH = on-hand inventory
OO = on-order inventory
SS = safety stock
$R\tau$ = requirements per future time period τ

The requirement for time period 1 is adjusted to account for the safety stock, on-hand and on-order quantities, as shown below:

$$r_1 = (R_1 + SS) - (OH + OO)$$

For the remaining months, the adjusted requirements are here denoted as $r\tau$ for future time period τ, and are obtained as follows:

$$r\tau = R\tau \quad \tau = 2, 3, \ldots.$$

In the event that the adjusted requirement for time period $\tau = 1$, r_1, is less or equal to zero, then no order quantity is needed for the item, and thereby, $Q^* = 0$.

For each future month, the order quantity Q is the sum of the cumulative requirements up to that month, and the cost, K, to order plus hold for that quantity is computed as shown below:

$$\begin{array}{lll} \tau = 1 & Q = [r_1] & K = Co + ch[0.5r_1] \\ \tau = 2 & Q = [r_1 + r_2] & K = Co + ch[0.5r_1 + 1.5r_2] \\ \tau = 3 & Q = [r_1 + r_2 + r_3] & K = Co + ch[0.5r_1 + 1.5r_2 + 2.5r_3] \end{array}$$

etc.

For each future month, the cost per unit is calculated as below:

$$K/Q = (\text{order} + \text{hold}) \text{ cost per unit}$$

Now for the planning horizon months, the quantity, Q, with the minimum cost per unit is selected as the quantity to buy since it has the least unit cost.

Example 3.22 Assume a component where the planning horizon is 6 months and the following data for the component is as follows:

c = \$100.00 = cost per unit
Co = \$50 = cost per order
h = 0.02 = monthly holding rate
$f\tau$ = 40, 30, 60, 20, 0, 40 = requirements for future months 1–6
OH = 20 = on-hand inventory

OO = 40 = on-order inventory
SS = 30 = safety stock

The requirements for future month $\tau = 1$ is computed as below:

$$r_1 = 40 + 30 - (20 + 40) = 10$$

Table 3.2 lists the computations for each of the future 6 months. Since the least unit cost is K/Q = \$3.75 and this occurs at $\tau = 2$, the economic quantity becomes Q* = 40.

3.11 Multi-Part Buy Quantity

Some suppliers, who are the source of N different parts, offer discounts to entice the customers to buy larger than normal quantities of the N parts to meet specified target levels. The target levels could be in dollars, pounds, weight, barrels, or whatever is appropriate for the type of units. The incentives are typically a combination of free freight and discounts on the total price across all of the parts.

When the target levels are stated in costs, the customer buy quantities on each of the N parts are labeled as: $[q_1, \ldots, q_N]$, and the corresponding cost per unit are as follows: $[c_1, \ldots, c_N]$. The total cost of the combined buy of the N parts is denoted as \$Q and is tallied as shown below:

$$\$Q = [c_1 q_1 + \ldots + c_N q_N]$$

The supplier incentives are demonstrated with the example that follows.

Example 3.23 Table 3.3 shows how the incentives might look for a situation where the targets are in dollars. If the total buy is \$1000 or more, the customer receives

Table 3.2 Least unit cost computations for Example 3.22

τ	R	r	Q	K	K/Q
1	40	10	10	60	6.00
2	30	30	40	150	3.75
3	60	60	100	450	4.50
4	20	20	120	590	4.91
5	0	0	120	590	4.91
6	40	40	160	1030	6.43

Table 3.3 Example of multi-part buy quantity terms

\$Q	incentive
\$1000	Free freight
\$2000	Free freight + 10 % discount
\$5000	Free freight + 20 % discount

free freight. If $2000 or more, the customer also gets a ten percent discount on the total buy, and if $5000 or more, a twenty percent discount is given.

When the customer is arranging a volume buy on the supplier, it behooves him/her to seek the advantage in higher buy quantities, and if so, how to nudge the raw buy quantities ($q_1, \ldots, q_N$) up to allow $Q to reach next target quantity.

Note four examples of the total cost below.

If Q$ = $300:	customer pays freight
If Q$ = $1200:	free freight & no discount
If Q$ = $2300:	free freight & 10 % discount
If Q$ = $5100:	free freight & 20 % discount

3.12 Initial Buy Quantity (IBQ)

A common problem in service parts distribution centers (spdc) occurs when a buy quantity for a part is in need and there is no prior demand history on the part to base the forecast. This situation originates in assembly when a new model or an engineering change occurs on a finished-good-item (FGI) and a new part or component is included in the bill-of-material of the FGI. Since the part is new, no prior demand history has occurred, yet the spdc is obliged to carry the part in the event there is a need for maintenance or repair at one of the dealerships, for which the spdc is the source. Often, experts give estimates on how much stock to carry for the new part based on their prior experience. This first time buy is called the initial-buy-quantity (IBQ).

To confront this situation, the notion of new-parts-of-the-past (npop) and new-part-of-the-future (npof) is introduced. The npof includes a part where a decision on the quantity to stock is needed now. As much as possible, this part should be identified with an attribute code, denoted as A, that identifies where the part is used on the FGI and its function. The attribute serves as a link between the npof and npop. The npop are the parts with the same attribute A and where the first year of demand history is available on the database.

For each of the eligible npop, the first 12 months of demand history are labeled as ($D_1,\ldots,D_{12}$). The average 1-month demand is denoted as d, and is computed by $d = \Sigma D_t/12$ ($t = 1$–12). Assuming n parts are in the npop set, the average month demands are labeled as $d_1,\ldots,d_n$. The average and standard deviation from these n parts are denoted as μ_d and σ_d; respectively, and the shape of the demands, d, typically follows a lognormal distribution. In the subsequent computations, the lognormal distribution is converted to the normal distribution.

Of interest now is to find ten demand points of d where each serves as the center of a range that contains ten percent of the values of d. The selected values are identified as d(i) (i = 1–10), where $P(d < d(i)) = \alpha(i)$ for i = 1–10, and $\alpha(i) = (0.05, 0.15, 0.25, 0.35, 0.45, 0.55, 0.65, 0.75, 0.85, 0.95)$.

To find the ten demand points of d(i), apply the following four steps:

1. Convert the demands d to y = ln(d) where ln is the natural log. The parameters of d, μ_d, σ_d, of the lognormal become μ_y and σ_y, respectively, of the normal distribution. The transformation is below:

$$\mu_y = \ln\left({\mu_d}^2/\sqrt{\sigma_d^2 + \mu_d^2}\right)$$

$$\sigma_y{}^2 = \ln\left(\left[{\sigma_d}^2 + {\mu_d}^2\right]/{\mu_d}^2\right)$$

2. Find z(i) from the standard normal distribution where $P(z < z(i)) = \alpha(i)$.
3. Find $y(i) = \mu_y + z(i)\sigma_y$.
4. Now $d(i) = \exp(y(i))$.

The results yield d(i) where $P(d < d(i)) = \alpha(i)$. Note, for attribute A, the ten demand points represent ten equally likey values of the average 1-month demands.

For the npof, let L represent the lead-time and c the cost-per-unit. In addition to the attribute, A, this is the only data available to determine the IBQ.

The lead-time L is used with each demand point to select ten candidate buy quantities, $q_1, \ldots, q_{10}$, where $q_i = d_{Li} = L \times d_i$. Note, each of the candidates is one of the ten lead-time demands. Ten candidate values of the IBQ are now obtained. These are denoted as q(i) where $q(i) = d_L(i)$ for i = 1–10. In this way, the IBQ for the part will be selected from one of these candidates.

Consider the data that is gathered and calculated for a new part. The new part has an attribute A and also a lead time L. The attribute is linked to the npop that also have the same attribute A. The data now available for the new part is the following:

A = attribute
L = lead time
c = cost per unit
d(i) = ten demand points i = 1–10
q(i) = ten candidate IBQ points i = 1–10.

3.12.1 Lead-Time Cost

To determine an economic quantity for the IBQ, a cost model is constructed. The ten possible IBQ values computed above are identified and the final choice will come from one of these. The model estimates the lead-time cost of the new part with each of the candidate quantities. The quantity associated with the minimum cost is selected as the economic IBQ to use.

Let K(q,d) = lead-time cost when the quantity is q and the average 1-month demand is d.

The cost model uses the following three parameters:

h = one-month holding rate.
Co = cost per order.
gw = goodwill rate

The cost K(q,d) is based on the difference between q and the lead-time demand, denoted as $d_L = (d \times L)$. The computations are below:

$$\text{If } q \geq d_L: \quad K(q,d) = (q - d_L)c \times h \times L$$
$$\text{if } q < d_L \quad K(q,d) = (d_L - q)c \times gw + Co$$

The relation below shows how to estimate the expected cost when the IBQ = q, regardless of the average 1-month demand:

$$K(q) = \Sigma_i K(q, d(i)) \times 0.10$$

Note in the above, each value of d(i) is selected with probability of 0.10. Hence, the sum for i = 1–10 yields the average when the candidate q is in consideration. With the ten candidate quantity points, it is now possible to determine the quantity that is associated with the minimum lead-time cost. This is q` where K(q`) = min{K[q(j)] j = 1–10}. And so, the economic quantity to use for the new part is IBQ = q`.

Example 3.24 Suppose a npof with an attribute A, lead-time L = 2 months and cost per unit c = $100. Suppose, five npop are found with the same attribute A. The average (first year) monthly demands for these five parts are: $(d_1,\ldots,d_5) = (1.7, 0.5, 0.8, 1.9, 10.5)$. The average and standard deviation from the demands are computed and become: $\mu_d = 3.08$ and $\sigma_d = 4.19$, respectively. Since the demands (d) are assumed to follow a lognormal distribution, the relations given earlier to find the statistics for the counterpart normal distribution (y) are now used; and yield $\mu_y = 0.601$ and $\sigma_y = 1.023$.

So now the ten demand points of d are found as described here. To accomplish, we use the ten probabilities α(i), i = 1–10 and the corresponding ten standard normal measures z(i), i = 1–10. The ten measures of y are obtained by: $y(i) = \mu_y + z(i)\,\sigma_y$ for i = 1–10. Finally, the ten demand points from the lognormal distribution are obtained by d(i) = exp(y(i)) for i = 1–10. See the list below.

i	α(i)	z(i)	y(i)	d(i)
1	0.05	1.645	−1.08	0.34
2	0.15	1.036	−0.46	0.63
3	0.25	0.674	−0.09	0.92
4	0.35	0.385	0.21	1.23
5	0.45	0.124	0.47	1.61
6	0.55	0.124	0.73	2.07
7	0.65	0.385	1.00	2.71
8	0.75	0.674	1.29	3.64
9	0.85	1.036	1.66	5.27
10	0.95	1.645	2.28	9.82

For simplicity in Table 3.4, the ten demand points and ten quantity points are listed in closest integers. So the integer demand points (d) are: 0, 1, 1, 1, 2, 2, 3, 4, 5, 10. The ten quantity points (q) become: 1, 1, 2, 2, 3, 4, 5, 7, 11, 20. Subsequently, in the computations, the fractional values are applied.

The parameters used in the example are: gw = 0.3, h = 0.02, and Co = 20. Recall L = 2 months and c = $100.

To illustrate,

if d = 1.61 and q = 1.84, $d_L = 3.22$:

$$K(1.61, 1.84) = (3.22 - 1.84)100 \times 0.3 + 20 = 61.5.$$

If d = 1.61 and q = 7.28, $d_L = 3.22$:

$$K(1.61, 7.28) = (7.28 - 3.22)(100 \times 0.02 \times 2) = 16.24$$

and so forth.

The ten (average monthly) demand points (d) are listed in the left column (in integers), and the ten candidate initial buy quantities (q) are given in the first row (in integers). Recall, d and q are both shaped like a lognormal distribution. For brevity, the table lists only the integer portion of d and q; but in the computation of the cost model, the corresponding fractional values are used throughout. The body of the table gives the values of K(q,d) = lead-time cost when the initial buy quantity is q and the average monthly demand is d. The bottom row gives the values of K(q), the expected lead-time cost when the initial buy quantity is q. Note. the minimum of K(q) is 52.5 and this occurs when q = 10.54. Thereby, IBQ = 11.

Table 3.4 K(q,d) for 10 average monthly demand points (d) and 10 candidate initial buy quantities (q) on a part with cost per unit = $100 and lead-time is L = 2 months

d/q	1	1	2	2	3	4	5	7	11	20
0	0.0	2.3	4.6	7.1	10.1	13.9	18.9	26.4	39.4	75.9
1	37.6	0.0	2.3	4.8	7.8	11.5	16.6	24.0	37.1	73.5
1	54.6	37.0	0.0	2.5	5.5	9.2	14.3	21.8	34.8	71.3
1	73.5	55.9	38.9	0.0	3.0	6.7	11.8	19.2	32.3	68.7
2	96.1	78.5	61.5	42.6	0.0	3.7	8.8	16.2	29.3	65.7
2	123.9	106.4	89.4	70.5	47.9	0.0	5.1	12.5	25.6	62.0
3	162.0	144.4	127.4	108.5	85.9	58.1	0.0	7.4	20.5	56.9
4	217.9	200.3	183.3	164.4	141.8	113.9	75.9	0.0	13.0	49.5
5	315.7	298.1	281.1	262.2	239.6	211.8	173.7	117.8	0.0	36.4
10	589.1	571.5	554.5	535.6	513.0	485.1	447.1	391.2	293.4	0.0
K(q)	167.0	149.4	134.3	119.8	105.5	91.4	77.2	63.7	52.5	56.0

3.13 Final Buy Quantity (FBQ)

A finished good item (FGI) coming off the assembly line is composed of a large number of parts (and components). The combination of parts on a FGI are often evolving over time due to engineering changes and model improvements. When a part is discarded and is no longer in the bill-of-material (bom) of the FGI currently coming off the assembly line, the assembly plant stops replenishing the part from the supplier. The service parts distribution center (spdc), however, is obliged to carry the part for any potential needs of maintenance and repair on a FGI that has the particular part in its bom. The obligation period for a part is often 5 or more years after the part is no longer included in the bom of the FGI. Because the plant demand has stopped and little demand comes from the spdc, the supplier may cease producing the part and notifies the spdc to order one final-buy-quantity (FBQ) of the part. Upon such notice, the spdc is faced with the task of forecasting the demands over the coming years and determining how many units to have in stock for this obligation time need.

3.13.1 Final Buy Forecast (FBF)

To generate the forecasts for the obligation years, some data is needed from the history of the part, and also from a hypothetical FGI that has the part in its bom. This includes the following:

t_o = number of years the part was in the bom of the FGI
T = number of years since the part was introduced on the bom of the FGI
j_o = number of years after assembly that the FGI is scrap free
r = annual probability an FGI is scrapped, after j_o years
t_1 = first year spdc history demands are known
d(t) ($t = t_1$ to T) = annual spdc demands known for the part
N = number of obligation years remaining after year T for the part at the spdc

Hence, t = (1 to t_o) are the years the part was in production in the assembly plant, where t = 1 is the introduction years, t = (1 to T) identifies the history years, and T is the current year. Note, the life of a FGI starts when the unit leaves the assembly line and enters the population of active units, and ends when the FGI is scrapped. The parameter states that the FGI will not be scrapped for the first j_o years, and thereafter the annual probability of a scrap is r. Note, (1−r) is the probability an active FGI in the population will not be scrapped in the following year. The spdc annual demands for the part are known for years t_1 to T.

So for an individual FGI, the probability it will fail in the j-th year after it begins its active life is as follows:

$$p(j) = \begin{cases} 0 & j = 1 \text{ to } j_o \\ r(1\text{-}r)^{(j\text{-}jo\text{-}1)} & j > j_o. \end{cases}$$

In this situation, j is a discrete index with $j \geq 1$ and where j_o serves as a location parameter. The probability is the same as a *geometric distribution* for the years $j > j_o$.

The cumulative probability a FGI unit is scrapped in year j or sooner is:

$$F(j) = \Sigma_{i=1,j} p(i)$$

The complementary probability, H(j), is obtained from $H(j) = 1 - F(j)$, representing the probability the FGI unit survives j years or longer.

3.13.2 Part Life Cycle

A generic representation of the part's annual life cycle can now be formulated. Let x(t) denote the shape of the life cycle when the parameters (t_o, r, j_o) are provided. For the years the part is in production, $t = 1$ to t_o,

$$x(t) = [H(1) + \ldots + H(t)] \quad t = 1 \text{ to } t_o$$

and for the years after the part is no longer in production, $t > t_o$,

$$x(t) = [H(t\text{-}t_o + 1) + \ldots + H(t)] \quad t = t_o + 1, t_o + 2, \ldots.$$

Since the known demands are in years $t = t_1$ to T, the sum of the known demands and the associated life cycle sum are as below:

$$\sum_{t=t_1}^{T} d(t) = \text{sum spdc part demands known for years } t_1 \text{ to T}$$
$$\sum_{t=t_1}^{T} x(t) = \text{ generic life cycle sum for years } t_1 \text{ to T}$$

The ratio of the above two sums yields:

$$a = \Sigma d(t) / \Sigma x(t).$$

This ratio is used to generate the forecast of demands for the N future obligation years beyond $t = T$. The forecasts become:

$$f(t) = a \times x(t) \quad \text{for } t = T + 1, \ldots, T + N$$

The final buy forecast, FBF, for the part over the future obligation years becomes:

$$\text{FBF} = \sum_{t=T+1}^{T+N} f(t)$$

Example 3.25 Assume a part where r = 0.20 (FGI annual scrap rate), $j_o = 3$ (FGI scrap free years), $t_o = 3$ (part production years), T = 10 (years since introduce part), and N = 8 (obligation years). The table below lists the probabilities p(j), F(j) and H(j) for the years j = 1–18, and also x(t), the generic life cycle curve. For simplicity, the table entries are rounded to two decimal places. Note the following:

$$\begin{aligned} x(1) &= H(1) \\ x(2) &= H(2) + H(1) \\ x(3) &= H(3) + H(2) + H(1) \\ x(4) &= H(4) + H(3) + H(2) \end{aligned}$$

and so forth.

j	p(j)	F(j)	H(j)	x(t)
1	0.00	0.00	1.00	1.00
2	0.00	0.00	1.00	2.00
3	0.00	0.00	1.00	3.00
4	0.20	0.20	0.80	2.80
5	0.16	0.36	0.64	2.44
6	0.13	0.49	0.51	1.95
7	0.10	0.59	0.41	1.56
8	0.08	0.67	0.33	1.25
9	0.07	0.74	0.26	1.00
10	0.05	0.79	0.21	0.80
11	0.04	0.83	0.17	0.64
12	0.03	0.86	0.14	0.52
13	0.03	0.89	0.11	0.42
14	0.02	0.91	0.09	0.34
15	0.02	0.93	0.07	0.27
16	0.01	0.94	0.06	0.22
17	0.01	0.95	0.05	0.18
18	0.01	0.96	0.04	0.15

The generic life cycle curve, x(t), for the part is listed in Table 3.5 below. Note the current year is T = 10 and the demand history of the part is known for the 3 years, t = 7, 8, 9. 10. The table notation of the years are the following: p = part in production years, d = part demand history years, and f = part demand forecast years. The sum of the part 4-year demands is 148, and the corresponding 4-year life cycle fit is 4.61. Thereby, the ratio of demands over life cycle is a = 148/ 4.61 = 32.10. This yields the forecast, f(t), for the part over the obligation years t = 11–18 from the relation f(t) = 32.10x(t). The final buy forecast on the part becomes:

$$\text{FBF} = \sum\nolimits_{t=11}^{18} f(\tau) = 85.38.$$

Table 3.5 Final forecast with t = years, p = production years, d = demand years, f = forecast years, x(t) = generic life cycle, d(t) = demands, f(t) = forecasts

t		x(t)	d(t)	f(t)
1	p	1.00		
2	p	2.00		
3	p	3.00		
4		2.80		
5		2.44		
6		1.95		
7	d	1.56	50	
8	d	1.25	42	
9	d	1.00	29	
10	d	0.80	27	
11	f	0.64		20.54
12	f	0.52		16.37
13	f	0.42		13.16
14	f	0.34		10.59
15	f	0.27		8.67
16	f	0.22		6.74
17	f	0.18		5.14
18	f	0.15		4.17

3.13.3 Final Buy Quantity

Two methods are given to find the final buy quantity, FBQ. A first method is by the service level method, and a second is by an economic method.

To begin, the standard deviation of the FBF is needed. To accomplish, an estimate of the coefficient-of-variation, cov, for each future year forecast is provided. With this estimate, the standard deviation for the forecast of future year t becomes:

$$\sigma(t) = \text{cov} \times f(t)$$

Example 3.26 Assuming cov = 0.40, the estimate of the standard deviations for each of the eight obligation year forecasts are listed in Table 3.6. Note at t = 12, for example, $\sigma(12) = 0.4 \times 16.4 = 6.5$. Assuming the forecasts are independent allows estimating the standard deviation of FBF as follows:

$$\sigma = \sqrt{\sum_{t=11}^{18} \sigma(t)^2} = 13.5$$

Recall, the sum of the forecasts is FBF = 85.38.

So now, FBF = 85.38 and σ = 13.5 are the estimate of parameters for the demands, x, over the obligation years. Table 3.7 gives ten equally likely values of x that range from 63.2 to 107.6. The entries are obtained with use of the standard normal distribution with cumulative probabilities of: 0.05, 0.15,. . ., 0.95. For each

Table 3.6 Future years, t; forecast, f(t); and standard deviation, σ(t)

t	f(t)	σ(t)
11	20.5	8.2
12	16.4	6.5
13	13.2	5.3
14	10.6	4.2
15	8.7	3.5
16	6.7	2.7
17	5.1	2.1
18	4.2	1.7

Table 3.7 Ten indices, i; probability, α(i); standard normal variate, z(i); equally likely FBF, x(i); and FBQ candidate, q(i)

i	α(i)	z(i)	x(i)	q(i)
1	0.05	−1.645	63.2	63.2
2	0.15	−1.036	71.4	71.4
3	0.25	−0.674	76.3	76.3
4	0.35	−0.385	80.2	80.2
5	0.45	−0.124	83.7	83.7
6	0.55	0.124	87.1	87.1
7	0.65	0.385	90.6	90.6
8	0.75	0.674	94.5	94.5
9	0.85	1.036	99.4	99.4
10	0.95	1.645	107.6	107.6

probability, the associated z entry is taken from Table 4.2. Note, for example, $P[z < -1.036] = 0.15$, and $x(2) = 85.38 - 1.036 \times 13.5 = 71.39$.

3.13.3.1 Service Level Final Buy Quantity

The service level final buy quantity is obtained from Table 3.7. In the event the service level is set to $SL = 0.95$, the final buy quantity is the q(i) where $\alpha(i) = 0.95$, and thereby, the rounded final buy quantity becomes:

$$FBQ = 108$$

This quantity is an estimate on how much stock is needed to ensure with 95 % confidence that the demands over the eight obligation years will be immediately filled.

3.13.3.2 Economic Final Buy Quantity

The economic final buy quantity seeks the quantity to buy now that yields the minimum cost. Using the notation, x = demand over the obligation years, and q = final buy quantity, the three major events that could occur are the following:

$$\begin{aligned} &\text{if } x > q, \text{ the quantity is short} \\ &\text{if } x < q, \text{ the quantity is in excess} \\ &\text{if } x = q, \text{ the quantity is exact} \end{aligned}$$

A cost model for this purpose is given here and the data needed to carryout is listed below:

$$\begin{aligned} &c = \text{cost-per-unit} \\ &Co = \text{order cost if short} \\ &gw = \text{goodwill rate if short} \\ &mu = \text{markup rate in cost per unit if short} \\ &sr = \text{scrap rate if excess} \end{aligned}$$

The current cost-per-unit is c.

When a short condition, the following three costs occur:

$$\begin{aligned} &\text{order cost:} && Co \\ &\text{goodwill cost per unit short:} && c \times gw \\ &\text{cost-per-unit short:} && c` = c(1 + mu) \end{aligned}$$

When an excess condition, the following costs occurs:

$$\text{scrap cost per excess unit} : c(1 - sr)$$

Assuming the above, the cost for each combination of x and q is listed below and denoted as K(x,q):

$$\begin{aligned} &\text{If } x > q: \quad K(x, q) = (x - q)[c \times mu + c \times gw] + Co \\ &\text{If } x < q: \quad K(x, q) = (q - x)[c \times (1 - sr)] \end{aligned}$$

Example 3.27 Assume the following: c = $50, mu = 1.00, gw = 0.50, Co = $50, and sr = 0.20, are used to compute the buy quantity using the economic method. Table 3.8 computes K(x,q) for each combination of x and q. The values on x are in column 1 and the corresponding candidates of q are in row 1. At x = 71.4 and q = 90.6,

$$K(71.4, 90.6) = [90.6 - 71.4][50(1 - 0.2)] = 768.$$

Also, at x = 90.6 and q = 71.4,

$$K(90.6, 71.4) = [90.6 - 71.4][50 \times 1.00 + 50 \times 0.50] + 50 = 1489.$$

Table 3.8 Body lists 100 combinations of Cost, K(x,q), when ten values of x = FBF in first column, and ten of q = FBQ in first row are in effect

x/q	63	71	76	80	84	87	91	94	99	108
63	0	329	524	681	822	955	1096	1253	1448	1777
71	667	0	196	352	493	627	768	924	1119	1448
76	1033	417	0	156	297	431	572	728	924	1253
80	1326	709	343	0	141	275	416	572	768	1096
84	1590	974	607	314	0	134	275	431	627	955
87	1842	1225	858	565	301	0	141	297	493	822
91	2106	1489	1122	830	565	314	0	156	352	681
94	2399	1782	1415	1122	858	607	343	0	196	524
99	2765	2148	1782	1489	1225	974	709	417	0	329
108	3382	2765	2399	2106	1842	1590	1326	1033	667	0
K(q)	1711	1184	925	762	654	591	565	581	659	888

The bottom row is the estimated cost when FBQ = q

Recall, each of the ten FBF candidates, x(i), are equally likely and with probability of 0.10. The bottom row gives the estimated expected cost, K[q(i)], when q is used regardless of x. This is computed by,

$$\mathrm{K[q(j)]} = \sum_{i=1}^{10} K[x(i)q(j)](0.10)$$

Note, the minimum cost, 565, occurs when q(j) = 91 and thereby the economic final buy quantity becomes:

$$\mathrm{FBQ} = 91$$

3.14 Summary

In most situations, the ideal order quantity will yield the minimum cost of hold plus order to the stocking facility. This quantity is called the economic order quantity. Another expenditure, called the stocking cost is a hidden expense due to the process of holding plus ordering the item at the stocking facility. The four ingredients that determine the economic order quantity are the following: the annual forecast, A; the cost per order, Co; the cost per unit, c; and the annual holding rate, h. Sensitivity relations are developed in the chapter between these ingredients and the economic order quantity. In some situations, constraints are imposed on the order quantity to use. The typical constraints are the following: price breaks, set-up-cost, min and max limits, multiple quantity, bin max and job quantity. Special care is needed to adjust the order quantity to conform to the constraints. In service part facilities, two other order quantities are of frequent need. These are the initial buy quantity and the final buy quantity.

Chapter 4
Safety Stock

4.1 Introduction

Safety stock is needed to attain a desired level of service when the demands exceed the forecasts over the lead-time duration. Four popular methods to generate the safety stock are in use: months supply, service level, percent fill and Lagrange. For each method, a safety stock parameter is assigned by the management to specify how to compute the safety stock. The months supply method is based on a specified number of future forecasts. The service level method is set so the probability of not having an out-of-stock condition during the lead-time is satisfied. The percent fill method seeks to accommodate the ratio of (demand fill)/(total demand) over the lead-time. The Lagrange method is a cost-benefit way to set the safety stock. The latter three methods are based on the distribution of forecast for the lead-time demand. The distribution for the lead-time demand can be normally distributed or not normally distributed. When normally distributed, the standard normal distribution is applied to set the safety stock. When not normally distributed, the truncated normal distribution is used to set the safety stock. The coefficient of variation for the lead-time demand is used to identify which of the two distributions to use.

The Lagrange method works well when the safety stock for a group of items is controlled simultaneously. The goal is to have a desired service level for the entire group at the total minimum cost. The safety stock depends on the lead-time provided by the supplier; but for suppliers who are late in the delivery of the replenish stock, adjustments to the safety stock is needed. Each supplier is measured as to its delivery score, and the safety stock adjustment for the supplier is based on his score. Sometimes loss sales occur when stock is not available to complete a customer demand. A way to measure and control the amount of loss sales is developed.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_4

4.2 Safety Stock Need

Safety stock is the amount of extra stock held in the inventory as precaution for the event the future demand may exceed the forecast. This situation is prevalent in retail and distribution centers where the demands for immediate need occur in a random manner without any prior notice. Safety stock is needed to provide a desired level of service to the customers, often referred as the service level. When demands occur and the stock is not available to send, the demand is either placed in a backorder status or is a lost sale. Neither case is desirable.

4.3 Safety Stock Methods

Various ways to determine how much safety stock to have are in use in industry. The most common methods are presented here. These are the following: months supply, service level, percent fill, and Lagrange. Sometimes the term service level is used to describe the percent fill method, and service level is sometimes used to define the service level method. The goal is to provide the minimum amount of safety stock to achieve a desired level of service to the customer's satisfaction. Each of the methods are controlled by the management who sets a parameter that specifies how to determine the safety stock for each item in the inventory.

4.4 Safety Stock Parameters

The simplest method is the months supply way of providing safety stock. The parameter is MS that states the months of future forecast to provide as the safety stock. The parameter could be a fraction or an integer and often is less than one.

The parameter for the service level method is labeled as SL, and is management's desired probability of not running out of stock. The method seeks the amount of safety stock needed to yield the probability of not running out-of-stock over the order cycle duration.

The parameter for the percent fill method is denoted as PF, and designates management's desire to control the ratio of (demand filled) over (total demand). Statistical methods are used to determine how much safety stock is needed accordingly.

The Lagrange method is used to control the inventory over a cluster of parts in the inventory all together. The parameter is LG and this is used with each part in the cluster to set the safety stock in a cost-benefit manner. The goal is to allocate the investment in safety stock for the cluster that gives the highest average service level for the entire cluster.

In summary, the parameters are the following:

Months supply:	MS = portion of future forecasts
Service level:	SL = probability inventory is available
Percent Fill:	PF = ratio of (demand filled/total demand)
Lagrange:	LG = cost benefit factor

4.5 Months Supply Method

The easiest way to set the safety stock is by the months supply method, whereby the safety stock is set to a portion of the future months forecasts. To compute, all that is needed is the forecast of the demands for the future months, and a parameter, denoted as MS, that is set by the management and specifies the number of future forecast months to use as the safety stock. The safety stock is obtained as described below.

When the monthly forecasts are relatively flat and f = average monthly forecast, the safety stock, SS, is computed as below:

$$SS = MS \times f$$

When the forecasts vary each month, and are denoted as: f_τ = forecast for future month τ ($\tau = 1, 2, \ldots$), the safety stock is computed as shown below:

$$\begin{array}{ll} \text{If } MS \leq 1: & SS = MS \times f_1 \\ \text{If } 1 < MS \leq 2: & SS = f_1 + (MS - 1)f_2 \\ \text{If } 2 < MS \leq 3: & SS = f_1 + f_2 + (MS - 2)f_3 \end{array}$$

and so forth.

This method applies when the forecast system does not keep track on the accuracy of the forecasts by measuring the month-to-month forecast errors, and thereby, no standard deviation of the forecast error is measured.

The concept is simple, for horizontal forecasts, the month to month measures of the safety stock is flat. For trending forecasts, up or down, the monthly measures of the safety stock would also trend in the same direction as the forecasts. Should the forecast be seasonal, the monthly measures of the safety stock will flow in the same way as the seasonal pattern of the forecasts.

Example 4.1 Consider a part where the forecast of future monthly demands is flat at ten pieces per month. If the months supply parameter is set to MS = 1.25 months, the safety stock becomes $SS = 1.25 \times 10 = 12.5$. Should the parameter be set to MS = 0.50 months, the safety stock would be $SS = 0.50 \times 10 = 5.0$ pieces.

Example 4.2 Suppose the forecasts for the coming 4 months are: 10, 12, 14, and 16, and the months supply parameter is set at MS = 2.3. The safety stock becomes:

$$SS = 1 \times 10 + 1 \times 12 + 0.3 \times 14 = 26.2$$

4.6 Normal Distribution

The normal distribution is perhaps the most used distribution in materials management. The notation for a normal variable x with mean μ and standard deviation σ is $x \sim N(\mu, \sigma^2)$. The primary use in inventory is when the future demands are not known and a forecast is generated to estimate the demands over the coming months. The actual demands for a month is assumed to follow a normal distribution with a mean of μ and a standard deviation of σ. In this way, the subsequent inventory decisions may be made using the normal probability distribution as a basis. The range of a normal variable is mostly from L to U, where $L = (\mu - 3\sigma)$ and $U = (\mu + 3\sigma)$, as depicted in Fig. 4.1.

4.7 Standard Normal

A companion distribution to the normal is the standard normal distribution with the variable denoted as z whose mean is zero and standard deviation is one. The notation for a standard normal variable z is N(0,1). The variables x and z are related as below:

$$z = (x - \mu)/\sigma$$
$$x = \mu + z\sigma$$

whereby the range of z is mostly from −3 to +3 as shown in Fig. 4.2.

4.7.1 Cumulative Probability Distribution

Consider a particular value k of the standard normal variable z, and the probability of z less or equal to k. This is a cumulative probability of $P(z \leq k)$ and is denoted as F(k). Table 4.1 lists a few representative values of F(k) from −3 to +3.

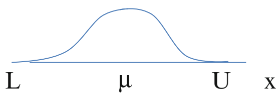


Fig. 4.1 The range of a normally distributed x is from L to U

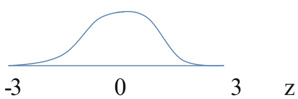


Fig. 4.2 The range of a standard normal variable z is from −3 to 3

Table 4.1 Some entries of the cumulative probability of z less or equal to k, F(k)

K	−3.0	−2.0	−1.0	0.0	1.0	2.0	3.0
F(k)	.001	.023	.159	.500	.841	.977	.999

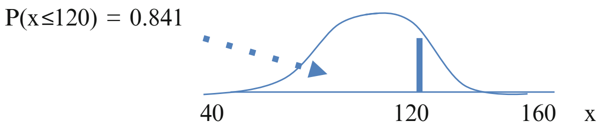


Fig. 4.3 Depiction of $P(x \leq 120)$ for normal variable x

Example 4.3 Assume a normal variable x with mean 100 and standard deviation 20. Of concern is to find the probability of x less or equal to 120, $P(x \leq 120)$. The corresponding standard normal value, k, of the normal x = 120, is k = (120−100)/20 = 1.00, and the cumulative probability of z less or equal to one is F(1) = 0.841. Hence the counterpart probability of x less or equal to 120 is the same, whereby, $F(1) = P(x \leq 120) = 0.841$. Figure 4.3 depicts this scenario.

4.7.2 *Standard Normal Table*

The common notation for a standard normal variable is z ~ N(0,1), indicating that the mean is zero and the variance is one, whereby the standard deviation is also set to one. The probability density of z is the following:

$$f(z) = \sqrt{1/(2\pi)}\, \exp -\left(z^2/2\right)$$

The density is symmetrical around zero and the range is mostly from −3.0 to 3.0.

Table 4.2 considers values of z = k ranging from −3.0 to 3.0. For each k, the associated probability density, f(k), the cumulative distribution, F(k), and the partial expectation, E(z > k) are listed. Note where f(k) is highest at k = 0.0, and is smallest at k = −3.0 and at k = 3.0.

The cumulative distribution represents the probability of z less or equal to k, whereby, $F(k) = P(z \leq k)$. Note at the edges: F(−3.0) = 0.001, and F(3.0) = 0.999, indicating almost all of the values of z lie between − 3.0 and 3.0. In the middle, at z = 0.0, F(0.0) = 0.500.

The partial expectation is needed to find the safety stock when the percent fill method is used. This is the average value of (z − k) for all values of z larger than k. This quantity is computed by the following:

$$\begin{aligned} E(z > k) &= \int_k^\infty (z - k) f(z) dz \\ &= f(k) - k[1 - F(k)] \end{aligned}$$

In the safety stock computations, k represents the safety factor, and the only values that are needed are when $k \geq 0$.

Table 4.2 The Standard Normal Distribution with: the probability density, f(k), the cumulative probability distribution, F(k), and the partial expectation, E(z > k)

k	f(k)	F(k)	E(z>k)	k	f(k)	F(k)	E(z>k)
−3.0	0.004	0.001	3.000	0.0	0.399	0.500	0.399
−2.9	0.006	0.002	2.901	0.1	0.397	0.540	0.351
−2.8	0.008	0.003	2.801	0.2	0.391	0.579	0.307
−2.7	0.010	0.003	2.701	0.3	0.381	0.618	0.267
−2.6	0.014	0.005	2.601	0.4	0.368	0.655	0.230
−2.5	0.018	0.006	2.502	0.5	0.352	0.691	0.198
−2.4	0.022	0.008	2.403	0.6	0.333	0.726	0.169
−2.3	0.028	0.011	2.304	0.7	0.312	0.758	0.143
−2.2	0.035	0.014	2.205	0.8	0.290	0.788	0.120
−2.1	0.044	0.018	2.106	0.9	0.266	0.816	0.100
−2.0	0.054	0.023	2.008	1.0	0.242	0.841	0.083
−1.9	0.066	0.029	1.911	1.1	0.218	0.864	0.069
−1.8	0.079	0.036	1.814	1.2	0.194	0.885	0.056
−1.7	0.094	0.045	1.718	1.3	0.171	0.903	0.046
−1.6	0.111	0.055	1.623	1.4	0.150	0.919	0.037
−1.5	0.130	0.067	1.529	1.5	0.130	0.933	0.029
−1.4	0.150	0.081	1.437	1.6	0.111	0.945	0.023
−1.3	0.171	0.097	1.346	1.7	0.094	0.955	0.018
−1.2	0.194	0.115	1.256	1.8	0.079	0.964	0.014
−1.1	0.218	0.136	1.169	1.9	0.066	0.971	0.011
−1.0	0.242	0.159	1.083	2.0	0.054	0.977	0.008
−0.9	0.266	0.184	1.000	2.1	0.044	0.982	0.006
−0.8	0.290	0.212	0.920	2.2	0.035	0.986	0.005
−0.7	0.312	0.242	0.843	2.3	0.028	0.989	0.004
−0.6	0.333	0.274	0.769	2.4	0.022	0.992	0.003
−0.5	0.352	0.309	0.698	2.5	0.018	0.994	0.002
−0.4	0.368	0.345	0.630	2.6	0.014	0.995	0.001
−0.3	0.381	0.382	0.567	2.7	0.010	0.997	0.001
−0.2	0.391	0.421	0.507	2.8	0.008	0.997	0.001
−0.1	0.397	0.460	0.451	2.9	0.006	0.998	0.001
				3.0	0.004	0.999	0.000

4.7.3 Lead-Time Demand

Three more methods to generate the safety stock are called the following: service level, percent fill, and Lagrange. All three of these methods use the standard deviation of the lead time forecast in the computations. Recall, the lead-time is the duration of time that begins when a stocking location sends a replenish order to the supplier and ends when the stock in received at the location. For convenience here, the lead-time is defined in monthly units and is denoted as L.

In the typical inventory control system, at the end of each month a new set of future monthly forecasts, f_τ $\tau = 1, 2, \ldots$, are generated for every item in stock, and at the same time, the system will compute an associated one month standard deviation, σ of the forecast error. The monthly forecasts, f_τ, are converted to lead time forecast, F_L as shown below:

$$\begin{aligned} &\text{If } L \leq 1, && F_L = L \times f_1 \\ &\text{If } 1 < L \leq 2, && F_L = f_1 + (L-1)f_2 \\ &\text{If } 2 < L \leq 3, && F_L = f_1 + f_2 + (L-2)f_3 \end{aligned}$$

and so forth.
When the monthly forecasts are rather flat with f the average monthly forecast, the lead time forecast becomes the following:

$$F_L = L \times f$$

The 1-month standard deviation, σ, is converted to a lead time duration, σ_L, as shown below:

$$\sigma_L = \sqrt{L}\sigma$$

Example 4.4 Suppose an item in a distribution center with a lead time of $L = 1.2$ months, has forecasts of ten pieces for month 1, 12 pieces for month 2, and the associated monthly standard deviation is $\sigma = 3$. Below shows how to convert the data from monthly to lead time durations.

$$\begin{aligned} F_L &= 10 + 0.2 \times 12 = 12.4 \\ \sigma_L &= \sqrt{1.2} \times 3 = 3.3 \end{aligned}$$

The estimates of the parameters for the lead-time demand is $F_L = 12.4$ for the mean and $\sigma_L = 3.3$ for the standard deviation.

4.7.4 Safety Stock from the Standard Normal

The normal distribution for the lead-time demand is commonly in use to generate the safety stock for the individual items in the inventory. This distribution is acceptable when the coefficient of variation for the lead-time demand, denoted as cov_L is 0.33 or less. The standard normal is still a good approximate when the cov_L is 0.50 or less. Below shows how to generate the safety stock for the three methods: service level, percent fill and Lagrange using the standard normal.

4.7.4.1 Service Level Method

A popular way to generate the safety stock is by the service level method. This method seeks to satisfy the probability that sufficient stock is available to fill all customer demands during the order cycle. Recall, the order cycle is the duration of time from one receipt of stock till the next receipt of stock from the supplier.

The data needed to carry out this method is the following:

f = average monthly forecast
σ = standard deviation of f
L = lead time (months)
SL = parameter = P(not out of stock)

Below shows the computations needed to find the safety stock. First the lead-time forecast, F_L, and the associated standard deviation, σ_L, are obtained.

$$F_L = L \times f = \text{lead time forecast}$$
$$\sigma_L = \sqrt{L}\sigma = \text{lead time standard deviation}$$

Next, the parameter SL is used as the cumulative distribution, F(k), of the standard normal distribution, whereby F(k) = SL, and a search of Table 4.2 yields the corresponding safety factor, k, to use. Figure 4.4 depicts the relation between F(k), k and z from the standard normal distribution. Some commonly used values of F(k) and k are listed in Table 4.3.

The safety stock is calculated as shown below:

$$SS = k \times \sigma_L = \text{safety stock}$$

A measure called the order point, OP, is needed to signal when to order new stock. Below shows how the safety stock is used along with the lead-time forecast to compute the order point.

$$OP = F_L + SS$$

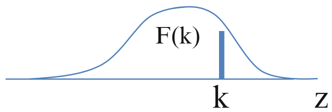


Fig. 4.4 Depiction of F(k) with k and z from the standard normal distribution

Table 4.3 Selected Values of F(k) and k from the Standard Normal Distribution

F(k)	k
.99	2.33
.98	2.05
.97	1.80
96	1.75
.95	1.65
.94	1.56
.93	1.48
.92	1.41
.91	1.34
.90	1.28

New stock is ordered when the sum of the on-hand, OH, plus on-order, OO, inventory is less or equal to the order point as described below:

$$\text{If } (OH + OO) \leq OP \quad \text{order new stock}$$

Example 4.5 Suppose an item in the inventory has an average monthly forecast of $f = 10$, and the corresponding standard deviation is $\sigma = 3$. Further, the lead-time is $L = 1.5$ months and the service level parameter is set as $SL = 0.95$. The lead-time forecast and standard deviation are obtained as below:

$$F_L = 1.5 \times 10 = 15$$
$$\sigma_L = \sqrt{1.5} \times 3 = 3.67$$

The cumulative distribution from the standard normal distribution becomes, $F(k) = SL = 0.95$. Using Table 4.3, the safety factor is $k = 1.65$. Finally, the safety stock and the order point are computed as below:

$$SS = 1.65 \times 3.67 = 6.06$$
$$OP = 15 + 6.06 \approx 21$$

New stock is ordered when the on-hand plus on-order inventory for the item is less or equal to 21 pieces.

4.7.4.2 Percent Fill Method

Generating the safety stock by the percent fill method is used when the management seeks to control the percent of the total demand that is filled from the stock in the inventory. Percent fill is essentially a ratio (not a probability) as defined below:

$$\text{Percent fill} = (\text{demand filled})/(\text{total demand})$$

The data needed to carryout this method is listed below:

PF = percent fill parameter
f = average forecast per month
σ = standard deviation of f
L = lead time in months
Q = planned order quantity

With the above data, the computations to find the safety stock are summarized below:

$$F_L = L \times f = \text{lead time forecast}$$
$$\sigma_L = \sqrt{L}\sigma = \text{lead time standard deviation}$$
$$E(z > k) = [1 - PF]Q/\sigma_L = \text{partial expectation}$$
$$\text{From } E(z > k) \text{ and Table 4.2, find the safety factor k}$$
$$SS = k\sigma_L = \text{safety stock}$$

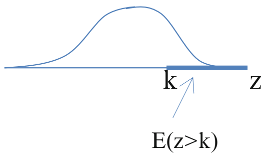


Fig. 4.5 The partial expectation E(z > k) with k and z from the standard normal distribution

Table 4.4 Selected Values of E(z>k) and k from the Standard Normal Distribution

E(z>k)	k
.40	0.003
.35	0.102
.30	0.217
.25	0.345
.20	0.493
.15	0.671
.10	0.902
.05	1.255
.001	2.600

Recall the partial expectation is a statistic from the standard normal distribution. For a given safety factor k, there is an associated value of $E(z > k)$ as listed in Table 4.2. Figure 4.5 depicts where the partial expectation is the average of all values of z that are greater than a given value of k. For brevity, Table 4.4 lists only a selected few values of $E(z > k)$. In the computations on the percent fill safety stock, the safety factor to apply is the k entry that corresponds with $E(z > k)$ in the standard normal tables.

As in the service level method the order point is computed as the sum of the lead-time forecast and the safety stock. A replenish quantity is called only when the on-hand plus on-order is less or equal to the order point.

Example 4.6 Consider a part held in a distribution center where the percent fill method of computing the safety stock is in effect. The part has a monthly forecast of $f = 10$, and a standard deviation of $\sigma = 3$. The lead-time is $L = 1.5$ months and the ideal order quantity is $Q = 8$. The management percent fill parameter is set as $PF = 0.95$.

The computations to find the safety stock are listed below. First the lead-time forecast, F_L, and standard deviations, σ_L, are computed.

$$F_L = 1.5 \times 10 = 15$$
$$\sigma_L = \sqrt{1.5} \times 3 = 3.67$$

Next the partial expectation is derived, $E(z > k)$. Using Table 4.2, the value of the safety factor, k, that corresponds to the partial expectation is located.

$$E(z > k) = (1 - 0.95)8/3.67 = 0.109$$
$$k \approx 0.856$$

Finally, the safety stock, SS, is computed.

$$SS = 0.856 \times 3.67 = 3.14$$

Having obtained the safety stock, the order point, OP, is measured as below, indicating a new replenishment of stock is needed whenever the on-hand plus on-order inventory is eighteen pieces or less.

$$OP = 15 + 3.14 \approx 18$$

4.7.4.3 Lagrange Method

The Lagrange method is used to control the investment in safety stock that spans a collection of parts that belong to a homogeneous group of some type; could be high volume parts, one supplier parts, parts from a sales division of the firm, so forth. Altogether, the group has N parts where i = 1, ..., N identifies the i-th part in the group. To compute the safety stock for an individual part in the group, the following data is needed:

LG = Lagrange parameter
F12 = 12 month forecast for part i
c = cost per unit for i
Q = plan order quantity for i
σ = standard deviation of one month forecast for i
L = lead time (month's) for i

The computations for each part are listed below. The lead-time forecast, F_L, and the corresponding standard deviation, σ_L, are calculated in the usual way.

$$F_L = L \times F12/12 = \text{lead time forecast}$$
$$\sigma_L = \sqrt{L}\sigma = \text{lead time standard deviation}$$

Next, the probability the part is not out of stock, F(k), is derived. Using Table 4.2, the safety factor, k, that corresponds to F(k) is obtained.
F`(k) = (1 − LG ×c×Q/F12) = Lagrange computed service level for the part.
Set F(k) = F`(k) and use Table 4.2 to find the safety factor k that corresponds to F(k).
Finally, the safety stock, SS, is computed.

$$SS = k \times \sigma_L = \text{safety stock}$$

Recall the cumulative probability is a statistic from the standard normal distribution. For a given safety factor k, there is an associated value of F(k) as listed in Table 4.2. Figure 4.4 depicts where the cumulative probability is the probability of z less than k. For brevity, Table 4.5 lists only a selected few values of F(k) with related values of k. In the computations to find the percent fill safety stock, the safety factor is obtained by finding the value of k that corresponds to F(k) in the standard normal tables.

Table 4.5 Selected values of F(k) and k from the Standard Normal Distribution

F(k)	k
.99	2.33
.95	1.65
.90	1.28
.85	1.04
.80	0.89
.75	0.67
.70	0.52
.65	0.38
.60	0.25
.55	0.12
.50	0.00

Example 4.7 Consider a group of parts that are to be controlled using the Lagrange method of generating the safety stock, and the Lagrange parameter for the group is set at LG = 0.02. Assume a part in the group with the following data: the 12-month forecast is F12 = 120, the cost per unit is c = \$10.00, the standard deviation is $\sigma = 3$, the lead-time is 1.5 months, and the plan order quantity is Q = 8 pieces.

Using the part data, the computations to find the safety stock for this part is summarized below. The computations show the safety stock is 8.07 or approximately SS = 8.

$$\begin{aligned}
&\sigma_L = \sqrt{1.5} \times 3 = 3.67\\
&F^{`}(k) = (1 - 0.02 \times 10 \times 8/120) = .987\\
&k \approx 2.2\\
&SS = 2.2 \times 3.67 = 8.07
\end{aligned}$$

4.8 Truncated Normal Distribution

The truncated normal distribution has a location parameter, k, that defines the range of the values of the standard normal variable, z, that are admissible, whereby, the truncated normal includes all values of z larger than k, as shown in Fig. 4.6.

The variable $t = (z - k)$ where $t > 0$ is defined as a truncated normal variable and yields the mean, $\mu_T(k)$, standard deviation, $\sigma_T(k)$, and coefficient-of-variation, $c_T(k) = \sigma_T(k)/\mu_T(k)$. These statistics are computed from various standard normal statistics as shown below:

$$\begin{aligned}
&E(z > k) = f(k) - kH(k)\\
&E(z > k)^2 = H(k)\left(1 + k^2\right) - kf(k)\\
&\mu_T(k) = E(z > k)/H(k)\\
&V_T(k) = E(z > k)^2/H(k) - \mu_T(k)^2\\
&\sigma_T(k) = [V_T(k)]^{0.5}
\end{aligned}$$

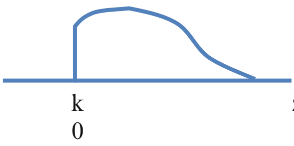


Fig. 4.6 Truncated normal distribution with Parameter k, and variable $t > 0$

Another way to express the truncated variable is by the following:

$$\begin{aligned} w &= [t - \mu_T(k)]/\sigma_T(k) \\ &= [z - k - \mu_T(k)]/\sigma_T(k) \end{aligned}$$

The truncated normal variable w is a counterpart to the standard normal z variable where the variables are located about the mean. The smallest value of w is $-1/c_T(k)$ and the largest is infinity.

Table 4.6 lists location parameter values of k ranging from −3.0 to 3.0. For each k, the associated cumulative distribution, F(k), truncated mean, $\mu_T(k)$, standard deviation, $\sigma_T(k)$ and coefficient-of-variation, $c_T(k)$ are also given.

4.8.1 Safety Stock from the Truncated Normal

Technically, the truncated normal distribution for the lead-time demand is proper when the coefficient of variation for the lead-time, cov_L, is larger than 0.33. In practicality, however, the truncated normal applies when $cov_L > 0.50$, since the truncated normal is close to the standard normal at $cov_L \leq 0.50$. Some may use the truncated normal only when $cov_L \geq 0.70$. Below shows how to generate the safety stock for the three methods: service level, percent fill and Lagrange using the truncated normal.

4.8.1.1 Service Level Method

The service level method seeks the safety stock needed to fill the demands over the order cycle with a probability of SL, as specified by the management. Below shows the data needed to achieve this goal using the truncated normal distribution.

cov_L = coefficient-of-variation for the lead-time
SL = desired service level
σ_L = standard deviation for the lead-time

With the data provided, the following six steps are taken:

1. From Table 4.6, set $c_T(k) = cov_L$ and find the truncated normal parameter k that corresponds to $c_T(k)$. Recall, $c_T(k)$ is the coefficient-of-variation for the truncated normal when k is the location parameter.

Table 4.6 The Truncated Normal Distribution with location parameter, k: cumulative probability distribution, F(k), the truncated mean, $\mu_T(k)$, truncated standard deviation, $\sigma_T(k)$, and coefficient-of-variation, $c_T(k)$

k	F(k)	$\mu_T(k)$	$\sigma_T(k)$	$c_T(k)$	k	F(k)	$\mu_T(k)$	$\sigma_T(k)$	$c_T(k)$
−3.0	0.001	3.004	0.993	0.331	0.0	0.500	0.798	0.603	0.756
−2.9	0.002	2.906	0.991	0.341	0.1	0.540	0.763	0.585	0.767
−2.8	0.003	2.808	0.989	0.352	0.2	0.579	0.729	0.568	0.778
−2.7	0.003	2.710	0.986	0.364	0.3	0.618	0.698	0.551	0.789
−2.6	0.005	2.614	0.982	0.376	0.4	0.655	0.669	0.534	0.799
−2.5	0.006	2.518	0.978	0.388	0.5	0.691	0.641	0.518	0.808
−2.4	0.008	2.423	0.972	0.401	0.6	0.726	0.615	0.503	0.817
−2.3	0.011	2.329	0.966	0.415	0.7	0.758	0.590	0.488	0.826
−2.2	0.014	2.236	0.959	0.429	0.8	0.788	0.567	0.473	0.834
−2.1	0.018	2.145	0.951	0.443	0.9	0.816	0.546	0.460	0.842
−2.0	0.023	2.055	0.942	0.458	1.0	0.841	0.525	0.446	0.850
−1.9	0.029	1.968	0.931	0.473	1.1	0.864	0.506	0.433	0.857
−1.8	0.036	1.882	0.920	0.489	1.2	0.885	0.488	0.421	0.863
−1.7	0.045	1.798	0.907	0.504	1.3	0.903	0.470	0.409	0.870
−1.6	0.055	1.717	0.894	0.520	1.4	0.919	0.454	0.398	0.876
−1.5	0.067	1.639	0.879	0.536	1.5	0.933	0.439	0.387	0.882
−1.4	0.081	1.563	0.863	0.552	1.6	0.945	0.424	0.376	0.887
−1.3	0.097	1.490	0.847	0.569	1.7	0.955	0.410	0.366	0.892
−1.2	0.115	1.419	0.830	0.585	1.8	0.964	0.397	0.356	0.897
−1.1	0.136	1.352	0.812	0.601	1.9	0.971	0.385	0.347	0.902
−1.0	0.159	1.288	0.794	0.616	2.0	0.977	0.373	0.338	0.906
−0.9	0.184	1.226	0.775	0.632	2.1	0.982	0.362	0.330	0.910
−0.8	0.212	1.168	0.756	0.647	2.2	0.986	0.351	0.321	0.914
−0.7	0.242	1.112	0.736	0.662	2.3	0.989	0.341	0.313	0.918
−0.6	0.274	1.059	0.717	0.677	2.4	0.992	0.332	0.306	0.921
−0.5	0.309	1.009	0.697	0.691	2.5	0.994	0.323	0.298	0.924
−0.4	0.345	0.962	0.678	0.705	2.6	0.995	0.314	0.291	0.926
−0.3	0.382	0.917	0.659	0.718	2.7	0.997	0.306	0.284	0.928
−0.2	0.421	0.875	0.640	0.731	2.8	0.997	0.298	0.277	0.929
−0.1	0.460	0.835	0.621	0.744	2.9	0.998	0.291	0.270	0.931
0.0	0.500	0.798	0.603	0.756	3.0	0.999	0.283	0.264	0.933

2. Using Table 4.6 with location parameter k, find the values of cumulative distribution, F(k); truncated mean, $\mu_T(k)$; and truncated standard deviation, $\sigma_T(k)$.
3. Now compute $F(z_o)$ fas follows:

$$F(z_o) = F(k) + SL[1 - F(k)]$$

4. With $F(z_o)$, use Table 4.2 to located the value of z_o that corresponds. Note the table lists k as z_o.

5. Compute the truncated normal safety factor as follows:

$$w_o = [z_o - k - \mu_T(k)]/\sigma_T(k)$$

6. The safety stock becomes: $SS = w_o \times \sigma_L$

Example 4.8 Assume a service part stocked at a dealer has a forecast of $f = 8$ pieces per month and the 1-month standard deviation is $\sigma = 4.0$. The lead-time from the supplier is $L = 0.5$ months and the management desires a safety stock that yields a service level of $SL = 0.95$.

In this situation, the lead-time forecast is $F_L = 0.5 \times 8 = 4.0$ pieces, and the associated lead-time standard deviation is $\sqrt{0.5} \times 4.0 = 2.83$. Thereby, the lead-time coefficient of variation is $cov_L = 2.83/4.00 = 0.707$. Because cov_L is high, the truncated normal distribution applies. The six steps to find the safety stock are listed below.

1. Set $c_T(k) = cov_L = 0.707$, and use Table 4.6 to find the truncated normal location parameter, $k = -0.4$.
2. Also using Table 4.6, the truncated statistics are the following:

$$\mu_T(k) = 0.962, \sigma_T(k) = 0.678 \text{ and } F(k) = 0.345.$$

3. The counterpart cumulative distribution from the standard normal is:

$$F(z_o) = 0.345 + 0.95[1 - 0.345] = 0.967.$$

4. From $F(z_o) = 0.967$, Table 4.2 yields $z_o \approx 1.85$.
5. The truncated safety factor becomes:

$$w_o = [1.85 - (-0.4) - 0.962]/0.678 = 1.90.$$

6. Finally, the safety stock is $SS = 1.90 \times 2.83 = 5.38$

4.8.1.2 Percent Fill Method

The goal of the percent fill method is to generate the safety stock that satisfies the ratio of demand filled over the total demand. Management sets a parameter, PF, that specifies the ratio desired. The data needed to find the safety stock when the truncated normal distribution applies is listed below:

cov_L = coefficient-of-variation for the lead-time
PF = desired percent fill parameter
σ_L = standard deviation for the lead-time
Q = planned order quantity
$E(w > w_o) = [1 - PF]Q/\sigma_L$ = partial expectation for truncated normal

For a subsequent need, note:

$$E(w > w_o) = E(z > z_o)/[H(k)\sigma_T(k)]$$

and thereby,

$$E(z > z_o) = E(w > w_o)[H(k)\sigma_T(k)]$$

With the data provided, the following six steps are taken:

1. From Table 4.6, set $c_T(k) = cov_L$ and find the truncated normal parameter k that corresponds to $c_T(k)$. Recall, $c_T(k)$ is the coefficient-of-variation for the truncated normal when k is the location parameter.
2. Using Table 4.6 with location parameter k, find the values of cumulative distribution, F(k); truncated mean, $\mu_T(k)$; and truncated standard deviation, $\sigma_T(k)$.
3. Now compute the corresponding partial expectation for the standard normal as below:

$$E(z > z_o) = E(w > w_o)\ \ [1 - F(k)\sigma_T(k)]$$

4. With $E(z > z_o)$, use Table 4.2 to locate the value of z_o that corresponds. Note the table lists k as z_o.
5. Compute the truncated normal safety factor as follows:

$$w_o = [z_o - k - \mu_T(k)]/\sigma_T(k)$$

6. The safety stock becomes: $SS = w_o \times \sigma_L$

Example 4.9 Assume a service part stocked at a dealer has a forecast of f = 8 pieces per month and the 1-month standard deviation is s = 4.0. The desired order quantity is Q = 6 pieces, the lead-time from the supplier is L = 0.5 months and the management desires a safety stock that yields a percent fill of PF = 0.95.

In this situation, the lead-time forecast is $F_L = 0.5 \times 8 = 4.0$ pieces, and the associated lead-time standard deviation is $\sqrt{0.5} \times 4.0 = 2.83$. Thereby, the lead-time coefficient of variation is $cov_L = 2.83/4.00 = 0.707$. Because cov_L is high, the truncated normal distribution applies. The partial expectation becomes:

$E(w > w_o) = [1-0.95]6/2.83 = 0.106$.

The six steps to find the safety stock are listed below.

1. Set $c_T(k) = cov_L = 0.707$ and use Table 4.6 to find the truncated normal location parameter, $k = -0.4$.
2. Also using Table 4.6, the truncated statistics are the following:

$$\mu_T(k) = 0.962, \sigma_T(k) = 0.678\ \text{ and }\ F(k) = 0.345.$$

3. The associated standard normal partial expectation is:

$$E(z > z_o) = 0.106(1 - 0.345)0.678 = 0.047.$$

4. Table 4.2 yields the corresponding standard variable as $z_o \approx 1.3$.
5. The truncated safety factor becomes:

$$w_o = [1.3 - (-0.4) - 0.962]/0.678 = 1.09.$$

6. Finally, the safety stock is SS = 1.09 × 2.83 = 3.08.

4.8.1.3 Lagrange Method

The Lagrange method to generate the safety stock seeks a cost-benefit balance between the cost of the safety stock and the percent fill for the part. The method controls all the parts in a specified group where the goal is to yield a combined service for the group at a cost of the group safety stock that is accepted by the management. LG is a parameter, provided by the management, to carryout the computations needed to find the safety stock for each part in the group. The data to generate the safety stock using the truncated normal distribution is listed below:

cov_L = coefficient-of-variation for the lead-time
LG = Lagrange parameter
σ_L = standard deviation for the lead-time
c = cost per unit
Q = plan order quantity
F12 = 12 month forecast
$F^`(k) = 1 - LG \times c \times Q/F12$ = Lagrange computed service level

For notational simplicity, set SL` = F`(k) as the service level to apply for the part. With the data provided, the following six steps are taken:

1. From Table 4.6, set $c_T(k) = cov_L$ and find the truncated normal parameter k that corresponds to $c_T(k)$. Recall, $c_T(k)$ is the coefficient-of-variation for the truncated normal when k is the location parameter.
2. Using Table 4.6 with location parameter k, find the cumulative distribution, F(k); truncated mean, $\mu_T(k)$; and truncated standard deviation, $\sigma_T(k)$.
3. Now seek $F(z_o)$ from the standard normal that corresponds to the computed service level SL`.

$$F(z_o) = F(k) + SL^`[1 - F(k)]$$

4. With $F(z_o)$, use Table 4.2 to locate the value of z_o that corresponds. Note the table lists k as z_o.

5. Compute the truncated normal safety factor as follows:

$$w_o = [z_o - k - \mu_T(k)]/\sigma_T(k)$$

6. The safety stock becomes: $SS = w_o \times \sigma_L$

Example 4.10 Assume a service part stocked at a dealer has a forecast of f = 8 pieces per month and the 1-month standard deviation is σ = 4.0. Further, the cost-per-unit is $20 and the plan order quantity is Q = 6. The lead-time from the supplier is L = 0.5 months and the management sets the Lagrange parameter for the group as LG = 0.04.

In this situation, the lead-time forecast is $F_L = 0.5 \times 8 = 4.0$ pieces, and the associated lead-time standard deviation is $\sqrt{0.5} \times 4.0 = 2.83$. Thereby, the lead-time coefficient of variation is $cov_L = 2.83/4.00 = 0.707$. Because cov_L is high, the truncated normal distribution applies. Using the Lagrange computations, the service level for the part is set at $SL^` = F^`(k) = 1 - [0.04 \times 20 \times 6]/96 = 0.95$. The six steps to find the safety stock are listed below.

1. Set $c_T(k) = cov_L = 0.707$ and use Table 4.6 to find the truncated normal location parameter, $k = -0.4$.
2. Also using Table 4.6, the truncated statistics are the following:

$$\mu_T(k) = 0.962, \sigma_T(k) = 0.678 \text{ and } F(k) = 0.345.$$

3. The associated standard normal cumulative distribution is:

$$F(z_o) = 0.345 + 0.95[1 - 0.345] = 0.967.$$

4. Table 4.2 yields $z_o \approx 1.85$.
5. The truncated safety factor becomes:

$$w_o = [1.85 - (-0.4) - 0.962]/0.678 = 1.90.$$

6. Finally, the safety stock is $SS = 1.90 \times 2.83 = 5.38$.

4.9 Safety Stock for a Group of Parts

The Lagrange method of generating the safety stock allows management to control all of the parts at one time that are associated with a group. The method seeks an acceptable balance between the combined percent fill for the group and the total investment in safety stock for the group. In a cost benefit manner, the investment in safety stock per part is allocated where some parts have a higher percent fill than others. In general, parts with higher forecasts and those that are less costly are assigned more safety stock and given a relatively higher percent fill. The safety stock is apportioned to the parts, one-by-one, so that the combined percent fill for the group meets the desired level, and the group's total safety stock investment is minimal.

Recall, the group consists of N parts that could include all or a subset of all the parts in the inventory. The Lagrange parameter LG is set at a certain value for the group. Each part in the group is processed one-at-a-time and upon completion of all parts, the essential data saved by part is the following:

$$
\begin{aligned}
&\mathrm{F12} = \text{the 12-month forecast}\\
&\mathrm{k} = \text{the safety factor}\\
&\sigma_L = \text{the lead-time standard deviation}\\
&\mathrm{SS} = \mathrm{k} \times \sigma_L = \text{safety stock}\\
&\mathrm{c} = \text{the cost per unit}\\
&\mathrm{Q} = \text{the plan order quantity}
\end{aligned}
$$

Further computations by part i yields: $\$SS_i$, $E(z > k)_i$, $\$short_i$ and $\$F12_i$, as defined below.

$$
\begin{aligned}
\$SS_i \quad &= c \times SS = \text{cost of safety stock for part i}\\
E(z > k)_i &= \text{partial expectation of part i from the standard normal table}\\
\$short_i \quad &= c \times E(z > k)_i \sigma_L F12/Q\\
&= \text{cost of expected pieces short per year for part i}\\
\$F12_i \quad &= c \times F12\\
&= \text{cost of 12 month forecast for part i}
\end{aligned}
$$

The above data is summed for all of the parts in the group as shown below to obtain the following:

$$
\begin{aligned}
\$SS \quad &= \sum_{i=1}^{N} \$SS_i = \text{cost of safety stock for all parts in the group}\\
\$short &= \sum_{i=1}^{N} \$short_i = \text{cost of 12-months of pieces short for the group}\\
\$F12 \quad &= \sum_{i=1}^{N} \$F12_i = \text{cost of 12-month forecast for the group}
\end{aligned}
$$

With the above, the cost weighted percent fill for the group is calculated as below:

$$PF = 1 - [\$short/\$F12]$$

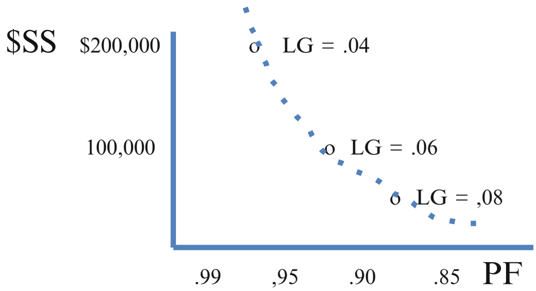
Example 4.11 Suppose N parts are those listed in Table 4.7 with data and tallies of the group investment in safety stock; the group 12-month forecast in dollar; and the estimate the cost of pieces short for 12 months. In the example with the Lagrange parameter set at LG = 0.08, the investment is safety stock is $53,000 and therefore the percent fill is PF = [1 − 140,000/1,200,00] = 0.883.

At the outset in seeking the Lagrange parameter to apply for the group, it is helpful to try at least three candidate parameter values. In the example, these are LG = 0.08, 0.06 and 0.04. The corresponding investment in safety stock and the percent fill is measured for each and an exchange curve is sketched as shown in the

Table 4.7 Worksheet when Lagrange parameter LG = 0.08, F12 = 12 months forecast, c = cost per unit, σ_L = lead-time standard deviation, Q = plan order quantity, $SS = safety stock cost, $short = short pieces cost, and $F12 = 12-month forecast at cost

	Part	F12	c	k	σ_L	Q	$SS	$short	$F12
	1	:	:	:	:	:	:	:	:
	:	:	:	:	:	:	:	:	:
	N	:	:	:	:	:	:	:	:
Sum							$53,000	$140,000	$1,200,000

Fig. 4.7 Exchange curve with $SS, PF and LG



example. See Fig. 4.7 which shows how $SS and PF vary as the parameter LG varies from 0.08 to 0.04. Note when LG decreases, more safety stock is called and the percent fill increases. The exchange curve allows the management to select the value of LG that appears just right for their system.

If the goal is to have PF = 0.95, then LG = 0.048 may be proper, and the investment in safety stock would be $\$SS \approx \$150,000$.

Once, LG is selected for the group, it is generally used for many months subsequently, as long as the inventory results seems appropriate for the parts in the cluster. It behooves the analyst to monitor the summary statistics for each group in the inventory so that the combined percent fill meets management's desired level and the total investment in safety stock is compliant with the management budget.

4.10 Safety Stock When Late Delivery

Consider a distribution center with 100,000 parts and over 2000 suppliers (vendors). Each vendor is the source to a large number of parts and typically publish a lead-time for each of its parts. Some vendors deliver the replenish parts on-time or shortly after the published lead-time. Others have more difficulty and are late in delivery. In a large automotive service part operation, known to the author, the actual replenishment delivery date was late on average 45 % of deliveries.

For notation sake, the supplier published lead-time is L (months) and the actual lead-time is w, where in most situations, $w \geq L$. The stock location computes the safety stock needs based on L and not w. Note, w is a variable and changes with each replenish delivery. Assume the percent fill method is used to generate the safety stock and the desired percent fill is PF. When the safety stock, SS, is computed and the actual lead-time, w, is larger than L, the effective percent fill will be lower than desired PF.

Thomopoulos (2012, pp. 165–172) shows how to overcome this situation by using a quantity called the safety time stock, STS. The effective safety stock becomes $SS_e = SS + STS$. The goal is to add the minimal amount of STS that yields the desired percent fill of the management. Below shows how this method is applied. The details are in the reference; and below shows how the method is used in practice.

4.11 Vendor Measures

Each vendor of the stock location has a performance scorecard that the stock location monitors and revises periodically. The measures by vendor are as listed below:

Po = probability the vendor delivery is on-time.
ML = average months late given the vendor is late.

Hence, for the vendor, the delivery time is L with probability Po, and the delivery time has an average of (L + ML) with probability (1 − Po). This yields the following conditional measures of the percent fill:

$$\begin{array}{ll} \text{When } w = L: & PF_{(w=L)} = PF \\ \text{When } w > L: & PF_{(w>L)} < PF \end{array}$$

The effective percent fill becomes:

$$PF_e = PF_{(w=L)}Po + PF_{(w>L)}(1 - Po)$$

where PF_e is less or equal to the desired PF, and is not known.

4.11.1 On-Time Delivery

A statistical analysis on the vendors is described here. The data measured for each vendor is the following:

PF = desired percent fill
Po = percent of replenishments on time
ML = average months late when a late replenishment

The data for each part of the vendor is below:

$$SS = \text{safety stock using the percent fill method}$$

The computations using the above data are followed:

1. With many samples across all parts, vendors, Po and ML, the average of the safety time factor, k, needed for each Po to achieve the desired percent fill is summarized below:

Po	k
0.0	1.03
0.1	0.98
0.2	0.92
0.3	0.84
0.4	0.78
0.5	0.69
0.6	0.57
0.7	0.45
0.8	0.33
0.9	0.20
1.0	0.00

2. Applying a quadratic equation to the above data yields:

$$k = 1.022 - 0.360Po - 0.642Po^2$$

3. The safety time stock (STS) becomes:

$$STS \approx k \times 1.25ML$$
$$k = \text{safety time factor}$$

4. The effective safety stock (SS_e) for a part with SS, and from a vendor with measures of: Po and ML, becomes the following:

$$SS_e = SS + STS$$

4.11.2 Vendor Safety Time

To overcome the late delivery so that the percent fill satisfies the desired PF, the safety time to use is computed in the following way. First a safety time factor k is computed, using the vendor Po, by the quadratic equation listed below:

$$k = 1.022 - 0.360Po - 0.642Po^2$$

Second, the vendor average months late, ML, is used to find the vendor safety time, ST, as follows:

$$ST = k \times 1.25ML$$

This vendor safety time is computed to provide the minimum amount of safety time stock for each of the replenishments that come from the vendor.

4.11.3 Part Data

For a particular item (part, product, sku) the vendor delivers, let F = the average monthly forecast and σ the associated standard deviation. The safety time stock, STS, becomes,

$$STS = ST \times F$$

and the effective safety stock is:

$$SS_e = SS + STS$$

Finally, the order point, OP, is computed by the following:

$$OP = F_L + SS_e$$

where F_L is the lead-time forecast.

Example 4.12 Consider a stocking location that is using the percent fill method to generate the safety stock and the desired percent fill is PF = 0.95. Assume a vendor with the following performance measures:

$$Po = 0.40 \text{ and } ML = 0.40 \text{ months}$$

For a particular part from this vendor, the data is the following:

$$F = 10$$
$$\sigma = 3$$
$$L = 2 \text{ months}$$
$$Q = 10$$

The safety stock is computed in the regular percent fill way and yields a safety stock of:

$$SS = 3.40$$

The stocking location will use the safety time stock with this part. The quadratic equations yields:

$$k = 1.022 - 0.360(0.40) - 0.642\left(0.40^2\right) = 0.776.$$

So now, the safety time for this part is:

$$ST = 0.776 \times [1.25ML] = 0.388$$

and the safety time stock is:

$$STS = 0.388 \times 10 = 3.88$$

Finally, the effective safety stock for the part becomes:

$$SS_e = 3.40 + 3.88 = 7.28$$

4.12 Loss Sales

A demand occurs when a customer notifies a stock holding entity (distribution center, dealer, store) on the desire to purchase an item and the item is currently available in stock or is out-of-stock. When available, the demand is filled immediately, and when not available, the demand is placed on backorder or is a loss sale. When on backorder, the demand is filled when the stock becomes available, and when a loss sale, the demand is gone forever. In most situations, the filled demands and the backorder demands are recorded and known to the entity, but the loss sales demands are not known or recorded. A way to estimate and control the loss sales demand is provided here.

Consider the following data and notation for the typical data known on each item in stock at the entity:

df = demand filled immediately
dbo = demand in backorder and filled later when stock becomes available
dr = df + dbo = recorded demand

With the above data, the percent fill, pf, for the item can be computed as follows:

$$pf = \text{ratio of}(\text{demand filled})/(\text{recorded demand})$$

Although not know, the notation for the loss sales demands is below:

$$dls = \text{loss sales demand}$$

whereby, notation for the unfilled demands becomes:

$$\text{du} = \text{dbo} + \text{dls}$$

Neither, dls or du are known.

Suppose the entity may be able to estimate the portion (0–1) of unfilled demands that become backorders, whereby, the following notation is stated:

pbo = portion of unfilled demand that is a backorder
pls = (1 − pbo) = portion of unfilled demand that is a loss sales

Observe, (pbo + pls) = 1.

Recall, dls is unknown, but note the relation below:

$$\text{dls}/\text{dbo} = \text{pls}/\text{pbo}$$

and thereby, an estimate of the loss sales demand is obtained by:

$$\text{dls} = \text{dbo}(\text{pls}/\text{pbo})$$

The effective demand, de, is typically unknown, but includes the sum of filled demands and unfilled demands as below:

$$\text{de} = \text{df} + \text{du}$$

So now, the effective percent fill can be defined as follows:

$$\text{pf}_e = \text{df}/\text{de}$$

Example 4.13 Assume a dealer with the following information on all of its parts.

df = \$10,000,000 = annual demand filled immediately
dbo = \$500,000 = annual demand on backorder prior to fill
dr = \$10,500,000 = annual demand recorded
pf = 0.952 = df/dr

Suppose the management estimates that of the unfilled demand, 40 % becomes backorders, therefore,

$$\text{pbo} = 0.40$$
$$\text{pls} = 0.60$$

With this new information, the estimate of the annual loss sales demand, dls\`, becomes:

$$\text{dls}` = \$500{,}000(0.60/0.40) = \$750{,}000$$

Further, an estimate of the annual effective demand, de`, is:

$$\text{de}^{`} = \text{df} + \text{dbo} + \text{dls}^{`} = \$11,250,000$$

With this information, an estimate of the effective percent fill, pfe`, can be computed as shown below:

$$\text{pfe}^{`} = \text{df}/\text{de}^{`} = 10,000,000/11{,}250{,}000 = 0.89$$

Example 4.14 Suppose the dealer management of Example 4.13 want to limit the amount of annual loss sales to dls` = $300,000, and wish to determine how to set their desired percent fill parameter to achieve. Note, de` = $11,250,000 is the effective demand. Since pbo = 0.4 and pls = 0.6, the estimate of potential backorder demands becomes:

$$\begin{aligned}\text{dbo}^{`} &= \text{dls}^{`}(\text{pbo}/\text{pls}) \\ &= 300,000(0.40/0.60) = \$200{,}000\end{aligned}$$

The filled demand becomes:

$$\begin{aligned}\text{df}^{`} &= \text{de}^{`} - \text{dls}^{`} - \text{dbo}^{`} \\ &= 11,250,000 - 300,000 - 200,000 \\ &= 10,750,000\end{aligned}$$

Hence, the estimate of the annual recorded demands becomes:

$$\begin{aligned}\text{dr}^{`} &= \text{df}^{`} + \text{dbo}^{`} \\ &= 10{,}750{,}000 + 200{,}000 \\ &= \$10{,}950{,}000\end{aligned}$$

The desired percent fill needed to achieve the dls` = $300,000 becomes;

$$\begin{aligned}\text{pf}^{`} &= \text{df}^{`}/\text{dr}^{`} = 10{,}750{,}000/10{,}950{,}000 \\ &= 0.982\end{aligned}$$

4.13 Summary

Four methods to generate the safety stock are described: months supply, service level, percent fill, and Lagrange. The latter three require the use of the normal distribution or the truncated normal distribution, whichever applies. The coefficient of variation, cov, for the lead-time demand is used to identify the distribution to use. The Lagrange method is used when the management wants to control the safety stock and service level for a group of parts, simultaneously, in a cost-benefit way. An adjustment to the safety stock is developed for an individual supplier based on its past record of on-time-delivery of replenish stock.

A way to measure and control the amount of loss sales, due to an out-of-stock condition is also described.

Chapter 5
Replenishments

5.1 Introduction

Replenishing the stock is one of the most important functions in the control of the inventory. The goal is to replenish the stock that properly covers the customer demands at minimum cost in inventory. To accomplish, two measures are regularly computed: the order point and order level. Each day, these are compared to the current on-hand plus on-order inventory to determine if a new replenishment is needed now, and if so, how much. Each replenishment must conform with any constraints provided by the supplier, as minimum buy quantity and multiple buy quantity. For analysis sake, the inventory is partitioned into cycle stock and safety stock, where the sum is the total stock. The ingredients that affect the stock levels are the percent fill, lead-time, coefficient of variation, and month-in-buy. The sensitivity of each with the stock levels and with the turnover is described. In many retail entities, the demand for each item is low, and thereby the Poisson distribution is used to determine the order point and order level. For entities with low demand items, table entries are provided to guide the management on how to set the order point and order level.

5.2 Order Point and Order Level

Replenishments of stock is perhaps the most important function in controlling the inventory. Management's typical goal is to yield a high level of service to the customers by having the suitable amount of stock available to meet the demands, and achieve this with the minimum investment in inventory. Replenishing the inventory at the right time and in the proper amounts is needed to accomplish this goal. Two guide-posts are used for this purpose: the order point (OP) and the order level (OL). They are used on a daily basis to monitor the

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_5

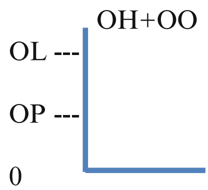


Fig. 5.1 Replenish guide-posts

amount of on-hand (OH) plus on-order (OO) inventory quantity that is available. See Fig. 5.1.

The order point and order level are revised each month when a new forecast is generated. This is when the forecasts are revised with the latest demands, and subsequently, the lead-time forecast, safety stock and plan order quantity are refreshed. The order point and order level are obtained as shown below:

$$\mathrm{OP} = \mathrm{F_L} + \mathrm{SS}$$
$$\mathrm{OL} = \mathrm{OP} + \mathrm{Q}$$

where,

$\mathrm{F_L}$ = lead time forecast
SS = safety stock
Q = plan order quantity

Example 5.1 Consider an item with the following data: $\mathrm{F_L} = 10$, SS = 4, Q = 8. The order point and order level become the following:

$$\mathrm{OP} = 10 + 4 = 14$$
$$\mathrm{OL} = 14 + 8 = 22$$

It is noted where the order point and order level also may need to be revised in the event the supplier alters one or more of the replenish terms for an item; such as the cost per unit, cost per order, or lead time.

5.2.1 When to Buy and How Much

Each day the inventory status, on-hand (OH) plus on-order (OO), of each item is monitored against the order point. Should (OH + OO) be higher than the OP, the status is proper and no replenish stock is needed. But when the (OH + OO) is equal or below the order point, a new replenish buy is called to bring the inventory status up to the order level (OL). The general rule is below:

$$\mathrm{If}[\mathrm{OH} + \mathrm{OO}] \le \mathrm{OP}: \quad \mathrm{Buy} = \mathrm{OL} - [\mathrm{OH} + \mathrm{OO}]$$

Example 5.2 Suppose a part with the following data: OP = 8, OL = 15, OH = 12, OO = 4. In this situation: Buy = 0.

Example 5.3 Assume, the part has the following data: OP = 8, OL = 15, OH = 2, OO = 4. In this situation: Buy = 15−[2 + 4] = 9

5.2.2 Rounding the Order Point and Order Level

In most computerized systems, the safety stock and order quantity are generated in fractional forms, and thereby the order point and order level are also initially in fractional form. Assume, OP\` and OL\` are the fractional form, and OP and OL are the integer counterparts. The conversion to integers is described below.

$$\text{OP} = \text{integer}[\text{OP}` + 0.5]$$
$$\text{OL} = \text{integer}[\text{OL}` + 0.5]$$

Example 5.4 Assume the data for a part is the following: $F_L = 7.3$, SS = 4.4, Q = 8.5.

Hence, the fractional form of the order point and order level become:

$$\text{OP} = [7.3 + 4.4] = 11.7$$
$$\text{OL} = [11.7 + 8.5] = 20.2$$

Rounding yields the following:

$$\text{OP} = \text{integer}[11.7 + 0.5] = 12$$
$$\text{OL} = \text{integer}[20.2 + 0.5] = 20$$

5.2.3 Adjusting Ties in Order Point and Order Level

Sometimes, the rounding process yields the same value for the order point and order level, whereby OP = OL. Because of this, an adjustment tie rule is applied that calls for the order point to be lowered by one unit. Ties often happen for low demand items where the forecasts are small fractional numbers. The general rule for a tie is listed below:

$$\text{If } \ \text{OL} = \text{OP}: \quad \text{OP} = \text{OL} - 1$$

Example 5.5 Assume a part with the following data: $F_L = 0.2$, SS = 0.0, Q = 0.2 The computations for the order point and order level are the following:

$$\text{OP}` = [0.2 + 0.0] = 0.2$$
$$\text{OL}` = [0.2 + 0.2] = 0.4$$

Rounding yields:

$$\begin{aligned}OP &= \text{integer}[0.2 + 0.5] = 0\\ OL &= \text{integer}[0.4 + 0.5] = 0\end{aligned}$$

Applying the tie rule:

$$\begin{aligned}OP &= 0 - 1 = -1\\ OL &= 0\end{aligned}$$

Hence, (OP, OL) = (−1,0).

5.2.4 When a Negative Order Point

A negative order point is a common occurrence for low demand items, indicating that a new replenishment is not needed unless a backorder is generated.

5.3 Replenish Stock

The data needed to monitor the stock level of each item in inventory is the following; the order point and order level, that are typically refreshed at the start of each month; and the on-hand and on-order inventory, that are subject to change each day. As stated earlier, the order point and order level are computed using the lead-time forecast, safety stock and plan order quantity. These ingredients are computed at the start of each month when the prior month demand data becomes available whereby the forecasts are revised accordingly.

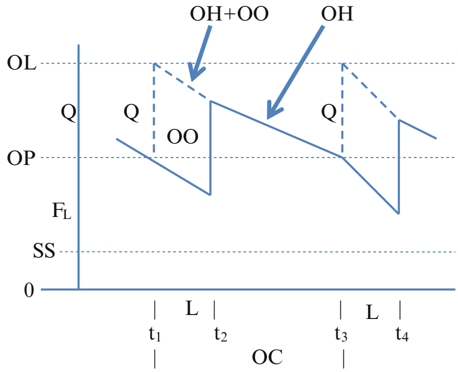
5.3.1 Replenish Rule

The replenish rules are fairly simple: each day the stock status (on-hand plus on-order inventory) is compared to the order point. A new replenish buy is called only when the stock status is equal or below the order point; whereby the buy quantity brings the inventory up to the order level. The replenish process is as follows:

The rule on when to buy and how much are the following:

$$\begin{aligned}&\text{Buy when:} && (OH + OO) \leq OP\\ &\text{Buy quantity:} && \text{buy} = OL - (OH + OO)\end{aligned}$$

Fig. 5.2 Replenish flow



5.3.2 Inventory Flow Over the Order Cycle

Figure 5.2 depicts the flow of inventory over time for an item in the stocking location. The vertical scale shows the order point and order level and their ingredients. The solid line represents the on-hand inventory and the dotted line shows the on-hand plus on-order inventory. The horizontal scale gives various times when events occur. At t_1, the stock status is at the order point and thereby a new replenish quantity, Q, is ordered to bring the stock up to the order level. At t_2, the stock arrives. At t_3, another replenish order is needed that brings the stock up to the order level. Finally, at t_4, the stock arrives.

Note, the lead-time is the duration from t_1 to t_2, and also from t_3 to t_4. The order cycle is the duration between t_1 and t_3, the times between ordering new stock. The order cycle OC could also represent the duration between t_2 and t_4, the time between two receipts of replenish stock.

5.4 Replenish Quantity Constraints

Sometimes the supplier places constraints on the buy quantities that it will accept. Recall, a buy quantity is needed whenever the on-hand plus on-order (OH + OO) inventory is at the order point, OP, or lower. The buy quantity brings (OH + OO) up to the order level, OL. When a supplier constraint is present, this quantity is temporary and in essence becomes an interim quantity, labeled here as q`, and the final quantity is q. In most supplier contracts, no constraints are called and thereby, the interim quantity is the same as the final quantity, q = q`. The computations for the interim buy quantity, q`, is below.

$$\text{if } (OH + OO) > OP: \quad \text{buy} = 0$$
$$\text{if } (OH + OO) \leq OP: \quad \text{buy} > 0$$
$$\text{If buy} > 0: \quad q` = OL - (OH + OO)$$

5.4.1 Min Quantity and Multiple Quantity

The most common supplier constraints are the following: the minimum quantity, q_{min}, and the multiple quantity, q_M. The minimum quantity is a restriction that stipulates the smallest quantity that is allowed in the buy. The multiple quantity requires all buys to be in units that coincide with the stated multiple.

Sometimes both of the constraints are stated in the supplier contract, whereby the minimum quantity must also be compliant with the multiple quantity. For example if $q_{min} = 10$ and $q_M = 6$, the two constraints are not compliant, and an effective q_{min} becomes 12.

The computations to comply with the minimum constraint is below:

$$\begin{array}{ll} \text{if } (OH + OO) > OP : & \text{buy} = 0 \\ \text{if } (OH + OO) \leq OP : & \text{buy} > 0 \\ \text{If buy} > 0 \ \& \ q_{min} > 0 : & q` = OL - (OH + OO) \\ & q = \max[q`, \ q_{min}] \end{array}$$

When a multiple constraint is in effect, two interim quantities are computed, q` and q``. The calculations to reach the final quantity, q, is below:

$$\begin{array}{ll} \text{if } (OH + OO) > OP : & \text{buy} = 0 \\ \text{if } (OH + OO) \leq OP : & \text{buy} > 0 \\ \text{if buy} > 0 : & q` = OL - (OH + OO) \\ \text{If } q` > 0 \ \& \ q_M > 0 : & q`` = \text{integer}[q`/q_M + 0.5]q_M \\ & q = \max[q``, \ q_M] \end{array}$$

Example 5.6 When the interim quantity is q` = 9 and the minimum quantity is $q_{min} = 10$, the final quantity is q = 10.

Example 5.7 If q` = 9, and the multiple is $q_M = 6$. In the calculations, a second interim buy quantity is needed here and is denoted as $q`` = \text{int}[9/6 + 0.5] \times 6 = 12$. The final quantity becomes q = max(12, 6] = 12.

Example 5.8 If q` = 9 and $q_M = 8$, and a decision is needed on whether to buy 8 or 16 units from the supplier to be compliant with the multiple quantity. A common rule is to round to the closest multiple, and as such, the final quantity becomes q = 8.

Example 5.9 If q` = 1 and $q_M = 5$, $q`` = \text{int}[1/5 + 0.5] \times 5 = 0$, and q = max[0, 5] = 5.

5.5 Cycle Stock, Safety Stock and Total Stock

For management planning purposes, it is helpful to consider a partition of the inventory into two segments, the cycle stock and the safety stock. The cycle stock (CS) is the stock needed to fill the forecast demands, and the safety stock

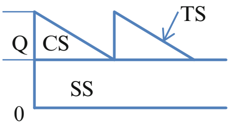


Fig. 5.3 Flow of cycle stock, safety stock and total stock

(SS) is required in the event the demand exceeds the forecast. The sum of the two segments is denoted as the total stock (TS). Figure 5.3 depicts the composition of the stock partitions and also depicts the flow over the replenish cycle.

5.5.1 Average Cycle Stock, Safety Stock and Total Stock

Note how the cycle stock ranges from a high of Q, the planned order quantity, at the start of a replenish cycle, and a low of zero at the end of the cycle. In the same way, the total stock ranges from (Q + SS) to SS over a replenish cycle. Because of this variation, it is helpful to consider the average cycle stock over the replenish cycle. This is $\overline{CS} = Q/2$, and so, the average total stock becomes $\overline{TS} = (Q/2 + SS)$.

Example 5.10 Suppose a part with the following data: F = 10, SS = 6 and Q = 20. The average cycle stock, safety stock and total stock become: $\overline{CS} = 20/2 = 10$, SS = 6, and $\overline{TS} = 16$, respectively.

5.5.2 Average Months of Cycle Stock, Safety Stock and Total Stock

It is also useful to consider the stock in months of supply in the following way. The average month's supply of cycle stock becomes $CS_M = \overline{CS}/F$ where F is the average forecast per month. In the same way, $TS_M = \overline{TS}/F$ is the average month's supply of total stock.

Example 5.11 Consider the part from Example 5.10 again where F = 10, SS = 6 and Q = 20; and also $\overline{CS} = 10$, SS = 6, and $\overline{TS} = 16$. Now find the corresponding measures of inventory in months of supply. Since F = 10, the average months of cycle stock is $CS_M = 10/10 = 1.0$, and the average months of safety stock is $SS_M = 6/10 = 0.6$. The average months supply of total stock becomes $TS_M = 16/10 = 1.6$.

5.5.3 Months-in-Buy

In the same manner, the plan order quantity is converted to month's and denoted as months-in-buy (MIB) in the following way: MIB = Q/F. Note also where the average months of cycle stock is merely MIB/2.

Example 5.12 For the part of Examples 5.10 and 5.11, the months supply in the order quantity, called the months-in-buy, becomes MIB = 20/10 = 2.0. The average months of cycle stock is $CS_M = MIB/2 = 2.0/2 = 1.0$.

5.6 On-Order Inventory

The on-order inventory is the quantity of units that have been purchased with the supplier and has not yet been received at the stocking location. In general, this stock will subsequently become part of the on-hand inventory.

5.6.1 Average On-Order Inventory

The on-order (OO) inventory is the stock ordered and not yet received. Sometimes it is of interest to compute the average on-order inventory, $\overline{OO}$. The average is obtained for the duration of an order cycle, the time between two receipts of stock from the supplier. Note that the average time of an order cycle is MIB (months-in-buy). For this period, the on-order inventory ranges from a low of zero to a high of Q. The on-order remains at a size of Q for a duration of L, and is zero for the duration of (MIB−L). Dividing by MIB, and using some algebra, the average on-order inventory becomes F_L, the lead-time forecast, as listed below.

$$\overline{OO} = [(\mathrm{MIB} - \mathrm{L}) \times 0 + \mathrm{L} \times \mathrm{Q}]/\mathrm{MIB} = \mathrm{F_L}$$

5.6.2 Average Months of On-Order Inventory

The average months supply of on-order is L, the lead-time months, as shown below.

$$\mathrm{OO_M} = \overline{OO}/\mathrm{F} = \mathrm{F} \times \mathrm{L}/\mathrm{F} = \mathrm{L}$$

5.7 Turnover

An important metric used in inventory is called the turnover. This gauges the number of inventory replenishments per year, and is usually measured on a group of items, and often in units of aggregate dollars.

The turnover can also be measured for an individual item in pieces. In this sense, the turnover, TO, is the ratio of the annual demand over the average on-hand inventory, and this is equivalent to the ratio of 12-months forecast over the average total stock. Below shows that with the use of some algebra, the turnover becomes the ratio of 12 over the average months of total stock, i.e., $TO = 12/TS_M$.

$$\begin{aligned} TO &= \text{turnover} \\ &= (\text{Annual demand/average on-hand inventory}) \\ &= (\text{12-month forecast/average total stock}) \\ &= F12/\overline{TS} \\ &= 12/TS_M \end{aligned}$$

where,

$F12 = 12 \times F = 12$ month forecast
$\overline{TS} = TS_M \times F =$ average total stock
$F =$ average monthly forecast
$TS_M =$ average months of total stock

Example 5.13 Find the turnover for a part with $F = 10$, $SS = 6$ and $Q = 20$. The average pieces of total stock becomes: $\overline{TS} = (20/2 + 6) = 16$, and thereby, the average months supply is $TS_M = 16/10 = 1.6$. Finally, the turnover for the part becomes, $TO = 12/1.6 = 7.5$.

5.8 Safety Stock Ingredients

The common methods (service level, percent fill, Lagrange) of generating the safety stock, SS, is by the product of the safety factor, k, and the lead-time standard deviation, σ_L as below:

$$SS = k\sigma_L$$

Recall also where σ_L is composed of σ (standard deviation) and L (lead-time in months) as below:

$$\sigma_L = \sigma\sqrt{L}$$

Note also where the coefficient-of-variation is $cov = \sigma/F$, with F the average monthly forecast. Below shows how the safety stock is derived from the ingredients: k, cov, L and F.

$$SS = k \times cov \times F\sqrt{L}$$

Dividing by the average monthly forecast, F, yields the months of safety stock which is composed of k, cov and L, as listed below.

$$SS_M = k \times cov\sqrt{L}$$

5.8.1 CS, SS and TS Needed for PF, Cov, L and MIB

The amount of inventory needed for an individual part depends on the combination of the following ingredients: percent fill (PF), coefficient-of-variation (COV), lead-time (L) and months-in-buy (MIB). Table 5.1 lists the partition of inventory in units of months supply by cycle stock (CS_M), safety stock (SS_M) and total stock (TS_M).

Table 5.1 Months of CS, SS, and TS, along with turnover, TO, from PF, cov, L and MIB

PF	cov	L	MIB	CS	SS	TS	TO
0.90	0.3	0.25	0.25	0.13	0.09	0.22	55.81
0.90	0.3	0.25	1.00	0.50	0.00	0.50	24.00
0.90	0.3	1.00	0.25	0.13	0.30	0.43	28.24
0.90	0.3	1.00	1.00	0.50	0.04	0.54	22.22
0.90	0.5	0.25	0.25	0.13	0.22	0.35	34.78
0.90	0.5	0.25	1.00	0.50	0.00	0.50	24.00
0.90	0.5	1.00	0.25	0.13	0.63	0.76	15.89
0.90	0.5	1.00	1.00	0.50	0.24	0.74	16.22
0.90	1.0	0.25	0.25	0.13	0.63	0.76	15.89
0.90	1.0	0.25	1.00	0.50	0.24	0.74	16.22
0.90	1.0	1.00	0.25	0.13	1.57	1.70	7.08
0.90	1.0	1.00	1.00	0.50	0.90	1.40	8.57
0.95	0.3	0.25	0.25	0.13	0.15	0.28	43.64
0.95	0.3	0.25	1.00	0.50	0.02	0.52	23.08
0.95	0.3	1.00	0.25	0.13	0.40	0.53	22.86
0.95	0.3	1.00	1.00	0.50	0.18	0.68	17.65
0.95	0.5	0.25	0.25	0.13	0.31	0.44	27.59
0.95	0.5	0.25	1.00	0.50	0.12	0.62	19.35
0.95	0.5	1.00	0.25	0.13	0.78	0.91	13.26
0.95	0.5	1.00	1.00	0.50	0.45	0.95	12.63
0.95	1.0	0.25	0.25	0.13	0.78	0.91	13.26
0.95	1.0	0.25	1.00	0.50	0.45	0.95	12.63
0.95	1.0	1.00	0.25	0.13	1.85	1.98	6.08
0.95	1.0	1.00	1.00	0.50	1.25	1.75	6.86
0.99	0.3	0.25	0.25	0.13	0.25	0.38	32.00
0.99	0.3	0.25	1.00	0.50	0.17	0.67	17.91
0.99	0.3	1.00	0.25	0.13	0.60	0.73	16.55
0.99	0.3	1.00	1.00	0.50	0.43	0.93	12.90
0.99	0.5	0.25	0.25	0.13	0.48	0.61	19.83
0.99	0.5	0.25	1.00	0.50	0.34	0.84	14.29
0.99	0.5	1.00	0.25	0.13	1.10	1.23	9.80
0.99	0.5	1.00	1.00	0.50	0.83	1.33	9.02
0.99	1.0	0.25	0.25	0.13	1.10	1.23	9.80
0.99	1.0	0.25	1.00	0.50	0.83	1.33	9.02
0.99	1.0	1.00	0.25	0.13	2.45	2.58	4.66
0.99	1.0	1.00	1.00	0.50	1.94	2.44	4.92

The table also lists the association turnover (TO). The combinations are: PF = 0.90, 0.95, 0.99; cov = 0.30, 0.50, 1.00; L = 0.25, 1.00; and MIB = 0.25, 1.00.

5.8.2 Sensitivity of the Inventory to PF

Recall PF is a parameter set by the management, indicating the portion of demand to be filled from the available stock. The higher the desired percent fill, the more need for safety stock. For example, when cov = 0.5, L = 1 and MIB = 1, and when PF is 0.90, 0.95 and 0.99, the corresponding months of safety stock becomes 0.24, 0.45 and 0.83.

5.8.3 Sensitivity of Inventory to Cov

The cov is unique to each individual part, depending largely on the distribution of the monthly demands. The more consistent the monthly demands, the smaller the cov, and the more wild the monthly demands, the larger the cov. For example, when PF = 0.95, L = 1 and MIB = 1, and when cov = 0.3, 0.5 and 1.0, the corresponding months of safety stock becomes 0.18, 0.45 and 1.25, respectively. The larger the cov, the more safety stock.

5.8.4 Sensitivity of the Inventory to L

The lead-time, L, is provided by the supplier of the part indicating the months needed from ordering the part till receiving the part. The larger L, the more safety stock is needed. For example, when L = 0.25 and L = 1, and when PF = 0.95, cov = 0.5 and MIB = 1, the corresponding months of safety stock becomes 0.12 and 0.45, respectively. As the lead-time increases, the safety stock gets larger.

5.8.5 Sensitivity of the Inventory to MIB

The month-in-buy, MIB, is unique to each part and depends on the economics on the size of the order quantity. MIB is typically larger for inexpensive parts, and smaller for expensive parts. For example, when MIB = 0.25 and MIB = 1, and when PF = 0.95, cov = 0.5 and L = 1, the corresponding months of safety stock becomes 0.78 and 0.45, respectively. The larger the MIB, the smaller the need for safety stock.

5.8.6 *Sensitivity of Turnover to PF, Cov, L and MIB*

The table also shows how the turnover, TO, changes as the combination of ingredients vary. Note also how the turnover directly decreases as the total stock, TS_M, becomes higher.

5.8.7 *Percent of Safety Stock Over Total Stock*

The table also illustrates how large the safety stock is compared to the total inventory. For example, when cov = 0.5, L = 1, MIB = 1, and PF = 0.90, safety stock is (0.24/0.74) = 32% of the total stock. At PF = 0.95, safety stock is (0.45/0.95) = 47% of the total stock, and at PF = 0.99, it is (0.83/1.33) = 62%. The final four columns of the Table 5.1 are the following: CS_M = average months of cycle stock, SS_M = months of safety stock, TS_M = average months of total stock, and TO = turnover.

5.9 Low Demand Items

Most items in a retail store are of the low demand type. These are items that have an average monthly demand of two or less. Could be pricey items like: televisions, refrigerators, power tools, furniture, mattresses, silverware, china, and so forth. Many other retail items have much competition, like toasters, razors, toothpaste, hair cream, and lamps. The more items in competition, the less demand per item. Many other items are by style with a variety of sizes, like: shoes, sweaters, shirts, trousers, dresses, coats, and so on. Each size of the style is a stock-keeping-unit (SKU). In most scenarios, the demand for an individual SKU is small and most are less than one per month. The month-to-month demands for low demand items are not normally distributed, and thereby the safety stock methods described earlier are not proper.

To illustrate why the demands are quite low in an inventory store, two examples from the shoe industry are sited. Consider a large shoe store with 500 styles and an average of 20 sizes per style, representing 10,000 SKUs. Assume for a normal month, the store sells 2000 pair. The average demand per SKU is thereby 0.20 pair. For a more moderate store, the number of styles is 300 and the average number of sizes per style is 15. In this scenario, the store carries 4500 SKUs. If the average number of sales per month is 1500 pair, the average demand per SKU is 0.33 pair.

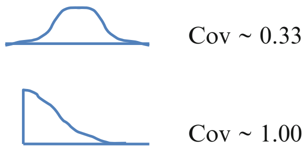


Fig. 5.4 The coefficient-of-variation and the demand distribution

5.9.1 Cov of the Normal Distribution

Recall, the coefficient-of-variation, cov, as applied in forecasting, is the ratio of the 1-month standard deviation of demand, σ, over the average monthly forecast, F. When the demands per month are normally shaped, cov ≤ 0.33. In the event the demands are lumpy, cov ≥ 1.00. Low demand items typically have high values of cov and thereby the monthly demands are not normally shaped. When the monthly demands are low, the Poisson distribution is more appropriate. See Fig. 5.4.

5.9.2 Poisson Distribution

The Poisson distribution is commonly used to describe the probability of a random variable with low outcomes. The random variable is x, and the mean is m. The probability of x occurring is denoted as P_x and is computed as below.

$$P_x = e^{-m} m^x / x! \quad x = 0, 1, 2, \ \ldots \ .$$

The expected value of x is $E(x) = m$, and the standard deviation is $\sigma = \sqrt{m}$. Thereby, the coefficient-of-variation becomes cov $= \sigma/m = 1/\sqrt{m}$.

5.9.3 Cov of the Poisson Distribution

As stated earlier, the average forecast for a retail item is small, where many have averages less than one. Assuming the monthly demands for these items are more like a Poisson distribution, the cov for these items can be analyzed as shown in Table 5.2. Note when the average is $m = 1$, cov $= 1.00$. When $m \leq 1$, cov ≥ 1.00. In general, when m is 4 or less, the cov is drifting away from a normal distribution. As m increases above 4, the cov of the Poisson starts trending towards a normal distribution.

Table 5.2 The mean, m, and Cov of the Poisson probability distribution

m	Cov
1	1.00
2	0.71
3	0.58
4	0.50
5	0.44
6	0.41
7	0.38
8	0.35
9	0.33
10	0.32

5.10 Inventory Flow for Low Demand Items

Consider Fig. 5.5 that depicts the flow of inventory for a low demand item. The example uses an item with an order point (OP) of 1, and an order level (OL) at 3. At time t_1, the order cycle begins when new stock is received, the oh-hand (OH) plus on-order (OO) rises to the OL. At time t_2, the (OH + OO) reaches the OP = 1, and this triggers a new order of size 2 pieces to bring the (OH + OO) up to OL = 3. A lead-time, L, later, at t_3, the stock comes in and brings the OH back up to 3 pieces. Note, the order cycle, OC, is the duration from t_1 to t_3, the time between two receipts of stock. In the depiction, the solid line is the OH, and the dotted line is the (OH + OO).

5.10.1 Percent Fill for Low Demand Items

Continuing with the above scenario, the percent fill is computed using the time duration over an order cycle, OC, as below.

$$\begin{aligned} \mathrm{Pf} &= (\text{demand filled in OC})/(\text{total demand in OC}) \\ &= \mathrm{df}/\mathrm{dt} \end{aligned}$$

where, df = demand filled in OC, and dt = total demand in OC.

The expected values of df and dt are obtained using in part the Poisson probability distribution, where x denotes the demands over the lead-time and F_L is the mean demand over the lead-time.

$$\begin{aligned} \mathrm{E(df)} &= (\mathrm{OL} - \mathrm{OP}) + \sum\nolimits_{x=0}^{OP} xP_x + \mathrm{OP} \times \mathrm{P_{x>OP}} \\ \mathrm{E(dt)} &= (\mathrm{OL} - \mathrm{OP}) + \mathrm{m} \end{aligned}$$

Note in the above, m and F_L are the same. With the formulation above, PF can now be calculated for any combination of: [F_L, OP, OL].

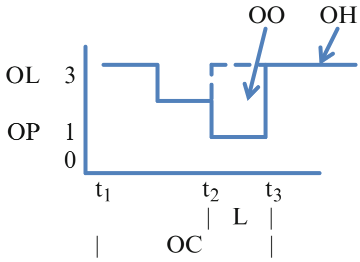


Fig. 5.5 Flow of inventory for a low demand item

5.10.2 Min OP and OL for F_L and PF

Table 5.3 pertains to low demand items and lists the minimum values of (OP, OL) needed for selected combinations of the lead-time forecast, F_L, and percent fill, PF. The values of F_L range from 0.1 to 4.0, and PF entries are [0.80, 0.85, 0.90, 0.95]. The results were obtained by way of computer processing.

For a low demand item, the replenishments of stock is fairly simple. If the [OH + OO] is less or equal to OP, a buy quantity is called to bring the [OH + OO] back up to OL.

Example 5.14 If $F_L = 1.5$ and PF = 0.90,(OP, OL) = (2,4).

Example 5.15 If PF = 0.95, F = 1.2 = average monthly forecast, and L = 0.5 months lead-time, $F_L = 0.6$ and (OP, OL) = (2,3).

Example 5.16 If F = 0.5, L = 0.2, $F_L = 0.1$, and PF = 0.90, (OP, OL) = (0,1).

5.11 Replenishments for SKUs

Consider a distribution center (DC) who stocks style good items and serves as the suppler to a cluster of retail stores. Each style good item, has a multiple number of sizes as in men's shirts, trousers, shoes, sweaters, so on. For each style item, the DC computes a monthly forecast, F, a corresponding standard deviation, σ, and an order quantity size, Q. For the individual sizes, a separate layer of computations are needed to determine when and how much to replenish by size. For notation sake, the N sizes represent N SKUs, and are labeled as i = 1 to N. From the past demands, the following data is known for each SKU:

p_i = portion of total demand of the style (i = 1 to N)

Below are the computations needed for each SKU, i = 1 to N:

Table 5.3 OP and OL for low demand items with F_L and Pf specified

OP, OL				
F_L	PF 0.80	PF 0.85	PF 0.90	PF 0.95
0.1	0,1	0,1	0,1	0,2
0.2	0,1	0,2	0,2	1,2
0.3	0,2	0,2	1,2	1,2
0.4	0,2	1,2	1,2	1,2
0.5	0,2	1,2	1,2	1,3
0.6	1,2	1,2	1,2	2,3
0.7	1,2	1,2	1,3	2,3
0.8	1,2	1,2	1,3	2,3
0.9	1,2	1,3	2,3	2,3
1.0	1,2	1,3	2,3	2,3
1.1	1,3	1,3	2,3	2,4
1.2	1,3	2,3	2,3	2,4
1.3	1,3	2,3	2,3	3,4
1.4	1,3	2,3	2,3	3,4
1.5	1,4	2,3	2,4	3,4
1.6	1,4	2,3	2,4	3,4
1.7	1,4	2,3	2,4	3,4
1.8	1,4	2,3	3,4	3,4
1.9	1,4	2,4	3,4	3,5
2.0	1,4	2,4	3,4	3,5
2.1	1,4	2,4	3,4	4,5
2.2	2,4	3,4	3,4	4,5
2.3	2,4	3,4	3,4	4,5
2.4	2,4	3,4	3,5	4,5
2.5	2,4	3,4	3,5	4,5
2.6	2,4	3,4	3,5	4,5
2.7	3,4	3,4	4,5	4,6
2.8	3,4	3,4	4,5	4,6
2.9	3,4	3,5	4,5	5,6
3.0	3,4	3,5	4,5	5,6
3.1	3,4	3,5	4,5	5,6
3.2	3,4	3,5	4,5	5,6
3.3	3,4	4,5	4,5	5,6
3.4	3,5	4,5	4,6	5,6
3.5	3,5	4,5	4,6	6,7
3.6	3,5	4,5	4,6	6,7
3.7	3,5	4,5	5,6	6,7
3.8	4,5	4,5	5,6	6,7
3.9	4,5	4,5	5,6	6,8
4.0	4,5	4,5	5,6	6,8

Table 5.4 Worksheet for Example 5.17

i	p	f	q	σ	σ_L	ss	op	ol
1	0.5	30	50	10.7	7.6	12.5	27.5	77.5
2	0.3	18	30	7.0	4.9	8.1	17.1	47.1
3	0.2	12	20	5.1	3.6	5.9	11.9	31.9

i SKU, *p* portion of demand, *f* forecast, *q* plan quantity, σ standard deviation, σ_L lead-time standard deviation, *ss* safety stock, *op* order point, *ol* order level

$$
\begin{aligned}
f_i &= p_i \times F = \text{monthly forecast for } i \\
\sigma_i &= [F \times p_i(1 - p_i) + p_i^2 \times \sigma^2]^{0.5} = \text{standard deviation for i} \\
q_i &= p_i \times Q = \text{plan order quantity for i} \\
f_{Li} &= L \times f_i = \text{lead time forecast for i} \\
\sigma_{Li} &= \sqrt{L} \times \sigma_i = \text{lead-time standard deviation for i} \\
k &= \text{safety factor (from safety stock method)} \\
ss_i &= k \times \sigma_{Li} = \text{safety stock for i} \\
op_i &= f_{Li} + ss_i = \text{order point for i} \\
ol_i &= op_i + q_i = \text{order level for i}
\end{aligned}
$$

Each day, the on-hand, oh_i, and on-order, oo_i, for SKU i is scanned, and used in the following computations:

$$
\begin{aligned}
&\text{if } (oh_i + oo_i) > op_i: \quad buy = 0 \\
&\text{if } (oh_i + oo_i) \le op_i: \quad buy = ol_i - (oh_i + oo_i)
\end{aligned}
$$

Example 5.17 Assume a style good item has a monthly forecast of F = 60, standard deviation, σ = 20, lead-time, L = 0.5 months, plan order quantity, Q = 100, and the safety stock will be computed using the service level method with SL = 0.95 as the parameter. Note, when SL = 0.95, the safety factor is k = 1.65. Suppose, the style good item has N = 3 SKUs with the portion of demands at: 0.5, 0.3, 0.2, for i = 1 to 3, respectively. Table 5.4 lists the computation results to find the order point and order level for each of the three SKUs.

5.12 Summary

The order point and order level of an item are computed using the lead-time forecast, safety stock and plan order quantity. These are updated regularly, usually monthly. The on-hand plus on-order inventory for an item is maintained daily and is compared to the order point to determine when a new replenishment is needed. The inventory of each item is partitioned into the cycle stock, safety stock and the sum is the total stock. The relation between the partitions and the turnover ratio is described, along with the sensitivity with the percent fill, lead-time, coefficient of variation and month-in-buy. The way to control the inventory for a low demand item is described using the Poisson probability distribution. Replenishments for style good items with various sizes are also described.

Chapter 6
Distribution Control

6.1 Introduction

For convenience here, a network (NW) is defined when two or more stock holding facilities are connected with one entity. This is a distribution system with two or more locations, could be a system with distribution centers (DC), or two or more retail outlets. The individual stocking locations are here referred as locations. The goal is to control the inventory for the NW and for each location. Typically, once a month at the NW, forecasts, standard deviation of the forecast error, and the planned order quantity are generated. Also, at each location, the forecast, standard deviation and order quantity are also needed to compute the order point and order level. At each location, the on-hand and on-order are observed and compared to the order point and order level to determine if a location needs a replenishment buy, and if so, how much. Sometimes the location buys directly with the supplier, and other times, the NW buys for the total system.

When the NW buys the stock, and the replenish quantity arrives from the supplier, a way to distribute the stock to the locations is needed. This process is called allocation, with a goal to allot the stock to the locations in a fair-share manner. Another important process at the NW is called transfer. Transfer occurs periodically, perhaps weekly, where the stock status of each location is measured, and if a location is low and another is high, a transfer quantity is computed. Share is another process that takes place on occasion. One location is selected as the source, and each of the parts at the location are analyzed in a share manner as follows: If the inventory status of a part is high at the share location, and it is low at another location, a share quantity is computed, whereby the quantity is transferred accordingly. Limit buy is another process that sometimes occurs in the NW. The total buy for a supplier cannot exceed a given limit; could be in dollars, pieces, so forth. Discount buy is when the total buy, all parts, must reach a goal to qualify for a discount; could be for a supplier discount, for a truckload min limit, so forth.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_6

6.2 Network Forecasts, Standard Deviation and Order Quantity

Some computations on the NW operations are needed to control the inventory at the individual locations. This includes the total NW forecast of demands for the future months, and the corresponding standard deviation. To accomplish, the aggregate demand history from the prior N months of the NW is gathered, denoted as $D_1, \ldots,$ D_N where t = 1 is the oldest month and t = N is the most current month. This history is used to generate forecasts for future month τ, denoted as Fτ for τ = 1 to N`, where N` is the number of future months in the planning horizon. The estimate of the standard deviation of the monthly forecast is labeled as σ. Also needed on the NW is the plan order quantity, Q. In many situations, the NW orders a quantity Q from the supplier for all of the locations to share. More on this is given subsequently.

6.3 Location Percent Demand

Consider a network with M locations labeled as (i = 1, . . . , M) where each location holds stock to fill the customer demands as they arrive. Assume the cumulative demands from several prior months are available on the database for each location, denoted as $(d_1, \ldots, d_M)$ for locations 1 to M, respectively. With this information, the percent of demand by location is estimated as:

$$p_i = d_i / \sum_i d_i \quad i = 1 \text{ to } M$$

where

$$\sum_i p_i = 1.0.$$

6.4 Location Forecast and Standard Deviation

To control the flow of inventory at each location requires the forecast of demands by location for the coming months, and also the corresponding standard deviation of the forecast. The typical way to accomplish is by the top-down (TD) forecast method as summarized below.

Using the top-down method, the forecasts for location i becomes:

$$f_{i\tau} = p_i F \quad \tau = 1 \text{ to } N`$$

Using F to represent the average aggregate monthly forecast, and applying the results from Chap. 2 on the standard deviation for SKUs, the standard deviation for location i is obtained as follows:

$$\sigma_i = \left[F \times p_i(1 - p_i) + (p_i \times \sigma)^2\right]^{0.5}$$

6.5 Location Order Quantity

Since Q is the plan order quantity for the total NW, the fair share portion for location i becomes:

$$q_i = p_i Q$$

6.6 Location Safety Stock

Assuming the percent fill (PF) method is used to generate the safety stock, the data needed to find the safety stock for location i is the following:

PF = percent fill parameter
σ_i = standard deviation for location i
L = lead-time
q_i = order quantity for location i

The lead-time standard deviation is needed and is computed for location i as follows:

$$\sigma_{Li} = \sqrt{L}\sigma_i$$

Continuing in the calculations, the partial expectation for the location becomes:

$$E = (1 - PF)q_i/\sigma_{Li}$$

As described in the safety stock chapter, a Table 4.2 lookup using the partial expectation, E, finds the associated safety factor k, denoted here as k_i for location i. Thereby, the safety stock for location i is obtained by:

$$ss_i = k_i \sigma_{Li}$$

6.7 Location Order Point and Order Level

The order point and order level for location i are denoted as op_i and ol_i, respectively. These are computed as follows:

$$op_i = f_{Li} + ss_i$$

where f_{Li} is the lead time forecast for location i, and,

$$ol_i = op_i + q_i$$

6.8 Location Buy

The above computations are typically revised one time, at the start of the month. However, the buy decisions for a location are needed each day of the month. Using the notation, oh_i = current on-hand at i, and oo_i = current on-order at i, the amount to buy at location i is based on the following:

$$\begin{array}{ll} \text{If } (oh_i + oo_i) > op_i: & buy_i = 0 \\ \text{if } (oh_i + oo_i) \le op_i: & buy_i = ol_i - (oh_i + oo_i) \end{array}$$

Example 6.1 Suppose a distribution system with four locations (1, 2, 3, 4) where the percent fill safety stock method is applied with parameter PF = 0.95. The total data for a part (over all locations) is the following: F = 100 is the average monthly forecast, σ = 40 is the standard deviation of F, L = 2 months is the lead-time, and Q = 50 is the plan order quantity. The cumulative prior 6-month demands for the locations are: 400, 300, 200, 100, for locations 1–4, respectively. The computations to find the order point and order level for each location are summarized below.

Consider location i = 1 and the computations shown below:

$$\begin{aligned} p_1 &= 400/1000 = 0.40 \\ f_1 &= 0.40 \times 100 = 40 \\ \sigma_1 &= \left[100 \times 0.40(1 - 0.40) + (0.40 \times 40)^2\right]^{0.5} = 16.73 \\ \sigma_{Li} &= \sqrt{2} \times 16.73 = 23.65 \\ q_i &= 0.40 \times 50 = 20 \\ E_i &= (1 - 0.95)20/23.65 = 0.042 \\ k_1 &\approx 1.34 (\text{via Table 4.2}) \\ ss_1 &= 1.34 \times 23.65 = 31.7 \end{aligned}$$

$$\begin{aligned} f_{L1} &= 2 \times 40 = 80 \\ op_i &= 80 + 31.7 = 111.7 \\ ol_i &= 111.7 + 20 = 131.7 \end{aligned}$$

A summary of the computations for the four locations is given in Table 6.1.

Table 6.1 Prior 6-month demand, d, percent demand, p, forecast, f, standard deviation, σ, order quantity, q, safety stock, ss, order point, op, and order level, ol, for locations 1–4 of Example 6.1

Locations	d	p	f	σ	q	ss	op	ol
1	400	0.40	40	16.73	20	31.7	111.7	131.7
2	300	0.30	30	12.84	15	24.5	84.5	99.5
3	200	0.20	20	8.94	10	17.2	57.2	67.2
4	100	0.10	10	5.00	5	10.0	30.0	35.0

6.9 Allocation

When a new replenishment is called for an item in stock, a request for an order quantity is sent to the supplier. When the stock is ready at the supplier, the quantity, Q, is shipped to the stocking location. Sometimes the shipment goes only to the master location who distributes to the stocking locations. Other times, the shipment is sent directly from the supplier to the individual locations. In the former situation, after the master location receives the replenish quantity Q, an allocation algorithm determines how much stock to send to each location. In the latter situation, an algorithm is needed to instruct the supplier how to divvy up the replenish quantity, Q, among the receiving locations. In either event, the receiving stock is allotted to the locations in a fair-share manner, as described below.

For computational ease, the monthly forecasts over the planning horizon are assumed to be relatively flat. The data needed to determine the allocation is listed below:

Q = replenish quantity
i = 1 to M = the network stocking locations
oh_i = current on-hand at i
oo_i = current on-order at i
f_i = average monthly forecast at i
ss_i = safety stock at i

At the outset, all the locations in the NW are set as active and are included in an active set of locations. Subsequently, in step 4, a location can be dropped from the active set of locations. The allocation computations are the following:

1. Do steps 2–5 for all active locations, i.
2. $F = \Sigma f_i = \text{NW average monthly forecast}$
   $OHOO = \Sigma[oh_i + oo_i] = \text{NW on-hand plus on-order}$
   $SS = \Sigma ss_i = \text{NW safety stock}$
3. $MS = [OHOO - SS + Q]/F = \text{NW months-supply}$
4. $ms_i = [oh_i + oo_i - ss_i]/f_i = \text{current months-supply for location i}$
   if $ms_i > MS: \quad q_i = 0$ and drop location i from the active set
5. If any i dropped from the active set (in step 4), repeat steps 1–4.
   If none of the active locations are dropped, do the following for each active location:

$$f_{MS} = MS \times f_i = \text{MS forecast for location i}$$
$$q_i = [f_{MS} - (oh_i + oo_i - ss_i)] = \text{raw allocation quantity for i}$$

Note: $\Sigma q_i = Q$. The set of raw allocation quantities, $(q_1, \ldots, q_M)$ should be converted to comply with integers and package multiple quantities. The process must ensure that the sum of converted allocation quantities still sums to the aggregate receiving quantity, Q.

Example 6.2 Recall the part in Example 6.1 once more with average monthly forecasts, f, and safety stock, ss, as listed in Table 6.2. Also, suppose the current on-hand plus on-order inventory for each location, ohoo, are those listed in the table. The supplier notifies the distribution center that a replenish quantity of Q = 50 pieces is ready for shipment and requests how to allot to the four locations. At the foot of the table are the NW sums, F, SS and OHOO.

Applying the new replenishment with this data, the NW months supply is computed below:

$$MS = [120 - 84 + 50]/100 = 0.86$$

In a corresponding way, the months supply for each location, ms_i, are now computed. For location i = 1, say, the months supply becomes:

$$ms_1 = [50 - 32]/40 = 0.45$$

Table 6.3 lists the same for all four locations.

Because the months supply of location 4 is larger that the aggregate MS, location i = 4 already has sufficient stock and should not be included in the allotment of the new replenish stock. With this situation, the computations starts over excluding location i = 4.

Without location i = 4, the revised aggregate sums becomes: F = 90, SS = 74 and OHOO = 100, whereby, the revised months supply is the following:

$$MS = [100 - 74 + 50]/90 = 0.844$$

The location months supply, for locations (1, 2, 3) are the same as listed in Table 6.3. Note all are now below the revised MS = 0.844.

Table 6.2 Location, i, monthly forecast, f; safety stock, ss; on-hand plus on-order; ohoo; and corresponding sums, F, SS, and OHOO

i	f	ss	ohoo
1	40	32	50
2	30	25	35
3	20	17	15
4	10	10	20
Sum	100	84	120

Table 6.3 Months supply, ms, by location

i	ms
1	0.45
2	0.33
3	–0.10
4	1.00

Table 6.4 The raw allot quantities, q`, and integer quantities, q

i	q`	q
1	15.76	16
2	15.32	15
3	18.88	19
4	0.00	0
Sum	49.96	50

The raw allot quantities, q`, are now computed. For location i = 1, say, the following data: oh, oo, f_{MS}, and ss are obtained:

$$oh_1 + oo_1 = 50$$
$$f_{MS} = MS \times f_1 = 0.844 \times 40 = 33.76$$
$$ss_1 = 32$$

So now, the raw allot quantity for i = 1 is derived below:

$$q_1 = [f_{MS} - (oh + oo - ss)] = [33.76 - (50 + 0 - 32)] = 15.76$$

Table 6.4 lists the raw allot quantities, q`, and the corresponding integer quantities, q. Note, the sum of the raw quantities rounds to Q = 50 and the integer sum is Q = 50.

In the event, package multiple quantities are specified, the final location alllot quantities are rounded to conform accordingly. For example, if the package quantity were M = 5, say, the allot quantities would be set as: 15, 15, 20, 0, for locations 1–4, respectively.

6.10 Transfer

When one location is short on stock and another has excess, it is good practice to have the excess location ship some stock to the short location. This process is called transfer, and in many stock holding networks, when two or more stocking locations, a transfer routine is run 1–4 times a month, perhaps each Friday of the month. With each run of the system, one part at a time is processed, and for the part, each location is checked to see if it is short or is in excess. The locations short, are called receivers, and the locations in excess are called senders. In the event a part has at least one receiver and at least one sender, the quantity to transfer from sender(s) to receiver(s) is determined. The transfer computation follows.

The parameters are below:

$$msL = \text{parameter of months supply low}$$
$$msH = \text{parameter of months supply high}$$

The data for each location, i, is listed below:

$$
\begin{aligned}
i &= \text{location} \\
oh_i &= \text{on-hand} \\
oo_i &= \text{on-order} \\
f_i &= \text{average monthly forecast} \\
ss_i &= \text{safety stock}
\end{aligned}
$$

The month's supply for each location i are computed as below:

$$ms_i = [oh_i + oo_i - ss_i]/f_i$$

if $ms_i < msL$: location i is a receive candidate
if $ms_i > msH$: location i is a send candidate

For any location that is a receiver, an ideal raw receive quantity, xr, is needed. A way to find the quantity is by seeking 1 month of forecast above the months supply low limit as shown below:

$$
\begin{aligned}
f_{[msL+1]} &= (msL + 1) \times f_i \\
&= \text{forecast for } (msL + 1) \text{ months}
\end{aligned}
$$

and thereby, the ideal receive quantity for location i becomes:

$$xr_i = f_{[msL+1]} - (oh_i + oo_i - ss_i)$$

For a sender location, the ideal raw send quantity is also needed. This is obtained by computing 1 month of forecast below the months supply high limit as shown below:

$$
\begin{aligned}
f_{[msH-1]} &= (msH - 1) \times f_j \\
&= \text{forecast for } (msH - 1) \text{ months}
\end{aligned}
$$

whereby, the ideal send quantity for location j becomes:

$$xs_j = (oh_i + oo_i - ss_i) - f_{[msH-1]}$$

In the event, there is one receiver, i, and one sender, j, the receive and send quantities, xr_i and xs_j, respectively, are converted to the integer of the minimum as follows:

$$q = \text{integer}\left[\min\left(xr_i,\ xs_j\right)\right]$$

If the part is encased in a package with multiple units, the transfer quantities have to be set into the same multiple units. When two or more receivers and or senders, the transfer algorithm is more confounding, since the sum of integer receive quantities

Table 6.5 Location data of f = forecast, ss = safety stock, oh = on-hand, oo = on-order and ms = months supply

i	f	ss	oh	oo	ms
1	40	32	61	0	0.72
2	30	25	63	0	1.27
3	20	17	38	0	1.05
4	10	10	54	0	4.40

must equal the sum of integer send quantities. Also, the transfer algorithm should seek senders and receivers where the minimum transfer distance occurs between locations.

Example 6.3 Consider the part in Example 6.1 with four locations and the average monthly forecasts, f, and safety stock, ss, as computed in the earlier example. Assume the part is in review for a transfer and the current data by location are those listed in Table 6.5 as: forecast, f, safety stock, ss, on-hand, oh, on-order, oo, and months supply, ms. Each location's months supply, ms, is calculated in the usual way. Suppose also the transfer parameters are set as msL = 1.0 month, and msH = 4.0.

Because $ms_1 < msL$, location i = 1 is a candidate for a receiving location, and the ideal receive quantity is the following:

$$xr = f_{[msL+1]} - (oh_i + oo_i - ss_i) = 80 - (61 + 0 - 32) = 51$$

Note also where $ms_4 > msH$, and thereby location i = 4 is a candidate for a sender location. The ideal send amount is computed as below:

$$xs = (oh_i + oo_i - ss_i) - f_{[msH-1]} = (54 + 0 - 10) - 30 = 14$$

Locations i = 2 and 3 are not candidates as receivers or senders, and so, a possible transfer can be made since there is at least one receiver and one sender location. The transfer amount is computed as below:

$$q = \min(51,\ 14) = 14$$

Thereby, location i = 4 will send q = 14 pieces to location i = 1.

6.11 Share

On some occasions, the management has a need to clear some or all the inventory out of one of the stock locations and send the units to the other locations in the network. They desire to do so in a fair-share manner. For notation sake, the location chosen to send its stock is labeled as i_o, and all other locations are labeled as i. One parameter is needed to specify how much stock to remove from the sender location. The parameter is the following:

$$\text{msS} = \text{months supply for the send location}$$

To accomplish, the data needed by location is the following:

$i = 1$ to M identifies the locations in the network
oh_i = current on-hand at i
oo_i = current on-order at i
f_i = average monthly forecast at i
ss_i = safety stock at i

As before, for computational ease, the monthly forecasts over the planning horizon are assumed to be relatively flat.

Using the parameter, msS, the send amount, denoted as Q, coming from location i_o is obtained as follows:

$$Q = (oh_{io} + oo_{io} - ss_{io}) - msS \times f_{io}$$

Note Q must be set as an integer.

At the outset, all the locations, not i_o, in the NW are set as active and are included in an active set of locations. Subsequently, in step 4 below, a location can be dropped from the active set of locations. The computations are the following:

1. Do steps 2–5 for all active locations, i, in the active set of locations.
2. $F = \Sigma f_i$ = aggregate average monthly forecast
   $OHOO = \Sigma[oh_i + oo_i]$ = aggregate on-hand plus on-order
   $SS = \Sigma ss_i$ = aggregate safety stock
3. $MS = [OHOO - SS + Q]/F$ = aggregate months-supply
4. $ms_i = [oh_i + oo_i - ss_i]/f_i$ = current months-supply for location i
   if $ms_i \geq MS$: $q_i = 0$ and drop location i from the active set.
5. If any i dropped from the active set (in step 4), repeat steps 1–4.
   If none of the active locations are dropped, do the following for each active location:

$$q_i = [MS \times f_i - (oh_i + oo_i - ss_i)] \quad = \text{raw share quantity for i.}$$

Note: $\Sigma q_i = Q$. The set of raw share quantities, $(q_1, \ldots, q_M)$ should be converted to integers and in package multiple quantities. The process must ensure that the sum of converted share quantities still sums to the aggregate send quantity, Q.

Example 6.4 Recall the part in Example 6.1 once more with average monthly forecasts, f, and safety stock, ss, as listed in Table 6.4. Also, suppose the current on-hand plus on-order inventory for each location, ohoo, are those listed in the table. Management wants to apply the share routine with location $i_o = 4$ as the send location, and the share parameter set as $msS = 2.0$.

With this information, the quantity to share from location $i_o = 4$ is computed as below:

Table 6.6 Location, i; monthly forecast, f; safety stock, ss; on-hand, oh; and on-order, oo

i	f	ss	oh	oo
1	40	32	80	0
2	30	25	65	0
3	20	17	40	0
4	10	10	55	0

Table 6.7 Receiving locations, i, with f, ss, oh and oo, along with their sums: F, SS, OH and OO

i	f	ss	oh	oo
1	40	32	80	0
2	30	25	65	0
3	20	17	40	0
4	–	–	–	–
Sum	90	74	185	0

$$Q = (oh_4 + oo_4 - ss_4) - msS \times f_4$$

$$= (55 - 10) - 2 \times 10 = 25$$

The month's supply for the aggregate of the three receive locations (1, 2, 3) is calculated below:

$$\begin{aligned} MS &= [Q + OH + OO - SS]/F \\ &= [25 + 185 + 0 - 74]/90 \\ &= 1.51 \end{aligned}$$

So now the month's supply for each of the three receiving locations is computed as below:

$$ms_i = [oh_i + oo_i - ss_i]/f_i$$

whereby,

$$\begin{aligned} ms_1 &= [80 + 0 - 32]/40 = 1.20 \\ ms_2 &= [65 + 0 - 25]/30 = 1.33 \\ ms_3 &= [40 + 0 - 17]/20 = 1.15 \end{aligned}$$

Since all are less than the $MS = 1.51$, the three locations remain active and will receive some stock for $i_o = 4$.

The amount of stock allotted to location i is computed by:

$$q_i = MS \times f_i - [oh_i + oo_i - ss_i]$$

Thereby,

$$q_1 = 1.51 \times 40 - [80 = 0 - 32] = 12.4$$
$$q_2 = 1.51 \times 30 - [65 + 0 - 25] = 5.3$$
$$q_3 = 1.51 \times 20 - [40 + 0 - 17] = 7.2$$

Note, the sum rounds to Q = 25. The true receive amounts would naturally be rounded to integers so that the sum is still Q = 25. Hence, the receive amounts become, (13, 5, 7), for locations 1, 2, 3, respectively. In the event the units of the part are package in a multiple quantity, the receive amounts would have to be converted to the nearest multiple, accordingly. For example, if the multiple was five, the share amounts would be set to: (15, 5, 5), respectively.

6.12 Sourcing

When a customer order is not entirely filled at a location, a search for another location to fill the remainder of the order is sometimes sought. This process is called sourcing. Not all customer orders are eligible by way of sourcing. The OEM generally establishes a set of rules to determine which orders to source. In a service parts NW, hazardous parts are not sourced, while orders that are classified as emergency or critical are eligible, and so forth. Upon accepting an unfilled customer order for sourcing, the locations that have enough stock to fill the remained of the order are sought in a systematic manner that is pre-established by the OEM. Each location in the NW is assigned a sourcing sequence. It specifies how to search for a location to fill the remainder of the customer order. Each sequence is arranged by distance between locations. The general rule is to find the closest location that can fill the total remainder of the unfilled order.

Table 6.8 shows how the sourcing sequence might be set for the five locations in a NW. Note, if location 1 cannot fill a customer order entirely, the search of another location to fill the remainder of the order begins. This is by first seeking location 3, then 5, 2 and 4 in that sequence. In the event none can help, the remaining quantity of the order is placed on backorder.

Table 6.8 Sourcing sequence for the locations in the NW

Location	Sourcing sequence
1	3, 5, 2, 4
2	4, 3, 5, 1
3	2, 4, 5, 1
4	3, 5, 2, 1
5	1, 3, 4, 2

6.13 Limit Buy

Suppose N parts are in a cluster, could be from a single supplier, where all are purchased at the same time and the management wants to limit the combined cost of the aggregate buy quantity to a specified amount called a max limit and denoted as, CQ_o. For notation, the parts are denoted as i = 1 to N, and the data available on each part is listed below. Note the raw buy quantity for each part is the quantity computed previously without regard to the max limit. For subsequent use, the method described here is denoted as the Max-Buy-Algorithm (MXBA). This method is seeking to limit the cost of the total buy, but could be simply converted to any other measure as: units, weight, barrels, volume.

i = 1 to N
c = cost per unit
f = average monthly forecast
ss = safety stock
oh = on-hand
oo = on-order
q` = raw buy quantity

The aggregate cost of the 1-month forecast for all of the N parts is the following:

$$\text{CF`} = \sum_{i=1}^{N} c_i f_i$$

The aggregate cost of the raw buy for all of the N parts is the following:

$$\text{CQ`} = \sum_{i=1}^{N} c_i q`_i$$

If CQ` $\leq$ $CQ_o$, the raw buy quantities $q`_i$, i = 1 to N, require no change.

If CQ` $>$ CQ_o, some or all of the buy quantities need to be reduced.

In the event of the latter situation, the following steps takes place. At the outset, all the parts are classified as active.

1. Below are cost aggregates for all active parts:
   $\text{CQ} = \sum c_i q`_i$ = raw buy for active parts
   Note: (CQ` − CQ) = raw buy for non-active parts
   $\text{CF} = \sum c_i f_i$ = forecast for active parts
   $\text{COHOO} = \sum c_i (oh_i + oo_i)$ = on-hand plus on-order for active parts
   $\text{CSS} = \sum c_i ss_i$ = safety stock for active parts
   MS = [CQ_o − (CQ` − CQ) + COHOO − CSS]/CF = aggregate month's supply at cost
2. For each active part, the month's supply is computed as below:

Table 6.9 The columns are the following: $c = $ cost per unit, $f = 1$ month forecast, $ss = $ safety stock, $oh = $ on-hand, $oo = $ on-order and $q' = $ raw order quantity

i	c	f	ss	oh	oo	q‘
1	100	4	3	5	0	2
2	90	21	8	11	0	40
3	50	33	10	22	0	50
4	30	55	20	26	0	70
5	10	42	22	18	0	100

$ms_i = [q`_i + oh_i + oo_i - ss_i]/f_i$
If $ms_i < MS$: change the part to non-active

3. If any part in this cycle is changed to non-active, repeat steps 1 and 2.
   Else, compute the revised buy quantity for each part, i, as follows:
   $q_i = MS \times f_i - (oh_i + oo_i - ss_i)$ for active parts
   $q_i = q_i`$ for non-active parts

Example 6.6 Suppose a retailer buys from a supplier who offers five parts, $i = 1$ to 5. The buy quantities on the parts is placed together in a combined buy, and the management wants to restrict the combined cost to $CQ_o = \$8000$. The data for the parts are listed in Table 6.9.

Cycle 1: The aggregate cost elements become the following:

$$CF` = 6010$$
$$CSS = 2340$$
$$COHOO = 3550$$
$$CQ = 8900$$

Because CQ > CQo, some of the raw buy quantities need to be reduced. The plan aggregate months supply is obtained below:

$$MS_o = [8000 + 3550 - 2340]/6010 = 1.53$$

The month's supply for part $i = 1$, is computed by:

$$ms_1 = [qi + ohi + ooi - ssi]/fi$$
$$= [5 + 2 + 0 - 3]/4 = 1.00$$

In the same way,

$$ms_2 = 2.04$$
$$ms_3 = 0.98$$
$$ms_4 = 1.38$$
$$ms_5 = 2.28$$

Since ms_1, ms_3 and $ms_4 < MS_o$, parts $i = 1$, 3 and 4 are set to non-active parts and the final buy for $i = 1$, 3 and 4 will be the same as the raw buy. Also because the other two parts have a months supply larger than MS_o, they remain as active parts and their buy quantity will be reduced.

Cycle 2: The process continues with two active parts and three non-active. The aggregate cost sums for the active parts are listed below:

$$CF = 2310$$
$$CSS = 940$$
$$COHOO = 1070$$
$$CQ` = 4600$$

Note, (8900 – 4600) = 4300 = aggregate buy cost for all non-active parts.

The aggregate months supply becomes:

$$MSo = [8000 - 4300 + 1070 - 940]/2310 = 1.66$$

The revised buy quantity for parts i = 2 and 5 are:

$$q_2 = 1.66 \times 21 - [11 - 8] = 31.9$$
$$q_5 = 1.66 \times 42 - [18 - 22] = 73.7$$

Recall, $q_1 = 2$, $q_3 = 50$ and $q_4 = 70$.

So now, the cost of the aggregate buy becomes:

$$\sum c_i q_i = \$8000$$

with the difference from $8000 due to rounding. Note also, the final buy quantities would be set to integers as: 2, 32, 50, 70, 74, for parts 1–5, respectively. In the event that package multiple quantities are also involved, further refinements to the buy quantities would be needed.

6.14 Discount Buy

Suppose N parts are in a cluster, could be from a single supplier, where all are purchased at the same time and the management wants to satisfy the combined load (pieces, weight, volume) of the aggregate buy quantity to a specified amount, WQ_o. For notation, the parts are denoted as i = 1 to N, and the data available on each part is listed below. Note the raw buy quantity for each part is the quantity computed previously without regard to the max limit.

$$\begin{aligned}
&\text{i} = 1 \ \text{to} \ \text{N} \\
&\text{w} = \text{load per unit} \\
&\text{f} = \text{average monthly forecast} \\
&\text{ss} = \text{safety stock} \\
&\text{oh} = \text{on-hand} \\
&\text{oo} = \text{on-order} \\
&\text{q`} = \text{raw buy quantity}
\end{aligned}$$

The aggregate load of the forecast for all of the N parts is the following:

$$\text{WF`} = \sum_{i=1}^{N} w_i f_i$$

The aggregate load of the raw buy for all of the N parts is the following:

$$\text{WQ`} = \sum_{i=1}^{N} w_i q`_i$$

If WQ` $\geq$ WQ_o, apply the Max-Buy-Algorithm (MXBA) with $c_i = w_i$ for part i.
If WQ` $<$ WQ_o, apply the Min-Buy-Algorithm (MNBA) described below.

In the event of the latter situation, the following steps takes place. At the outset, all the parts are classified as active.

1. Below are load aggregates for all active parts:
   $\text{WQ} = \sum w_i q`_i = \text{raw buy for active parts}$
   Note: (WQ`−WQ) = raw buy for non-active parts
   $\text{WF} = \sum w f_i = \text{forecast for active parts}$
   $\text{WOHOO} = \sum w_i (oh_i + oo_i) = \text{on-hand plus on-order for active parts}$
   $\text{WSS} = \sum w_i ss_i = \text{safety stock for active parts}$
   MS = [WQ_o−(WQ`−WQ)+WOHOO−WSS]/WF = aggregate month's supply at cost
2. For each active part, the month's supply is computed as below:

$$\begin{aligned}
&ms_i = [q`_i + oh_i + oo_i - ss_i]/f_i \\
&\text{If } ms_i > \text{MS: change the part to non-active}
\end{aligned}$$

3. If any part in this cycle changed to non-active, repeat steps 1 and 2.
   Else, compute the revised buy quantity for each part, i, as follows:

$$\begin{aligned}
&q_i = \text{MS} \times f_i - (oh_i + oo_i - ss_i) \ \text{for active parts} \\
&q_i = q_i` \ \text{for non-active parts}
\end{aligned}$$

Example 6.6 Suppose a retailer buys from a supplier who offers five parts, i = 1 to 5. The buy quantities on the five parts is placed together in a combined buy, and the management wants to restrict the combined cost to WQ_o = 9000. The data for the five parts are listed in Table 6.10.

Table 6.10 The columns are: w = load, f = 1 month forecast, ss = safety stock, oh = on-hand, oo = on-order and q‘ = raw order quantity

i	w	f	ss	oh	oo	q‘
1	10	100	30	80	0	100
2	30	70	20	100	0	100
3	50	40	20	50	0	10
4	80	30	10	20	0	60
5	100	10	5	5	0	20

Cycle 1: The aggregate load elements become the following:

$$\begin{aligned} &WF` = 8500 \\ &WSS = 3200 \\ &WOHOO = 8400 \\ &WQ` = 11300 \end{aligned}$$

Note, WQ` = load all parts, WQ = load active parts, (WQ` − WQ) = load non-active parts, and at he outset, WQ = WQ`.

Because WQ’ > WQ_o, some of the raw buy quantities need to be reduced. The aggregate month’s supply is obtained below:

$$MS_o = [9000 + 8400 - 3200]/8500 = 1.67$$

The month’s supply for part i = 1, is computed by:

$$\begin{aligned} ms_1 &= [q_i + oh_i + oo_i - ss_i]/f_i \\ &= [100 + 80 - 30]/100 = 1.50 \end{aligned}$$

In the same way,

$$\begin{aligned} ms_2 &= 2.57 \\ ms_3 &= 1.00 \\ ms_4 &= 2.33 \\ ms_5 &= 2.00 \end{aligned}$$

Since ms_1 and ms_3 are less than MS_o, the buy quantities for parts 1 and 3 remain as is, and those of parts 2, 4 and 5 need to be converted.

Cycle 2: The process continues with four active parts and one non-active. The aggregate cost sums are listed below:

$$\begin{aligned} &WF = 5500 \\ &WSS = 1900 \\ &WOHOO = 5100 \\ &WQ = 9800 \end{aligned}$$

Note, 9800 is the aggregate load for all active parts, and 1500 is the load for the non-active parts.

The aggregate month's supply becomes:

$$MSo = [(9000 - 1500) + 5100 - 1900]/5500 = 1.945$$

The revised buy quantity for part i = 2, 4 and 5 are:

$$q_2 = 1.945 \times 70 - [100 - 20] = 56.2$$
$$q_4 = 1.945 \times 70 - [20 - 10] = 48.4$$
$$q_5 = 1.945 \times 10 - [5 - 5] = 19.4$$

Recall, $q_1 = 100$, $q_3 = 10$.

So now, the load of the aggregate buy becomes:

$$\sum w_i q_i = \$9000$$

with the difference from 9000 due to rounding. Note also, the final buy quantities would be set to integers as: 100, 56, 10, 48, 19, for parts 1–5, respectively. In the event that package multiple quantities are also involved, further refinements to the buy quantities would be needed.

6.15 Summary

A network is here defines as an inventory system with two or more stocking locations. An important management goal is to have inventory at each location that fills the customer demands and accomplishes with a minimum NW cost. Forecasts, standard deviations and planned order quantities are computed for the aggregate of the NW and for each location. When the inventory is low at a location, a replenish buy quantity is called. Sometimes the location orders from the supplier, and other times, the NW places a replenish order. In the latter case, when stock comes in from the supplier, an allocation algorithm determines how to allot the stock to the individual locations. On a once a week basis, a scan of the inventory at each location determines if one location is low and another is high on a part's inventory, so that, a transfer of stock can take place. Share is an option that occurs when one location distributes its high inventory with the other locations. Limit buy is when the NW total buy on all parts of a supplier cannot exceed a given limit. Discount buy is when the NW buy on all parts to a supplier must meet a goal to gain a discount of some type.

Chapter 7
Manufacturing Control

7.1 Introduction

The manufacturing plant consists of an assortment of raw materials, components, machines of various type, and an variety of skilled workers, with the goal to produce goods to a higher level, and some to finished-good-items. The management's task of coordinating all this activity is difficult indeed. A first concern of the management is to periodically generate a production plan for the coming planning horizon. This plan pertains to the aggregate of all items in the plant and yields the volume of production for all. The plan depends on the type of items to produce: make-to-stock, make-to-order, or a combination of make-to-stock and to order. For each finished-good-item, a master production plan is generated to coordinate the schedule that satisfies the inventory status and customer demands. This schedule also yields the available-to-promise quantity that is a vital tool to the sales force. The concept of raw load and level load by future time periods for various production centers in the plant are described. To ensure the schedule of all items are doable, a rough-cut capacity planning analysis is calculated for each of the production centers in the plant. When the capacity does not meet the load, adjustments are required. For every item to be produced, a bill-of-material is used to identify each of the parts and components that are required in the build. From here, a materials requirement planning set of computations determines the build schedule for every item.

7.2 Manufacturing Pioneer

In 1797, Eli Whitney, mass-produced muskets for the U.S. Government. He introduced standard parts, whereby all the parts of the muskets were produced in advance and with the same engineering tolerance. Thus each part could be used in any musket. With his success, manufacturers began using machine tools and jigs

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_7

to produce the standard parts and components with specified tolerances. This ability allowed the manufacturers to hire less-skilled workers in their production operations, and at the same time achieve an increase in the units produced.

7.3 Production Planning

The production plan is a strategic tool developed by the top management at a manufacturing facility to provide an overall directive to the operations staff on the resource needed (manpower, raw materials, hours) over the near future months. The plan is based on input from various sources in the enterprise: production, finance, marketing and accounting. The units of the plan differ depending on the products in production; this could be pieces, dollars, tons, barrels, gallons, so forth. The production staff is responsible to carryout the plan. Although the plan spans several months, the first future month is the primary target, and the subsequent months are projections needed for advance preparations. Oftentimes, the plan is revisited and revised each month as new data and conditions arise. For convenience, the description that follows assumes the facility operates on a monthly basis over a 6 month planning horizon.

Many versions of a production plan are possible, depending on the needs of the manufacturing facility. Three basic models are described here: the make-to-stock production plan, the make-to-order production plan, and the mix of make-to-stock and make-to-order production plan.

7.3.1 Make-to-Stock

Make-to-Stock (MTS) defines a system where a manufacturing facility produces products based on demand forecasts, and stocks the products in a warehouse or distribution center awaiting subsequent customer demands. The accuracy of the forecast is a significant key to preventing excess inventory and out-of-stock conditions. The products could be various models of washing machines, stoves, refrigerators, and so forth, where each unit of a model is built exactly the same.

The notation and data are the following:

$t = 0$ is the current month
$t = 1$ to 6 are the planning months
F_t = monthly aggregate forecasts for all products at $t = 1$ to 6
I_0 = current aggregate inventory
I_t = management's desire of ending inventory for $t = 1$ to 6
P_t = production plan for $t = 1$ to 6

The units of the forecast, inventory and production must be the same depending on what is meaningful with the products in aggregate terms. The production plan is generated with the forecast and inventory data and stipulates the aggregate production output for each of the planning months. The production plan for a month represents the aggregate compilation of all the products produced in the plant. The operation management uses the production plan as a directive to plan their activities with little deviation above or below.

Example 7.1 Assume a plant produces N items on a make-to-stock basis and the planning horizon is 6 months. The sales department provides aggregate monthly forecasts on all N items as listed in the table below. The current aggregate inventory is obtained from the inventory data files and measures at $I_0 = 200$. The top management specifies how much inventory is ideal for the future months, and this is listed in the table for the six planning horizon months.

Months (t)	0	1	2	3	4	5	6
Forecast of sales (F)		100	100	120	120	120	110
Desired ending inventory (I)	200	180	180	190	200	200	190

The production plan for each month is computed as shown below, as well as the corresponding results for the 6 months.

$$\begin{aligned}
P_t &= F_t + [I_t - I_{t-1}] \\
P_1 &= 100 + [180 - 200] = 80 \\
P_2 &= 100 + [180 - 180] = 100 \\
P_3 &= 120 + [190 - 180] = 130 \\
P_4 &= 120 + [200 - 190] = 130 \\
P_5 &= 120 + [200 - 200] = 120 \\
P_6 &= 110 + [190 - 200] = 100
\end{aligned}$$

A summary of the monthly production plan that conforms with the forecasts and desired ending inventories is listed below:

Months (t)	0	1	2	3	4	5	6
Production plan (P)		80	100	130	130	120	100

7.3.2 *Make-to-Order*

Make-to-Order (MTO) is a production system that depends on orders already available from customers, and usually for expensive items like: tractors, trucks, boats and aircraft. When a customer order comes in, it is placed on an order-board as a backlog with a desired due date. The customer order specifies the exact features wanted on the finished good item, and often no two items are the same. The production facility is obliged to produce the items on or before the due dates.

The notation and data to develop the production plan are the following:

$t = 0$ is the current month
$t = 1$ to 6 are the planning months
F_t = monthly forecasts of coming sales for $t = 1$ to 6
B_0 = current aggregate backlog
B_t = management's desire of ending backlog for $t = 1$ to 6
P_t = production plan for $t = 1$ to 6

The units of the forecast, inventory and production must be the same, and whatever units are meaningful with the products in aggregate terms. The production plan is generated with the forecast and inventory data and stipulates the aggregate production output for each of the planning months. The production plan for a month represents the aggregate compilation of all the products produced in the plant. The operation management uses the production plan as a directive to plan their activities with little deviation above or below.

Example 7.2 An assembly plant produces trucks on a make-to-order basis and the planning horizon is 6 months. Each truck assembled at the plant is for a particular customer who specifies the features wanted and the desired delivery date. The sales department provides aggregate monthly forecasts on all trucks as listed in the table below. The current aggregate backlog is obtained from the order-board and is currently set at $B_0 = 6000$. The top management specifies how many backlog units is ideal for the future months, and this is listed in the table for the six planning horizon months.

Months (t)	0	1	2	3	4	5	6
Forecast of sales (F)		4000	4200	4400	4400	4600	4600
Desired ending inventory (B)	6000	6200	6200	6400	6400	6600	6600

The production plan for each month is computed as shown below, and the corresponding results for the 6 months.

$$P_t = F_t - [B_t - B_{t-1}]$$
$$P_1 = 4000 - [6200 - 6000] = 3800$$
$$P_2 = 4200 - [6200 - 6200] = 4200$$
$$P_3 = 4400 - [6400 - 6200] = 4200$$
$$P_4 = 4400 - [6400 - 6400] = 4400$$
$$P_5 = 4600 - [6600 - 6400] = 4400$$
$$P_6 = 4600 - [6600 - 6600] = 4600$$

A summary of the monthly production plan that conforms with the forecasts and desired ending backlogs is listed below. In this situation, the production plan is in units and gives the count on the number of trucks to produce on the assembly line for a given month. In the first future month of 22 workdays, the average number of trucks to assemble per day is $3800/22 = 172.7$.

Months (t)	0	1	2	3	4	5	6
Production plan (P)		3800	4200	4200	4400	4400	4600

7.3.3 Make-to-Stock and Make-to-Order

The combination of make-to-stock and make-to-order (MTS-MTO) is a production system where some products are standard and others are specialty made for the customers. The standard units are produced to replenish the inventory in the warehouse and awaiting customer demands, while the specialty made are built to satisfy the specific customer orders. In this way, the plant has a combination of make-to-stock and make-to-order system of production. This could be an auto manufacturing plant where some of the units produced are standard and others have customer orders with specified due dates and features assigned. The standard units are produced to replenish the inventory, and the specialty units are produced from the backlog on the order board.

The notation and data to develop the production plan are the following:

$t = 0$ is the current month
$t = 1$ to 6 are the planning months
F_t = monthly forecasts of coming sales for $t = 1$ to 6
I_0 = current aggregate inventory
I_t = management's desire of ending inventory for $t = 1$ to 6
B_0 = current aggregate backlog
B_t = management's desire of ending backlog for $t = 1$ to 6
P_t = production plan for $t = 1$ to 6

The units of the forecast, inventory, backlog and production must be the same, and whatever is meaningful with the products in aggregate terms. The production plan is generated with the forecast, inventory, and backlog data and stipulates the aggregate production output for each of the planning months. The production plan for a month represents the aggregate compilation of all the products produced in the plant. The operation management uses the production plan as a directive to plan their activities with little deviation above or below.

Example 7.3 An assembly plant produces automobiles on a combination of make-to-stock and make-to-order basis and the planning horizon is 6 months. Some of the autos are standard for inventory replenishment and others are specialty with customer features and due dates specified. Forecasts are provided for the standard autos and the backlog on the order board monitors the customer orders as they arrive. The current warehouse inventory is set at $I_o = 16{,}000$ and the current backlog is $B_o = 3000$. The forecast for the standard plus specialty units is listed in the table below as F_t for t = 1 to 6. The management gives the desired inventory and backlog levels for the future 6 months as I_t and B_t, respectively.

Months (t)	0	1	2	3	4	5	6
Forecast of sales (F)		10,000	11,000	11,000	10,000	10,000	9000
Desired ending inventory (I)	16,000	17,000	17,000	16,000	16,000	15,000	15,000
Desired ending backlog (B)	3000	3500	4000	4000	3500	3000	3000

The production plan for each month of the future 6 months is computed as shown below.

$$
\begin{aligned}
P_t &= F_t + [I_t - I_{t-1}] - [B_t - B_{t-1}] \\
P_1 &= 10{,}000 + [17{,}000 - 16{,}000] - [3500 - 3000] = 10{,}500 \\
P_2 &= 11{,}000 + [17{,}000 - 17{,}000] - [4000 - 3500] = 10{,}500 \\
P_3 &= 11{,}000 + [16{,}000 - 17{,}000] - [4000 - 4000] = 10{,}000 \\
P_4 &= 10{,}000 + [16{,}000 - 16{,}000] - [3500 - 4000] = 10{,}500 \\
P_5 &= 10{,}000 + [15{,}000 - 16{,}000] - [3000 - 3500] = 9500 \\
P_6 &= 9000 + [15{,}000 - 15{,}000] - [3000 - 3000] = 9000
\end{aligned}
$$

A summary of the monthly production plan that conforms with the forecasts, inventory and desired ending backlogs is listed below. In this situation, the production plan is in units and gives the count of the number of autos to produce on two assembly lines for a given month. In the first future month of 22 workdays, the average number of autos to assemble per day is 10,500/22 = 477.

Months (t)	0	1	2	3	4	5	6
Production plan (P)		10,500	10,500	10,000	10,500	9500	9000

7.4 Master Production Schedule

For each end-item in production at the plant, a master production schedule (MPS) is generated for the planning horizon, usually for 1–2 months and in weekly time durations. This MPS could be revised each week to account for the latest information. The MPS is a directive to the manufacturing management on when to begin production on each end-item and how many to produce. The MPS takes into consideration many important factors: the future actual demands, the forecast for production demands, the forecast for service demands, the current available inventory, the safety stock, lead time, ideal lot size, and multiple quantity. The MPS seeks to have the minimum stock needed to meet all demands. The MPS is also an important tool to the sales management since it gives vital information on when they can give promise delivery dates to customers on the end-item. The MPS is seldom altered during the first few weeks, but as conditions change, with weekly revisions of the MPS, revisions could occur in the latter weeks of the schedule.

The notation and data needed to generate the MPS is listed below:

$$
\begin{aligned}
&t = 0 \text{ is the current week} \\
&t = 1 \text{ to } 6 \text{ are the future planning weeks} \\
&L = \text{lead time (weeks)} \\
&AI = \text{current available inventory} \\
&SS = \text{safety stock} \\
&Q = \text{desired lot size} \\
&M = \text{multiple quantity} \\
&AD_t = \text{actual demand for week t} \\
&FP_t = \text{forecast of production needs for week t} \\
&FS_t = \text{forecast of service needs for week t} \\
&R_t = [AD_t + FP_t + FS_t] \text{ is requirements for week t}
\end{aligned}
$$

Example 7.4 Consider a part where the MPS is needed for the coming 6 weeks. The current data is the following: available inventory is $AI = 255$, safety stock is $SS = 30$, lead-time is $L = 1$-week, ideal lot size is $Q = 200$, and the multiple quantity is $M = 50$. Further data for each week of the 6-week horizon is the following: AD = actual demand, FP = forecast of demand for production, FS = forecast of demand for service, whereby R = projected requirements for future weeks 1–6. For clarity sake, the available inventory is the amount of stock available prior to the receipt of the MPS quantity, and the on-hand, OH, is the inventory that includes the MPS quantity. Since the lead-time is 1-week, the MPS will start 1 week prior to the due data.

The data for the part is listed below:

$$
\begin{aligned}
&AI = 255 \\
&SS = 30 \\
&L = 1 \text{ week} \\
&Q = 200 \\
&M = 50
\end{aligned}
$$

w	0	1	2	3	4	5	6
AD		85	70	55	35	28	10
FP		15	30	45	65	72	100
FS		20	0	0	0	0	20
R		120	100	100	100	100	130
OH	255						

7.4.1 MPS Computations

The MPS computations take place 1 week at a time. The computations give the MPS that is due for the week. At the start of the week, the available inventory (AI) is computed as the on-hand minus the requirements. If the AI is at or below the safety stock, an MPS is due for the week. The MPS will need to begin a lead-time prior to the MPS due week. The on-hand of the week becomes the available inventory plus the corresponding MPS.

$$
\begin{aligned}
&\text{At } w = 0: \quad OH_0 = AI \\
&\text{At each } w: \quad AI = OH_{w-1} - R_w \\
&\qquad \text{if } AI > SS: MPS_w = 0 \\
&\qquad \text{if } AI \leq SS: MPS_w = Q - AI(\text{round to } M) \\
&\qquad OH_w = AI + MPS_w
\end{aligned}
$$

7.4.2 MPS for Weeks 1–6

Below shows how the MPS is computed for weeks 1–6.

$w = 0$ $\quad OH_0 = 255$

$w = 1$
$$
\begin{aligned}
&AI = 255 - 120 = 135 \\
&\text{since } AI > 30: MPS_1 = 0 \\
&OH_1 = 135 + 0 = 135
\end{aligned}
$$

$w = 2$
$$
\begin{aligned}
&AI = 135 - 100 = 35 \\
&\text{since } AI > 30: MPS_2 = 0 \\
&OH_2 = 35 + 0 = 35
\end{aligned}
$$

$w = 3$
$$
\begin{aligned}
&AI = 35 - 100 = -65 \\
&\text{since } AI \leq 30: MPS_3 = 200 - (-65) = 265(\text{round})250 \\
&OH_3 = -65 + 250 = 185
\end{aligned}
$$

$w = 4$
$$
\begin{aligned}
&AI = 185 - 100 = 85 \\
&\text{since } AI > 30: MPS_4 = 0 \\
&OH_4 = 85 + 0 = 85
\end{aligned}
$$

$w = 5$
$$
\begin{aligned}
&AI = 85 - 100 = -15 \\
&\text{since } AI \leq 30: MPS_5 = 200 - (-15) = 215(\text{round})200 \\
&OH_5 = -15 + 200 = 185
\end{aligned}
$$

$w = 6$
$$
\begin{aligned}
&AI = 185 - 130 = 55 \\
&\text{since } AI > 30: MPS_6 = 0 \\
&OH_6 = 55 + 0 = 55
\end{aligned}
$$

7.4.3 MPS Worksheet

Below is a worksheet on the MPS data and calculations so far. The table gives the MPS due weeks and the associated MPS start weeks. Because the lead-time is 1 week, the start week is 1 week prior to the corresponding due week.

w	0	1	2	3	4	5	6
AD		85	70	55	35	28	10
FP		15	30	45	65	72	100
FS		20	0	0	0	0	20
R		120	100	100	100	100	130
OH	255	135	35	185	85	185	55
MPS_{due}		0	0	250	0	200	0
MPS_{start}		0	250	0	200	0	0

7.4.4 Available to Promise Computations

The available to promise (AP) is an important output from the MPS computations since it allows the sales force to quote delivery dates to their customers. The AP quantities only pertain when current or new inventory is available. These happen during the current week, and any future week when an MPS is due. The AP does not interfere with the actual demand (AD), but only with for forecast demands for production and service. The AP algorithm is listed below:

At $w = 0$ $\quad MPS_0 = OH_0$

At each w: $\quad$ If $MPS_w > 0$: $\quad w_1 = w$ & $w_2 =$ week of next $MPS > 0$

$$AP_w = MPS_w - \sum_{w=w_1}^{w_2-1} AD_w$$

IF $MPS_w = 0$: $\quad AP_w = 0$

7.4.5 Available to Promise for Weeks 0–6

Below shows how the available to promise is computed for weeks 0–6.

$w = 0$ $\quad MPS_0 = 255$

since $MPS_0 > 0$: $\quad w_1 = 0$ & $w_2 = 3$

$AP_0 = 255 - [0 + 85 + 70] = 100$

$w = 1$ $\quad$ since $MPS_1 > 0$: $\quad AP_1 = 0$

$w = 2$ $\quad$ since $MPS_2 > 0$: $\quad AP_2 = 0$

$w = 3$ $\quad$ since $MPS_3 = 250$: $\quad w_1 = 3$ & $w_2 = 5$

$AP_3 = 250 - [55 + 35] = 160$

Table 7.1 MPS worksheet for Example 7.1

w	0	1	2	3	4	5	6
AD		85	70	55	35	28	10
FP		15	30	45	65	72	100
FS		20	0	0	0	0	20
R		120	100	100	100	100	130
OH	255	135	35	185	85	185	55
MPS_{due}		0	0	250	0	200	0
MPS_{start}		0	250	0	200	0	0
AP	100	0	0	160	0	162	0

w = 4 since $MPS_4 = 0$: $AP_4 = 0$

w = 5 since $MPS_5 = 200$: $w_1 = 5 \ \& \ w_2 > 6$

$AP_5 = 200 - [28 + 10] = 162$

w = 6 since $MPS_6 = 0$: $AP_6 = 0$

7.4.6 MPS and AP Worksheet

The input data and the final results from the MPS and AP computations are listed in Table 7.1. An MPS is needed in weeks 3 and 5, and AP is quoted for weeks, 0, 3 and 5.

7.5 Rough Cut Capacity Planning

Recall, the MPS by product is based on the total requirements over the planning horizon, and does not consider the resources (labor hours, machining, materials), needed to carryout this schedule. Each of the products are competing with other products that also have a schedule from MPS and may use the same resources. For a given resource, the rough-cut capacity plan (RCCP) determines if the schedules on all the competing products conform to the capacity of the resources. If a conflict is detected, the resource may call for the need of overtime, third-party help, or a shift in the schedules. The goal of RCCP is to ensure the plant can meet the product requirements and the plant can operate in an efficient way.

Example 7.5 A machine shop in the XYZ Corporation processes four components: 111, 122, 158 and 167. The shop typically has ten workers who each are available 40 h a week. Thus, the shop has a capacity of 400 man-hours per week. The MPS of scheduled units for the four components over the coming 6 weeks, and the man-hours needed per unit, HU, are listed in Table 7.2.

In this situation, the load is in units of man-hours, MH, and Table 7.3 gives the load per component and per week, along with the sum load on the shop. Also listed is the sum load and the comparison with the capacity for each of the 6 weeks. The

Table 7.2 Man hours per unit, HU, and hours needed by component and week

Component	HU	Weeks					
		1	2	3	4	5	6
111	0.4	0	250	0	200	0	0
122	0.5	100	100	100	100	100	100
158	0.8	300	400	200	0	500	400
167	0.3	200	200	200	200	200	200

Table 7.3 Load computations for parts 111, 122, 158 and 167

Component	HU		1	2	3	4	5	6
111	0.4	Units	0	250	0	200	0	0
		MH	0	100	0	80	0	0
122	0.5	Units	100	100	100	100	100	100
		MH	50	50	50	50	50	50
158	0.8	Units	300	400	200	0	500	400
		MH	240	320	160	0	400	320
167	0.3	Units	200	200	200	200	200	200
		MH	60	60	60	60	60	60
Sum		MH	350	530	270	190	510	430
		Capacity	400	400	400	400	400	400
Variance		MH	50	−130	130	210	−110	−30
		Cumulative	50	−80	50	260	150	120

variance is the weekly difference between the load and the capacity. Note, the shop is short 130 man-hours in week 2, 110 in week 5, and 30 h in week 6. The cumulative variance over the 6 weeks is also listed, and shows 80 short man-hours in week 2 and all other weeks in excess of capacity. The cumulative variance shows the shop requires overtime work in weeks 1 and/or 2, or the use of temporary manpower in weeks 1 and/or 2, or a change in the MPS schedule. Further, the shop has excess capacity and could shift some of the workforce to other duties in the plant.

7.6 Raw Load and Level Load Production Scheduling

For sake of simplicity, the description here is by way of an example. Suppose the ABC Corporation, with N = 127 products, runs their MPS system and the results are summarized in Table 7.4. For each part, i, M = multiple quantity, Lu = load per unit, w = weeks 0–6, where w = 0 is the current week, R = average requirement per week, Q = release quantity per week, and OH = ending on-hand per week. At the foot of the table is the sum of the load for weeks 1–6. For notation sake, the load

Table 7.4 Raw load for parts i = 1 to 127, M = multiple, Lu = load per unit, and w = weeks 0–6. R = requirement per week, Q = release quantity per week, and OH = ending on-hand per week

i	M	Lu	w	0	1	2	3	4	5	6
1	20	15	R		40	40	40	40	40	40
			Q		0	0	20	0	40	20
			OH	186	146	106	86	46	46	26
:										
127	10	5	R		25	25	25	25	25	25
			Q		0	20	20	0	30	0
			OH	194	169	164	159	134	139	114
Sum		Raw load			5491	6048	3737	3406	5980	5498

computed is here called the raw load. The weekly capacity for the plant is 5500, whereby the table shows a wide difference between the capacity and the raw load over the 6 future weeks. The goal of the level load algorithm is to even out the weekly loads so that they are closer to the capacity of 5500.

Note that the raw load is obtained from the aggregate of the MPS production schedules derived one product at a time; without regard to the schedules of the N products. In the level load method, the production schedule for each product is generated simultaneously with the production schedules from all of the N products.

Below is the algorithm to generate a level load for the N products. At week w = 1, the following steps (1–5) are run for each of the N products. The steps may have to be run several times to reach the schedule needed for week w = 1. After week w = 1 is completed, the steps now are applied to week w = 2 in the same way. Then week w = 3, and so forth till w = 6.

The data needed to apply the level load method is listed below. First is the capacity of the plant for each week w. In most situations, the weekly capacities are the same. Second is the following data per product i: load per unit, average weekly requirements, and opening on-hand for the week.

Data:

$$C_w = \text{plant capacity for week } w = 1 \text{ to } 6$$
$$R_w = \text{requirement for week } w$$
$$Q_w = \text{quantity for week } w$$
$$i = \text{product index } (i = 1 \text{ to } N)$$
$$OHW_0 = \text{on-hand at week } w = 0 \text{ for } i$$
$$Mi = \text{multiple for } i$$
$$Lu_i = \text{load per unit for } i$$
$$R_i = \text{weekly requirements for } i$$

Each week, the on-hand for product i is revised as below:

$$OHW_w = OHW_{w-1} + Q_{w-1} - R_w$$

and thereby, the on-hand for i at week w is denoted as:

$$\mathrm{OH_i = OHW_w = opening\ \ inventory\ \ for\ \ i\ \ at\ \ week\ \ w}$$

For simplicity in the computational description here, the weekly requirements are set to an average, but in reality, they may vary by week. In processing each product, an index identifies whether the product requires further computations for the week. The index is denoted as a_i with settings of (0 or 1). The index is set to 1 when the product is active in week w, and is set to 0 when non-active.

The level load algorithm of steps 1–4 is below:

1. Do the following at the start of each week, w = 1–5.
   Set $a_i = 1$ and $Q_i = 0$ for each of the N products.
   Apply steps 2, 3, and 4 for the given week, w.
2. $\mathrm{MS} = \left[\mathrm{C_w} + \sum_{i=1}^{N} a_i Lu_i(OH_i)\right] / \sum_{i=1}^{N} a_i Lu_i\ \ R_i$ is months supply for week w
3. $\mathrm{ms_i = OH_i/R_i}$ is the current months supply for i
   if $\mathrm{ms_i > MS}$: set $a_i = 0$ and $Q_i = 0$
   if $\mathrm{ms_i \leq MS}$, compute x and Q_i as below:
   $\mathrm{x = [MS \times R_i - OH_i]}$ is the stock needed for product i to conform to MS
   $\mathrm{Q_i}$ = integer of $\mathrm{[(x + 0.5)/M_i] \times M_i}$ is the stock for i rounded to the multiple $\mathrm{M_i}$
4. If any i has its index changed to non-active in step 3, the cycle (2–3) needs to be repeated.
   Else, apply Q_i i = 1 to N for week w, and proceed to the next week.

Table 7.5 lists the schedule results that are obtained upon applying the level load algorithm. Note the sum of the weekly loads is close to the weekly capacities of 5500. The load per week w is computed from:

$$\sum_{i=1}^{127} Lu_i Q_i$$

Table 7.5 Level load for parts i = 1 to 127, M = multiple, Lu = load per unit, and w = weeks 0–6. R = requirement per week, Q = release quantity per week, and OH = ending on-hand per week

i	M	Lu	w	0	1	2	3	4	5	6
1	20	15	R		40	40	40	40	40	40
			Q		0	0	0	20	40	40
			OH	186	146	106	66	46	46	46
:										
:										
:										
127	10	5	R		25	25	25	25	25	25
			Q		0	0	30	30	0	0
			OH	194	169	144	149	154	129	104
Sum		Level load			5472	5502	5512	5485	5512	5484

7.7 Bill of Materials

A bill-of-materials (BOM) is list of all the parts, components, and materials needed to produce one unit of an end-item, sometimes referred as a finished-good-item (fgi), and the quantity per unit. This is a handy way to identify the structure of each end-item, and is needed in the plants when planning the manufacturing process. The BOM is also often included in the operating instructions provided to customers when they purchase the item.

In Example 7.6 the end-item is X and the components and materials are denoted as A, B, C, d, e, f, g, and h. Four units of C are needed for one unit of X, and two units of f are needed for one unit of B. Figure 7.1 shows how the BOM is displayed in a graphical way, and Table 7.6 gives the same information in a tabular manner. The items in the BOM are sometimes referred to their tiers, whereby, X is in tier 0; A, B and C are in tier 1; and the others (d, e, f, g, h) are in tier 2. Tier 0 is the location of the end-item X, and tier 1 holds the items that are used to produce the item located in tier 0. In the same way, the items in tier 2 are needed to produce the items in tier 1.

Example 7.6 Graphical bill-of-materials for end-item X, with components (A, B, C), materials (d, e, f, g, h), and quantity per unit is in Fig. 7.1

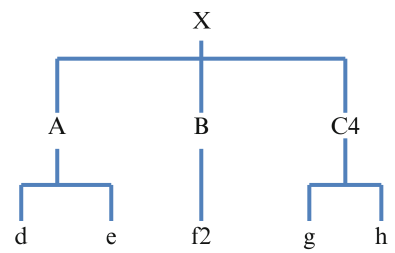


Fig. 7.1 Graphical bill-of-material for end-item X, with components (A, B, C) on the first tier, and materials (d, e, f, g, h) on the second tier; and where four units of C and two units of f are needed

Table 7.6 Tabular bill-of-material for end-item X, with components (A, B, C), raw materials (d, e, f, g, h), and the quantity per unit, q

Tier			
0	1	2	q
X			1
	A		1
		d	1
		e	1
	B		1
		f	2
	C		4
		g	1
		h	1

Table 7.7 Product structured table for end-item X, with components (A, B, C), materials (d, e, f, g, h); indented tiers; quantity per unit, q; source, lead-time (weeks); L, and multiple, M

Item	q	Source	L	M
X	1	S100	1	
A	1	S150	1	5
d	1	ABC	1	1
e	1	XYZ	2	1
B	1	S190	1	1
f	2	ABC	2	1
C	4	S200	2	20
g	1	AAA	2	5
h	1	XYZ	3	1

7.8 Product Structured Record

A product-structured record for an end-item contains the bill-of-materials data and any other information needed in the production process. This data is housed on the database for every end-item in the plant. Besides naming the components and materials for each end-item and the quantity of each, the record typically includes the description, the source, lead-time and lot size multiple-quantity.

Example 7.7 Table 7.7 shows how the product structured record would look using the same end-item from Example 7.6. Components A, B, C are indented to the right of end-item X indicating they are in the first tier, and materials, d, e, f, g, h are indented to the right of their respective components and they are in second tier. The example also lists the source for each item, identifying a shop in the plant (S100, S150, S190, S200), or a vendor (ABC, XYZ, AAA). Included also is the lead-time (weeks), L, of receive time to obtain each item, and the multiple quantity, M. For simplicity in this example, all of the lead times are in weekly buckets.

7.9 Materials Requirement Planning

Materials requirement planning (MRP) concerns the simultaneous inventory and production scheduling for each component and material included in the end-item BOM. The goal is to systematically arrange to achieve the production quantity of the end-item with the minimum throughput time and the least work-in-process inventory. The production quantity of the end-item often is generated from the MPS system. MRP determines the start and end dates for the end-item and for each component and material, including the corresponding source and quantity.

Two examples are shown. The first is when one end-item is in production. This example illustrates the basic needs and the methodology to achieve. The second is when two end-items are in the plant and there is some interaction between the components. The second example illustrates the method and computations that take place. The latter example can be extended to include a multiple number of

Table 7.8 Materials requirement planning for a production schedule of 100 units of end-item X with due date 05.22.15; quantity per unit, q; current on-hand, OH; multiple, M; raw quantity, Q`; rounded quantity, Q; lead-time weeks, L; start date; end date; and source

Item	q	OH	M	Q`	Q	L	Start	End	Source
X					100	1	051,815	052215	S100
A (X)	1	11	5	89	90	1	051115	051515	S150
B (X)	1	5	1	95	95	1	051115	051515	S190
C (X)	4	28	20	372	380	2	050415	051515	S200
d (A)	1	31	1	59	59	1	050415	050815	ABC
e (A)	1	7	1	83	83	2	042715	050815	XYZ
f (B)	2	17	1	173	173	2	042715	050815	ABC
g (C)	1	74	5	306	310	2	042015	050115	AAA
h (C)	1	53	1	327	327	3	041315	050115	XYZ

end-items, for which the computations become quite large, but are doable in a systematic manner. A common problem with MRP is that the method is highly depended on the integrity of the data of scheduling, on-hand, resource capacity, and lead-times.

Example 7.8 Suppose a plant calls for producing 100 units of end-item X on 05.22,15 (May 22, 2015) in shop S100, and is seeking the scheduling of all the items needed to accomplish. The results are summarized in Table 7.8. The table lists the items, and their immediate predecessor and quantity per unit, q. Also listed is the on-hand, OH, multiple quantity, M, and the raw production quantity, Q`. The final production quantity rounds Q` up to Q using M. The lead-time, L, is assumed in weekly durations, to allow for simplicity in description. The start date is when production begins, and is on a Monday for this example; and the end date is on a Friday. Finally is the source that identifies either a shop in the plant or a vendor.

Note where 100 units of X is due on 05.22.15 and the lead time is 1 week, indicating that the components A, B, C need to complete their production 1 week earlier, 05.15.15. The production of X begins on Monday 05.18.15 and ends of Friday 05.22.15. In the same way, all the start and end dates are listed. All the start dates are on Mondays, and the end dates are on Fridays. To complete the schedule for X on 05.22.15, item h, for example, has to begin its operation on Monday, 04.13.15, which is 5 weeks before X begins production.

Example 7.9 Assume the plant seeks the scheduling of two end-items, X and Y, where X is the same as the earlier example, and the product structured record for Y is listed in Table 7.9. Note Y requires items, A, D, d, e, g and k. Some of the items needed are common with X and some are unique.

The MRP results for the two end-items are summarized in Table 7.10. where the schedule calls for 100 units of X and 50 units of Y, all due on 05.22.15. The table lists all the items needed for the production process and the BOM precedence relationships are in parenthesis. Note, for example, item A is needed on X and Y. Item A must produce enough units (150) for both X and Y. Because A has

Table 7.9 Product structured record for end-item Y, with components (A, D), materials (d, e, g, k), indented tiers; quantity per unit, q; source, lead-time (weeks), L; and multiple, M

Item	q	Source	L	M
Y	1	S100	1	
A	1	S150	1	5
d	1	ABC	1	1
e	1	XYZ	2	1
D	4	S200	2	20
g	1	AAA	2	5
k	1	XYZ	3	1

Table 7.10 Materials requirement planning for a production schedule of 100 units of end-item X, 50 of end-item Y; with due date 05.22.15; quantity per unit, q; current on-hand, OH; multiple, M; raw quantity, Q\`; rounded quantity, Q; lead-time weeks, L; start date; end date; and source

Item	q	OH	M	Q\`	Q	L	Start	End	Source
X					100	1	051815	052215	S100
Y					50	1	051815	052215	S100
A (X,Y)	1	11	5	139	140	1	051115	051515	S150
B (X)	1	5	1	95	95	1	051115	051515	S190
C (X)	4	28	20	372	380	2	050415	051515	S200
D (Y)	4	69	20	131	140	2	050415	051515	S200
d (A)	1	31	1	109	109	1	050415	050815	ABC
e (A)	1	7	1	133	133	2	042715	050815	XYZ
f (B)	2	17	1	173	173	2	042715	050815	ABC
g (C,D)	1	74	5	446	450	2	042015	050115	AAA
h (C)	1	53	1	327	327	3	041315	050115	XYZ
k (D)	1	53	1	87	87	3	041315	050115	XYZ

11 units on-hand, they only need $(150-11)=139$ units. But because the multiple on A is $M=5$ pieces, the quantity to produce in $Q=140$. Note the end dates for X and Y is 05.22.15 (Friday), and since the lead times is 1 week, the start dates for both is Monday 05.18.15. The end date for A becomes Friday, 05.15.15. Note, items h and k begin their processing on 04.13.15 which is 5 weeks prior to the start dates for X and Y.

7.10 Summary

Controlling the production operations in a plant is not easy. To assist the management, a series of quantitative tools are used in the planning and control decisions of the manufacturing process in the plant. The production plan is used to determine the aggregate volume over the near future time horizon for all of the items in the plant. The plan depends on whether the products are: make-to-stock, make-to-order, or a combination of both. For each finished-good-item, a master production plan is computed. The output yields the schedule for the planning horizon, and the available-to-promise quantities. To ensure the sum of all schedules conforms to the capacity by production center, a series of rough-cut-capacity calculations are developed. For each component part to be produced, a bill-of-material is recorded, and this is a key ingredient in computing the material requirement plan for the component.

Chapter 8
Just-in-Time

8.1 Introduction

Just-in-time (JIT) is a philosophy of production based on the concept of adding value and eliminating waste. JIT and lean manufacturing have very similar goals. Value is added only by work performed on the product, and waste is anything other than a minimal amount of necessary resources—material, manpower, and the capital equipment—that is required for production, and does not add value to the product. The process called Kanban is a system where cards are used between send and receive stations in a way so that the stations produce only the necessary quantity of goods at the necessary time. JIT examples are presented where the components to a product are received from a supplier shortly after the customer order arrives. The relation between lean manufacturing and JIT is described. Smaller batch sizes are preferred in production, and this is accomplished as the setup time at a production process is reduced. The smaller the setup time, the lower the economic batch size. The safety stock to achieve a service level to the customer depends largely on the lead-time of replenish time from the supplier. As the lead-time becomes smaller, the amount of safety stock needed is lowered accordingly. The management should also seek to level the week-to-week aggregate production loads to avoid excess cost of overtime, backorders and outsourcing. When the finished good items are on a make-to-order basis, the strategy of postponement reduces the lead-time to the customers, and also eliminates much of the complication in the assembly.

8.2 Just-in-Time Pioneer

In the late 1940s, Taiichi Ohno of the Toyota Corporation in Japan, began implementing a cost-saving program that became known as the just-in-time method. With his method, parts are produced only as they are needed and are

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_8

delivered to the production line just prior to their use. His innovation helped Toyota rise from a bankrupt company to the third largest automobile maker in the world.

8.3 Kanban

Kanban is a production scheduling system that started in Japan shortly after World War II in the late 1940s. Kanban is the Japanese word for card. Taiichi Ohno is credited with the development and this occurred while he was working for Toyota. The system is one of the early applications in just-in-time inventory management, and has since been implemented successfully by many Japanese companies. The US started adopting the system in the 1970s. It is in contrary to the MRP production scheduling system.

The method uses two cards for each component: withdrawal card and production card. A card is assigned to each component in the plant and applies to two connected workstations. One workstation is the source producer of the component, and another workstation uses the component in its production process. When the withdrawal card is sent to the source workstation, it specifies the component number and quantity to send to the production workstation; and when the production card is sent to the source workstation, it specifies the component number and quantity to produce. One or more of such cards could be circulating at one time. In this way, the management controls the amount of inventory needed in the plant for each of the components. This system requires minimal inventory status revisions, production scheduling and computer systems.

Example 8.1 Consider workstations A and B, where A produces component X and serves as the source to B, who consumes X in its production process. Suppose three containers are in use to stock the units of X. In this situation, the card data is listed on the storage containers of component X where the quantity per container is set at ten units, say. Every unit that A produces is placed in a container and when the container is full, A stops production of X, and the container is sent to B. During the production process, B removes one unit of X at a time until the container is empty; thereupon the empty container is sent back to A. Receiving the empty container at A is a signal for A to produce ten more units of X. Should the forecast for X increase, the quantity per container or the number of containers increases; and if the forecast of X decreases, the quantity per container or the number of containers will decrease, accordingly. In this way the inventory and production process flows in harmony. Note, the updated inventory status of X is not needed, no paperwork is involved, and no computer application is run.

To illustrate how this system runs, consider the events ($t = 1$ to 10) listed below. The number of units of X in container i is listed as C_i.

At event $t = 1$, A has container 1 with zero units of X and B has containers 2 and 3 with ten units each. Since $C_1 = 0$, A starts producing X up to ten units. During the production process at B, the X units in container 2 are consumed one-at-a-time.

At $t = 2$, A has produced X and filled up container 1 with ten units ($C_1 = 10$). Also, B has consumed the ten units from container 2 ($C_2 = 0$), and has a full container 3 ($C_3 = 10$) in reserve.

At $t = 3$, the empty container 2 ($C_2 = 0$) is sent from B to A, which is a trigger for A to send the full container 1 ($C_1 = 10$) to B.

At $t = 4$, A receives the empty container 2 ($C_2 = 0$) which acts as a signal for A to produce ten more units of X. B begins using the X units in container 3 in its production process.

The events continue in the same way. Note where event $t = 10$ is the same as event $t = 1$, whereby a full cycle has transpired.

t

1. A has $C_1 = 0$
   B has $C_2 = 10$ and $C_3 = 10$
2. A has $C_1 = 10$
   B has $C_2 = 0$ and $C_3 = 10$
3. B sends $C_2 = 0$ to A
   A sends $C_1 = 10$ to B
4. A has $C_2 = 0$
   B has $C_3 = 10$ and $C_1 = 10$
5. A has $C_2 = 10$
   B has $C_3 = 0$ and $C_1 = 10$
6. B sends $C_3 = 0$ to A
   A sends $C_2 = 10$ to B
7. A has $C_3 = 0$
   B has $C_1 = 10$ and $C_2 = 10$
8. A has $C_3 = 10$
   B has $C_1 = 0$ and $C_2 = 10$
9. B sends $C_1 = 0$ to A
   A sends $C_3 = 10$ to B
10. A has $C_1 = 0$
   B has $C_2 = 10$ and $C_3 = 10$

8.4 Just-in-Time Manufacturers

The aim of JIT in manufacturing is to produce end items with low cost in inventory of components and with efficiency in meeting the customer needs. This process entails that the end item is not manufactured until a customer order arrives and the unit is paid for. The goal is to eliminate waste of all types, such as: minimizing component inventory, minimizing end item inventory, and minimizing the throughput-time from customer order to customer delivery. In essence, JIT seeks to have the input components delivered from the suppliers just prior to their use in production.

The JIT manufacturer often has a small number of suppliers, who are nearby and can deliver components with short notice and with quick delivery. Often, the suppliers are willing to hold the inventory at their location, reducing the need at the JIT manufacturer's plant. The JIT manufacturer also emphasizes the need for preventive maintenance on its machinery to minimize any downtime occurrence.

JIT works best when a good forecast of customer demands is available to project the flow of customer orders coming into the system. Also needed is high quality in the component parts that are used in the assembly of the end item, since there is little time to reorder another component in the event of a fault. The production operation requires the assembly tasks to be handled efficiently with no defects in parts or faults in machinery.

Some examples of JIT operations are described below.

8.4.1 Gift Baskets

A small entity receives orders on gift baskets that offer a list of items to include in the basket. The customer sends in an order with specific ingredients to include, and the entity has access to a handful of suppliers who quickly fulfills the call on the ingredients. The entity has no inventory and quickly receives the ingredients and places into a nice basket arrangement that satisfies the customer base. The customer order is filled efficiently within 1 day from order to delivery. This is a JIT application with little to no inventory at the entity.

8.4.2 Fast Food Grill

A fast food grill lists a menu with many items that require several ingredients (precooked eggs, bacon, and sausage, cheese, mustard, pickle, so on) to complete the order. When a customer arrives and places an order, a worker assembles the ingredients and quickly processes as needed to produce the item for the customer. The grill carries no inventory on the menu item, but only on the ingredients. In this way, the grill is applying JIT methods to service their customers.

8.4.3 Florist

A small shop florist can take orders over the phone or by the internet with specifications on the arrangement that is needed. Could be for celebrations, weddings, funerals, so on. The florist has little or no inventory and visits a floral mart who carries all the items needed. Upon purchase, the florist returns to his/her location and assembles a nice floral arrangement as specified by the customer. Within hours from receiving the order, the arrangement is completed and ready for delivery to the customer. The florist runs the business in a JIT manner.

8.4.4 Book Publisher

It is common today to order a book from a book publisher who currently has no inventory, but has a digital file and a book binding system that is capable of producing the book, with cover, on a print-on-demand basis. When a customer orders one or more books, the publisher triggers the system to produce the book (s) as called. With this capability, a book ordered today, can be delivered tomorrow. This practice is JIT to its fullest.

8.4.5 Computer Firms

Some computer firms have perfected the practice of JIT to the point so that when they receive an online customer order for a personal computer with specific features, the order triggers a call to the component supplier(s) to immediately deliver the components needed, whereby the assembly of the unit takes place shortly afterwards. With this practice, the firm has little to no inventory on the components, and yet is able to deliver the final product of the customer in 1 or 2 days.

8.5 Lean Manufacturing

Lean manufacturing's goal is to reduce all waste and create efficiency in production and be primarily responsive to what is needed now to produce the products for the current customer orders. Efficiency is needed in the people, equipment and the work stations of the production process. Any function in the manufacturing process that does not add value to the product is a waste and should be eliminated.

Henry Ford's adaption of an assembly line with a moving conveyor was a great addition to the production process. But, only one product was produced and when demand was low, the units were stored in inventory awaiting customer demands. Lean manufacturing emphasizes the pull-system where production is triggered by customer demands and inventory is held to a minimum. Lean manufacturing calls for flexibility so that the production process can fabricate a wide variety of units as specified by the customers with little conversion time of the machinery.

To apply these concepts smoothly requires zero defects in the materials used in the products, and also zero faults in the production machinery that does the processing. Preventive maintenance of the machinery is vital for the machinery, and statistical process control with six-sigma quality is the goal for the components.

8.5.1 Identify Waste

Lean manufacturing seeks to find any waste in the production process using a lean team of experts. The lean team lays out a map that follows the flow of the material and processing throughout the plant, including the movement of material and work-in-process till the units are completed and sent to shipping for delivery to the customer. The lean team seeks the root cause of any function(s) that cause waste, and are charged to find a solution on how to correct the fault.

8.5.2 Lean Tools

The goal of lean manufacturing is to run the production process with zero waste, high quality of products, and minimum cost. To achieve this, the lean team uses JIT as much as possible to minimize the need to carry inventory and to reduce the throughput time from the customer order to delivery. They also utilize the spirit of Kanban to control the production at the stations in the plant. Zero defects is a requirement for all components and material in use. Flexibility in machining that allows the system to switch the production of models frequently with minimal time from one batch of products to another is also a requirement. Standard parts are emphasized to simplify the switch from one model in production to another.

8.6 Setup Time Reduction

A key strategy in just-in-time is to reduce the setup time on each machining operation in the plant as the products are sequenced for processing in batch sizes, one after the other. The lower the setup time, the smaller economic batch size is needed, and the smaller the on-hand inventory.

8.6.1 Batch Size

In a typical manufacturing center, each machine is assigned a schedule that lists the parts and quantity to process for the coming time horizon. The quantity is in batch sizes, where one part is run after another, and small batch sizes are favored to reduce inventory. But to achieve small batch sizes that are economic, the setup time between batches must also be small. As the setup time decreases, the batch size becomes smaller, the lead-time to the customers reduces, and the work-in-process inventories drops. All of these are benefits to the firm.

8.6.2 Setup Time

When the machine completes production on a part, a setup process takes place before the next part can begin. The time duration between processing the two parts is called the setup time. The setup of the machine requires some adjustments to process the next part; and also requires the gathering of certain tools, fixtures and materials. All of these adjustments affect the setup time. The economic batch size for each part is directly related to the setup time.

8.7 Economic Batch Size

The economic batch size is directly related to the economic order quantity. Let Q = the economic batch size. To find the value of Q, the following ingredients are needed:

$\mathrm{A} = \mathrm{annual\ demand} = 12\ \mathrm{times\ the\ current\ month\ forecast}$
$\mathrm{Co} = \mathrm{cost\ per\ order} = (\mathrm{setup\ hours}) \times (\mathrm{cost\ per\ hour})$
$\mathrm{c} = \mathrm{cost\ per\ unit} = \mathrm{part\ standard\ cost\ per\ unit}$
$\mathrm{h} = \mathrm{annual\ holding\ rate} = \mathrm{annual\ (storage + interest\ costs)\ per\ dollar\ of\ inventory}$

Note, the cost per order, Co, is a function of the setup time. With the above data, using a bit of calculus yields the economic batch size as below:

$$\mathrm{Q} = \sqrt{(2ACo)/(ch)}$$

Should the cost per order decrease, a related reduction occurs in the batch size, Q. Recall Q is a function of Co; and if another cost per order is in use, say Co`, the corresponding order quantity becomes Q`. The relation between Q` and Q is below:

$$\mathrm{Q}` = \mathrm{Q}\sqrt{Co`/Co}$$

or,

Table 8.1 Change in (Q'/Q) with changes in (Co'/Co)

Co'/Co	1.00	0.90	0.80	0.70	0.60	0.50	0.40	0.30	0.20	0.10
Q'/Q	1.00	0.95	0.89	0.84	0.77	0.71	0.63	0.55	0.45	0.32

$$Q'/Q = \sqrt{Co'/Co}$$

Table 8.1 below shows how much the economic batch size decreases with each reduction in Co. When the cost per order is reduced in half, Co' = 0.5Co: Q' = 0.71Q, and so forth. The table shows the benefits that are gained when the setup time is reduced.

It behooves the management to find ways to lower the setup times. One way is to arrange the sequence of parts on a machining process to have a minimal adjustment in setup from part to part. Another way is to start as much of the setup change as possible while the prior part is still processing. This may require gathering the materials, components, tools and fixtures so they are available prior to when needed. Some companies prefer cellular arrangement of the machines so that the inventory, tools and fixtures are near each machine, and also where the operator has all of the items needed on the next part nearby.

8.8 Lead Time Reduction

The lead-time is the duration of time that begins when a stock location places an order with a supplier for new stock, and ends when the stock is received.

When the source is a manufacturing plant, any way to lower the throughput time is helpful in reducing the lead-time to customers. This includes: reducing setup times, processing times, transfer times, delay times, and defective components. In make-to-order products, a popular strategy is to postpone some of the assembly until the specific customer order comes in with the exact feature and option combinations.

In manufacturing plants, when the lead-time decreases, the benefits to the plant also increase in the following ways: less time for a fault to occur in the system; better on-time delivery; less inventory needs, less time for a change in customer orders; and an increase in customer service.

The supplier lead-time is important to distribution centers, retailers and dealers where the demands are not known in advance. In these locations, an extra layer of inventory, called safety stock, in needed to provide a desired level of service to the customers. The lead-time is used in the calculations to determine how much safety stock is required. The lower the lead-time, the less inventory.

8.9 Service Level Method of Safety Stock

Chapter 4 describes the four common methods that are used to compute the safety stock (month's supply, service level, percent fill and Lagrange) for the individual parts. In the following discussion, the service level method is selected to demonstrate how the lead-time affects the inventory for a part number. Below is a short review on the service level method.

Data required:

F = average one month forecast
σ = standard deviation for one month forecast error
L = lead time in months
Q = order quantity
SL = desired service level

The lead-time forecast and standard deviation are obtained as below:

$$F_L = L \times F$$
$$\sigma_L = \sqrt{L}\sigma$$

The parameter setting of the service level, SL, is the same as the cumulative distribution function, F(k), from the standard normal distribution where k is a particular value of the variable z. Using F(k), Table 4.2 is scanned to find the associated value of k, the safety factor. With k now known, the safety stock is computed in the following way:

$$SS = k\sigma_L = k\sqrt{L}\sigma$$

Recall, the order point and order level for the part are obtained using the safety stock, SS, lead-time forecast, F_L, and the order quantity, Q, as shown below:

$$OP = F_L + SS$$
$$OL = OP + Q$$

The average on-hand inventory is denoted as $\overline{OH}$, and average on-order inventory is $\overline{OO}$. These are obtained as follows:

$$\overline{OH} = Q/2 + SS$$
$$\overline{OO} = F_L$$

The turnover for the part is computed as below:

$$T/O = 12 \times F/\overline{OH}$$

8.10 Service Level Method in Months Supply

For analysis purpose, it is useful to convert the safety stock computation in a generic way that uses months of supply. The description follows.

Data required:

$$\begin{aligned} &\mathrm{F} = \text{average one month forecast} \\ &\sigma = \text{standard deviation for one month forecast error} \\ &\mathrm{L} = \text{lead time months} \\ &\mathrm{Q} = \text{order quantity} \end{aligned}$$

Data in months supply:

$$\begin{aligned} &\mathrm{cov} = \sigma/\mathrm{F} = \text{coefficient of variation} \\ &\mathrm{M} = \mathrm{Q}/\mathrm{F} = \text{months of order quantity} \\ &\mathrm{SL} = \text{desired service level} \end{aligned}$$

Computations in months supply

$$\begin{aligned} &\mathrm{SS}` = \mathrm{k}\sqrt{\mathrm{L}}\mathrm{cov} = \text{months of safety stock} \\ &\mathrm{M} = \mathrm{Q}/\mathrm{F} = \text{months of order quantity} \\ &\mathrm{OP}` = \mathrm{L} + \mathrm{k}\sqrt{L}\mathrm{cov} = \text{months of order point} \\ &\mathrm{OL}` = \mathrm{M} + \mathrm{L} + \mathrm{k}\sqrt{L}\mathrm{cov} = \text{months of order level} \\ &\mathrm{OH}` = M/2 + \mathrm{k}\sqrt{L}\mathrm{cov} = \text{average months of on-hand} \\ &\mathrm{OO}` = \mathrm{L} = \text{average months of on-order} \\ &\mathrm{T/O} = 12/\mathrm{OH}` = \text{turnover} \end{aligned}$$

Example 8.2 Suppose a part where M = 1.00, cov = 0.30 and SL = 0.95. Using Table 4.2, the safety factor becomes k = 1.65. Table 8.2 shows the measures of months of safety stock, order point, order level, average on-hand, average on-order and turnover when the lead-time varies from 4.00 to 0.25 months.

Table 8.2 Worksheet for Example 8.2, when SL = 0.95, L = lead time (months); SS` = months of safety stock, M` = months of order quantity, OP` = months order point, OL` = months of order level, OH` = average months of on-hand, OO` = average months of on-order, T/O = turnover

L	SS`	M`	OP`	OL`	OH`	OO`	T/O
4.00	0.99	1.00	4.99	5.99	1.49	4.00	8.05
3.00	0.86	1.00	3.86	4.86	1.36	3.00	8.82
2.00	0.69	1.00	2.69	3.69	1.19	2.00	10.08
1.00	0.50	1.00	1.50	2.50	1.00	1.00	12.00
0.75	0.43	1.00	1.18	2.18	0.93	0.75	12.90
0.50	0.35	1.00	0.85	1.85	0.85	0.50	14.12
0.25	0.25	1.00	0.50	1.50	0.75	0.25	16.00

Note when the lead-time of L = 2.00 months is cut in half to L' = 1.00, the average on-hand months change from 1.19 to 1.00, and the turnover from 10.08 to 12.00. If the lead-time is L = 1.00 months and is reduced to 0.50 months, the average months on-hand changes from 1.00 to 0.85 months and the turnover from 12.00 to 14.12, and so forth. The table demonstrates the advantages of reducing the lead-time with respect to the on-hand and turnover metrics in the inventory.

8.11 Aggregate Scheduling

An aggregate schedule pertains to the planning of the total production activities for a group of the items in the plant. The plan considers the future time horizon of weeks or months. The goal is to schedule the production so that the periodic fluctuations in the requirements are satisfied, and as much as possible seeks to minimize the associated costs. The management pursues a workforce for the group of items that is relatively constant over the planning horizon. As the requirements fluctuate over the time periods, additional production costs may be needed; such as overtime and outsourcing, and sometimes backorders occur. A cost analysis determines the efficiency of each plan.

Two example plans are described below where the planning horizon is in weekly buckets and are for the future 6 weeks. The notation of data and the computations are listed below:

Required data:

$$R_w = \text{requirements for week w with w} = 1 \text{ to } 6$$
$$OH_{E0} = \text{current (ending) oh-hand inventory at w} = 0$$

Plan scheduling data:

$$Q_{Sw} = \text{standard quantity at w}$$
$$Q_{OTw} = \text{overtime quantity at w}$$
$$Q_{OSw} = \text{outsource quantity at w}$$

Computed data:

$$OH_{Bw} = \text{beginning on-hand at w}$$
$$OH_{Ew} = \text{ending on-hand at w}$$
$$BO_w = \text{backorders at w}$$

Management scheduling decisions for weeks w = 1 to 6 are applied to satisfy the weekly requirements. Sometimes a buildup is needed in an earlier week to fill a subsequent week's requirement. The number of units for week w from the standard workforce is a constant for all weeks and is denoted as, Q_{sw}. Should a higher quantity be needed, overtime is called and the quantity is denoted as Q_{OTw}. This added quantity may not be sufficient since the overtime quantity often has a limit in size. If necessary, outsourcing may be called and the quantity is Q_{OSw}. In the event, the sum of the three production quantities for week w is still short of the requirements, the amount short for the week is put on backorder, BO_w, and is filled in a subsequent week.

Below shows how the on-hand at the beginning and ending of each week are computed, and also how the weekly backorders are determined.

$$\text{If } [R_w + BO_{w-1}] \le [OH_{Bw} + Q_{Sw} + Q_{OTw} + Q_{oSw}]:$$
$$OH_{Ew} = [OH_{Bw} + Q_{Sw} + Q_{OTw} + Q_{OSw}] - [R_w + BO_{w-1}]$$
$$BO_w = 0$$

$$\text{If } [R_w + BO_{w-1}] > [OH_{Bw} + Q_{Sw} + Q_{OTw} + Q_{oSw}]:$$
$$OH_{Ew} = 0$$
$$BO_w = [R_w + BO_{w-1}] - [OH_{Bw} + Q_{Sw} + Q_{OTw} + Q_{OSw}]$$

$$OH_{Bw+1} = OH_{Ew}$$

Example 8.3 Suppose a plant where the planning horizon is 6 weeks and the weekly requirements are those listed in Table 8.3. The table also lists the production schedule for each of the 6 weeks. For convenience, the schedule is called Plan 1, and an analysis of this plan follows.

The plan begins by calling for a standard workforce that can produce 30 units per week. Overtime is called in weeks 3, 4 and 5, and the overtime limit for a week is 6. Outsourcing of 20 units is also called for weeks 3 and 4. Since there is still a short quantity for week w = 4, a backorder of size two is recorded for the week. There also is a shortage in week w = 5 of four units, and thereby, the backorders swell to six units. All the backorders are filled in week w = 6.

The right-hand column of Table 8.3 is the 6-weeks sums and these are listed below:

Table 8.3 Plan 1 schedule for 6 week requirements, R; standard quantity, Q_S; overtime quantity, Q_{OT}; outsource quantity, Q_{OS}; begin on-hand, OH_B; end on-hand, OH_E; and backorders, BO

Weeks	0	1	2	3	4	5	6	Sum
R		20	24	60	80	40	20	244
Q_S		30	30	30	30	30	30	180
Q_{OT}		0	0	6	6	6	0	18
Q_{OS}		0	0	20	20	0	0	40
OH_B		10	20	26	22	0	0	78
OH_E	10	20	26	22	0	0	4	72
BO		0	0	0	2	6	0	8

$$\sum Q_S = 180$$
$$\sum Q_{OT} = 18$$
$$\sum Q_{OS} = 40$$
$$\sum OH_B = 78$$
$$\sum OH_E = 72$$
$$\sum BO = 8$$

The average sum on-hand is denoted as ΣOH_A and is obtained as below:

$$\sum OH_A = \left[\sum OH_B + \sum OH_E\right]/2 = 75$$

Below are the unit costs for the production, backorders and inventory:

Standard labor cost per unit	\$100
Overtime rate per unit	\$150
Outsource rate per unit	\$200
Backorder rate per unit	\$80
Inventory rate per unit	\$20

Finally, the costs for the sum of weeks 1–6 are below:

Standard:	$\Sigma Q_S \times 100 = 180 \times 100 = \$18,000$
Overtime:	$\Sigma Q_{OT} \times 150 = 18 \times 150 = 2700$
Outsource:	$\Sigma Q_{OS} \times 200 = 40 \times 200 = 8000$
Backorder:	$\Sigma BO \times 80 = 8 \times 80 = 640$
Inventory:	$\Sigma OH_A \times 20 = 75 \times 20 = 1500$

The total cost for 6-weeks is: \$30,840.

Example 8.4 Consider the same requirement data as listed in Example 8.3, but now the management is considering another plan, called Plan 2.

The standard workforce is increased to reduce the use of overtime and outsourcing. The standard quantity is now 43 units per week, and the overtime and outsourcing quantity is zero for the 6 weeks. Backorders are also reduced to two units over the 6-week duration. The right-hand column gives the sums for the 6 weeks (Table 8.4).

Table 8.4 Plan 2 schedule for 6 week requirements, R; standard quantity, Q_S; overtime quantity, Q_{OT}; outsource quantity, Q_{OS}; begin on-hand, OH_B; end on-hand, OH_E; and backorders, BO

Weeks	0	1	2	3	4	5	6	Sum
R		20	24	60	80	40	20	244
Q_S		43	43	43	43	43	43	258
Q_{OT}		0	0	0	0	0	0	0
Q_{OS}		0	0	0	0	0	0	0
OH_B		10	33	52	35	0	1	131
OH_E	10	33	52	35	0	1	24	145
BO		0	0	0	2	0	0	2

The unit costs are the same as the prior example, and the sum of the 6-week costs is computed below. Note, the average on-hand is 138 pieces. The total costs for Plan 2 is $28,720.

$$\begin{array}{ll} \text{Standard:} & 258 \times 100 = \$25,800 \\ \text{Overtime:} & 0 \times 150 = 0 \\ \text{Outsource:} & 0 \times 200 = 0 \\ \text{Backorder:} & 2 \times 80 = 160 \\ \text{Inventory:} & 138 \times 20 = 2760 \end{array}$$

8.12 Level Scheduling

Note the costs for Plan 2 is near $2000 below the same for Plan 1, indicating it is the favored plan. Another important advantage of Plan 2 is that the production schedule is constant for all of the 6-weeks. The constant schedule is also ideal throughout the plant and with the outside sources who are the suppliers of the components and materials needed in the production process. The steady schedule allows an even inflow of components and raw materials to meet the production.

8.13 Postponement Strategy

Postponement is a strategy that can take place when the finished-good-items (fgi) are of the make-to-order type. In make-to-order, the OEM offers the items in various combinations of features and options, such as in trucks, automobiles, construction equipment, and farm vehicles. The customer selects the specific option per feature for the individual order. This is a production process where each item going down the assembly line is partially unique because of the feature option combinations.

8.13.1 Make-to-Order

An order board is maintained that includes all the customer orders that are in a backlog status. Each customer order has a promised delivery date and the specified combination of feature and options. N is the number of units the assembly line is scheduled to produce each day. N customer orders are selected for assembly several days prior to the assembly date. The orders selected are those nearest to the due delivery date, and when all of its required parts and components are available for assembly. The sequence of the units going down the line is generated in advance in a way to allow a smooth flow of work by station. On the assembly day, the sequence is posted at each station along the line with every customer order identifying the specific options and parts that are required at the station. Each station must be stocked in advance with the proper parts for the combination of options required.

No doubt, make-to-order production requires coordination in the plant with the parts and components needed, with the instructions, and with the assembly process. As the synchronization in the system increases, the lead-time increases.

8.13.2 Postponement

When postponement is applied in the fullest way, no order board is needed, and as the customer orders come in with the combination of features and options, the unit for the customer is selected from a pool of standard (bland) units that have most of the assembly already done. The assembly for the specific options is finalized, and the finished-good-item is delivered to the customer. The lead-time to the customers is greatly reduced and the synchronization of the system is minimized.

To accomplish this system, at the plant, the basic units are assembled in a standard form without any of the options included. This way, the assembly line is run as a single-model line, where all units going down the line are the same. This is much easier on the workforce and simpler to manage than the make-to-order line. The finished products are sent to a holding location awaiting the future customer orders with the exact options listed. This holding location must be stocked with the variety of components and parts required to fulfill the variety of options. Also needed is a team of operators who are trained to accomplish the final assembly. A final test of the unit is also required to ensure all is working properly.

Sometimes postponement cannot be employed fully as described above. This happens often with the color of the vehicles Forecasts are used to estimate the color proportions that the customers favor. In the assembly process, the units are produced with the forecast mix of colors. These assembled units are sent to the warehouse waiting for the customer orders with specification wanted, including the color. When a customer order comes in, a unit from the warehouse with the requested color is selected as the unit, and the other specifications are finalized for the customer accordingly.

Example 8.5 Suppose a plant with make-to-order items that has eight features and the following option combinations:

feature	options
1	3
2	5
3	2
4	6
5	3
6	4
7	2
8	5

In the make-to-order plant, the number of option combinations becomes:

$$3 \times 5 \times 2 \times 6 \times 3 \times 4 \times 2 \times 5 = 21,600.$$

In the postponement system, the number of parts that have to be available at the warehouse is the following;

$$3 + 5 + 2 + 6 + 3 + 4 + 2 + 5 = 30.$$

The postponement strategy has several advantages:

- The lead-time to the customer is reduced significantly.
- The assembly at the plant is simplified.
- If the component suppliers are nearby and the warehouse orders the components when the customer order arrives, the warehouse has minimal component inventory.

8.14 Summary

In a production facility, just-in-time is a notion that seeks to reduce the time, materials and resources needed to satisfy the demands on its products. Kanban is a system where work-stations in a plant produce only the quantity of goods that is necessary. JIT examples are presented where the components of a final product are ordered from the suppliers just after a customer order is received. The relation with lean manufacturing is also described. Inventory is reduced as the batch size is lowered, and this is accomplished as the setup time at the work stations decrease. When the supplier lead-time is lowered, the stocking facility needs less safety stock to accomplish the level of service desired for the customers. It behooves the plant to generate aggregate schedules where the load on all products is relatively the same from 1 week to the next. This avoids undue costs of confusion, overtime, backorders and outsourcing. When make-to-order finished-good-items are produced, the concept of postponement eliminates complications in assembly and reduces the lead-time to the customers.

Chapter 9
Assembly

9.1 Introduction

A variety of assembly lines are applied in industry. For the smaller type of products, the assembly lines are often of the single model type, where one model of the product is produced on the line. The work elements to assemble one unit are gathered, and set in a precedence diagram showing the feasible work relations between the elements. The number of operators needed is computed and the work elements are assigned to each operator in a fair-share way. This latter function is called line balancing.

For larger products that are stocked in a warehouse or distribution center waiting for subsequent customer demands, (washing machines, refrigerators, so forth), the assembly is called a mixed model make to stock line. Two or more models of the product are assembled on the line at the same time. The plan schedule to produce the units is stated by the management, and the total time to complete all the work elements is tallied. The number of operators needed to conform to the schedule is computed, and the work elements are then assigned to the operators, via line balancing. Because two or more models are mixed on the line, another task of the management is to determine the sequencing of the models down the line.

On the higher price items, (trucks, tractors, so on), the units produced on the line each have a variety of features and options assigned by the customers. This is a make-to-order line. Because of this, each unit going down the line is unique. Each day, a different schedule of units, called jobs, is schedule to be produced. The work elements are assigned (via line balancing) to the operators in a way to satisfy an average day of production. The sequencing of the units down the line is arranged so the operator times are efficient as the units roll down the line from one unit in the sequence to the next unit.

In the latter two assembly lines described, robots are often scattered on the line and perform major tasks of the units, like in the paint booth, or in lifting of heavy

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_9

components as takes place in the body shop of the assembly line. The line balancing and sequencing tasks are still in need to accommodate the robots.

9.2 Assembly Pioneer

In 1902, Henry Ford founded the Ford Motor Company and started to build various models of vehicles. In 1909, he introduced the assembly line with moving conveyer, whereby the workers did not have to move from their stations. This novelty increased the production from a handful of units a day to 1000 per day. By 1923, the assembly operation increased to 1,800,000 per year.

9.3 Assembly

To produce one unit of a product on an assembly line, all of the parts and components on the bill-of-material are gathered, and are installed in the unit one at a time until the product is completed, whereby it becomes a finished good item. A series of station operators are assigned to complete all of the tasks required on the units. The assembly line is a work area that is set up to allow the units to flow from operator to operator along production stations where each operator performs one or more assigned tasks on each unit and at the last operator, the unit becomes the finished good item. A variety of lines are in use and the more common are the following: single model, mixed model make-to-stock, and mixed model make-to-order. Planning and control of the lines is not easy. Various quantitative tools are used to assist the management in the control of the lines, and these are explained below.

For brevity sake, the number of work elements used in these examples is smaller than most real situations, but adequate in size to allow the reader to follow the calculations, and sufficiently large enough to illustrate the methodology. The same is true with the number of parts in the assembly process. To assemble an outboard motor requires 500 parts including the nuts and bolts; an automobile requires over 2000 parts.

9.4 Single Model Assembly

Single model assembly is when the units in production are all the same, such as a specified model of a home coffee maker, or in the assembly of a lawn mower model.

9.4.1 Data

The data to plan and control the line is listed below. This includes Ne work elements labeled as e = 1 to Ne; the work element time, denoted as t_e; the immediate predecessor elements, p_e; and any part installed by element e, h_e. Also needed in planning is the number of units to produce, N, and the number of operators (work stations), n, on the line.

$$
\begin{aligned}
Ne &= \text{number of work elements} \\
e &= \text{elements } (e = 1 \text{ to } Ne) \\
t_e &= \text{time for element e} \\
p_e &= \text{predecessor(s) to e} \\
h_e &= \text{part installed by e} \\
N &= \text{number of units to produce} \\
n &= \text{number of operators (stations)}
\end{aligned}
$$

9.4.2 Computations

Below shows the early computations to apply using the data provided. First is the sum of the time to complete one unit of product, Σt_e, called the work content time. Second, the average work time per operator on the line is measured and is denoted as $\overline{c}$, and is computed as below:

$$
\begin{aligned}
\Sigma t_e &= \text{work content time} \\
\overline{c} &= \Sigma t_e / n = \text{average time per operator}
\end{aligned}
$$

9.4.3 Line Balance

The next step is for the line foreman to assign the Ne work elements to the n operators where the operator times are relatively close to the average time, $\overline{c}$, and where none of the predecessor elements are violated. This process is called line balancing. The times assigned to each operator is denoted as c_i for i =1 to n. Note the sum of the operator times is the same as the work content time.

$$
\begin{aligned}
i &= \text{operator number } (i = 1 \text{ to } n) \\
c_i &= \text{assigned time for operator i} \\
\Sigma c_i &= \Sigma t_e
\end{aligned}
$$

9.4.4 Line Measures

The maximum of the operator times is denoted as c, and this acts as the bottleneck of the line, since the units cannot be produced faster than the largest operator time. Thereby, c measures the time between units coming off the end of the line. The total time to produce N units can now be measured using N and c, and is labeled as T. An efficiency measure for the line assignments is called the line efficiency and is the ratio of $\overline{c}$ over c. The closer the ratio is to 1.00 (100 %), the better. The three measure are listed below:

$$c = \max(c_1, \ldots, c_n) = \text{cycle time}$$
$$T = c \times N = \text{timespan to complete N units}$$
$$E = \overline{c}/c = \text{line efficiency}$$

Example 9.1 Consider the assembly plant of the Keri Corporation with a single model line where the number of work elements is $N_e = 12$. Table 9.1 lists the elements, their times in minutes, the predecessor elements, and the associated parts. The total work content time is $\Sigma t_e = 31.5$ min, and the plan number of units to produce is N = 100. Note also, the end item consists of ten parts. The number of operators is set as n = 4; whereby, the average time per operator is $\overline{c} = 31.5/4 = 7.875$ min.

9.4.5 Precedence Diagram

Figure 9.1 depicts the relationship of the elements showing all possible ways to perform the element tasks. Note the predecessor elements listed in Table 9.1 where elements e = 2 and 3 cannot be performed until element e = 1 is completed, and so forth.

The shift foreman has the responsibility to assign the elements to each operator in a way where the operator times are as even as possible, and where the predecessor elements are satisfied. This process is called line balancing. The results are in Table 9.2, that lists the operators, the elements assigned, the time per element, the associated parts, and the operator times. With the operator assignments, the cycle

Table 9.1 Example 9.1 data, e = element; t_e = time per element; p_e = predecessor; and h_e = part(s)

e	t_e	p_e	h_e
1	1.2		1
2	4.0	1	
3	3.0	1	2
4	2.4	2, 3	3
5	1.3	3	4
6	3.0	3	5, 6
7	1.2	5, 6	
8	1.2	6	7
9	3.4	6	
10	1.1	4	8
11	2.6	8, 9	9
12	7.1	7, 10	10

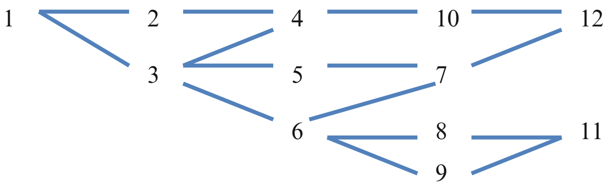


Fig. 9.1 Precedence diagram for the single model line of Example 9.1

Table 9.2 Element assignment by station, i; with elements, e; element times, t_e; parts, h_e, and operator times, c_i

Operator	Elements	Time	Parts	Operator time
i	e	t_e	h_e	c_i
1	1	1.2	1	
	2	4.0		
	3	3.0	2	8.2
2	4	2.4	3	
	5	1.3	4	
	6	3.0	5, 6	
	7	1.2		7.9
3	8	1.2	7	
	9	3.4		
	11	2.6	9	7.2
4	10	1.1	8	
	12	7.1	10	8.2

time for the line can now be measured, and represents the time between completed units coming off at the end of line. The cycle time is:

$$c = \max(8.2, 7.9, 7.2, 8.2) = 8.2 \ \text{min}$$

The timespan is a measure of the total time needed to complete the N = 100 units. This is:

$$T = c \times N = 8.2 \times 100 = 820 \ \text{min.}$$

A final measure is the line efficiency:

$$E = 7.875/8.2 = .960(96.0 \ \%)$$

9.5 Mixed Model Make-to-Stock Assembly

When a firm has two or more models of a product and produces them together on the same line, the system is called a mixed model make-to-stock assembly line. Upon completion, the units are placed in a warehouse awaiting future customer orders, An example is an appliance firm with, say, three models of a clothes washing machine, that is produced on one line at the same time and the finished units are stored in a nearby warehouse. Each unit of the individual model is exactly the same. There also is often some similarity between the models. Some of the assembled parts are unique to a model, and others are common between two or more of the models. Each day, the schedule mix of the models could vary, depending on the stock status of the models in the warehouse and the arrival of customer demands for the models. The planning and control of this type of line has two main steps: line balancing and sequencing. The line balance step is based on the average schedule mix of the models per shift. Assigning the work elements to the station operators is also based on an average mix of the models per shift. The sequencing of the units down the line, however, differs daily and depends on the specific schedule mix of the day.

9.5.1 Data

The data needed to carryon the planning for the line is listed below. The number of models is denoted as Nj, and j = 1 to Nj identifies the various models. The shift time is T, the average number of units of model j to produce each shift is N_j, and the total number of units to produce is N. There are Ne work elements and the elements are identified as e = 1 to Ne. The time per element is t_e and the associated predecessor elements are labeled as p_e. Since not all models use each of the elements, another data index is needed, a_{ej}, which is set to one if model j is used on element e and is zero otherwise. A final data source is when element e uses the part h_{ej} on model j.

$$
\begin{aligned}
Nj &= \text{number of models} \\
j &= 1 \text{ to } Nj = \text{model identification} \\
N_j &= \text{average number of units of j per shift} \\
N &= \Sigma N_j = \text{number of units to produce each shift} \\
T &= \text{time per shift} \\
Ne &= \text{number of work elements} \\
e &= \text{work element identification } (e = 1 \text{ to } Ne) \\
t_e &= \text{time per element} \\
p_e &= \text{predecessor element(s) to e} \\
a_{ej} &= (0, 1) \;\text{ index where : } (1 \text{ if e is used on j}, 0 \text{ if not}) \\
h_{ej} &= \text{part used for j on e}
\end{aligned}
$$

9.5.2 Computations

The data needed prior to assigning the work elements to the station operators are listed below. Since each element is not always used on all models, the average time for the element per unit is measured for subsequent use. This average is denoted as

$\bar{t}_e$ for element e, and the work content time by model is also obtained and is here denoted as T_j. With the data gathered and computed, the minimum number of station operators needed, n`, is calculated. Finally, the integer number of operators, n, is stated for the line.

$$\bar{t}_e = t_e\left[\Sigma_i a_{ej} N_j/\mathrm{N}\right] = \text{average time of e per unit}$$
$$T_j = \Sigma_e a_{ej} t_e = \text{work content time on j}$$
$$n` = \Sigma_j N_j T_j / T = \text{minimum number of operators needed}$$
$$n = \text{ceiling}\left[n`\right] = \text{integer number of operators}$$

9.5.3 Line Balance

Using the average daily schedule mix of the models, the assignment of elements is applied to the n station operators. Each operator is assigned work elements whose average element times sum as close as possible to the average operator time, $\bar{c}$, computed below. The element assignments along the stations must also adhere to the associated precedence elements. The operator times are denoted as c_i for i = 1 to n. The max of the operator times is a measure of the time between units coming off the line, and becomes the cycle time. The line efficiency, E, can now be measured as the ratio of the average operator time over the cycle time.

$$i = \text{operator identification} \quad (i = 1 \text{ to } n)$$
$$i : \left\{e, \bar{t}_e, h_{ej}\right\} = \text{elements assigned to i, with average time, and associated parts}$$
$$c_i = \Sigma_e \bar{t}_e = \text{sum of elements assigned to i} \quad (i = 1 \text{ to } n)$$
$$\bar{c} = \Sigma_i c_i / n = \text{average time per unit per operator}$$
$$c = \max(c_1, \ldots, c_n) = \text{cycle time}$$
$$E = \bar{c}/c = \text{line efficiency}$$

9.5.4 Sequencing

Each day, a new schedule mix on the line is possible and thereby, the sequencing of the models down the line has to be reestablished each day. The daily number of units for model j is denoted as N(j), and the sum of units over all models is N. The sequence of models for the day is denoted as j(s) for s = 1 to N. The sequencing algorithm for this make-to-stock line is listed below in partitions labeled as (input, algorithm, output):

```
Input:
N(j) = schedule of model j for the shift (j = 1 to Nj)
Algorithm:
N = ∑_j N(j) = number of units to produce in the shift
w(j) = N/N(j)    j = 1 to NJ
a(j) = w(j)/2    j = 1 to Nj
```

```
for s = 1 to N
   jo = j with min [a(1), ..., a(Nj)] = next model in sequence
   j(s) = jo = identifies the model in sequence number s
   a(jo) = a(jo) + w(jo)
next s
Output:
j(1), ...., j(N) = sequence of models for the day
```

Example 9.2 The Zante Manufacturing plant produces models A, B, and C and has one line dedicated to the assembly of three models. The shift time is T = 450 min, and the average shift schedule for the three models is $N_A = 50$, $N_B = 30$ and $N_C = 20$, and thereby, N = 100 units per shift. The number of work elements is Ne = 62 and some information on the elements is listed in Table 9.3. Note where element e = 1 is used on all three models and thereby the average time per unit for the element is the same as the element time of 2.4 min. Since element e = 2 is only used on models A and B, or an average of (50 + 30) = 80 of the 100 units per shift, the average time of the element per unit becomes: $(80/100) \times 1.0 = 0.8$ min. In this way, the average time of all the elements are computed for an average mix of the models. The sum of the average element times is also listed at the foot of the table as 105.4 min. This is the average work content time per unit.

9.5.5 Number of Station Operators

Using the data listed above, the number of station operators is found by the computations below. First, the minimum number of operators needed becomes n` = 23.4, and thereby, the integer number of operators is n = 24.

$$\begin{aligned} T &= 450 \text{ min} \\ N &= 100 \\ n` &= N \times \Sigma \bar{t}_e / T = 100 \times 105.4/450 = 23.4 \\ n &= \text{ceiling}(23.4) = 24 \end{aligned}$$

Table 9.3 Elements, e; element times, t_e; element assignments by model, a_{ej}; and average element time per unit, $\bar{t}_e$

e	t_e	a_{eA}	a_{eB}	a_{eC}	$\bar{t}_e$
1	2.4	1	1	1	2.4
2	1.0	1	1	0	0.8
3	3.6	1	0	1	2.5
4	2.6	0	1	1	1.2
5	5.1	1	0	0	2.6
...					
62	4.0	1	1	1	4.0
Sum					105.4

9.5.6 Line Measures

First, the average operator time is measured as $\bar{c} = 4.39$ min. Second, the assignment of the 62 elements to the 24 operators is applied seeking to give each operator an average time close to 4.39 min, without violating any of the precedence constraints. The maximum of the operation times is called the cycle time and in the example, c = 5.00 min, say. With these results, the line efficiency is measured as E = 0.878 or 87.8 %.

$$\bar{c} = \Sigma\bar{t}_e/n = 105.4/24 = 4.39 \text{ min}$$
$$c = \max(c_1, \ldots, c_{24}) = 5.00 \text{ min}$$
$$E = \bar{c}/c = 4.39/5.00 = 0.878 = 87.8\,\%$$

9.5.7 Sequence

Recall, each day, a new mix of models is called where the sum is still N = 100 units. Assume for a given day, the mix calls for 50, 29 and 21 units for models A, B and C, respectively. Table 9.4 lists the three models, the schedule and the algorithm measures for model j: w(j) and a(j), as explained earlier in the sequence algorithm. Table 9.5 shows how the measures apply in seeking the sequence of the models down the line. Note where s identifies the sequence number, and for each model, the current value of a(j) is computed. The model with the minimum a(j) is selected as the next model in the sequence. At s = 1, a(1) = 1.00 is the min, and thereby, j(1) = A as the first unit in the sequence. At s = 2, a(2) = 1.72 is the min, and j(2) = B. This process continues until all 100 units in the sequence are determined.

Table 9.4 A day's schedule, N(j) for model j, with sequence indices, w(j) and a(j)

j	N(j)	w(j)	a(j)
A	50	2.00	1.00
B	29	3.44	1.72
C	21	4.76	2.37
sum	100		

Table 9.5 The first five accumulations of a(j) at sequence s, and the model chosen j(s)

s	A	B	C	j(s)
1	1.00	1.72	2.37	A
2	3.00	1.72	2.37	B
3	3.00	5.16	2.37	C
4	3.00	5.16	7.13	A
5	5.00	5.16	7.13	A
. . .				
100	–	–	–	–

9.6 Mixed Model Make-to-Order Assembly

Mixed model for make-to-order assembly is when the units going down the line are from customers orders that specify the options wanted on each feature offered on the units. Each unit on the line is unique with the combination of features and options and often no two units are the same.

9.6.1 Features and Options

Below is the data associated with the features and options in make-to-order assembly. Nf is the number of features and f identifies each feature where f = 1 to Nf. Each feature has a null option and Nk(f) non-null options. The notation is k = 0 for the null option, and k = 1 to Nk(f) for the non-null options of feature f. The firm saves the choice of features and options and p(f,k) is the portion of option k selected on feature f.

$$\mathrm{f} = \text{features}$$
$$\mathrm{Nf} = \text{number of features}$$
$$\mathrm{k} = \text{option of feature f}$$
$$\mathrm{k} = 0 \text{ is null option}$$
$$\mathrm{Nk(f)} = \text{number of non-null options of feature f}$$
$$\mathrm{k} = 0 \text{ to } \mathrm{Nk(f)} = \text{all options of f}$$
$$\mathrm{p(f,k)} = \text{portion of option k on feature f } [\mathrm{f} = 1 \text{ to } \mathrm{Nf}, \mathrm{k} = 0 \text{ to } \mathrm{Nk(f)}]$$

9.6.2 Jobs

Each customer order specifies the option wanted for each feature offered on the units. In the discussion here, the orders are called jobs and each job has a desired delivery due date. As the order comes in from the customers, they are placed on an order board in a backlog status. The assembly line is structured to produce N jobs per daily shift. Every day, the backlog of jobs on the order board are scanned to identify N jobs with due dates close to the current date, and where the stock inventory has all the needed parts and components available. These N jobs are selected as a group several days prior to assembly. Subsequently, the sequence arrangement of the jobs for the group with be determined. The number of jobs to assemble for the shift is Nj, where j identifies the job and j = 1 to Nj. Note also where Nj = N. The data on these jobs are listed below.

$$\mathrm{j} = \text{jobs}$$
$$\mathrm{Nj} = \text{number of jobs per shift}$$
$$\mathrm{j} = 1 \text{ to } \mathrm{Nj} = \text{job index}$$
$$\mathrm{k} = \mathrm{kj(j,f)} = \text{option k of feature f on job j}$$

9.6.3 Work Elements

The data for the work elements is listed below. The number of work elements is Ne, and the element identification is e where e = 1 to Ne. The standard time of element e on each unit is t_e and the predecessor elements to e are listed with p_e. One or more elements may be associated with one of the Nf features, and if so, f_e identifies the feature of element e. Finally, the part needed (if any) to perform the element e is denoted as h_e.

$$
\begin{aligned}
e &= \text{elements} \\
Ne &= \text{number of elements} \\
t_e &= \text{element time per unit} \\
p_e &= \text{predecessor(s) to e} \\
f_e &= \text{feature of element e (blank if none)} \\
h_e &= \text{part(s) associated with e}
\end{aligned}
$$

9.6.4 Computations

Below are the computations needed to determine the number of station operators and the assignment of elements to the operators. The shift time is denoted as T, and the number of jobs per shift is N. The average time per element per unit is determined as follows. The average time is the same as the standard time if the element is not associated with a feature (f_e = blank). For any element associated with a feature, the element time is adjusted based on the portion of null options stated for the feature. The sum of the average element times gives the average work content time for an arbitrary unit coming down the line.

$$
\begin{aligned}
&T = \text{shift time} \\
&N = \text{number of jobs per shift} \\
&\text{Note} : p(f_e, 0) = \text{the portion of null option for feature } f_e \\
&\text{if } f_e = \text{blank} : \quad \bar{t}_e = t_e \\
&\text{if } f_e \neq \text{blank} : \quad \bar{t}_e = t_e[1 - p(f_e, 0)] = \text{average element time per unit} \\
&\Sigma\bar{t}_e = \text{sum of all element average times}
\end{aligned}
$$

9.6.5 Operators

Below shows how to determine the minimum number of station operators, n`, and the integer number of operators, n, for the line. The operators are denoted by i where i = 1 to n. The average time to assign per operator is also calculated and denoted as $\bar{c}$, as shown below:

$$n' = [\Sigma\bar{t}_e \times N]/T = \text{minimum number of operators}$$
$$n = \text{ceiling}(n') = \text{integer number of operators}$$
$$i = \text{operator} \quad (i = 1 \text{ to } n)$$
$$\bar{c} = \Sigma\bar{t}_e/n = \text{average time per operator}$$

9.6.6 Line Balance

Using the number of operators, n, along with the average time per element, $\bar{t}_e$, and the precedence elements, p_e, the next task is to assign the elements to the operators. This is the line balancing stage. The role of line balancing is to assign the work elements to the stations in a way where the station work times are relatively even, and where all the precedence restrictions are satisfied. The line balance results generally stay in place, and only need adjustments when a major change takes place in the feature-option combinations or in the number of units, N, to produce in a shift. Below shows the notation and data from line balancing. The operators are denoted as i, and for each operator, the elements assigned, e, are listed along with the average element time, $\bar{t}_e$, and the parts, h_e, associated with the elements. The total time assigned to operator i is listed as c_i.

$$i = \text{operator identification} \quad (i = 1 \text{ to } n)$$
$$i\text{: } \{e, \bar{t}_e, h_e\} = \text{elements assigned to i, with average time, and associated parts}$$
$$c_i = \Sigma_e\bar{t}_e = \text{sum of element times assigned to station i } (i = 1 \text{ to } n)$$

9.6.7 Line Efficiency

Below shows how the efficiency of the line is measured. As before, the average time per operator is denoted as $\bar{c}$, and the maximum of the operator times is called the cycle time and labeled as c. The line efficiency can now be measured as the ratio of $\bar{c}$ over c and is denoted as E.

$$\bar{c} = \sum\nolimits_i c_i/n = \text{average time per unit per operator}$$
$$c = \max(c_1, \ldots, c_n) = \text{cycle time}$$
$$E = \bar{c}/c = \text{line efficiency}$$

9.6.8 Daily Shift Schedule

Each day, a new group of jobs are placed on the schedule. The number of jobs is Nj, the same as N, and the jobs are labeled as j = 1 to N. Each job, j, has a unique set of options to apply by feature, and denoted here as kj(j,f) for f = 1 to Nf. This is the data needed to determine the sequence for the day.

$$\begin{aligned} \text{Nj} &= \text{number of jobs per shift} \\ \text{j} &= \text{jobs} \;\; (\text{j} = 1 \text{ to Nj}) \\ \text{kj(j, f)} &= \text{option k of feature f for job j} \end{aligned}$$

9.6.9 Feature-Option Count for the Shift

A first step to determine the sequence for the day is to count the number of options by feature on the day's group of Nj jobs. This is denoted as N(f,k) for f = 1 to Nf. Recall k = 0 is a null option.

$$\text{N(f, k)} = \text{tally of option k on feature f for the Nj jobs of the shift}$$

9.6.10 Sequencing

Each day, a new cluster of Nj jobs are scheduled for assembly, and have a new mix of features and options. Thereby, the sequencing of the jobs down the line has to be reestablished each day. The sequencing algorithm for this make-to-order line is listed below, and the computed sequence of jobs for the day is denoted as j(s) for s = 1 to N. The sequencing algorithm is partitioned below as: Input, Algorithm and Output.

```
Input:
N(f,k) = tally of option k on feature f
Nj = number of jobs for the shift
Nf = number of features
Nk(f) = number of non- null options of feature f
k = 0 is null option
k = 0 to Nk(f) = null and non-null options of feature f
f = 1 to Nf = features
kj(j,f) = option k of feature f for job j
Algorithm:
s = 0
  for f = 1 to Nf
   for k = 0 to Nk(f)
     x(f,k) = [0.5 × Nj/N(f,k)]
   next k
  next f
for s = 1 to Nj
  for j          [select j not yet assigned]
   for f = 1 to Nf
    k = kj(j,f)
    y(j,f) = [x(f,k) + Nj/N(f,k)]
   next f
   Y(j) = ∑f y(j,f)
```

```
  next j
  Y(jo) = min{Y(j)}      [jo = min of all j not yet assigned]
  j(s) = jo              [j(s) = job in sequence number s]
  for f = 1 to Nf
    ko = kj(jo,f)
    x(f,ko) = [x(f,ko) + Nj/N(f,ko)]
  next f
next s
Output:
[j(1), ...., J(Nj)]   [sequence of jobs for the day's shift]
```

Example 9.3 Assume the Keri Trucking Company produces models on a make-to-order basis each day. The shift time is T = 450 min, and the average shift schedule is N = 72 units per shift. The number of features is Nf = 11 and Table 9.6 lists the average portion of option k that is used for feature f. Note at feature f = 1, 75 % of the customers, on average, call for a null option, and 25 % use option k = 1. Feature f = 4 has a null option and seven non-null options. Below lists the data and computations that apply in assigning the elements to the operators.

Ne = 204 (number of work elements)
N = 72 (number of units to assemble per shift)
T = 450 (shift time in minutes)
Nf = 11 (number of features)

9.6.11 Computations

Each element time is converted (if necessary) to the average element time and the sum of the average times over the Ne = 204 elements is 262 min. The plant schedule calls for N = 72 units per shift and the shift time is T = 450 min. The

Table 9.6 Percent usage for option k on each of the eleven features, f

f/k	0	1	2	3	4	5	6	7
1	75	25						
2	90	10						
3	85	15						
4	10	2	2	2	70	2	2	10
5	99	1						
6	99	1						
7	99	1						
8	99	1						
9	99	1						
10	35	40	15	10				
11	50	50						

computations below shows that the minimum number of operators becomes n` = 41.92 and the integer number of operators is n = 42, whereby, the average time per operator becomes, $\bar{c} = 6.24$ min.

$$
\begin{aligned}
\Sigma\bar{t}_e &= 262 \\
n` &= \Sigma\bar{t}_e \times N/T = 262 \times 72/450 = 41.92 \\
n &= \text{ceiling}(41.92) = 42 \\
\bar{c} &= \Sigma\bar{t}_e/n = 262/42 = 6.24
\end{aligned}
$$

Upon the line balancing stage, the maximum of the operator times is c = 7.10 min, say, as shown below.

$$c = \max(c_1, \ldots, c_{42}) = 7.10$$

Note the following computations for the line efficiency is measured as E = 87.9 %, and the number of units completed in T = 450 min is 63.4, far short of the plan of N = 72. Thereby, more operators are needed.

$$
\begin{aligned}
E &= c/c = 6.24/7.10 = 0.879 \\
N` &= T/c = 450/7.10 = 63.4
\end{aligned}
$$

Assuming an efficiency of 0.879, the number of operators needed is:

$$n = n`/E = 41.92/0.879 = 47.6 = 48$$

Rebalancing the line with 48 operators yields,

$$c = \max(c_1, \ldots, c_{48}) = 6.20$$

The number of units that now can be completed in T = 450 min is:

$$N = T/c = 450/6.20 = 72.6$$

which conforms with the plan schedule of the management.

9.6.12 Jobs for the Day

Every day a cluster of N = 72 jobs are selected from the order board for assembly. Each job lists the options specified for every one of the Nf features. Recall option k = 0 is a null option on the feature. Table 9.7 lists a sample of the jobs and options for a particular day. Note on job j = 1, option k = 4 is called on feature f = 4; and k = 1 is called on each of features f = 10 and 11. All other options are null.

Table 9.7 A day's list of 72 jobs, j, with option specified for each of the eleven features, f

j/f	1	2	3	4	5	6	7	8	9	10	11
1	0	0	0	4	0	0	0	0	0	1	1
2	0	0	0	4	0	0	0	0	0	0	0
3	0	0	0	4	0	0	0	0	0	1	1
4	0	0	0	4	0	0	0	0	0	3	1
5	0	1	1	7	0	0	0	0	0	2	0
. . .											
72	0	1	0	4	0	0	0	0	0	0	0

Table 9.8 Features, f, with the current day's tally of each of the options, k

f/k	0	1	2	3	4	5	6	7
1	60	12						
2	64	8						
3	62	10						
4	10	0	0	0	48	0	1	13
5	71	1						
6	72	0						
7	69	3						
8	72	0						
9	71	1						
10	25	30	10	7				
11	35	37						

9.6.13 Feature-Option Tally

Table 9.8 is a tally on the number of options by feature for the 72 jobs scheduled for the day. Note at feature f = 1, 60 of the options are null, k = 0, and 12 are for option k = 1. The sum of options per feature is 72.

9.6.14 Sequence

Applying the sequencing algorithm listed earlier, the sequence of the Nj = 72 jobs is applied. Table 9.9. lists the first 16 of the Nj = 72 jobs in the sequence. The algorithm seeks to spread the use of each option by feature as far apart in the sequence as possible. Note, the first job in the sequence is j = 16, the second is j = 50, and so forth. Also note option k = 7 on feature f = 4 occurs on 13 of the 72 jobs. Ideally this option should be placed each 72/13 = 5.5 units in the sequence. The sequence results shows that the jobs with option k = 7, for example, is spread and falls in sequence numbers, s = 2, 7 and 11 of the first 16 jobs.

Table 9.9 The sequence for the day with jobs, j, and a list of the options for each of the eleven features, f

j/f	1	2	3	4	5	6	7	8	9	10	11
16	0	0	0	4	0	0	1	0	0	1	1
50	0	0	1	7	0	0	0	0	0	0	0
6	1	0	0	0	0	0	0	0	0	0	0
46	0	1	0	4	0	0	0	0	0	0	0
11	0	0	0	4	0	0	0	0	0	2	1
4	0	0	0	4	0	0	0	0	0	3	1
10	0	0	0	7	0	0	0	0	0	1	0
2	0	0	0	4	0	0	0	0	0	0	0
1	0	0	0	4	0	0	0	0	0	1	1
12	1	0	0	0	0	0	0	0	0	0	0
24	0	0	1	7	0	0	0	0	1	0	0
3	0	0	0	4	0	0	0	0	0	1	1
35	0	0	0	4	0	0	0	0	0	2	1
72	0	1	0	4	0	0	0	0	0	0	0
7	0	0	0	4	0	0	0	0	0	1	1
47	1	0	0	4	0	0	0	0	0	0	0
...											

9.7 Robotic Assembly

After World War II there were many low paying workers in Japan, and as such, some of the US firms moved their assembly operations to Japan to reduce the cost of the producing the products. As time moved on and the Japanese wages grew, the assembly operations slowly shifted to South Korea, then to China, Taiwan, and Vietnam. But today, with the advances taking palace in robots, the assembly operations are slowly shifting back to the US.

The auto industry is by far the main user of industry robots. This takes place in the stamping press, the body assembly and in the paint shop of the plant where the robots perform almost all of the assembly tasks. However, in the final assembly of the product, there are no robots and the workload remains totally by humans. In the body shop, the robots are programmed to stamp out and assemble the entire upper body of the automobile one step at a time. One robot lifts the roof structure and another attaches the sides. Then the doors, front and rear body are attached by robots. This all requires heavy handling and movement of the body parts, welding, and other tasks to complete the body structure. With completion, the body moves into the paint booth where the robots paint spray the body to the color specified for the individual unit. Later, the body is attached to the frame and the assembly continues mainly by human workers.

A final stage in assembly is testing the functionality of the unit. Most of the steps in the testing stage is performed by the humans, and some is by robots that apply lasers to determine if the body parts are properly installed. Any unit detected with a flaw is noted and set aside for adjustment as needed. After the adjustment on the fault, the unit then repeats the testing stage.

With technology advancing, the use of robots are also in a variety of other industries like: trucks, buses, farm machinery, industrial machinery, motor homes, prefab buildings, furnaces, ovens, electrical equipment, electronic equipment, air conditioning and heating equipment.

Because of physical and mental fatigue in performing the repetitive assembly tasks, the assembly workers require short breaks for relaxing, refreshments and with body needs. But the robots continue to work without these intermittent needs. They perform their tasks flawlessly hour after hour without a break. With today's advances, the robots are programmed to perform a multitude of tasks from routine to complex, like: welding, serial part numbering, labeling, cutting, spraying, painting, grinding, molding, material removal, material movement, milling, polishing, and water spraying.

Industrial robots are expensive to purchase and to integrate into the production system. Each robot requires people to program, monitor and maintain them. They require special care in safety rails and cages to protect the people, inventory and other equipment that are nearby. But the robots do not require wages, benefits, vacations, insurance, severance pay, pensions, or vacation time, and in the long run, they are very cost effective

9.8 Summary

On single model assembly lines, the management assigns the work elements to the operators on the line in a fair share manner. The precedence relation between the elements is listed and the goal is to evenly spread the work to the operators. This function is called line balancing. For mixed model make to stock lines, two or more models of the product are assigned to the line. The production schedule of each model is stated, and the management assigns the work elements to the operators, and determines the sequence of the models down the line in a way where the flow from station to station is efficient. On mixed model make to order lines, each unit (job) going down the line is for an individual customer who specifies the features and option combination wanted. The management is obliged to find the assignment of work elements to each operator and to determine the sequence of jobs for each day. Robots are now widely used in the assembly of automobiles, trucks and for units that are of the heavier type.

Chapter 10
Statistical Process Control

10.1 Introduction

A process can be defined as a combination of components, tools, people and machines that together produce an item, like a part that subsequently is inserted into a finished good item. Although, the output of the process may vary, the final customer requires the product to satisfy the standards specified by the engineers. Management's goal is to control the output units of a process so that it conforms to the standards set by the engineers. The output measures are of two types: attribute or variable. Attribute is when the output measure is defective or not defective. Variable is when some measure is taken from each sampled unit.

Four attribute type control methods are described and labeled as the following: p-chart, np-chart, c-chart and u-chart. A pre-study is setup for each of the methods. N subgroups are studied at intermittent time intervals with a number of samples taken from each subgroup, and the number of defectives tallied, if any. Statistics on the capability of the process is measured after the pre-study, and if acceptable, lower and upper control limits are computed. The process is monitored in the same way for the future subgroups to ensure the process is performing with high standards.

Four variable type control methods are described and labeled as the following: x-bar and R-chart, median and R-chart, individual-chart, and x-bar and s-chart. A pre-study is arranged with N subgroups at systematic time intervals, where sample data is taken. Specification limits are provided from the engineers, and these are compared with the sample output data from the samples. The goal is for the output to be well within the specification limits, and ideally within six-sigma limits. Measures of process potential and process capability are computed using the results gathered in the pre-study. If these are acceptable, a set of statistical coefficients is used to compute control limits for the particular testing method. Thereupon, the process is monitored over future subgroups to ensure the output measures remain within the control limits.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_10

10.2 Statistical Process Control Pioneers

The first use of Statistical Process Control (SPC) was in the early 1920s by Walter Shewhart of Bell Laboratories. A few years later, Wm. Johnson, Harold Dodge and Harry Romig of ATT applied and added to the method. After World War II, W. Edward Deming was hired by the Japanese industry to help apply SPC to their products; and because of the success in Japan, the US and world industry leaders took notice and began applying SPC to all sort of applications.

10.3 Statistical Control

Statistical process control has two stages that are defined as: pre-study stage, and monitor stage. In both stages, a cluster of data is gathered at periodic time intervals in a series of subgroups. The pre-study stage is run to collect data and if necessary, to adjust the process until it conforms to engineering standards. Control limits are calculated and used subsequently in the monitoring stage. The monitor stage seeks to ensure the process remains stable and in conformance with the engineering standards. Using the pre-study results, the average (central line), standard deviation and control limits are computed for use in the monitoring stage. The monitoring subgroups are run in durations the same as in the pre-study stage, and the points from the output measures are plotted accordingly.

Over time, the plot is observed to assess whether or not the process continues in control. When out-of-control, the process is stopped, the cause is found and corrected, and the process continues. The Western Electric rules on determining an out-of-control condition are listed below:

- If a point falls outside of the control limits.
- If 2 of 3 consecutive points are near one of the control limits.
- If 4 of 5 consecutive points are beyond one σ of the central line.
- If 6 consecutive points form a trend going up or going down.
- If 8 consecutive points are all on one side of the central line.
- If 6 consecutive points are trending up or down.
- If 14 consecutive points are all altering above and below the central line.

10.4 Data Type

The output data from the process under study falls into two main categories: attribute and variable. Attribute type data measures are of two categories: defective or not defective. Variable type data lie in a continuous range of values, like temperature, weight, strength, so on. Four methods to control the process of attribute type data, and four methods for variable type data are described in this chapter.

10.5 Attribute Type Data

Attribute data is when a unit inspected is identified as: conforming or non-conforming; acceptable or non-acceptable; good or bad; non-defective or defective; and so forth. Four categories of units with attribute type data are denoted as the following: p-chart, np-chart, c-chart, and u-chart. A short description on each is given below:

The p-chart applies when n observations are sampled per subgroup to determine a count on how many observations are of the defective type. The number of observations per subgroup is not always the same.

The np-chart is selected when the number of observations per subgroup is constant and a count of defective units in each subgroup is noted.

The c-chart occurs when only one unit is observed in a subgroup and zero, one or more defectives can occur per unit.

The u-chart is proper when one or more units are observed per subgroup and zero, one or more defectives can occur per unit.

10.5.1 Process Review Decisions

Prior to the pre-study phase for a particular process under review, the management determines which of the four attribute types conforms with the process. Once the attribute type is identified, the timing between collecting subgroup data is selected, and also the reasonable number of observations per subgroup is stipulated. The management specifies the approximate number of subgroups to study in the pre-study stage; and also the duration between subgroups (hours, days, so forth).

10.5.2 Pre-Study Subgroup Data

For each subgroup, select n observations from the process and count the number of defectives. Repeat this procedure in accordance with the timing frequency specified by the management. Halt the pre-study after data from an agreed number of subgroups have been collected. The pre-study data is summarized below:

$$N = \text{number of subgroups}$$
$$i = (1, \ldots, N) \ \text{identifies the subgroups}$$
$$n_i = \text{observations in subgroup i}$$

10.5.3 P-Chart

The p-chart is in use when n observations are sampled by subgroup and x is the subgroup count of non-conforming units. The sample size per subgroup is typically 50 or more. The observation size is not necessarily the same from subgroup to subgroup. The goal of the pre-study stage is to measure the process capability and deem whether it is acceptable with the management acceptable standards. If acceptable, a monitoring stage will follow the pre-study stage. If not acceptable, the production process needs to be analyzed to determine what is causing the unacceptable fraction of defective units. The fault must be corrected and a new pre-study collection of data is rerun.

10.5.3.1 Data Collection

The data gathered for each of N subgroups is the number of observations and the number of defectives. This data is listed below:

$$N = \text{number of subgroups}$$
$$n_i = \text{number of observations in subgroup i}$$
$$x_i = \text{number of defectives in subgroup i}$$

10.5.3.2 Computations

Upon completion of collecting data from the N subgroups, aggregate summations on the number of observations and the number of defectives are measured. These summations are used to compute the average number of observations per subgroup, and the fraction of defectives. The calculations are listed below:

$$\Sigma n = \text{sum of } n_i \text{ in N subgroups}$$
$$\Sigma x = \text{sum of } x_i \text{ in N subgroups}$$
$$\overline{n} = \Sigma n / N = \text{average observations per subgroup}$$
$$\overline{p} = \Sigma x / \Sigma n = \text{fraction defective}$$

10.5.3.3 Process Capability

The process capability is computed below:

$$(1 - \overline{p}) = \text{fraction not defective}$$

This is a measure on the quality of the units being produced, identifying the fraction of production output units that are non-defective. This measure must satisfy the standards set by the management. If the process capability is not acceptable, it behooves the process operator to find the cause, fix it, and rerun the collection of data for N subgroups. If the process capability meets the acceptable standards, the production control limits are computed as shown below.

10.5.3.4 Control Limits

Upon approval of the process capability, more measures are computed. First is the average fraction of defective units, and this represents the central line on a p-chart. Next, is the standard deviation of the fraction defective, and last are the lower and upper limits on the fraction defective. The notation and computations for the measures are listed below:

$$\begin{aligned} \overline{p} &= \text{central line} \\ \sigma_p &= \sqrt{\overline{p}(1-\overline{p})/\overline{n}} = \text{standard deviation} \\ \text{LCL} &= \overline{p} - 3\sigma_\text{p} = \text{lower control limit} \\ \text{UCL} &= \overline{p} + 3\sigma_\text{p} = \text{upper control limit} \end{aligned}$$

10.5.3.5 Monitor Stage for the p-chart

After the process capability meets the acceptable standard set by the management, the monitoring of the production begins. The purpose of the monitor stage is to ensure the process continues with the same standard quality of process capability. The frequency of subgroup timing remains the same as in the pre-study stage; and the number of observations per subgroup continues similar as before. The data for each subgroup is used to measure the fraction of defective units for the subgroup.

10.5.3.6 Subgroup Data

The subgroup data is listed below:

$$\begin{aligned} &\text{n} = \text{number observations} \\ &\text{x} = \text{number of defectives} \\ &\text{p} = \text{x}/\text{n} = \text{fraction defective} \end{aligned}$$

10.5.3.7 Monitor the Subgroup

With the subgroup data and computations, the next step is to determine if the subgroup results are acceptable or not. Acceptable is when the fraction defective is within the lower and upper control limits, as shown below:

Table 10.1 Pre-study results listing the sample size and defects per subgroup

Subgroup i	1	2	3	4	5	Sum
n_i	53	52	48	60	52	265
x_i	1	0	0	2	1	4

$$\text{If } \text{LCL} \leq \text{p} \leq \text{UCL}: \quad \text{acceptable}$$

In the event the fraction defective is outside of the limits, the process operator is obliged to stop the process, find and fix the fault, and proceed accordingly.

Example 10.1 Assume a process where the p-chart applies and the pre-study results are those listed in Table 10.1. The number of subgroups is N = 5 and the summations are **Σ**n = 265 and Σx = 4. The average fraction of defects is computed as: $\overline{p} = 4/265 = 0.015$, and thereby, the process capability becomes: $(1 - \overline{p}) = 0.985$.

Suppose the management approves the estimate of the process capability and moves on to monitoring the production stage. The subgroups will be formed so their timing is the same as in the pre-study stage, and the sample size will remain near 50 units. The central line is $\overline{p} = 0.015$; and so now, the control limits are computed as below. The average number of samples per subgroup is obtained from $\overline{n} = 265/5 = 53.0$, and the standard deviation of the fraction defective is computed as below:

$$\sigma = \sqrt{(0.015 \times 0.985)/53.0} = 0.017.$$

Finally, the control limits are obtained as follows:

$$\text{UCL} = 0.015 + 3 \times 0.017 = 0.066$$
$$\text{LCL} = 0.015 - 3 \times 0.017 = -0.036 \rightarrow 0.000$$

Because, LCL is negative, it is reset to LCL = 0.000.

10.5.4 Np-Chart

The np-chart applies when the production process yields data of the attribute type, and the number of observations in each subgroup is relatively large. Because of large samples, the analysis assumes the normal distribution, the number of observations, n, per subgroup must be 50 or more. A pre-study stage is run to collect data and obtain some measurements needed in the subsequent production stage. Upon completion of the pre-study, a quality measure, called the process capability, is obtained to determine if the process is acceptable with the engineering standards. When the process is accepted, the parameters for the monitor stage are computed. Below describes the pre-study stage and then the production monitoring stage.

10.5.4.1 Pre-Study Stage

The pre-study stage consists of data collection over N subgroups, followed by some computations, and a measure of the process capability. Should the management not approve of the capability, the cause has to be found, corrected and the pre-study is run again.

Data Collection

The data collected during the pre-study stage is the following:

$$
\begin{aligned}
N &= \text{number of subgroups} \\
i &= 1 \text{ to } N = \text{subgroups} \\
n &= \text{a constant number of observations per subgroup} \\
x_i &= \text{number defectives per subgroup}
\end{aligned}
$$

Computations

Upon collecting the data from the N sub-groups in the pre-study, the following computations are applied:

$$
\begin{aligned}
n\bar{p} &= \Sigma x_i / N = \text{average defectives per subgroup} \\
\bar{p} &= n\bar{p}/n = \text{average fraction defective}
\end{aligned}
$$

Process Capability

With the above computations, the quality of the process is measured as below:

$$(1 - \bar{p}) = \text{average fraction non-defective}$$

10.5.4.2 Monitor Stage

In the event the capability of the process conforms with the management standards and is accepted, the data from the pre-study stage is used to compute the parameters needed to monitor the process in the subsequent monitor stages. The variable is the number of defectives in a subgroup, denoted as x, and also is commonly denoted as np. The parameters are the central line, and the control limits on the number of defectives for a subgroup. To begin, the standard deviation of np is measured.

Standard Deviation of Np

The standard deviation of the number defectives in a subgroup is now computed. The standard deviation is obtained as follows:

$$\sigma_{np} = \sqrt{(n\overline{p})\,(1-\overline{p})}$$

Control Limits

With the data and computations noted above, the parameters to monitor the production process are now obtained. These are the central line, and the control limits as shown below:

$$\begin{aligned} n\overline{p} &= \text{central line} \\ \text{LCL} &= n\overline{p} - 3\sigma_{np} = \text{lower control limit} \\ \text{UCL} &= n\overline{p} + 3\sigma_{np} = \text{upper control limit} \end{aligned}$$

The lower control limit is always set to zero in the event it measures as a negative value.

Example 10.2 Suppose a production process where the np-chart applies, and in the pre-study stage N = 8 subgroups are run with n = 60 observations per subgroup. Assume a total of 20 defectives are detected. The data and computations that pertain are below:

$$\begin{aligned} \Sigma x_i &= 20 \\ n\overline{p} &= 20/8 = 2.5 \quad (\text{central line}) \\ \overline{p} &= 2.5/60 = 0.042 \\ (1-\overline{p}) &= (1-0.042) = 0.958 \end{aligned}$$

Assuming the management approves of the process capability (0.958), the computations needed for the monitor stage follows. This entails, first computing the standard deviation of np; and second, the parameters ($n\overline{p}$, LCL, UCL) to monitor the subgroups. The standard deviation and the control limits are computed below:

$$\begin{aligned} \sigma_{np} &= \sqrt{2.5(1-0.042)} = 1.55 \\ \text{LCL} &= 2.5 - 3 \times 1.55 = -2.15 \rightarrow 0.00 \\ \text{UCL} &= 2.5 + 3 \times 1.55 = 7.14 \end{aligned}$$

10.5.5 C-Chart

The c-chart applies when one unit is observed during each subgroup, and the number of defectives per unit, denoted as c, could be zero, one or more. This might be a windowpane that could be without any flaws, or could have one or more flaws detected. At each subgroup, the number of flaws is noted, and after N subgroups, the average number per subgroup is computed. Data of this type follows a Poisson probability distribution, allowing the estimate of the mean and standard deviation on the number of defectives per subgroup. The management specifies an upper limit on the average number of defectives per unit, and if the pre-study shows the process to be acceptable, the results are used to compute the parameters for monitoring the system during the subsequent stage. Below describes the steps to take in the pre-study stage and beyond.

10.5.5.1 Pre-Study Stage

The pre-study stage for the c-chart consists of data collection for each on N subgroups, the computations, and the process capability. Should the process be deemed not acceptable, the process is reexamined to find the cause, fix the problem, and then, rerun the pre-study. When the process is acceptable, the parameters for the monitor stage are computed.

Data Collection

The data collected during the pre-study stage is the number of defects per subgroup. Below is a list of the notation and data.

$$\begin{aligned} N &= \text{number of subgroups} \\ i &= 1 \text{ to } N = \text{subgroups} \\ n &= 1 \text{ unit observed per subgroup} \\ c_i &= \text{number of defects detected in the unit of subgroup i} \end{aligned}$$

Computations

From the data collected, the following computations finds the sum of the defects in N subgroups and the average number of defects per unit observed.

$$\begin{aligned} \Sigma c_i &= \text{number of defects in N subgroups} \\ \overline{c} &= \Sigma c_i / N = \text{average number defects per unit} \end{aligned}$$

Process Capability

To determine the process capability, the management must specify an upper limit on the number of defects per unit c that is acceptable, denoted as c_o. If $\overline{c} > c_o$, the process is not acceptable, and the cause must be found, corrected and the pre-study is rerun. If $\overline{c} \leq c_o$, the process is deemed acceptable.

$$c_o = \text{acceptable upper limit on c}$$

10.5.5.2 Monitor Stage

The monitor stage begins with computing the parameters for the control charts. This includes, the central line, and the standard deviation on the number of defectives for a subgroup. With this information, the lower and upper control limits on the number of defectives in a subgroup are calculated. The lower control limit is not allowed to fall below zero.

Standard Deviation of c

The standard deviation of c is computed as shown below, assuming c follows a Poisson probability distribution. This is needed to compute the control limits of c.

$$\sigma_c = \sqrt{\bar{c}} = \text{standard deviation of c}$$

Control Chart Parameters

The parameters (central line, and control limits) to develop the control chart for the c-chart are shown below.

$$\bar{c} = \text{central line}$$
$$LCL = \bar{c} - 3\sigma_c$$
$$UCL = \bar{c} + 3\sigma_c$$

In the monitor stage, the process is deemed stable as long as the number of flaws per subgroup falls within the control limits. In most situations, the lower control limit is not applied.

Example 10.3 Assume a production process where floor tiles are the products, and a blemish in the color is considered a fault. Each tile inspected could have zero or more faults. A pre-study of N = 100 subgroups is planned where one tile is inspected per subgroup. The acceptable limit on the number of faults per tile is 0.15. Below shows the data and computations for the pre-study.

$$N = 100$$
$$\Sigma c_i = 10$$
$$\bar{c} = 10/100 = 0.10$$
$$c_o = 0.15$$

Since, $\bar{c} \leq 0.15$, the process is accepted and the production monitoring stage follows. The parameters for the monitor stage are the central line and the control limits. The standard deviation to c is needed to determine the control limits. The lower control limit is restricted to zero or larger. The computations are below:

$$\text{central line} = 0.10$$
$$\sigma_c = \sqrt{0.10} = 0.32$$
$$LCL = 0.10 - 3 \times 0.32 = -0.86 \rightarrow 0.00$$
$$UCL = 0.10 + 3 \times 0.32 = 1.06$$

10.5.6 U-Chart

The u-chart applies when one or more units are observed during each subgroup, and the number of defectives per subgroup, denoted as c, could be zero or larger. This might be a kitchen cabinet door that could be without any defects, or could have a multiple number detected. At each subgroup, the number of units inspected, and the number of defects is noted. After N subgroups, the average number of units inspected per subgroup, and the average number of defects per unit is computed. Data on the number of units defective per unit, u, follows a Poisson probability distribution. The estimate on the mean and standard deviation on the number of defects per unit is calculated. The management specifies an upper limit on the average number of defects per unit, and if the pre-study results show the process is acceptable, the results are used to compute the parameters for monitoring the system during the monitor stage. Below describes the steps to take in the pre-study stage and in the monitor stage.

10.5.6.1 Pre-Study Stage

The pre-study stage for the u-chart consists of data collection for each of N subgroups, with one or more units inspected per subgroup. After all the subgroup results are in, the subsequent computations are measured, and the process capability is computed. In the event the process is deemed not acceptable, the process is reexamined to find the cause, fix the problem, and then, rerun the pre-study. When the process is acceptable, the parameters for the monitor stage are computed.

Data Collection

The data collected during the pre-study stage is the number of units inspected, and number of defects per subgroup. Below is a list of the notation and data.

N = number of subgroups
$i = 1$ to N = subgroups
n_i = number of units observed in subgroup i
c_i = number of defectives detected in subgroup i

Computations

From the data collected, the following computations finds the sum of the units observed, the sum of the defects in N subgroups, also the average number of units observed per subgroup, and the average number of defects per unit.

$$\Sigma n_i = \text{number of units observed in N subgroups}$$
$$\Sigma c_i = \text{number of defects in N subgroups}$$
$$\bar{n} = \Sigma n_i / N = \text{average number units observed in a subgroup}$$
$$\bar{u} = \Sigma c_i / \Sigma n_i = \text{average number defects per unit}$$

Process Capability

To determine the process capability, the management specifies an upper limit, denoted as u_o, on the number of defects per unit, u, that is acceptable. If $\bar{u} > u_o$, the process is not acceptable, and the cause must be found and corrected, whereupon, the pre-study is rerun. If $\bar{u} \leq u_o$, the process is deemed acceptable.

$$u_o = \text{acceptable upper limit on u}$$

10.5.6.2 Monitor Stage

The monitor stage begins with computing the parameters for the control charts. This includes, the central line, and the standard deviation of the defectives for a subgroup. With this information, the lower and upper control limits on the number of defectives in a subgroup are calculated. The lower control limit is not allowed to fall below zero.

Standard Deviation of u

The standard deviation of u is computed as shown below, assuming u follows a Poisson probability distribution. This is needed to compute the control limits of u.

$$\sigma_u = \sqrt{\bar{u}/\bar{n}} = \text{standard deviation of u}$$

Control Chart Parameters

The parameters (central line, and control limits) to develop the control chart for the u-chart are shown below.

$$\bar{u} = \text{central line}$$
$$LCL = \bar{u} - 3\sigma_u$$
$$UCL = \bar{u} + 3\sigma_u$$

In the monitor stage, the process is deemed stable and in statistical control as long as the number of flaws per subgroup falls within the control limits.

Example 10.4 Assume a production process where cement bricks are the products, and a crack in the brick is a defect. Each brick inspected could have zero or more faults. A pre-study of N = 50 subgroups is planned where one or more bricks are inspected per subgroup. The acceptable limit on the number of defects per brick is 0.16. Below shows the data and computations for the pre-study.

$$N = 50$$
$$\Sigma n_i = 120$$
$$\Sigma c_i = 15$$
$$\bar{n} = 120/50 = 2.40$$
$$\bar{u} = 15/120 = 0.125$$
$$u_o = 0.16$$

Since, $\bar{u} \leq 0.16$, the process is accepted and the production monitoring stage follows. The parameters for the monitor stage are the central line and the control limits. The standard deviation to u is needed to determine the control limits. The lower control limit is restricted to zero or larger. The computations are below:

$$\text{central line} = 0.125$$
$$\sigma_u = \sqrt{0.125} = 0.354$$
$$LCL = 0.125 - 3 \times 0.354 = -0.937 \rightarrow 0.00$$
$$UCL = 0.125 + 3 \times 0.354 = 1.187$$

The process is stopped, if during a production subgroup, the average number of defects per unit falls outside of the control limits.

10.6 Variable Type Data

Variable type data applies when a sample output from a production process is a quantitative measure. The measure is compared to specification limits to assess whether the unit is acceptable. The samples are taken in subgroups at prearranged intervals of time. The variable type data are of four different kind and are labeled as below:

- x-bar and R chart
- Median and R chart
- Individual chart
- x-bar and s chart

10.6.1 Pre-study Stage

At the outset on the analysis of a process, a series of subgroups are set and with each subgroup, the process is stopped at constant intervals of time and some output units from the process are observed and measured for quality. After several subgroups have passed, the data gathered are summarized to determine if the process is conforming with the specifications set by the engineers. If not, the process is stopped, the cause is found, corrected, and the pre-study stage of data gathering is rerun. If the process is conforming with the specifications, the data gathered is used to set control limits for future monitoring of the process.

10.6.2 Specification Limits

The quality standard for the output units, from the process, is given by way of upper and lower specification limits. As long as the output measures fall within the limits, the units are in conformance.

10.6.3 Six Sigma

Six Sigma is a quality goal that seeks zero defects in the output from any process in manufacturing or service. It is a data analysis method that aims to eliminate defects by striving to spread six standard deviations between the average measure of a process and the closest specification limit. Although this is management's desire for the process, it is sometimes difficult to attain. Statistical means on how close this goal can be reached for a process is measured and are labeled as process potential and process capability.

10.6.4 Process Potential

The process potential is a measure between the variability of a process and the limits given by the management. These limits specify the acceptable range of quality from the outputs of the process and are labeled as (LSL and USL) and are defined as lower specification limit, and upper specification limit, respectively. This measure assumes the central line is perfectly situated half way between the specification limits. The goal is for all output products from the process to fall between these limits. The variability of the process is one of the measures of the pre-study stage and is denoted as σ, the standard deviation. With the variability of the process measured, the best capability possible for the process is called the process potential, labeled as C_p. When C_p is one or larger, the better the quality of the process. At C_p of two or larger, the six sigma goal is possible. When C_p falls below one, defectives in the process are expected.

10.6.5 Process Capability

The process capability measures how the variability of the process is actually related to the specification limits. The location of the central line is paramount, and the measure is based on where the central line actually falls with respect to the specification limits. The process capability is measured from an index labeled as C_{pk}. The larger C_{pk}, the better. If C_{pk} is larger than one, almost no defects should occur in the process. When C_{pk} is two or larger, six sigma perfection is achieved. In the event C_{pk} is below one, defectives can be expected. If C_{pk} is too low, the cause of the faults should be investigated, corrected, and the pre-study stage repeated.

10.6.6 Monitor Stage

Upon acceptance of the process capability, the data gathered from the pre-study stage is used to compute control limits for monitoring the process over the future production. Upper and lower control limits are now computed for the output measures that pertain. In the future, subgroups are arranged as in the pre-study where samples from the process output are observed, and the measures of the output are obtained as before.

10.6.7 Statistical Coefficients

Table 10.2 lists a series of statistical coefficients that are used for three of the variable type data: x-bar and R chart, median and R chart, and individual charts. The data gives the coefficient values when the sample size, n, is 2–10. These coefficients are applied subsequently to estimate the standard deviation of the output measure, and to compute the control limits on the output measures.

10.6.8 X-Bar and R Chart

The x-bar and R-chart method is used when measures from the output units of the process are taken to access the quality of the process. The average of the units and the range are computed for each subgroup.

Table 10.2 Coefficients for the x-bar and R chart, median and R chart, and individual chart, when n = 2–10

n	A_2	D_3	D_4	d_2
2	1.88	–	3.27	1.13
3	1.02	–	2.57	1.69
4	0.73	–	2.28	2.06
5	0.58	–	2.11	2.33
6	0.48	–	2.00	2.53
7	0.42	0.08	1.92	2.70
8	0.37	0.14	1.86	2.85
9	0.34	0.18	1.82	2.97
10	0.31	0.22	1.78	3.08

10.6.8.1 Pre-study Stage

In the pre-study stage of an x-bar and R chart process, a fixed number of subgroups are scheduled where the time interval (hours, days, so forth) between the subgroups are set in advance. At each subgroup, the process is stopped and a fixed number of units are observed and measured for quality in a data collection manner. After a number of subgroups are run, computations of the data follow to establish the various parameters related to the process output.

Data Collection

A total of N subgroups are run in the pre-study stage. At each subgroup, n output units are selected and a measure of the quality is taken. The average of the n samples and the range from high to low is measured for each subgroup, as shown below:

$$\begin{aligned} \mathrm{N} &= \text{number of subgroups} \\ \mathrm{i} &= 1 \text{ to } \mathrm{N} = \text{subgroups} \\ \mathrm{n} &= \text{samples per subgroup} \\ (x_{1i}, \ldots, x_{ni}) &= \text{n samples of subgroup i} \\ \bar{x}_i &= \text{average of subgroup i} \\ R_i &= \text{range of subgroup i} \end{aligned}$$

Computations

After the data from N subgroups have been collected, a series of computations follow to estimate the parameters that are measured for future use. The parameters are the average, range, and standard deviation of the output. To measure the standard deviation, a coefficient from Table 10.2, denoted as d_2, is needed. The data collected is below.

$$\Sigma \bar{x}_i = \text{sum of N subgroup averages}$$
$$\Sigma R_i = \text{sum of N subgroup ranges}$$
$$\bar{\bar{x}} = \Sigma \bar{x}_i / \text{N} = \text{average of N subgroup averages}$$
$$\bar{R} = \Sigma \text{R}_\text{i} / \text{N} = \text{average of N subgroup ranges}$$
$$\sigma = \bar{R} / \text{d}_2 = \text{standard deviation of } \bar{x}$$

Process Potential and Process Capability

The next measure concerns the capability of the process with respect to the engineering standards. Specification limits on the output measure is provided from engineering and are listed as below:

$$\text{LSL} = \text{lower specification limit}$$
$$\text{USL} = \text{upper specification limit}$$

The process potential, denoted as C_p, is measured as follows:

$$C_p = [\text{USL} - \text{LSL}]/6\sigma$$

The process capability, labeled as C_{pk}, is computed below:

$$\text{kU} = \left[\text{USL} - \bar{\bar{x}}\right]/3\sigma$$
$$\text{kL} = \left[\bar{\bar{x}} - \text{LSL}\right]/3\sigma$$
$$C_{pk} = \min[\text{kU}, \text{kL}]$$

10.6.8.2 Control Limits

When the process is deemed capable with respect to the engineering standards, the next step is to measure control limits on: $\bar{x}$ and R, that are used to monitor the process in the future. The computation of the control limits requires the use of coefficients, A_2, D_3, D_4, from Table 10.2.

$$\text{LCL}_{\bar{x}} = \bar{\bar{x}} - A_2\bar{R}$$
$$\text{UCL}_{\bar{x}} = \bar{\bar{x}} + A_2\bar{R}$$
$$\text{LCL}_\text{R} = D_3\bar{R}$$
$$\text{UCL}_\text{R} = D_4\bar{R}$$

Example 10.5 Assume the ABC firm is analyzing a machining process using the x-bar and R chart method. The analysis will cover $N = 8$ subgroups, one at every 4 h, and the number of samples per subgroup will be $n = 5$. To illustrate, the output and measures from the first subgroup ($i = 1$) is listed below:

$$\begin{aligned} i &= 1 \\ (x_{11}, \ldots, x_{51}) &= (94, 95, 95, 96, 94) \\ \bar{x}_1 &= [94 + 95 + 95 + 96 + 94]/5 = 94.8 \\ R_1 &= (96 - 94) = 2 \end{aligned}$$

After the N = 8 subgroups, the computations are summarized. The example assumes the sum of the subgroup x-bars is 770, and the sum from the ranges is 18. Note also, the coefficient from Table 10.2 is $d_2 = 2.33$.

$$\begin{aligned} \Sigma\bar{x}_i &= 770 \\ \Sigma R_i &= 18 \\ \bar{\bar{x}} &= \Sigma\bar{x}_i/N = 770/8 = 96.25 \\ \bar{R} &= \Sigma R_i/N = 18/8 = 2.25 \\ \sigma &= \bar{R}/d_2 = 2.25/2.33 = 0.966 \end{aligned}$$

Suppose the specification limits (LSL, USL) from engineering are as listed below. With these, the process potential, $C_p = 1.72$ is measured, indicating the process could be in conformance if the central line is proper. The process capability is also measured and becomes $C_{pk} = 1.29$, which indicates very few defects will occur, but is not in the quality class of six-sigma.

$$\begin{aligned} LSL &= 90 \\ USL &= 100 \\ C_p &= [100 - 90]/[6 \times 0.966] = 1.72 \\ kL &= [96.25 - 90]/[3 \times 0.966] = 2.16 \\ kU &= [100 - 96.25]/[3 \times 0.966] = 1.29 \\ C_{pk} &= \min[1.29, 2.16] = 1.29 \end{aligned}$$

Assuming the management accepts the quality of the process, the next step is to compute the control limits for the subgroup output measures, $\bar{x}$ and R. In the monitoring stage, the output measures are plotted and compared to the limits to ensure the process output is stable and in conformance with the standards set by engineering.

$$\begin{aligned} LCL_{\bar{x}} &= 96.25 - 0.58 \times 2.25 = 94.94 \\ UCL_{\bar{x}} &= 96.25 + 0.58 \times 2.25 = 97.56 \end{aligned}$$

$$\begin{aligned} LCL_R &= 0.00 \times 2.25 = 0.00 \\ UCL_R &= 2.11 \times 2.25 = 4.75 \end{aligned}$$

10.6.9 Median and R Chart

In the pre-study stage of a median and R-chart process, a fixed number of subgroups are scheduled where the time interval (hours, days, so forth) between the subgroups is set in advance. Somewhere in each subgroup, the process is stopped and a fixed number of outputs are observed and measured for quality in a data collection manner. After a number of subgroups are run, computations of the data follow to establish the various parameters related to the process output.

10.6.9.1 Data Collection

A total of N subgroups are run in the pre-study stage. At each subgroup, n output units are selected and a measure of the quality is taken. The median of the n samples and the range from high to low is measured for each subgroup, as shown below:

$$\begin{aligned} \mathrm{N} &= \text{number of subgroups} \\ \mathrm{i} &= 1 \text{ to } \mathrm{N} = \text{subgroups} \\ \mathrm{n} &= \text{samples per subgroup} \\ (\mathrm{x}_{1\mathrm{i}}, \ldots, \mathrm{x}_{\mathrm{ni}}) &= \mathrm{n} \ \text{ samples of subgroup i} \\ \tilde{x}_i &= \text{median of subgroup i} \\ \mathrm{R}_\mathrm{i} &= \text{range of subgroup i} \end{aligned}$$

10.6.9.2 Computations

After the data from N subgroups have been collected, a series of computations follow to estimate the parameters that are measured for future use. The parameters are the median, range, and standard deviation of the output. To measure the standard deviation, a coefficient from Table 10.2, denoted as d_2, is needed.

$$\begin{aligned} &\sum \tilde{x}_i = \text{sum of N subgroup medians} \\ &\sum \mathrm{R}_\mathrm{i} = \text{sum of N subgroup ranges} \\ &\overline{\tilde{x}} = \sum \tilde{x}_i/\mathrm{N} = \text{average of N subgroup medians} \\ &\overline{R} = \sum \mathrm{R}_\mathrm{i}/\mathrm{N} = \text{average of N subgroup ranges} \\ &\sigma = \overline{R}/\mathrm{d}_2 = \text{standard deviation of } \tilde{x} \end{aligned}$$

10.6.9.3 Process Potential and Process Capability

The next decision concerns the capability of the process with respect to the engineering standards. Specification limits on the output measure is provided from engineering and are listed as below:

$$\begin{aligned} \mathrm{LSL} &= \text{lower specification limit} \\ \mathrm{USL} &= \text{upper specification limit} \end{aligned}$$

The process potential, denoted as C_p, is measured as follows:

$$C_p = [USL - LSL]/6\sigma$$

The process capability, labeled as C_{pk}, is computed below:

$$kU = \left[USL - \overline{\widetilde{x}}\right]/3\sigma$$
$$kL = \left[\overline{\widetilde{x}} - LSL\right]/3\sigma$$
$$C_{pk} = \min[kU, kL]$$

10.6.9.4 Control Limits

When the process is deemed capable with respect to the engineering standards, the next step is to measure control limits on: $\widetilde{x}$ and R, that are used to monitor the process in the future. The computation of the control limits requires the use of coefficients, A_2, D_3, D_4, from Table 10.2.

$$LCL_{\widetilde{x}} = \overline{\widetilde{x}} - A_2\overline{R}$$
$$UCL_{\widetilde{x}} = \overline{\widetilde{x}} + A_2\overline{R}$$
$$LCL_R = D_3\overline{R}$$
$$UCL_R = D_4\overline{R}$$

Example 10.6 Assume the Acme Corp. is analyzing a machining process using the Median chart method. The analysis will cover N = 10 subgroups, one each day, and the number of samples per subgroup will be n = 4. To illustrate, the output and measures from the first subgroup (i = 1) is listed below:

$$i = 1$$
$$(x_{11}, \ldots, x_{41}) = (10.33, 10.35, 10.34, 10.32)$$
$$\widetilde{x}_1 = [10.34 - 10.33]/2 = 10.335$$
$$R_1 = \left(10.35 - 10.32\right] = 0.03$$

After the N = subgroups, the computations are summarized. The example assumes the sum of the subgroup medians is 103.20, and the sum from the ranges is 0.36. Note also, the coefficient from Table 10.2 is $d_2 = 2.06$.

$$\sum \widetilde{x}_i = 103.20$$
$$\sum R_i = 0.36$$
$$\overline{\widetilde{x}} = \sum \widetilde{x}_i/N = 103.20/10 = 10.32$$
$$\overline{R} = \sum R_i = 0.36/10 = 0.036$$
$$\sigma = \overline{R}/d_2 = 0.036/2.06 = 0.017$$

Suppose the specification limits (LSL, USL) from engineering are as listed below. With these, the process potential is $C_p = 4.90$, indicating the process could reach six sigma quality if the central line is proper. The process capability is also measured and becomes $C_{pk} = 3.52$, which signifies the process is in the quality class of six sigma.

$$
\begin{aligned}
\mathrm{LSL} &= 10.00 \\
\mathrm{USL} &= 10.50 \\
C_p &= [10.50 - 10.00]/[6 \times 0.017] = 4.90 \\
\mathrm{kL} &= [10.32 - 10.00]/[3 \times 0.017] = 6.27 \\
\mathrm{kU} &= [10.50 - 10.32]/[3 \times 0.017] = 3.52 \\
C_{pk} &= \min[6.27, 3.52] = 3.52
\end{aligned}
$$

The next step is to compute the control limits for the subgroup output measures, $\widetilde{x}$ and R. In the monitoring stage, the output measures $(\widetilde{x}, R)$ are plotted and compared to the limits to ensure the process output is stable and in conformance with the standards set by engineering.

$$
\begin{aligned}
\mathrm{LCL}\widetilde{x} &= 10.32 - 0.73 \times 0.036 = 10.29 \\
\mathrm{UCL}\widetilde{x} &= 10.32 + 0.73 \times 0.036 = 10.35
\end{aligned}
$$

$$
\begin{aligned}
\mathrm{LCL_R} &= 0.00 \times 0.036 = 0.00 \\
\mathrm{UCL_R} &= 2.28 \times 0.036 = 0.08
\end{aligned}
$$

10.6.10 Individual Chart

In the pre-study stage of an Individual chart process, a fixed number of subgroups are scheduled where the time interval (hours, days, so forth) between the subgroups are set in advance. During each subgroup, the process is stopped and an output is observed and measured for quality in a data collection manner. After a number of subgroups are run, computations of the data follow to establish the various parameters related to the process output.

10.6.10.1 Data Collection

A total of N subgroups are run in the pre-study stage. At each subgroup, one output unit is selected and a measure of the quality is taken. The data from the N subgroups is below:

$$
\begin{aligned}
\mathrm{N} &= \text{number of subgroups} \\
\mathrm{i} &= 1 \text{ to } \mathrm{N} = \text{subgroups} \\
\mathrm{n} &= \text{one sample per subgroup} \\
x_i &= \text{sample of subgroup } i(i = 1 \text{ to } N) \\
R_i &= |x_i - x_{i-1}| = \text{range of subgroup } i(i = 2 \text{ to } N)
\end{aligned}
$$

10.6.10.2 Computations

After the data from N subgroups have been collected, a series of computations follow to estimate the parameters that are measured for future use. The parameters are the average, range, and standard deviation of the output. To measure the standard deviation, a coefficient from Table 10.2, denoted as d_2, is needed.

$$\begin{aligned}
\Sigma x_i &= \text{sum of N subgroup individuals} \\
\Sigma R_i &= \text{sum of N} - 1 \text{ subgroup ranges} \\
\bar{x} &= \Sigma x_i / N = \text{average of N subgroup individuals} \\
\bar{R} &= \Sigma R_i / (N - 1) = \text{average of } (N - 1) \text{ subgroup ranges} \\
\sigma &= \bar{R} / d_2 = \text{standard deviation of } \bar{x}
\end{aligned}$$

10.6.10.3 Process Potential and Process Capability

The next decision concerns the capability of the process with respect to the engineering standards. Specification limits on the output measure is provided from engineering and are listed as below:

$$\begin{aligned}
\text{LSL} &= \text{lower specification limit} \\
\text{USL} &= \text{upper specification limit}
\end{aligned}$$

The process potential, denoted as C_p, is measured as follows:

$$C_p = [\text{USL} - \text{LSL}]/6\sigma$$

The process capability, labeled as C_{pk}, is computed below:

$$\begin{aligned}
kU &= [\text{USL} - \tilde{x}]/3\sigma \\
kL &= [\tilde{x} - \text{LSL}]/3\sigma \\
C_{pk} &= \min[kU, kL]
\end{aligned}$$

10.6.10.4 Control Limits

When the process is deemed capable with respect to the engineering standards, the next step is to measure control limits on: x and R, that are used to monitor the process in the future. The computations of the control limits require the use of coefficients, A_2, D_3, D_4, from Table 10.2. The sample size of $n = 2$ is assumed.

$$\begin{aligned}
LCL_x &= \bar{x} - (3/d_2)\bar{R} \\
UCL_x &= \bar{x} + (3/d_2)\bar{R} \\
LCL_R &= D_3\bar{R} \\
LCL_R &= D_4\bar{R}
\end{aligned}$$

Example 10.7 Assume the Acme Corp. is analyzing a machining process using the individual chart method. The analysis will cover N = 20 subgroups, and the number of samples per subgroup is one. To illustrate, the output and measures from the first two subgroups (i = 1 and 2) are listed below:

$$\begin{aligned} i &= 1 \ \ (x_1 = 201.35) \\ 2 &= 2 \ \ (x_2 = 200.98) \\ R_2 &= |200.98 - 201.35| = 0.37 \end{aligned}$$

After the N = 20 subgroups, the computations are summarized. The example assumes the sum of the subgroup x′s is 4028.45, and the sum from the ranges is 8.16. Note also, the coefficient from Table 10.2 is $d_2 = 1.13$.

$$\sum x_i = 4028.45$$

$$\sum R_i = 8.16$$

$$\bar{x} = \sum x_i/N = 4028.45/20 = 201.42$$

$$\bar{R} = \sum R_i/(N - 1) = 8.16/19 = 0.43$$

$$\sigma = \bar{R}/d_2 = 0.43/1.13 = 0.38$$

Suppose the specification limits (LSL, USL) from engineering are as listed below. With these, the process potential, $C_p = 1.75$ is measured, indicating the process is not in six sigma quality. The process capability is also measured and becomes $C_{pk} = 1.25$, which signifies the process is not in the quality class of six sigma.

$$\begin{aligned} LSL &= 200 \\ USL &= 204 \\ C_p &= [204 - 200]/[6 \times 0.38] = 1.75 \\ kL &= [201.42 - 200]/[3 \times 0.38] = 1.25 \\ kU &= [204 - 201.42]/[3 \times 0.38] = 2.26 \\ C_{pk} &= \min[1.25, 2.26] = 1.25 \end{aligned}$$

If the quality is accepted, the next step is to compute the control limits for the subgroup output measures, x and R. In the monitoring stage, the output measures are plotted and compared to the limits to ensure the process output is stable and in conformance with the standards set by engineering.

$$\begin{aligned} LCL_x &= 201.42 - (3/1.13) \times 0.43 = 200.28 \\ UCL_x &= 201.42 + (3/1.13) \times 0.43 = 202.56 \end{aligned}$$

$$\begin{aligned} LCL_R &= 0.00 \times 0.43 = 0.00 \\ UCL_R &= 3.27 \times 0.43 = 1.41 \end{aligned}$$

Table 10.3 Coefficients for the x-bar and s chart, when n = 2–20

n	A_3	B_3	B_4	c_4
2	2.69	–	3.26	0.79
3	1.95	–	2.56	0.88
4	1.62	–	2.26	0.92
5	1.42	–	2.08	0.94
6	1.28	0.03	1.97	0.95
7	1.18	0.11	1.88	0.95
8	1.09	0.18	1.81	0.96
9	1.03	0.23	1.76	0.96
10	0.97	0.28	1.71	0.97
11	0.92	0.32	1.67	0.97
12	0.88	0.35	1.64	0.97
13	0.85	0.38	1.61	0.97
14	0.81	0.40	1.59	0.97
15	0.78	0.42	1.57	0.98
16	0.76	0.44	1.55	0.98
17	0.73	0.46	1.53	0.98
18	0.71	0.48	1.51	0.98
19	0.69	0.49	1.50	0.98
20	0.68	0.51	1.49	0.98

10.6.11 Statistical Coefficients

Table 10.3 lists a series of statistical coefficients that are used for the variable type data: x-bar and s chart. The data gives the coefficient values when the sample size, n, is 2–20. These coefficients are applied to estimate the standard deviation of the output measure, and to compute the control limits on the output measures.

10.6.12 X-Bar and s Chart

The x-bar and s-chart method is used when measures from the output units of the process are taken to access the quality of the process. The average of the units and the standard deviation are computed for each subgroup.

10.6.12.1 Pre-study Stage

In the pre-study stage of an x-bar and s chart process, a fixed number of subgroups are scheduled where the time interval (hours, days, so forth) between the subgroups are set in advance. Somewhere in each subgroup, the process is stopped and a fixed number of outputs are observed and measured for quality in a data collection manner. After a number of subgroups are run, computations of the data follow to establish the various parameters related to the process output.

Data Collection

A total of N subgroups are run in the pre-study stage. At each subgroup, n output units are selected and a measure of the quality is taken. The average of the n samples and the standard deviation are measured for each subgroup, as shown below:

$$N = \text{number of subgroups}$$
$$i = 1 \text{ to } N = \text{subgroups}$$
$$n = \text{samples per subgroup}$$
$$(x_{1i}, \ldots, x_{ni}) = \text{n samples of subgroup i}$$
$$\overline{x}_i = \text{average of subgroup i}$$
$$s_i = \sqrt{\sum \frac{(x_{ji} - \overline{x}_i)^2}{n-1}} = \text{standard deviation of subgroup i where } j = 1 \text{ to } n$$

Computations

After the data from N subgroups have been collected, a series of computations follow to estimate the parameters that are measured for future use. The parameters are the average, average of standard deviation, and standard deviation of x-bar. To measure the standard deviation of x-bar, a coefficient from Table 10.3, denoted as c_4, is needed.

$$\sum \overline{x}_i = \text{sum of N subgroup averages}$$
$$\sum s_i = \text{sum of N subgroup standard deviations}$$
$$\overline{\overline{x}} = \sum \overline{x}_i / N = \text{average of N subgroup averages}$$
$$\overline{s} = \sum s_i / N = \text{average of N subgroup standard deviations}$$
$$\sigma = \overline{s}/c_4 = \text{standard deviation of } \overline{x}$$

Process Potential and Process Capability

The next decision concerns the capability of the process with respect to the engineering standards. Specification limits on the output measure is provided from engineering and are listed as below:

$$LSL = \text{lower specification limit}$$
$$USL = \text{upper specification limit}$$

The process potential, denoted as C_p, is measured as follows:

$$C_p = [USL - LSL]/6\sigma$$

The process capability, labeled as C_{pk}, is computed below:

$$kU = \left[USL - \bar{\bar{x}}\right]/3\sigma$$
$$kL = \left[\bar{\bar{x}} - LSL\right]/3\sigma$$
$$C_{pk} = \min[kU, kL]$$

10.6.12.2 Control Limits

When the process is deemed capable with respect to the engineering standards, the next step is to measure control limits on $\bar{x}$ and s, that are used to monitor the process in the future. The computations of the control limits require the use of coefficients, A_3, B_3, B_4, c_4, from Table 10.3.

$$LCL_{\bar{x}} = \bar{\bar{x}} - A_3\bar{s}$$
$$UCL_{\bar{x}} = \bar{\bar{x}} + A_3\bar{s}$$
$$LCL_s = B_3\bar{s}$$
$$UCL_s = B_4\bar{s}$$

Example 10.8 Assume the ABC firm is analyzing a machining process using the x-bar and s-chart method. The analysis will cover N = 10 subgroups, one each 4 h, and the number of samples per subgroup will be n = 15. To illustrate, the output and measures from the first subgroup (i = 1) is listed below:

$$i = 1$$
$$(x_{11}, \ldots, x_{151}) = (1.02, 1.01, 1.01, 1.00, 1.01, 1.00, 1.02, 1.01, 1.01, 1.01, 1.02, 1.00, 1.02, 1.00, 1.01)$$
$$\bar{x}_1 = 15.15/15 = 1.01$$
$$s_1 = \sqrt{\sum_i (x_i - 1.01)^2/14} = 0.0076$$

After the N = 10 subgroups, the computations are summarized. The example assumes the sum of the subgroup x-bars is 10.16, and the sum from the standard deviations is 0.085. Note also, the coefficient from Table 10.3 is $c_4 = 0.98$.

$$\sum \bar{x}_i = 10.16$$

$$\sum s_i = 0.085$$

$$\bar{\bar{x}} = \sum \bar{x}_i/N = 10.16/10 = 1.016$$

$$\bar{s} = \sum s_i = 0.085/10 = 0.0085$$

$$\sigma = \bar{s}/c_4 = 0.0085/0.98 = 0.0087$$

Suppose the specification limits (LSL, USL) from engineering are as listed below. With these, the process potential, $C_p = 1.53$ is measured, indicating the process could be in conformance, and not in six sigma quality. The process capability is also measured and becomes $C_{pk} = 0.61$, which signifies many defects could occur, and this is because the central line is not positioned near the center of the specification limits.

$$
\begin{aligned}
\text{LSL} &= 1.00 \\
\text{USL} &= 1.08 \\
C_p &= [1.08 - 1.00]/[6 \times 0.0087] = 1.53 \\
\text{kL} &= [1.016 - 1.00]/[3 \times 0.0087] = 0.61 \\
\text{kU} &= [1.08 - 1.019]/[3 \times 0.0087] = 2.34 \\
C_{pk} &= \min[0.61, 2.34] = 0.61
\end{aligned}
$$

Assuming the management accepts the quality of the process, the next step is to compute the control limits for the subgroup output measures, $\bar{x}$ and s. In the monitoring stage, the output measures are plotted and compared to the limits to ensure the process output is stable and in conformance with the standards set by engineering.

$$
\begin{aligned}
\text{LCL}_{\bar{x}} &= 1.016 - 0.78 \times 0.0085 = 1.009 \\
\text{UCL}_{\bar{x}} &= 1.016 + 0.78 \times 0.0085 = 1.022
\end{aligned}
$$

$$
\begin{aligned}
\text{LCL}_s &= 0.42 \times 0.0085 = 0.0036 \\
\text{UCL}_s &= 1.57 \times 0.0085 = 0.0133
\end{aligned}
$$

10.7 Summary

Four methods to study and monitor attribute type data are described. These are the p-chart, np-chart, c-chart and u-chart. A pre-study is arranged with N subgroups where samples are taken from each group to determine whether each unit is defective or not. At the end of the N subgroups, statistics are computed to determine whether the process is capable, and if so, future subgroups are run to ensure the process stays in an acceptable zone. Also, four methods are described for processes where the output measures are of the variable type: x-bar and R, median and R, individual, and x-bar and s. A pre-study is arranged for the process and samples are taken from each. Statistical measures of the process potential and process capability are calculated. The ideal goal is to have a process whose output measures are within a six-sigma range from the specification limits. If the process is capable, control limits are computed and future subgroups are run. In the monitoring stage, as long as the output falls within the control limits, the process remains in statistical control and is capable.

Chapter 11
Distribution Network

11.1 Introduction

In a typical original equipment manufacturing (OEM) system, the suppliers ship stock to one or more distribution centers, that serves as the source to a series of retailers (dealers, stores), and the retailers sell the products to the customers. This could be a network that holds service parts for an automotive corporation. The dealers sell the autos to the customers and when repair or maintenance is needed, the customer seeks service from the dealer. The dealer carries a limited supply of parts for this purpose, and relies on the DC to have a full set of parts as needed. The OEM is faced with holding the minimal amount of stock needed to properly service the demands from the dealers. In the ideal situation, the DC system consists of a network of locations that are strategically near the dealers and to some extent, near to the suppliers. One location is the master stock location that serves as the headquarters for the entire system. The other locations are branch locations. The master location is often a much larger location and holds more variety of stock. Sometimes a small location is provided and is called a 2-level-service location. This location holds minimal stock and is supplied by the master stock location. The parts that have a high amount of demand are stocked in most locations and are called distributed parts. The low demand parts are labeled as non-distributed, and are mostly stocked only from the master stock location.

Periodically, the management runs an analysis to seek what is the best arrangement of the DC network. The analysis seeks answers to: how many locations, where should the locations be sited, and how to allot the stock by locations. A series of parameters are provided in a 'what if' arrangement. The management can run a variety of options with various setting of the parameters and the system provides summary results for each option.

Sometimes the parts supplier does not package the parts as needed by the DC system, and thereby sends the units produced to a packager who places the parts in containers with appropriate labeling. The packager then ships the packaged parts to

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_11

the DC network. Cross-docking is also a strategy where the suppliers ship in large vehicles to a cross-dock facility and the goods are taken off the larger vehicles, sorted and placed in smaller vehicles for transport directly to the stocking facility (DC or dealer). Back at the DC network, the integrity of the inventory count is vital to the efficiency of the service to customers. It behooves the DC network to have a continual cycle counting system in place to maintain the count of the amount of units in stock for each part. Three cycle counting methods are popular: random, ABC, and Pareto. Queuing theory is introduced showing how it has a wide variety of applications in the operation of a network.

11.2 Network

A distribution network is an intermediary system between suppliers of goods and the end users. This typically consists of a series of warehouses that receive stock from the suppliers and houses the inventory until they are redistributed to dealers, stores, retailers, and sometimes straight to the customers. The warehouse is often called a distribution center and is denoted as DC. In the typical situation, more than one DC is assigned to the network. The total customer region is partitioned where each DC serves the retailers in its assigned region. The DC's may have much space for storage to receive large quantities of stock at regular intervals. They serve as the source to the retailers. The retailers desire is to stock a wide variety of the products, but have limited space and thereby they have minimal backup inventory. The lead-time from the suppliers to the DC's may take one or more months, while the lead-time from the DC to the retailers is often one or more days.

11.2.1 Master and Branch Locations

The distribution network system consists of a number of locations, called branch locations. One of the branch locations is the master location, denoted as ML. The master location is typically centrally located and is the leader, as well as the largest of the locations. This frequently is where the headquarters for the network is housed. Often the suppliers will only ship to the ML, and the ML then redistributes to the other locations in the network. The total territory is partitioned into regions, and the ideal goal is for each branch to fill the demands of the retailers in its region. But this goal is overridden for low demands items as is described subsequently.

11.2.2 2-Level-Service Locations

Another type of location that is sometimes found in a distribution network is a 2-level-service location, denoted as 2LS. These locations have a small customer territory assigned to them, and is housed in a smaller facility with limited storage space. The ML is the source for the 2LS location; hence, because it takes two levels of replenishment for stock to reach the 2LS location—from supplier to ML to 2LS—this location is called a 2-level-service location. When stock runs low at the 2LS location, a replenish order to the ML is called. The replenish size is relatively smaller than from a typical branch location. In this way, the ML has the dual role of serving as the source to all the 2LS locations.

11.2.3 Distributed and Non-distributed Parts

Not all parts are stocked in all of the branch locations. Often the management has a rule on how to separate the parts between high and low demands. The low demand parts are assigned as non-distributed parts, and all others are distributed. The low demand parts have small demands, whereby it is not economical to stock these parts in all of the branches, and instead, it is only in the ML. This way, all the branch demands of this type are directed to the ML and the ML ships to the customers accordingly. This is another reason why the ML requires a much larger building for storage than the branches.

11.2.4 Stock Locations

Not all of the locations will carry stock on a part even though the part is classified as a distributed part. Sometimes there is enough demand on a part to warrant a distributed class. But, the demands may not be high in each location, whereby a location with low demand is not assigned the part. In this situation, only the locations with sufficient demands will stock the part, and all other locations will direct their demands to the ML who has the responsibility to fill the demands. This way, the ML has to have enough stock to also fill the demands of this type as well. The locations that stock the part are called stock locations for the part and those who don't stock the part are non-stock locations.

11.3 Network Design

The distribution network of a firm entails the flow of stock from suppliers to the stock locations, whereupon the stock subsequently flows to the customers. A smooth running network is needed to provide a high level of service to the customers. Costs incur when receiving stock from suppliers, when holding stock at the locations, and when shipping stock to the customers. The goal of a distribution network is to allow flexibility and efficiency in stocking the goods so that the service to the customers is maximized, and the operating costs are minimized. A network algorithm is described here to illustrate a way to analyze the location design of the total system.

As the customer base shifts, the management may review the design of their network distribution system, with the possibility of altering the number of stocking facilities and also the locations of the facilities. More frequently, the management may consider changing the stocking policy on their parts. Various software tools have been developed to assist the management in the decision making of this type. One such network algorithm is described here.

The system requires data on the candidate network locations under consideration. Also needed is a file on all the parts in the system, including the demands they incur, the cost per unit, and so forth. Other data on each part is the demands that come in from different demand locations. A limited number of demand locations are named and the closest network location to each demand location is found for each. This way, the part demands that would occur at each network location can be estimated. Each part is run through the algorithm and various stocking policy parameters are used to determine which parts are stocking and non-stocking; and which are distributed and non-distributed. The system also determines which locations will hold stock on a part and how much. The algorithm estimates the demand by network location, as well as how much stock is needed by location. Further calculations estimate how many times in a year the location will receive stock from the supplier. The aggregate of these statistics from all the parts are summarized for subsequent use to determine the costs for operating the overall system under review.

Below describes how the network algorithm works. This includes a list of the data needed, and a description on how the algorithm computes the statistics for each network location. Further down is an example output from the algorithm.

11.3.1 Network Locations

Table 11.1 identifies the candidate network locations and names the one that will be the master location. For each location, the map coordinates (latitude and longitude) are listed. Note, the coordinates are not always needed in the algorithm; since they are used only when the demand locations do not specify their assigned network

Table 11.1 Candidate network locations with master location, and coordinates

NW locations	Master location	Latitude	Longitude
Toledo	Yes	41.66	−83.56
Portland		36.17	−86.78
Nashville		45.52	−122.67

Table 11.2 Demand locations (DL) with coordinates, and closest Network Location if known, where NDL = (number of demand locations)

DL	Name	Latitude	Longitude	NW_Loc
1	Chicago	–	–	–
...				
NDL	Seattle	–	–	–

location. In the example, the candidate network locations are: Toledo, Portland and Nashville, where Toledo is named the master location. The number of network locations is denoted as Nw and the network locations are labeled as w = 1 to Nw.

11.3.2 Demand Locations

Table 11.2 lists all of the demand locations, denoted as j = 1 to Nj, where Nj is the number of demand locations. The coordinates (latitude and longitude) are given with each demand location. Also listed is the assigned network location (NW_Loc) if it is known. In the event the network location is blank, the algorithm uses the coordinates of the demand location and the same for the network locations to assign to the shortest distance. Either way, the network location, denoted as w, that is assigned to the demand location, j, is denoted as w(j). In total, w(j) is found for each demand location, j = 1 to Nj.

11.4 Parameters

Table 11.3 houses the parameters (SL, h12, Co) used in the algorithm to compute the order quantity and safety stock for each part number. Other parameters (NS, ND, NLS) are used to specify the stocking policy to use. More detail on the definition and use of the parameters is below.

11.4.1 Stock and Non-stock Parts

The rule on determining whether to stock a part in the distribution system uses the parameter NS, and the part measure n12, as they are defined below:

Table 11.3 Example network parameters: SL (service level), h12 (12 month holding rate), Co (cost per order), NS (number of 12 month lines to stock part), ND (number of 12 month lines to distribute part), NLS (number of 12 months lines at location to stock part)

SL	h12	Co	NS	ND	NLS
0.95	0.20	20	2	21	4

$$\text{NS} = \text{min of 12-month part lines to stock the part}$$
$$\text{n12} = \text{number of lines in 12-months for a part}$$

The stocking rule is below:

$$\text{if n12} \geq \text{NS, the part is stocked in the distribution system}$$
$$\text{If n12} < \text{NS, the part is not stocked, and any demand is placed in backorder}$$

For example, if NS = 2, any part with one or less lines in a year is not stocked in the network. Should any demand occur for the part, the demand is placed in a backorder status.

11.4.2 Distributed and Non-distributed Parts

All stock parts are classified as distributed or non-distributed. A part is distributed when the number of lines is sufficient to stock the part in two or more network locations. The part is non-distributed when the number of lines is low and the part is only stocked in the master location. The parameter for this policy is ND and is defined as below:

$$\text{ND} = \text{min of 12-month part lines to distribute the part.}$$

The rule is as follows:

$$\text{if n12} \geq \text{ND, the part is set as a distributed part}$$
$$\text{If n12} < \text{ND, the part is non-distributed}$$

For example, if ND = 21, any part with 20 or less lines in a year is only stocked at the ML.

11.4.3 Distributed and Non-distributed Part Locations

All distributed parts are further scrutinized to determine which locations will stock the part. A part is stocked at network location w if the 12-month lines at that location, n12(w), has a sufficient number of lines. The parameter, NLS is used to specify the system policy as follows:

n12(w) = part 12-month lines at network location w.
NLS = min of 12-month location lines to stock the part.

For example, if NLS = 4, any NW location of a distributed part that has three or less lines in a year will not stock the part, and the ML fills the demands.

11.5 Part Order Quantity

Each part is processed one at a time in the network algorithm. One of the measures computed is the part order quantity, Q. This is obtained in the same way as described in Chap. 3. The data needed to find Q for each part is the following:

d12 = 12-month demand
c = cost per unit
h12 = 12-month holding rate
Co = cost per order

The quantity, d12 is obtained from the part data listed in Table 11.4. Note in the table, the part data includes:

mh = months of history
dh = mh months of demand

The 1-month demand for the part, denoted as d1, is computed by: d1 = dh/mh, and thereby the 12-month demand is:

$$d12 = 12 \times d1$$

Note also, the number of lines in 12 months, denoted as n12, is obtained by,

$$n12 = 12 \times nh/mh$$

Table 11.4 PN (part numbers), c (cost per unit), mh (months history), nh (number lines in mh months), dh (number demands in mh months), cov (coefficient of variation), L (lead time months), dh_j (demand at demand location j)

							dh_j
PN	c	mh	nh	dh	cov	L	1·· ·.. .· · ·NDL
1							
. . .							
NPN							

Hence, the order quantity for the part is obtained as below:

$$Q = \sqrt{2 \times d12 \times Co/(c \times h12)}$$

11.6 Part Safety Stock

The safety stock is computed for the part using the parameter SL from Table 11.3, along with the coefficient of variation, cov, and the lead-time in months, L, from Table 11.4. The service level yields the safety factor, k, as shown in Chap. 4. Finally, the standard deviation and the safety stock for the part are computed as below:

$$\sigma = cov \times d1$$
$$ss = k\sqrt{L}\sigma$$

11.7 Part Cycle Stock and Total Stock

The average cycle stock for the part becomes:

$$cs = Q/2$$

and thereby the average total stock is:

$$ts = cs + ss$$

The number of receives in 12 months for the part is the following:

$$nr12 = d12/Q$$

11.8 Location Measures

The part data in Table 11.4 also includes the demands at each demand location, j = 1 to NDL. Each demand location, j, is associated with one of the network locations, w(j), and the sum of the demands, dh_j, for each location w, yields the demands at location w, denoted as dhw_w, for w = 1 to Nw. Thereby, the portion of demands at location w becomes:

$$p_w = dhw_w/\Sigma dhw_w \quad w = 1 \text{ to } Nw$$

and the portion sum for all locations is,

$$\Sigma p_w = 1$$

The above portions are revised depending on the stocking status at the location.

If the part is classified as non-distributed, the portion for the master location is set to 1.00, and at all other locations, it is set ot 0.00.

If the part is distributed, but location w is not distributed, the portion for w is added to the master location's portion, and the portion at w is set to zero.

The cycle stock, safety stock, total stock and 12 month receives at the individual locations are computed as below:

$$cs_w = p_w cs$$
$$ss_w = p_w ss$$
$$ts_w = cs_w + ss_w$$
$$nr12_w = nr12$$

These measures are cumulated and upon processing all the parts, a table summarizes the results for the candidate network locations and parameter options. The cycle stock, safety stock, total stock and 12-month demands are converted to dollars using the cost per unit of each part.

11.9 Network Summary Table

An example of the main output from the Network Algorithm is summarized in Table 11.5. The data is from an actual run for a construction equipment firm with 82,000 part numbers. In the real analysis, the firm ran many different options each with candidate locations, and with altering the stocking policy parameters. Upon comparing all the output results, the firm was able to select the network configuration to their liking and also the policy parameters to apply.

The columns are the following:

NPN = the number of part numbers
NL12 = number of lines in 12 months
D12 = demand at cost in 12 months
CS = average cycle stock at cost
SS = safety stock at cost
TS = average total stock at cost
NR12 = number receives in 12 months

Table 11.5 Network tallies for: non-stock parts, stock parts, NDP (non-distributed parts), and DP (distributed parts), with estimates of: NPN (number of part numbers), NL12 (12 month lines), D12 (12 month demands), CS (cycle stock), SS (safety stock), TS (total stock), and NR12 (12 month receives)

	NPN	NL12	D12	CS	SS	TS	NR12
	1000	1000	$1,000,000	$1,000,000			1000
Non-Stock Parts	14	13	5				
Stock Parts	69	1413	160	27	39	66	118
Sum	82	1426	165	27	39	66	118
Stock parts:							
NDP	54	337	65	15	22	36	50
DP	15	1076	94	12	17	30	68
Sum	69	1413	160	27	39	66	118
Stock locations:							
Toledo (ML)	69	1123	134	24	33	57	76
Portland	12	167	14	2	3	5	22
Nashville	11	123	12	1	3	4	20
Sum		1413	160	27	39	66	118

The rows are the following:

Non-Stock = parts with low demand that are not stocked in the system
Stock = parts that are stocked in the system
Sum = Non-Stock plus Stock parts

Stocked parts is below:

NDP = non-distributed parts
DP = distributed parts
Sum = NDP + DP parts

Stock locations is below:

Toledo
Portland
Nashville
Sum = aggregate of the stock locations

11.10 Packagers

The packager is another link in the distribution network between a distribution center of an original-equipment-manufacturer (OEM) and a parts supplier (PS). The role of the packager occurs when the OEM sends a replenish order to a PS with a quantity and due date specified; and the PS does not place the goods in the final containers ready for binning and display. The PS fabricates the quantity and upon completion, ships the goods in bulk to a packager designated by the OEM. A packaging bill-of-material (bom) states the type of container with the quantity of units and the labeling instructions. Upon completion, the packager ships the units to the DC of the OEM. The units are now available for proper binning and display at the DC and retail level.

The packager's role is not easy. Shipments of bulk goods come in to the packager each day unannounced with quantities and due dates stated. The packager is obliged to quickly complete its task meeting the due date specified. The OEM keeps a record on the delivery dates to judge the service of the packager. The packager must carry packaging components to meet the quantity and bom needs of the items that they receive. If the packaging components are not available, the packaging process cannot be completed and the units are not shipped to the DC on-time as needed. The components are often not unique, but are common on a variety of parts. The packager receives replenish components from a component supplier (CS).

It behooves the packager to have a forecast and replenish system to control all of the component inventory needed in its operation. To assist the packager in this respect, the OEM should notify the packager when an order goes to the parts supplier with quantity and due date. This way, the packager has a forecast on what items are coming and the quantity of each. It is also important for the packager to have a packaging bom on the items so that it has the inventory needed to complete the order on time.

11.11 Cross-Docking

Cross-docking is another link in the distribution network that involves transporting products that are unloaded from inbound vehicles (trucks, rail cars) and loaded onto smaller vehicles with little to no storage in-between. The cross-dock facility is often of a long linear shape with 150–200 loading docks on the sides and a large space in-between where the products are handled (unloaded, sorted, allocated, moved, loaded). The goods come in from one or more suppliers and are destined to one or more customers, as described below.

When the inbound shipment comes from one supplier, the vehicle is often fully loaded in a truckload (TL) status with goods going to a variety of customers. Upon cross-docking, the goods are placed in smaller vehicles of less-than-truckload (LTL) for delivery to the customers.

On other scenarios, the customer orders the goods from a variety of suppliers and in the cross-docking operation, the goods coming off the inbound vehicles, are sorted, combined and placed on one pallet to go to a single customer.

For retailers with many stores, the retailer buys products from one or more suppliers in larger quantities and the suppliers ship to the cross-dock facility. At the cross-dock, the products are unloaded and are separated for the various retailer outlets and in quantities specified. The products for an individual store are gathered together and placed on pallets for shipment.

Without cross-docking, the products from the supplier are shipped to a warehouse where they are unloaded from a truck, placed in storage, and subsequently are picked and loaded onto another vehicle for delivery to the customer.

With cross-docking, much of the cost with labor and storage is reduced. The warehousing costs are also lowered since the goods are stored in the transport vehicles from manufacturer to retailer, and there is no need for warehouse space. Further, the lead-time is shortened since the timing is mainly from manufacturer to the cross-dock facility and to the customer. With no warehousing, the time from manufacturer to retailer is decreased, and the need for safety stock is minimized.

An example on a different scale is in the produce business in a city center, where large produce trucks arrive early each day at a produce market and unload fruits and vegetables, some fresh from the farms, with others from foreign states or countries. The store representatives arrive early in the morning to buy the produce of their choice; and load them on their smaller trucks to display for sale in their stores. This is what my father-in-law did for over 30 years while managing his grocery store. The trip to the market goes on several times a week to continually keep a fresh batch of produce on display at the stores.

11.12 Cycle Counting

Cycle counting is an auditing method where a small number of items in the inventory is selected for counting on a particular day. The purpose is to compare the physical count with the recorded count on a database, and if a difference, adjust the database. In this way, a number of items are counted each day (or week), and over time, all or most of the items in the inventory are counted. A main purpose of counting is to optimize the business operations to achieve a high level of customer service.

Various methods are in use to establish the counting system. Three of the most common methods are listed here, and are the following: Random sampling, ABC, and Pareto. A description on each is provided below.

11.12.1 Random Sampling Cycle Count

Management chooses to randomly sample n items a day, and either of two ways are used to select the items to count: the constant population count or the diminishing population count.

- When the constant population count is used, n items from the population of N items are selected at random and these are the items to count for the day.
- When the diminishing population count is used, n items that have not yet been counted are selected at random for the day's count. When all N items in the population have been counted, the process starts again.

Example 11.1 Assume a DC has N = 20,000 parts and chooses to cycle count n = 100 parts a day.

- When the constant population count is employed, 100 parts from the 20,000 are randomly selected each day for the count. In a year of about 250 work days, 100 $\times 250 = 25,000$ parts are counted. Some parts are counted one time, others more than once, and some are not counted at all.
- When the diminishing population count is used, 100 parts that have not yet been counted are randomly selected for the day. Note in this system, N/n = 20,000/100 = 200 days will elapse before all the 20,000 parts are counted. Thereupon, the sample process starts anew.

11.12.2 ABC Cycle Count

The ABC cycle count method is used when the management first categorizes the items by demand volume, and second, determines how to sample count the items by category. The A category includes the items with high dollar demands, while the B and C items have lower dollar demands. For example, suppose the relation between the percent of items in the inventory and the percent of annual dollar demands are as listed below; and the categories of A, B, C are as shown.

category	%parts	%$ demand
A	20	80
B	30	15
C	50	5

Note, the A category includes 20 % of the items and 80 % of the dollar demand; the B category has 30 % of the items and 15 % of dollar demand; and the C category has 50 % of the items with 5 % of the dollar demand.

The management chooses to count the items as follows:

category	count per year
A	n_A
B	n_B
C	n_C

The A items are counted n_A times per year, B and C are counted, n_B and n_C times, respectively per year.

Example 11.2 Assume a DC has N = 20,000 parts and $10,000,000 of demands a year. The number of parts by categories, A, B, and C is listed below, along with the annual dollar demand. Also listed is management's desire of frequency counts per year for each part by category; and the number of counts per year by category.

Category	No. parts	$ demand	Count per year	No. counts per year
A	4000	8,000,000	5	20,000
B	6000	1,500,000	2	12,000
C	10,000	500,000	1	10,000
Sum	20,000	10,000,000		42,000

Note, the total number of counts in a year is 42,000. With approximately 250 workdays in a year, the average number of counts per workday is 42,000/250 = 168.

11.12.3 Pareto Cycle Count

When the Pareto cycle count method is used, the fraction of annual demand for item i, p_i, is computed at the outset, as follows:

$$p_i = c_i \times d12_i / \Sigma[c_i \times d12_i]$$

where c_i is the cost per unit for item i, and $d12_i$ is the 12 month demand in pieces for i. If n is the number of items to count per day, the number of days between counts for item i becomes:

$$n_i = 1/[n \times p_i]$$

Example 11.3 Assume a DC has $10,000,000 of demands a year; and desires to count n = 100 parts per day.

Part i = 1 has c_1 = $5.00 and $d12_1$ = 1000, whereby,

$$p_1 = 5 \times 1000/10{,}000{,}000 = 0.0005$$

Hence, the number of days between counts for part 1 is:

$$n_1 = 1/[100 \times 0.0005] = 1/0.05 = 20.0 \text{ days.}$$

Part i = 2 has c_2 = $0.10 and $d12_2$ = 500, whereby,

$$p_2 = 0.10 \times 500/10{,}000{,}000 = 0.000005$$

and, the number of days between counts for part 2 is:

$$n_2 = 1/[100 \times 0.000005] = 1/0.0005 = 2000.0 \text{ days.}$$

With approximately 250 working days in a year, part 1 is counted 250/20 = 12.5 times a year, while part 2 is counted 250/2000 = 0.125 times per year, or once in 8 years.

11.13 Queuing Applications

Queuing theory is successfully applied in all sorts of applications including: industry, government, and military. The theory measures the probability associated with a facility that includes a waiting line of almost any kind. The waiting line has one or more service facilities and a stream of customers who randomly seek service from the facility.

In a network system, that includes a warehouse, distribution center or dealer locations, the service facility could be a loading dock, a lift truck, an employee, or any such resource that is used to apply service to an arriving need. When the service is completed, the facility is available to service a next unit. When one or more of the facilities are open, the arrival unit enters and begins service. When all the facilities are busy, the arrival unit will wait until a facility is open. The time between arrivals and the time for service are assumed exponential. The average time between arrivals is denoted by τ_a, and the average time for service is τ_s. A measure of the use of the system is called the utilization ratio and is denoted by ρ, and measured as follows:

$$\rho = \tau_s/\tau_a$$

For notation sake, N is the number of service facilities that are available to the system, and n is the number of units currently in the system where n is zero or larger.

The probability of n units in the system is denoted as P(n) and is computed as follows:

$$P(n) = \begin{cases} P(0)\ \rho^n/n! & n = 0 \text{ to } N-1 \\ P(0)\ \rho^n/\left(N!N^{n-N}\right) & n \geq N \end{cases}$$

where,

$$P(0) = 1/\left(\Sigma_{(n<N)}\ \rho^n/n! + \rho^N/[(N-1)!(N-\rho)]\right)$$

The probability that a new arrival will not have to wait prior for service is called the service level, SL; and is computed as follows:

$$SL = \sum_{n=0}^{N-1} P(n)$$

Note, when $n \geq N$, new arrivals will need to wait before being serviced.

Table 11.6 lists results from the waiting line system for $\rho = 0.1$–700, and SL = 0.85, 0.90, 0.95 and 0.99.

Example 11.4 A parts distribution center requires lift trucks to pick orders for large items. A truck is needed on average ten times per hour and the average time to fill an order is 20 min. The management seeks how many trucks are necessary to obtain a service level of SL = 90 %.

Note, the average time between arrivals is $\tau_a = 1/10$ h = 6 min, and the average service time is $\tau_s = 20$ min, whereby, $\rho = 20/6 = 3.33$. Table 11.6 is searched with $\rho = 3.33$ and SL = 0.90 to find N = 7 lift trucks are needed.

Example 11.5 The distribution center of the Alpha Company has a large number of parts, and a pool of order pickers on hand to pick an average 2000 lines a day. For each customer order, one or more parts are included, and the order is called a line. The order picker finds the bin of the part, gathers the pieces and places them on a cart. The average time to locate and pick a line is 10 min. The orders steadily arrive to the center and the management desires a service level of SL = 0.85. This way, the lines are filled fairly quickly and do not cause an unduly delay in delivery to the customers. The management wants to know how many order pickers are needed to yield the desired service level. To accomplish, note the average time between the arrival of a line in a 480 min day is $\tau_a = 480/2000 = 0.24$ min. Also, note the average service time (to process a line) is $\tau_s = 10$ min, and so, the utilization ratio becomes $\rho = 10.0/0.24 = 41.66$. With this information, Table 11.6 is searched at $\rho = 41.66$ and SL = 0.85 to yield the need for $N \approx 51$ order pickers.

Example 11.6 Assume trucks arrive to the receive loading docks at the Beta Corp. distribution center for unloading at a rate of 2/h; and the average time to unload a truck is 15 min. The management wants to know how many loading crews are needed to attain a service level of SL = 0.90.

The service level is the probability an arriving truck does not have to wait to begin unloading. Note, the time between arrival of trucks is $\tau_a = 30$ min, and the average time to unload a truck is $\tau_s = 15$ min. Hence, the utilization ratio becomes $\rho = 15/30 = 0.50$. A search of Table 11.6 shows when $\rho = 0.50$ and SL = 0.90, N = 2 loading crews are needed.

Table 11.6 The minimum number of service facilities, N, when the utilization ratio is ρ and the service level is SL

	SL			
ρ	0.85	0.90	0.95	0.99
0.1	1	1	2	2
0.2	2	2	2	3
0.3	2	2	2	3
0.4	2	2	3	3
0.5	2	2	3	4
0.6	2	3	3	4
0.7	3	3	3	4
0.8	3	3	4	4
0.9	3	3	4	5
1.0	3	3	4	5
1.2	3	4	4	5
1.4	4	4	5	6
1.6	4	4	5	6
1.8	4	5	5	7
2.0	5	5	6	7
2.5	5	6	6	8
3.0	6	6	7	9
3.5	7	7	8	10
4.0	7	8	9	10
4.5	8	9	9	11
5	9	9	10	12
10	15	16	17	19
15	21	22	23	26
20	26	27	29	32
25	32	33	35	39
30	38	39	41	45
35	43	44	47	51
40	49	50	52	57
45	54	56	58	63
50	60	61	64	66
55	65	67	69	74
60	70	72	75	80
65	76	78	80	86
70	81	83	86	92
75	86	88	91	97
80	92	94	97	103
85	97	99	102	109
90	102	105	108	114
95	108	110	113	120
100	113	115	119	125
150	166	168	173	181

(continued)

Table 11.6 (continued)

	SL			
ρ	0.85	0.90	0.95	0.99
200	218	221	226	235
250	270	274	279	289
300	322	326	331	343
350	374	378	384	396
400	425	429	436	449
450	477	481	488	502
500	528	533	540	555
550	579	584	592	607
600	631	636	644	660
650	682	687	696	712
700	733	739	747	765

11.14 Summary

The locations that hold the stock waiting for customer demands are of various type. The master stock location is the main DC location and holds the larger variety of items. The other locations in a DC are the branch locations, and sometimes a 2-level-service location (2LS). The 2LS location is typically a smaller facility that is replenished by the master stock location. The higher demand parts are stocked in most locations, and low demand parts are mainly in the master stock location. An analysis to seek the best arrangement of the DC network is periodically run by the management. The analysis inquires on where should the locations be situated, and how should the stock be allotted across the locations. Sometimes packagers are needed to place the manufactured goods in containers as needed for stocking and distribution to the customers. When the goods are transported from the suppliers, they often pass through a cross-dock facility. The integrity of the part count is maintained by a cycle count system. Queuing theory is demonstrated as a useful tool in network applications.

Chapter 12
Supply Chain Management

12.1 Introduction

The supply chain management team in a manufacturing firm is concerned with the flow of raw materials, basic goods, components, work-in-process and finished goods from suppliers to warehouses, plants, distribution centers, retailers, and finally to the customers. The goal is to minimize throughput times and expenditure costs, while achieving a high level of service to the customers with a clean environment. To accomplish, the team remains vigilant on the use of all technology available to them. Data sharing with the customers and throughout the operation is needed so all can work efficiently. The team applies the technology of scan based tracking, electronic data interchange, bar codes, data matrix barcodes, quick response codes, and radio frequency identification where and when appropriate. In the inventory operations, vendor managed inventory and continuous replenishments are applied when suitable. In transportation, incoterms are applied along with tracing and tracking of the shipments; and whenever feasible, outsourcing and third-party providers are employed. Five examples are sited: a heavy-duty make-to-order manufacturer, a network of men's shoe stores, an offshore automotive dealership, a replenishments of inventory at retail stores for seasonal style good, and a delivery system to stores for a popular bakery item.

12.2 Supply Chain Management Team

An ever-increasing role in corporations has emerged in the planning and control of the flow of inventory as it enters the system from suppliers and eventually leaves as delivery to the customers. This function is assigned to a supply chain management team whose charge is to seek a high level of service to the customers, with high quality products, on-time delivery, and minimal cost to the customers and corporation. The team determines the best way to accomplish all the functions that are

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_12

needed along the way. The team helps to determine the suppliers to use, where to store the incoming materials, and when an in-house function can better be served by a third-party-service-provider. Cooperation and coordination is needed with all the parties involved. As much as possible, the demand and inventory data should be shared with suppliers, manufacturing, distribution, and sales, so all can strive for a smooth transition of the inventory.

In a manufacturing firm, the flow begins with the purchase of basic goods, and also with the replenishment of the parts and components from suppliers. The inventory is placed in storage until needed for various forms of processing, fabricating and assembly as required, where value is added as the inventory transforms to various stages of work-in-process, and finally to finished-good-items. Make-to-order products are delivered to the customers, and make-to-stock products are sent to distribution centers and later to retail outlets.

The supply chain team seeks the best way to move and store the inventory in each stage with minimal cost and throughput time. This is crucial to the financial well-being of the corporation. With global competition increasing for all products, the need to provide top service, high quality and low prices is essential to satisfy the demanding customer. Supply chain management uses a variety of methods and tools to accomplish their charge, and some are described in the following sections.

12.3 Vendor Managed Inventory (VMI)

VMI is an agreement between a retailer and a vendor allowing the vendor to provide inventory of its products with the retailer at an approved level of service and with friendly purchasing terms. The retailer supplies information to the supplier on the progress of the sales, and the vendor has the responsibility to replenish the stock as needed.

A valuable catalyst to VMI is a scan-based trading system (SBT) that uses electronic data interchange (EDI). To function properly, SBT requires the store employee to scan each item upon its sale. This allows the supplier to track the sales of each product, by store and date. Along with the replenish data, the supplier is able to control the inventory at the level agreed with the retailer.

The VMI terms can vary in many ways: One arrangement is where the retailer will buy the units for inventory, and should they not sell well, the vendor will repurchase the units. In another arrangement, the vendor offers the stock on consignment, and thereby the retailer only pays the vendor as the units sell. Any unsold units are returned to the vendor.

A vendor representative often helps to setup the units at the retailer and furnishes advertising displays to help in the sales. The representative also advises the sales force on the features of the products so they have knowledge to assist the customers. Should any parts or maintenance be required, the representative advises accordingly. Many large retail stores use VMI, like WalMart, Dollar General, and Home Depot, to stock their products.

Example 12.1 ABC has 20 games that it produces and has an arrangement with CToys to sell their games on a consignment basis. CToys allows a fixed shelf space for ABC to display its products. ABC receives a check once a month based on a percentage of the sales of its products. With this information, ABC visits the store regularly to replenish the stock. Should the sales turn low on one or more products, CToys may notify ABC to remove the remaining stock of those products. With this VMI arrangement, CToys has access to the 20 products and has no investment in the inventory.

12.4 Scan-Based Trading (SBT)

Scan-based trading is a system where the supplier keeps ownership of inventory that is in the retailer's possession until the units are scanned at a point-of-sale register. This way, the supplier owns the products even while it is on the store shelves until it is scanned. The retailer does not buy the inventory and merely provides shelf space to the suppliers to stock their products. Typically, SBT uses an EDI solution to carryout its functions.

In simple terms, the system flows as follows: The supplier ships its products to a warehouse of the retailer. The retailer transfers the units to the stores, and the stores carry the inventory for sale to the customers. Upon each sale at the store, the supplier is notified via scanning and EDI. The supplier keeps tract on the inventory at the store, and provides replenishment to the warehouse as needed. This system allows the retailer to carry inventory at no cost in procurement.

12.5 Electronic Data Interchange (EDI)

Electronic Data Interchange is an electronic system that provides standards when exchanging data via electronic machines. With use of the same standards, two entities (originator and recipient) can exchange information by any electronic method. The documents could be purchase orders, billing, invoices, shipping orders, and so forth. The National Institute of Standards and Technology defined electronic data interchange between an originator and a recipient as follows: computer-to-computer interchange of strictly formatted messages that represent documents other than monetary instruments. The parties exchanging information with EDI are referred as trading partners.

The EDI staff provides a list of common EDI transaction codes for use in Electronic Data Interchange. Many of the codes are unique to an industry, and others are common among a variety of industries. The list is partitioned into a variety of series as follows:

- order
- material handling

- warehousing
- government
- manufacturing
- delivery
- engineering management
- miscellaneous
- mortgage
- quality and safety
- student
- transportation
- reference

The list also is categorized by various applications. Supply Chain is one of the applications, and the list contains over 75 codes for Supply Chain applications alone. A sample of the supply chain codes is provided below:

- EDI 940—Warehouse Shipping Order
- EDI 943—Warehouse Stock Transfer Shipment Advice
- EDI 944—Warehouse Stock Transfer Receipt Advice
- EDI 945—Warehouse Shipping Advice
- EDI 947—Warehouse Inventory Adjustment Advice

The EDI 940 code, for example, requires data like: ship-to address, bill-to address, product numbers and quantities of each.

12.6 Bar Code

Bar codes were first used commercially in 1966, and soon after, it became apparent a standard system would need to be developed. The Uniform Product Code (UPC) was invented in 1973 and remains today as the standard for bar codes. The first UPC scanner was installed in 1974 at a supermarket in Ohio. The scanner, however, is only effective if the items also have a code printed on their package to scan. The first product to include a bar code on the package was the Wrigley Company (Fig. 12.1).

Almost every item you purchase from a drug store, grocery store, or any store has a Universal Product Code (UPC) bar code printed on it. The use of UPC bar codes started in the 1960s to help grocery stores at the checkout counter. Each time

Fig. 12.1 Barcode

an item was scanned at the register, the item was identified and the quantity was noted as a sale and deducted from the inventory count, giving the store management data on demand and the inventory status of the item. This was a great tool to help control the inventory in the store. Soon, UPC bar codes were being applied in all types of products and industries.

The bar code is 12 digits long, and identifies the manufacturer of the item and the item number. The item number of a bottle of Schlitz beer is different than the item number of a six pack of Schlitz, and so forth. 12 digit bar code works like this:

Digits 1–6 = manufacturer's identification number
Digits 7–11 = item identification number
Digit 12 = check number

The check digit is meant to verify that the register reads all the digits correctly. If not, a fault at the register is signaled and the worker must scan the item again. The check digit is obtained as below:

1. N1 = sum of the odd number digits: 1,3,5,7,9,11
2. N2 = sum of the even number digits: 2,4,6,8,10
3. $N3 = 3 \times N1 + N2$
4. N4 = the smallest number to add to N3 so that (N3 + N4) is a multiple of ten
5. N4 is the check digit

Example 12.2 Assume the 12 digits of the bar code is: 563413001427. The following information is noted:

$$\begin{aligned}&\text{Manufacturer identification number} = 563413\\&\text{Item identification number} = 00142\\&\text{check digit} = 7\end{aligned}$$

Note, the check digit is computed as below:

$$\begin{aligned}&N1 = 5 + 3 + 1 + 0 + 1 + 2 = 12\\&N2 = 6 + 4 + 3 + 0 + 4 = 17\\&N3 = 3 \times 12 + 17 = 53\\&N4 = 7, \text{ since } (53 + 7) = 60\end{aligned}$$

Therefore, the check digit is 7.

12.7 Data Matrix Barcode

Data Matrix barcode is the two-dimension version of a barcode. It was developed in 1989 and can house 2335 alphanumeric characters, considerably more than the counterpart, one-dimension barcode, and has gained wide use in the US and Europe. Because it is two dimensions, the matrix needs to be scanned by an image camera or a

Fig. 12.2 Data matrix

smartphone device. Scanning involves reading and decoding the data. The data is then passed to an information system for further processing. In healthcare, the data matrix information on an item may include: the global identification number, the expiration date, and the batch number. In food items, the information may be: plant of production, production date, batch number, expiration date, and ingredients. The batch number is important in traceability applications. In mobile marketing, a Data Matrix symbol is printed on a product, and upon scanning the matrix, the user is directed to an online web site or a blog with detailed information on the product (Fig. 12.2).

12.8 Quick Response Code (QR)

The Quick Response Code is another version of two-dimensional barcode and became popular in the automotive industry in Japan. This method was developed because the Data Matrix barcode did not include Kanji characters as used in Japan and China. As a result, in 1994, the Japanese developed an enhanced version, QR Code, that holds 4296 characters, including those needed for the Japanese. This version is very adaptable to alphanumeric and Chinese characters (Kanji), and is gaining favor around the globe as the two-dimension barcode to use. The QR Code allows users with a smart phone to easily read and retrieve all sorts of information, like: item identification, time tracking, instructions, specifications, marketing promotions, so forth. When limited space is available in the print form, the QR code opens the user to unlimited online information (Fig. 12.3).

12.9 Radio Frequency Identification (RFID)

Radio frequency identification is a system that uses tags and readers. The tags are attached to objects to be identified, and the readers are devices that send a signal to the tag and can read its response. The tags can be passive or active. A passive tag has no energy attached and receives it energy from the reader. An active tag has a battery attached and is triggered when near a reader. The tag could be read only, or read and write. Read only is when the tag has a serial number installed by the

Fig. 12.3 Quick response barcode

manufacturer as a key to the database. A read and write tag has information from the manufacturer and can add new data when it passes a reader.

When the tag is in the range of the reader, the reader transmits a radio signal to interrogate the tag. The tag then receives the signal and responds with its serial number and other information. When there are many objects and all have tags, the reader can interrogate them all very quickly. The objects do not have to be seen, they could be in a crate or stacked on a pallet with the tags inside.

A standard way to store data on a tag is via the Electronic Product Code (EPC). The tag has a 96-bit string of data, arranged as follows:

- bits 1–8 = identifies the protocol
- bits 9–36 = organization number
- bits 37–60 = object class (product type)
- bits 61–96 = serial number (a key for a global database)

RFID technology is applied ubiquitously. RFID is in use in many retail stores to provide location of the merchandise, and also as a self-checkout service to the customers. Inventory systems are using RFID tags to locate and count the individual items in the warehouse. Manufacturing uses the RFID tags to monitor the progress of a batch order as it progresses through the plant, all the way to shipping. The US Department of Defense and many retailers require their suppliers to place RFID tags on all shipments to aid in the supply chain management. Tags (transponders) are placed on vehicle windshields so they can be interrogated by a reader at toll-booths to allow the vehicles to pass without the need to stop. Identification badges for individuals have tags that allow the holder to pass through areas of secure locations. Manufacturers place tags on their products to track which item has been sold at the retailer.

12.10 Incoterms

Incoterms (International Commercial Terms) are standards that are used globally for domestic and international contracts for the sale of goods. They are published by the International Chamber of Commerce and are applied in international commercial transactions and procurement processes. The terms are denoted in three letters and pertain to the rules of tasks, costs, and risks corresponding with transportation and delivery of goods. The terms are legally binding and are set in sales contracts. They originate from 1936 and have been slowly revised over the years. A brief on the most common commercial terms are listed below:

- EXW—Ex Works (named place)
- FCA—Free Carrier (named place of delivery)
- CPT—Carriage Paid To (named place of destination)
- CIP—Carriage And Insurance Paid To (named place of destination)
- DAT—Delivered At Terminal (named terminal and place or port of destination)
- DAP—Delivered At Place (named place of destination)
- DDP—Delivered Duty Paid (named place of destination)
- FAS—Free Alongside Ship (named port of shipment)
- FOB—Free On Board (named port of shipment)
- CFR—Cost and Freight (named port of destination)
- CIF—Cost, Insurance and Freight (named port of destination)
- DAF—Delivered at Frontier (named place of delivery)
- DES—Delivered EX Ship
- DEQ—Delivered Ex Quay (named port of delivery)
- DDU—Delivered Duty Unpaid (named place of destination)

12.11 Tracking and Tracing

Tracking and tracing is the process of gathering data on the location (past and most recent) of a specific item. When an item is placed in a container for shipment to a destination, domestic or global, the role of track and trace is to note the date and arrival of the item at various locations along the way to the final destination. This is accomplished with the aid of barcodes, data matrices, or RFID technology. When a recording of the item is taken at a location on the way to the destination, the data is recorded in a global database. This database allows the sender or receiver to inquire on the database for the latest status of the specific item. In the event of a recall in industries like: food, pharmacy, toy, so forth. Tracing the location of the item is vital.

12.12 Continuous Replenishment (CR)

Continuous replenishment is a way to replenish stock from supplier (vendor) to retailer. A prearranged agreement allows the supplier to have full access of the demands at the stores, and also the inventory status of the sku's at the retailer's location. Each day, the supplier is notified of the demands by sku; whereby the supplier keeps track, on a database, of the corresponding inventory status. With this information, the supplier determines the sku replenish needs at the level agreed with the retailer. The supplier ships the replenish quantities to the retailer on a regular basis, often daily. Point-of-sale registers are often used in CR arrangements, allowing the supplier immediate feedback on the sku demands.

12.13 Logistics

Logistics involves all aspects of a firm that concerns the flow of items from the point of origin to the final destination to satisfy the needs of the customer. It involves many activities such as: make-or-buy decisions, supplier selection, purchase order contracts, transportation, material handling, production, packaging, warehousing, inventory storage, picking and packing, security throughout, and delivery to customers. Logistics is also responsible for properly disposing of any waste in the total process, and to abide by all ecological requirements. The common goal is to have a high level of service to the customer and provide all of this with the minimal cost and time duration. Throughout production, the goal is to add value and to eliminate waste at each process step; thus, provide the right product with the right quantity, with good quality, and at the right time.

12.14 Third Party Providers (3PP)

Third party providers companies are firms that offer services that allows companies to outsource some of their supply chain activities such as: inbound freight delivery, warehousing, distribution, and outbound freight shipping. Carrier firms like: FedEx, UPS and DHL, provides such a service to deliver goods from source to destination, usually overnight, reducing the need to warehouse the goods and with the spirit of just-in-time delivery. Although 3PP allows the firm to provide logistic services in a cost saving manner, they are obliged to conform to the standards of the firm they are representing.

12.15 Outsourcing

Outsourcing is when a company, X, hands over a portion of their business operations to another firm, Y, to perform services that could otherwise be performed internally. The incentive is to save the expense of providing a staff and all the associated wages, benefits, training, equipment, and space required. Often, the outsource company, Y, is in a foreign country and specializes in the services with lower labor rates. The specialized company, Y, also maintains up-to-date know how and the equipment to perform the tasks at a top level. Company X is not burdened with cost of performing these tasks and can concentrate on their main activities.

The outsource work includes a wide range of activities: delivery, payroll, telecommunication, assembly, engineering design, information technology, so on.

12.16 Make or Buy

The management is sometimes faced with a decision on whether to produce a component in-house or purchase from a supplier. The decision depends mostly on the comparison of cost and quality.

The factors that favor a buy are listed below:

- cost is less to buy
- equipment to produce is not available in-house
- expertise is low in-house
- customer demand is low
- capacity is low in-house

The factors that tilt to a make in-house are the following:

- cost is less in-house
- quality of the component is ensured in-house
- fill idle capacity in-house
- secure proprietary information in-house

12.17 Warehouse Management

Warehouse management's mission is difficult indeed. They are required to synchronize all of the activities concerning the inventory, space, equipment and people in their facility with a goal to minimize the cost and throughput time to process orders. They are responsible for all the movement of stock within a warehouse including all the paperwork involved. The duties involve receiving replenishment stock from incoming vehicles and placing in the appropriate bins; and picking stock to fill the steady flow of customer orders and placing in outgoing vehicles.

The management duties include optimizing the space utilization and organizing the infrastructure to meet the needs of the operation. They are required to utilize automatic identification systems to capture essential data on the status and movement of stock for database storage. This includes full use of barcodes, scanners, RFID and various wireless systems. The database is the catalyst to developing reports that are essential to guide the management of the entire operation.

In addition, the management is required to provide an inventory system that minimizes the cost and maintains a high level of service to the customers. They must be vigilant to identify any parts that are in oversupply, and also any that are in short supply. The inventory should be set at a level to properly fill the customer orders, while surplus inventory, backorders and lost sales are low.

The management is ever seeking ways to minimize the workers time needed to place the stock in the bins, and also to pick the customer orders. One strategy is to identify parts that are frequently ordered together, and place them in bins that are close to each other. They also seek to assign the parts to proper storage locations of: bins, shelving and floor space. Low demand parts are placed in the upper locations, and high usage parts in the middle locations, and so forth. Small part are in small bins, and large parts are set on the floor space.

12.18 Material Handler

The warehouse workers are generally called material handler's. These people are assigned a variety of duties that are essential to allow the operation to run smoothly. A short list on sum of the duties is below:

- unload freight from incoming vehicles
- run forklift trucks and pallet jack equipment
- check for damaged goods of incoming items
- verify count of incoming quantities with purchase order
- verify and sign bill of lading
- stack incoming stock in appropriate storage locations
- fill customer orders from storage
- maintain a clean and orderly working environment

12.19 Waste Management

The primary goal of the waste management team in a firm is the protection of human health and the environment. The source of the waste from a firm generally comes from surface-water, ground-water or the air, and the minimum amount of waste, the better. There is much diversity in the composition and type of industrial waste. The individual states, and local governments are responsible for legislating, monitoring, and implementing the programs to ensure proper handling of industrial waste. The waste management team is obliged to work side-by-side with the appropriate government agencies.

12.20 Supply Chain Examples

Below are some examples of the flow of products within a companies system beginning with the source and ending with the customer. The examples include the production and sales of a heavy duty item, the distribution of men's shoes in the stores and in the distribution center, the flow of service parts in a large automotive dealer in a foreign land, the replenishment of seasonal style goods to retailers, and the delivery of a popular bakery item to stores.

12.20.1 Make-to-Order Heavy-Duty Item

A demand driven supply chain is when the customer order determines the production needs of the firm. The customer orders a product with the exact specifications desired with a desired due date, and the manufacturer is obliged to produce the item, and deliver as near the due date as possible. The components for the unit is needed prior to assembly. Some of the components are already available in the inventory and others have to be ordered from the suppliers. When all the components are available, the unit is scheduled for assembly as soon as possible. Subsequently, the assembly takes place and the unit becomes a finished product. Finally, the manufacturer arranges for shipment to the customer. A summary of the events in the supply chain are listed below:

- Customer order arrives at the sales department for a heavy-duty item with desired due date and specified features and options.
- Planning department determines the component parts needed for the customer order.
- Purchasing department receives the component part needs and checks inventory currently available, and need to purchase from the component suppliers. A purchase order goes to component suppliers for the components needed.
- Inventory department subsequently receives the components and stores them until needed. When all the components are available for the customer order, the production department is notified.
- Production department schedules the customer order for assembly as-soon-as-possible. On the assembly date, the components are gathered and placed in the assembly area where needed. When the item is completed, the unit is tested and placed in storage awaiting transportation to the customer.
- Transportation determines the best way to delivery the item to the customer. A metric is reported on how efficient the delivery date was with respect to the customer due date.

12.20.2 Men's Shoe Stores

A large shoe company has over 1000 stores, and has a CR arrangement between the stores and the central distribution center (DC). For each store and every sku (style-size-width) that they carry, a prearrange order point (op) and order level (ol) is computed. With point-of-sales registers, the DC has access to each sale of an sku at every store. The DC also keeps track of the on-hand (oh) and on-order (oo) for every sku at the store. All this information is monitored on the DC's database. When an sku's (oh + oo) reaches the op or below, a replenishment of size q up to the ol is called; and the DC ships the replenishment to the store accordingly.

The DC also accumulates the demands on each sku (style-size-width) over all of the stores. With this information, the total demand by style (for all size and width) is gathered and a monthly forecast is generated. This style forecast is partitioned back down to the size and width based on the fraction of history demand be size-width. With these sku forecasts, and the level of service desired, the order point (OP) and order level (OL) for each sku is generated at the DC to service all 1000 stores.

The DC receives replenish stock from a plant that produces the shoes by style, size-width. When the on-hand (OH) plus on-order (OO) at the DC falls to the OP or lower, the DC sends a replenish order to the plant to bring the OH + OO up to the OL.

In the production process, the plant produces 12 shoes in a batch, where all the shoes are from the same style and as close as possible in size-width. This is a requirement to gain efficiency in the production process. When the orders from the DC are sufficient for this need, the raw materials are gathered and the batch is assigned to a plant worker. Upon completion, the shoes are shipped from the plant to the DC. A summary of the events is below:

- Each store's inventory is based on a quota on number of sku's to have depending on the size of store and customer base. An order point (op) and order level (ol) is established for each style-size-width (sku) to conform with the store's quota.
- A point-of-sale register is used with each sale allowing the distribution center (DC) to be aware of each sale at every store by sku.
- The DC has a vender-managed inventory arranged with each store. When the on-hand (oh) plus on-order (oo) for an sku falls to the op or lower at a store, a replenish order to bring the stock up to the ol is triggered at the DC.
- The DC is the supplier to a large number of stores, and based on a desired level of service, and monthly forecasts for each sku, the DC computes an OP and OL for each sku. When the on-hand (OH) plus on-order (OO) at the DC falls to the OP, a replenish order is sent to the plant for a quantity of Q = OL − (OH + OO).
- Using monthly forecasts, the plant projects the demands for the various sku's over the coming months. The planning department determines the raw materials needs to fulfill the forecasts.
- Based on the forecasts, and raw material currently available, the shortage of raw material is noted and a purchase order goes out to the raw material suppliers, accordingly.
- When all the raw materials and the plant capacity is available, the plant schedules the production of the sku's by style-size-width to meet the replenish orders plus the forecasts over the future months.

- Transportation determines how to ship the replenish orders from the plant to the individual stores as best as possible. Each pair of shoes heading to a DC is placed in a shoe-box with the identification of the style-size-width, and with an accompanying barcode.

12.20.3 Offshore Automotive Dealership

An original equipment manufacturer has a plant, and a distribution center in the U.S. and is the supplier to a large number of dealers in the U.S. and some globally. A particular dealer is in Columbia, say, and has ten locations scattered throughout the country. A master location at the dealer is the supplier to all of the other nine-branch locations. The distribution center in the U.S. is the supplier to the Columbia dealership, and the plant in the U.S. is one of the suppliers to the distribution center. The summary of the events for this supply chain is listed below:

- A dealership in Columbia has ten locations, one large master location (ML) and nine smaller branch locations (BL) that stock service parts for truck maintenance and repair. The ML is the supplier to the BL's. The ML receives its replenishments from a distribution center (DC) of an original equipment manufacturer (OEM). The DC is located in the U.S.
- The OEM plant in the U.S. produces many of the parts using a master production schedule system with level load capability, whereby the plant schedules an even flow of work over the time horizon.
- The plant receives replenish orders from the distribution center (DC) of the OEM. The DC is the supplier to many dealerships in the U.S. and globally.
- The DC receives a continuous flow of replenish orders from the ML of the Columbia dealership. As each replenish order arrives, it is picked and placed on a pallet in a container awaiting shipment to Columbia. Each pallet in the container has an RFID tag that identifies the content and the destination of all items on the pallet.
- Every 2 weeks, the pallets are lashed and bounded; with dunnage inserted to prevent slippage; and the container is shut. The DC orders a truck carrier to take the container to a cargo terminal for loading onto a cargo carrier for Columbia, whereby, the container is on the way to the dealership.
- At Columbia, the container is unloaded and placed on a truck for transport to the ML of the dealership.
- The ML has much storage space and is the supplier to all of its BL's. The ML has a computer system that monitors the stock status of each sku at every BL, knowing the on-hand (oh) and on-order (oo) quantity. An order point (op) and order level (ol) is generated for each sku at every BL and the information is known at the ML.
- When a BL has an sku where (oh + oo) reaches the op, a replenish quantity is generated for the ML to replenish.

- Using a continuous replenish system, each day, a truck from the ML is filled with all the replenish orders for the BL's. The truck stops at each BL and drops off the replenish quantities.
- The ML also runs a transfer program once a week to determine if any BL has excess stock on a part and another BL is short. If so, the truck picks up the excess stock from the excess BL, and delivers to the BL that is short of stock.

12.20.4 Replenishing Seasonal Style Goods

A large retailer has N stores and buys seasonal style items from an OEM at a discount in a large quantity. The purchase size for an individual item is denoted as Q. The selling season is 10 weeks, whereby the OEM ships the total quantity, Q, to the retailer's distribution center (DC) prior to selling season. The DC allots a total of $Q_w = Q/10$ each week to the stores. This way, the allotment is evenly split among the weeks and the store inventory is at minimal levels. The stores have point-of-sale registers and the DC has access to the sales and also the on-hand inventory of the product as each week passes. With this information, the DC determines a fair-share allotment of Q_w to each store every week. The store allotment for each week is denoted as q_j for store j, where $\Sigma q_j = Q_w$. The DC fills a truck with the replenish stock and the truck stops at a cross-dock facility where the stock is separated by retail location. The individual location's stock is placed on N smaller vehicles for delivery to the location accordingly. At the end of the 10-week selling season, the entire stock of size Q has been allotted and shipped to the N stores. Below is a summary of the steps to carryout this process.

- The retailer of N stores buys Q units from an OEM for a particular product. The product has a selling season of 10 weeks.
- Prior to the selling season, the OEM ships the Q units to the retailer's DC. The DC will ship a total of Q_w units each week to the stores, where $Q_w = Q/10$.
- The DC has estimates of each store's portion of total sales, denoted as p_j for store j. The initial allotment to store j becomes: $q_j = p_j \times Q_w$. The allot quantities are set to integers.
- The DC has access to each store's on-hand (oh_j) and sales to date (s_j) and uses this information to determine the fair-share allotment of stock to each store every week after the first week. The allotment to store j is denoted as q_j where $\Sigma q_j = Q_w$, at each week.
- Note the estimate of week's-of-supply (denoted as W_s) for all of the stores at the start of week w is $W_s = [\Sigma oh_j + Q_w]/\overline{S}$, where $\overline{S} = \Sigma \overline{s}_j/(w-1)$ is the average sales per week from the prior (w-1) weeks. The average weekly sale's at store j is denoted as $\overline{s}_j$.
- The fair share for store j after (w-1) weeks becomes, $q`_j = [W_s \times \overline{s}_j - oh_j]$. If $q`_j < 0$, set $q`_j = 0$. Since, some quantities may be negative, the adjusted

replenish quantities become $q_j = q`_j/\Sigma q`_j Q_w$ for store j. The replenish quantities have to be set to integers.

- Note, after the selling season of 10 weeks, the total purchase quantity per product has been distributed to the N stores, little by little, and in a fair-share manner.

12.20.5 Delivery of a Popular Bakery Item to Stores

A popular bakery item, B, is packaged at the plant and has a shelf life of 25 days. The units are produced in the plant and shipped to a central warehouse and held in stock until they are picked and a truck delivers them to stores all across the country. Because of the limited shelf like, B is delivered to 5000 stores that are accessible with the minimum delivery time. Many other retail outlets are left out since the delivery distance is out of reach. The delivery expense from warehouse to stores is costly and represents 36 % of the total revenue on B.

The management knows it has a popular item but is constrained since the delivery expense is too high. The firm seeks a way to alter the large delivery expense for the B item. They hire a research firm to find a way to increase the shelf life of the item. With success, in ingredients and processing adjustments, the shelf life increases to 65 days, with no loss in taste and no decrease in health issues. With this modified product and the increase in the shelf life, the management now can stock B in a series of regional warehouses throughout the country that are much closer to the final sales outlets. With longer shelf life, the number of stores to deliver increases from 5000 to 10,000. With shorter delivery distances, the delivery cost from warehouse to stores drops to 16 %. The product remains very popular and is now profitable for the firm to produce and distribute to the customer base.

12.21 A Supply Chain

Figure 12.4 depicts how the locations in a supply chain are typically related to one another. The chain begins with the raw good supplier and ends with the customers of the finished good item (FGI).

The raw good supplier provides raw materials from the earth, like cattle, ore and lumber. The material is purchased by basic good suppliers who convert the raw goods to a higher level of product, like cotton to cloth. The component and part supplier takes the basic goods and produces the items on a bill-of-material of a FGI. These items are sent to the plant that manufactures the FGIs. The plant sends the units to a finish good DC who is the supplier to dealers. Upon a dealer order, the units are shipped in truckloads (TL) to a cross-dock facility where the units are unloaded and transferred to a smaller vehicle for delivery to the final dealer. The service part DC is the supplier of components and parts to the dealer for

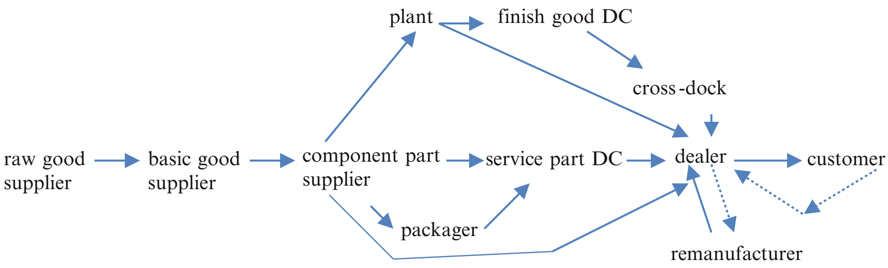


Fig. 12.4 A supply chain

maintenance and repair of customer FGIs. When the component and parts supplier sends the units in bulk, the bulk units go to a packager who places the units in containers with appropriate insertions and markings suitable for shelving and customer display. The dealer is the source to the customers for sale, maintenance and repair of the FGIs. The customer is the buyer of the FGIs. When a customer returns or replaces a FGI at the dealer, the dealer sends the unit to a remanufacturer who rebuilds the unit to like-new status and sends back to the dealer for sale.

A quick summary of the locations in the supply chain is provided below:

- Raw good suppliers are farms, ranches, mines, oceans that supply cotton, fish, so on.
- Basic good suppliers convert raw goods to a higher level, like lumber to (2 × 4)s.
- Component and parts suppliers take basic goods and produce parts from the bill-of-material of FGI.
- Plants manufacture the FGI.
- The DC stocks the FGI awaiting dealer demands.
- Cross-dock facility unloads truckload packages to smaller vehicles for deliver to final destination.
- Packager places bulk stock into containers for shelving display to customer.
- Service part DC is the supplier of parts and components for dealer needs.
- Dealer is the source for customer sale, maintenance and repair of FGI.
- Dealer gets return or replace units from customers and sends to remanufacturer.
- Remanufacturer rebuilds units like new and sends to dealer.

12.22 Summary

The supply chain management team is charged with using all the techniques available to them to achieve their goal of minimal cost and customer satisfaction throughout. Electronic data interchange, bar codes, data matrix barcodes, quick response codes, radio frequency identification, incoterms and tracking and tracing are useful tools available to them. Third-party providers, outsourcing, vendor managed inventory and continuous replenishments are techniques employed whenever they are cost effective to the total operation. Examples are provided on a make-to-order heavy-duty product, the distribution for a men's shoe store, the control of inventory in an offshore dealership, the replenishment of inventory at retail stores for seasonal style goods, and delivery of a popular bakery item to stores.

Chapter 13
Transportation

13.1 Introduction

Transportation pertains to the movement of items from one location to another, could be from a plant to a distribution center to a retailer. Shipping categories is a way to classify the type of goods: household, express, parcel or freight. Transport modes are the access ways to transport: rail, road, air, water, or pipes. Transport vehicles are the trucks, trailers, barges, aircraft, and cargo ships. Cargo handling is by ports, container terminals, shipping containers, forklift trucks, cranes, pallets, and dunnage. The shipment of goods also concern dispatchers, bill of lading, manifest, truckload shipping, less than truckload shipping, parcel carriers, bulk cargo and break bulk cargo. Some industrial trading terms are: free on board, carriage and freight, carriage, insurance and freight, and best way.

13.2 Shipping Categories

The items to ship are categorized as follows: household goods, express, parcel, and freight. Household goods are TV's, tables, refrigerators, couches, so on. Express are small light items like envelopes, and are shipped by ground for short distances, and air for long distances. Parcel is goods in cardboard cartons or wooden boxes of less than 110 lb, and is shipped by truck or train. Freight includes the heavier items (over 110 lb) per unit or in the shipping batch, and is transported by truck, train or barge.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_13

13.3 Transport Modes

Cargo refers to commercial goods shipped from origin to destination by boat, barge, air, train or truck. The modes of transport is the different type of facilities that are used to transport the cargo. The primary modes are the following: road, air, rail, water, and pipeline, and a description on each is below.

13.3.1 Road

A road is a paved path that accommodates a variety of motorized vehicles that travel from one location to another. By far, the most common vehicle to carry cargo is the truck. Buses and automobiles are sometimes used, but they are limited in cargo space. A common classification of trucks is: light trucks, medium trucks, heavy trucks, very heavy trucks and transporters. Transporters that move oversize loads require an escort and permits, and the drivers of these vehicles require a special driving license.

13.3.2 Air

The air vehicles (aircraft) that handle cargo delivery are: single engine, commercial jets and helicopters. To operate efficiently, they requires a network of airports with available crews to load and unload the cargo and to refuel, inspect and maintain the vehicle. The helicopters are more versatile since they can land and takeoff at a variety of locations. The advantage of aircraft is the speed of transport; and the disadvantage is the higher cost of delivery.

13.3.3 Rail

A railway (railroad) consists of the two parallel steel rails that are continuously connected from one location to another. The train is a series of connected vehicles that travel on the railway. The locomotive serves as the power to a large number of non-powered vehicles (freight cars). The freight cars are classified as follows: boxcars, containers and dedicated cars. The boxcars require manual loading and unloading; containers need little handling; and dedicated cars are for special cargo like: cattle, liquid, refrigeration, coal, so forth. When there is much cargo to deliver, railway is usually the most cost efficient.

13.3.4 Water

Another mode to transport cargo is by water using a marine craft of the following type: barge, boat or ship. The mode could be a river, canal, lake, sea or ocean. This method of transport is slow, but for non-fragile cargo, it is a cost effective way to move large quantities of stock.

13.3.5 Pipeline

A pipeline is a connected tubular structure that is continual along a path going from a source location to a destination location and is designed to haul a specific type of cargo. The typical cargo is: oil, natural gas, sewage, slurry, water, so forth.

13.4 Transport Vehicles

The vehicles of transport are trucks, freight trains, barges, aircraft and ships. A review on each is below.

13.4.1 Trucks

Trucks of various types are the vehicles that are used to transport cargo. Light trucks are much like an automobile or an SUV and are gasoline or diesel driven with automatic transmissions; while the heavier trucks use diesel engines. Small trucks are the pickup type that also run on gasoline and have automatic or manual transmissions. They are used mostly to transport goods short distances. The larger trucks are designed to tow one or more trailers or semi-trailers, and are used to haul cargo long distances. The large trucks have diesel engines and are trending towards automatic transmissions to ease the task of the driver.

The common category of trucks is the following: box truck, vans, medium duty trucks, medium standard trucks, platform trucks, and flatbed trucks. The trailers are of the following type: dry van, flatbed and refrigerated. The dry van is the most common, the flatbed is for heavy loads, and the refrigerated is when temperature control is needed. The typical length of a trailer is 53 ft, the length with the tractor is 75 ft. The height limit is 14 ft, and the width is 8.5 ft.

13.4.2 Freight Trains

Freight trains are commonly used to haul large quantities of cargo between railway locations as needed. The different types of freight cars are designed to accommodate the variety of cargo. This transport method is often favored since it is cost and energy efficient with environment benefits. A limitation is the lack of flexibility when there is absence of tracks at the original source or the final destination; whereby some additional transshipment is needed both in loading and unloading the goods from the train.

The freight cars are the following type: boxcars, containers, tanker cars, refrigerator cars, stock cars, gondolas, and flat cars.

- Boxcars are rectangular in shape and have been common in use. They can haul a large variety of items like: barrels, drums, bundles, boxes, and palletized cargo. The boxcar was favored since the contents are protected from any weather condition.
- Containers have replaced the boxcars for the non-bulk cargo since they require less manual labor to load and unload; and also since they transfer easily between trucks and barges using cranes.
- Tanker cars are cylindrical in shape and are designed to transport liquid products like chemicals, fuels, cooking oil, so forth. A poster is placed on the car to identify the contents inside.
- Refrigerator cars are used to haul perishable goods like: cheese, fruit, vegetables, poultry, meat and milk. Refrigerator cars can also come by way of containers.
- Gondola cars are also rectangular in shape, but have the top open for easy load and unload of the cargo. These cars are used to carry a wide variety of goods like: coal, scrap metal, stone, soil, so forth.
- Stock cars are used to haul live animals. They are like boxcars with openings on the sides for ventilation. They can be one level for cattle or horses, two or three levels for sheep, pigs and poultry.
- Flat cars are open on all four sides and are used to carry large goods that will not easily fit in a car with the sides. The goods are like: lumber, large rocks, iron beams, tractors, engines, and so on.

13.4.3 Barge

A barge is a flat-bottomed boat, built mainly for river and canal transport of heavy goods. Some barges are self-propelled, while others have no power and are pulled by tugboats, or pushed by towboats. In calm upstream or downstream waters, self-propelled barges are efficiently run; but in upstream rough waters, a tugboat is used to assist.

Barges are used to haul heavy and bulky items like: containers, coal, crushed stones, automobiles, coke, timber, iron ore, and so on. The cost to haul the goods by barge is low in comparison to other transport means. The typical barge is 195 by 35 ft and can haul very heavy loads. Some barges are open on the top and others are covered to protect the cargo from weather conditions. Other barges are specially designed to carry liquid products. Several barges have cranes onboard to help in the loading and unloading process. Two or more unpowered barges are sometimes connected to form a train-barge.

The barge is capable to carry a wide variety of items, that are classified as follows: automobiles; break bulk are items on pallets; bulk cargo are items like salt, scrap metal; neo-bulk are individual units as boxes; containers are the most popular; and project cargo are very big or heavy items.

13.4.4 Aircraft

A cargo aircraft is a fixed wing aircraft that is designed to haul items calling for quick delivery. This aircraft (also called freight aircraft or freighter) is often converted from a passenger aircraft with the passenger amenities removed. The aircraft is mostly operated by cargo airlines and are often equipped to handle a unit-load-device (ULD). The ULD is a container or pallet where a large amount of cargo is bundled together allowing easier loading and unloading of the aircraft. Each ULD has a manifest (packing list) attached to identify the contents and allows tracking.

Cargo aircraft are structured with strong floors, no windows for passengers, a wide top-hinged door on the main body for ease in loading and unloading the cargo. The body of the plane is set wide and the wings are raised to allow more cargo space; and the tail is raised high to allow vehicles to be driven in and out of the aircraft. The plane is also equipped with extra wheels to ease landing of the aircraft.

13.4.5 Modal Rating

In most situations, a firm planning to ship cargo to a destination has a narrow choice on the mode of transport. Below is a list of the characteristics of the modes and the typical selected shipment mode:

• accessible	– truck
• transit time	– truck and air
• reliability	– truck and air
• safety	– air and truck
• cost	– rail and water

When two or more modes are possible, the following modal features are considered in the selection process:

- cost
- speed
- cargo type
- cargo value
- cargo damage rate
- accessible route
- safety
- security
- mode equipment available

13.4.6 Cargo Ships

A cargo ship (also called a freighter) is a vessel that hauls commercial goods and materials from one port to another. These are the vast number of vehicles that are busy each day in international trade to transport the world's goods across the oceans on a daily basis. The cargo ships are often specially equipped for the type of cargo they carry. Many have cranes that are designed to load and unload the cargo. Cargo ships are often classified in various ways. Some of the common classifications are listed below:

- General ships carry packaged goods.
- Tankers carry liquid cargo, e.g., petroleum.
- Dry bulk ships carry items in loose form like coal, and grain.
- Multi-purpose ships can carry general, liquid, and dry bulk at the same time.
- Reefer is a refrigerated ship that carries foods of various types.
- Container ships are designed to carry containers.
- Liner ships are assigned a fixed location and time schedule and carry general cargo.
- Tramp ships have no fixed timetable and are hired to haul cargo between locations.

13.5 Cargo Handling

Cargo is the goods shipped from one place to another by way of some transportation vehicle. This process requires various facilities to carryout the transportation. Some of the facilities are the following: ports, container terminals, maritime security, shipping containers, forklift trucks, cranes, pallets, dunnage, and security. A description on each is below.

13.5.1 Ports

Ports are on the shoreline of oceans, lakes or rivers where boats, barges and ships can load and unload their cargo. The port is located in a spot that is rather free of foul wind or waves, and is readily accessible by land and by water. Importing and exporting of foreign goods is a major user of ports, and a large portion of the cargo is housed in containers. Maritime shipping is a popular and cost efficient way to ship the cargo between countries.

13.5.2 Container Terminals

Container terminals are entities that have the equipment to readily transfer containers from one mode of transportation to another mode. The terminal is located in or near large cities alongside a port, or at the end of a railroad line, or at a trucking center. A maritime container terminal is a facility that transfers containers from a ship or from a barge to a train or truck; or vice versa. An inland container terminal is a facility where the container transfer takes place between a train and a truck. In either event, the terminal also provides space for the containers to be stored, where some are empty and some are still loaded. The empty containers are stacked awaiting a next assignment, and the loaded containers are waiting for a call to be shipped to their final destination.

13.5.3 Shipping Containers

A shipping container (also called freight containers) is an enclosed steel structure of standardized dimensions and strong enough for shipping, storage and handling that is capable for transfer from one mode of transportation to another without the need to unload and load the content inside. All the handling is done mechanically with cranes and special forklift trucks. They are used globally as storage and transport units for shipping goods from one location to another. The containers are reusable and are designed for transport on container ships, barges, rail transport flatcars, and semi-trailer trucks. Each container is numbered allowing ease of tracking via computerized systems.

Besides shipping, containers are widely used for transport of cargo by way of trucks and trains. The standard sizes of containers are: 20, 40, 45, 48 and 53 ft in length, and the height standard is 8.5 ft. Manufacturers are now designing their products to fit neatly inside of containers. Containers are now regularly scheduled for transport by ship, truck and railway, and this allows parts and component suppliers to steadily replenish their customers in a flow of smaller quantities. The containers are mostly equipped with wood floors. The containers are numbered and are readily tracked by way of computers. The cost of transporting cargo by containers is cost efficient. Prior to loading, the hold area is swept clean and the dunnage is prepared for the cargo. Then, the cargo is loaded, with the big items one at a time.

13.5.4 Forklift Trucks

Forklifts are vehicles designed especially for applications in the operations of warehouses and distribution centers. This vehicle, with two blades extended in front can carry heavy loads at a short distance, both inside and outside. They are driven by electric or gasoline power, and those used indoors are mostly electrical. The vehicle is often called to go down a narrow aisle to set or retrieve a pallet of goods in a bin space whose height could range from low to high. When designing the bin space of a storage facility, the maneuvering role of the forklift needs to be considered. The forklift blades can be lowered or raised, and can be tilted forward or backwards, as needed. There is a wide range of forklifts in use in industry. On the smaller end, a typical warehouse forklift can control loads from 1 to 5 t. On the higher end, the large and heavy-duty forklifts can control items (like shipping containers) up to 50 t.

13.5.5 Cranes

A crane is a machine used in many industries that is designed to lift and lower items and move them in various directions. Their role in transportation is to load and unload freight from barges, ships, trains and trucks. Four of the common cranes are called: mobile, gantry, overhead and deck. The latter three are more commonly used in the transportation industry. A short description on each is below:

- Mobile cranes are mounted on a movable vehicle, such as: crawlers, tractors, or trucks; and some are self-propelled. These cranes are transported to the location where needed and require minor adjustments to begin its operation.
- Gantry cranes are widely used in container terminals to load and unload the heavy cargo on the ships or barges. These cranes are set on rails and move along the shoreline to the spot where it can handle the cargo as needed. They are capable of handling loaded containers, large machinery, steel beams, lumber, and so on.
- Overhead cranes have a rail set up high off the ground (called a bridge) that has a lifting component that travels along the rail. This crane is used in steel mills to receive, move and pour the hot material as needed. In auto assembly, the cranes are used to move some of the larger and heavier components from a subassembly area to the main line.

- Deck cranes are installed on some ships, boats and barges, and are used to load and unload cargo when no shoreline service is available.

13.5.6 Pallets

A pallet is a flat structure that serves as the base for goods that are transported or are in storage. On a flat floor, a pallet, with a heavy load of goods loaded on top, can easily be lifted, moved and stacked by a forklift vehicle or from a hand-drawn pallet jack. In transporting, the goods are sometimes secured by strapping, stretch wrapping or shrink wrapping. The pallets are mostly made of wood, but some are built from metal, plastic, recycle materials, and even paper. The use of pallets has reduced the cost of handling and storage of goods. The various type of pallets is below.

- Wooden pallet are structured with three or four wooden planks called stringers, and with wooden deck-boards on top. The common pallet size is 40 by 48 in. where the first number is the stringer length and the second is the deck-board length. The standard North American pallets are two-way and are lifted by the deck-boards.
- Heavier loads use four-way pallets that are built sturdier with hardwood, plastic or metal materials. The wooden pallets are often discarded upon use; while the more sturdy pallets are reused or are returned to the sender.
- Plastic pallets are made of recycled plastic bottle and box containers and are favored for the heavy loads. They are capable of reuse on many trips.
- Corrugated pallets are economical and are favored for lighter loads that can be moved by hand pallet trucks.
- Paper pallets are used for the lighter loads and are environmentally friendly for recycling.
- Steel pallets are used for moving and storing heavy loads. They are favored for durability, bug resistant, splinter resistant, sanitation, and are recyclable. However they are heavy, pricy and could rust.
- Aluminum pallets are favored since they are lighter than steel, weather resistant, do not wilt like plastic, do not corrode like steel, and do not rot like wood. They are often used in air and ship cargo and in outdoor storage applications.

13.5.7 Dunnage

Dunnage is an inexpensive material that is loosely placed between the cargo on a transport vehicle to prevent shifting of the cargo caused by the transportation motions. Dunnage is used on ships, barges, trucks, trains and aircraft; and is also used as padding in a shipping container. Inflatable plastic air bags are commonly used as dunnage, since they are flexible, reusable and easy to install in the shipping vehicle.

13.6 Maritime Security

With the large volume of commercial traffic at the ports and terminals, and since containers are of high use in maritime shipping, security is a major concern in fear of theft, sabotage or terrorist attacks. Even out on the seas and oceans, another fear is that of pirate hijackings. In the U.S., the Coast Guard and Customs and Boarder Protection is constantly on the alert for any such happenings at the ports, terminals and on the vessels. The coast guard is responsible for security at the ports and terminals, and in the waters for the safety of the barges, ships and boats. The customs and boarder protection regulates all international trade, with a mission to prevent any form of: terrorism, drug trade, illegal goods, theft, agriculture pests, unlawful animals and harmful diseases.

A big concern of governments is the risks of terrorism to the country from cargo entering at its borders. In the U.S., the Container Security Initiative (CSI) has been created for this concern. With seven-million containers entering the U.S. each year, representing about 90 % of all cargo, container security is the focus of the CSI efforts. CSI works with the foreign trade partners to help secure the incoming containers. They coordinate intelligence to identify containers that are at possible risk. The exporting country screens the risk containers at their home port. The arriving containers are scanned using detection technology with a Gamma Ray system. The system yields images of good quality that can identify the cargo inside the container and compare the image to the manifest. It also allows the operator to identify whether the loaded cargo looks suspicious and requires further inspection.

13.7 Dispatchers

A truck dispatcher coordinates the activities of all the vehicles with respect to the pickup and delivery assignments. They assign vehicles to service calls, manage the transportation schedules, and arrange the delivery of materials. For long distance runs, they maintain contact with the driver and keep track on the location and status of the vehicles.

A train dispatcher schedules and coordinates the activities of all the trains in an assigned area. He/she is aware of the delivery needs of all the cargo from origin to destination locations, and assigns the trains accordingly. The dispatcher also monitors the location of each train in his area and synchronizes their runs to minimize any interference between trains.

13.8 Bill of Lading

A bill of lading (BoL) is a document that lists the goods contained and gives title of the shipment content to a specified party. It is used in air, water, train and truck modes of transportation. The BoL is a legal document that lists the details on the goods transported from a seller to a recipient. It includes the type of goods, the number of items, and the destination address. The BoL is signed by the seller at time of shipping, then by the shipping carrier, and finally by the recipient.

BoL is used in international trade to guarantee that exporters receive payment and importers receive the goods. The BoL assigns title to the importer, certifies the merchandise is on a carrier, gives the condition of the merchandise upon loading, and requires the carrier to release the merchandise to the holder of the title or to a named representative.

13.9 Freight Bill

The typical freight bill includes the following:

- carrier's invoice
- origin of shipment
- destination of shipment
- supplier address
- consignee address
- cargo list
- total weight
- total cost

13.10 Freight Claims

This is a document that is filed with the carrier to recover any money lost due a fault by the carrier in not protecting the cargo.

The carrier is not at fault, however, in the event of the following:

- national disaster
- military attack
- government seizure
- freight not packed adequately
- fragile or perishable goods

13.11 Manifest

A manifest is a document that lists the cargo, passengers, and crew for a transport vehicle for the use of customs. A passenger manifest lists only the passengers along with vital information that pertains; and a cargo manifest, lists only the cargo with type and quantity. The passenger manifest is used to ensure the passengers who entered the vehicle at the origin are the same as those who depart at the destination.

13.12 Full Truckload (FTL)

Full truckload shipping is when a load is large enough to fill a semi-trailer or a intermodal container. The semi-trailers are generally 26–53 ft in length, and require for a full load of cargo to be profitable.

The shipper calls a truckload carrier to send a trailer for loading of the goods. The goods are generally from one customer and the cargo is typically homogeneous. The goods are loaded at the origin, and they stay intact without any intermediate handling, until they reach their final destination and are unloaded.

The full truckload carrier brings a semi trailer to a shipper who loads the trailer with freight for one destination. Thereupon, the driver returns to the shipper to collect the documents: bill of lading, invoice, and customs' forms, and then departs with the trailer on its way to the destination. The driver may go all the way, or may turn the trailer to another driver to finish the assignment. Full truckload drivers are restricted to a rate of 47 miles per hour over the entire trip.

Full truckload carriers typically specialize in shipping specific type of goods, like: food, grain, oil, iron, lumber, so on. Certain goods may require different equipment in shipping and may have alternative insurance needs. The carrier is also aware of constraints that restrict the type of goods that can be loaded together in a container.

13.13 Less Than Truckload Shipping

Less than truckload shipping (LTL) is the process of transporting the smaller loads of freight, whereby the load weights between 150 and 20,000 lb. LTL loads can share the space on a truck with other loads, and payments are usually based on the portion of space needed on the truck for the individual load. The typical concern in LTL shipping is tracking the load and seeking the delivery of the load to be claim free.

The opposite of LTL is parcel carriers who typically have shipments of one pallet that may include several cartons shrink-wrapped together to form a single unit; thereby reducing handling costs and damage of the content. They coordinate with shippers on a daily basis who notify them of a pickup need at their terminal.

13.14 Parcel Carriers

Parcel carriers are organizations that transport individual packages, parcels, or high value mail. The postal systems provide this service, along with express mail, and private package delivery carriers as: United Parcel Service, DHL Express, Federal Express, so on. Parcel deliveries are the main carriers of online shopping, and the volume of deliveries is now a big source of their activities.

13.15 Carrier Rating

The firms that frequently ship cargo and employ a variety of carriers, rely on a rating system to gauge the efficiency of the carriers. The typical features in the rating system are the following, arranged from the most important on down:

- on-time delivery
- loss or damage rate of cargo
- accuracy of the bill
- carrier equipment condition
- customer service

13.16 Bulk Cargo

Bulk cargo is goods that are transported in unpackaged form and in large volume, like: coal, sand, grain, oil, stones, cement, steel, so on. This is the cargo that is not boxed, bagged or hand stowed; or is in a large tank. They are transported on trucks, barges, ships, and trains. The cargo is either liquid or dry and is typically dropped or poured into a shipping unit's hold by a spout or a shovel bucket.

13.17 Break Bulk Cargo

Break bulk cargo is items to transport that are big in size and weight and too large to fit in containers, and are loaded individually. These are items like: industrial tractors, large engines, steel beams, so forth. They could also be transported in crates, drums, bags, boxes, barrels, and even on pallets. The break-in bulk is the place where the goods are transferred from one mode to another, as in the docks where the goods go between ship and truck, and large cranes are needed in the loading and unloading process. Transferring the goods from truck to ship is often as follows: Because of timing, the goods often arrive several days in advance by truck, where they are unloaded and stored in a warehouse until the ship arrives for loading.

13.18 International Trading Terms

Some of the common terms used in international trading are listed below. The terms define the responsibility in costs and security between the exporter, the importer, and the shipper. These are the following: Free on Board; Carriage and Freight; Carriage, Insurance and Freight; and Best Way.

13.18.1 Freight on Board (FOB)

The term free-on-board defines the terms in non-container sea freight and identifies the transfer of cargo ownership. The owner is responsible for any damage or loss during transport. The typical terms are listed below, partitioned by the owner of goods in transit, and who pays the carrier for shipment.

Buyer owns goods in transit:

• FOB origin with freight collect	– buyer pays
• FOB origin with freight prepaid	– seller pays
• FOB origin with freight prepaid & charged back	– seller pays & adds cost on invoice

Seller owns goods in transit:

• FOB destination with freight prepaid	– seller pays
• FOB destination with freight collect	– buyer pays
• FOB destination with freight collect & allowed	– buyer pays & deducts from freight cost

The seller is charged with delivering the goods to a specified destination in good condition. The costs include securing the goods on the carrier upon loading; with proper lashing (straps) of the items to prevent any unraveling. Special care is taken so that the goods are set in the carrier's hold in a way where they are stowed and secured to prevent any tipping or sliding.

13.18.2 Carriage and Freight

Carriage and Freight is when the importer pays the insurance costs and the exporter pays all the expenses on transportation from origin to destination.

13.18.3 Carriage, Insurance and Freight

Carriage, Insurance and Freight is when the exporter pays the insurance costs and the all the expenses on transportation from origin to destination.

13.18.4 Best Way

Best Way is a term used when the shipper determines the best rate for the transport of goods from origin to destination. This includes the costs of shipping, insurance and the travel timing.

13.19 Summary

Shipping categories are household good, express, parcel and freight. The goods are shipped on rail, roads, water, air and pipes, and the vehicles are by truck, rail, cargo boats and aircraft. The cargo handling is at ports, and container terminals. The goods are shipped using shipping containers, forklift trucks, cranes, pallets, and dunnage. Transportation of goods includes, dispatchers, manifest, bill of lading, bulk cargo and break-bulk cargo. Common industrial terms are free on board, carriage and freight, carriage, insurance and freight, and best way.

Chapter 14
Reverse Logistics

14.1 Introduction

Reverse logistics is the process of managing the operations concerned with any returned goods to the manufacturer. On average, near 5 % of the goods sold are returned to the original manufacturer for a variety of reasons as: worn out goods, damaged goods, unsold goods, recall goods, and so forth. The manufacturer is obliged to receive and process the return goods seeking any revenue that can be gained, or arranging for proper disposal. The role of processing return goods has expanded ever more as the environmental mandate of industrial growing green has gained strength. In the typical forward logistic way, the goods final destination is with the customer. As the goods become old, they may be replaced and returned to the manufacturer to begin the reverse logistics cycle. The returned goods are mostly one-model-at-a-time occurring in a disjointed manner. The return process is costly, perhaps 10 % of the total cost of the original sales price, requiring the manufacturer to seek as much value from the item as possible. In the typical situation, the returned goods are cleaned, tested for worthiness, and if accepted, are repaired and refurbished for resale. When not accepted, they are disposed in the proper manner.

The return goods are often from various reasons: worn out and no longer usable goods; damaged goods from shipment or use; new goods that are unsold at the retailer; goods ordered on the internet with cash-on-delivery terms and payment refusal when delivered; recalled goods; discontinued goods; and so forth. The returned items are also from warrantee claims, some from fraudulent claims, and some are stolen goods. Other return goods are of the reusable type whereby they are recycled after cleaning and repair. These include: pallets, containers, packages, bottles, cans, cylinders, and the like.

Two examples are described. The first concerns appliances where a customer purchases an appliance, and later in time, the unit does not operate properly and the defective product is sent back to a return facility that refurbishes the product as new for future use. The second pertains to the automotive industry from the truck

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0_14

industry where worn out and discarded engines, brakes, or transmissions are removed from the vehicle and are remanufactured to like-new status.

14.2 Revenue from Returned Products

Return products pose a variety of costs on the manufacturer. They are often dropped off at a dealer, from where subsequent bookkeeping and transportation costs occur to arrange for shipment to the manufacturer's core location. The location is a warehouse setting that requires bin and floor space for the variety of cores. The products arrive on a one-at-a-kind basis and identification tags are needed to keep track of the inventory, their status, and where they are located. The eventual disposition of the products is costly as well, no matter whether by refurbishment, remanufacturing, redistribution, donation or landfill. With all the costs that occur, it behooves the management to seek out any revenue benefits from the return products. Often third party service providers offer their services on this end as well. Below is a brief description on some of the common revenues streams available to the manufacturer of the returned goods.

14.3 Refurbishment

Some manufacturers and third-party enterprises gain revenue from returned products that they refurbished and resell at discount prices. Electronic products are typically of this type. The products are tested and repaired for functionality and become defect free, whereupon, they are certified to operate as new equipment.

On older products, the inside components may not be available to purchase from the component source any longer when the product is old and thereby, the components removed from the product become valuable items in the service parts market.

14.4 Remanufacturing

Another revenue source is from cores that are replaced and discarded main components of finished good items. The cores are like engines, gear-boxes, clutches, and so on. They are purchased by manufacturers and some by third-party enterprises and stored in inventory for future need. Eventually, as demands call, the cores are disassembled, where all inner component parts are cleaned, tested, repaired or replaced and the core is reassembled. The remanufactured cores are sold with full new warranty and with discount prices.

When the main core is not deemed reusable, the inner components may still be removed and refurbished as renewed components. They become valuable service parts, especially if the component vendor stops supplying the component when the demand is low.

14.5 Redistribution

Sometimes the returned products to a distributor come from retailers who are unable to sell the items at their location. The products could have been on consignment at the retailer who starts with a large inventory at no cost. These items cover valuable retail space that is needed to stock a new set of merchandise. The distributor can send the items to another location or to discount stores for sale. These products are still new and in good working condition. This situation occurs frequently in the clothing industry where the products, (shoes, shirts, sweaters, so on), may be placed in various retail outlet stores.

14.6 Donation

When the returned items to a manufacturer are not worthy to refurbish, remanufacture of redistribute, the manufacturer may decide to donate the items to a charitable organization. Could be clothing, used cars, computers, refrigerators and the like. The manufacturer receives some revenue with a tax return.

14.7 Discard

Sometimes the returned item has no value and does not apply to any of the revenue ventures described. As a last resort, the manufacturer is obliged to remove any dangerous material and discard the item in an environmentally prescribed manner.

14.8 Remanufacturing in the Appliance Industry

A customer purchases an appliance, and as time moves on, the unit does not operate properly and the customer notifies the dealer or the original equipment manufacturer (OEM) about the fault and seeks help on how to repair or replace the unit. Return logistics is the process on the flow of the defective product back to a return facility that arranges to replace the product to the customer, and also to refurbishes the product as new for future use.

14.8.1 Finished Good Item (FGI)

A finished good item is like an appliance: washing machine, dishwasher, microwave, and so forth. The FGI has various parts that give the functionality, like a motor or a printed circuit board. Each such part has a series of imbedded components attached that produce the functionality of the part; control board, relays, capacitors, resistors, flat cable, so on.

14.8.2 OEM and 3PP

The OEM either has its own return facility or has an agreement with a third-party provider (3PP) who has full responsibility to process all return parts from the OEM. Each returned part is called a core. For a given OEM, a finite number of parts are included as components of all returnable items, and these are the main parts of the FGI units from the OEM.

14.8.3 Engineering Analysis

For each of the candidate parts that can be returned, an engineering team analyzes the core and generates an instruction guide for any disassembly, inspection, assembly and repair that will be needed in the future repair, refurbish or remanufacture of the core. The engineers also devise the tools, fixtures and testing devices needed for each such part.

14.8.4 Receiving Cores

In the typical repair of a FGI, the defective part is removed from the unit by a repairman and replaced with a non-defective part allowing the unit to operate in working order again. The faulty part is shipped to a warehouse facility that processes the retuned goods, called cores. The receiving location identifies the part of the core and records the date, location and cause for its return. The core may be tagged with a barcode, and possibly wrapped to prevent further damage.

14.8.5 Core Inventory

Each returned core is placed in the warehouse along with the other cores of the same part in a common warehouse bin. The core inventory will stay in the warehouse, sometimes for months, until a future date when they accumulate to a volume that is feasible for remanufacture in a batch of an acceptable lot size.

14.8.6 Remanufactured Parts

Eventually the cores will be refurbished or remanufactured into defect free working units, and thereby are called remanufactured parts whose warranty is the same as a new part. These parts are also labeled as rebuilt parts.

14.8.7 Demand Forecasts

In the meantime, the demand for each part is monitored on the database of the return facility. This demand occurs when a FGI with the part requires a repair. For the newer FGIs, the part demands are typically low since few parts become defective in the early life cycle stage. As time goes on, the parts begin to deteriorate and the replacement demands start to rise.

14.8.8 Lot Size

The operation's management determines a cost effective lot size, N, for the various parts. This lot size for a model is the quantity of units that is deemed economic to process as a batch in production.

14.8.9 Schedule Production

The schedule to refurbish or remanufacture a part depends on information in the database. A call for production is triggered when the forecast of demands become large compared to the facility's FGI inventory. The production, however, cannot take place until the core inventory on the part is of an adequate size, typically the lot size N or larger.

14.8.10 Workstations

Should the core inventory on the part be adequate, prior to the remanufacture process, the engineering guide sheet is used to project the quantity of replaced components and material that will be needed for the batch size of units. After the number of workstations is noted, the floor supervisor arranges the workload by station to be as equally balanced as possible. The workstations are arranged in a linear or in a cellular arrangement so that the units can flow readily from one station to another. Each workstation is provided with the appropriate material, components, fixtures, tools, and testing equipment as needed. Guide sheets are also placed at each workstation so the operator's can properly carryout their assignments.

14.8.11 Refurbish

Some core parts have components that are easy to access and observe. These units do not need disassembly in processing and the components are readily accessible to the operator to process as needed. The components of each unit is inspected, cleaned, repaired, replaced, and tested as assigned in the workstation guide sheets. A latter station does a final test on the unit to ensure it is defect free.

In the event a unit does not pass the final test, the defective unit is sent to another workstation for engineering reexamination. Upon subsequent analysis, the engineers may alter the tasks and guide sheet on refurbishing the units.

14.8.12 Disassembly

Some core parts require disassembly of the inside components in the remanufacturing process. A disassembly line is setup and the workstations are assigned their tasks in a balanced manner. At the start of the disassembly, N cores are gathered for processing. Along the line, the detached components are cleaned, repaired, and placed in separate containers for future need in the subsequent assembly. Upon removal of all components of the core unit, the core base itself is cleaned and inspected for adequacy in future use.

14.8.13 Assembly

The components of the disassembled N core units have been cleaned, refurbished and gathered for the subsequent assembly of N units. In the preassembly stage, any component units tossed out in disassembly are replaced with new or rebuilt like units. With an adequate supply of components available, the assembly of N units can begin. As before, the operations management sets the number of workstations and provides guide sheets per station for the tasks to perform by part. The tasks per station are assigned so that the station times are balanced as much as possible. The tools, devices and component inventory needed for each workstation is provided as needed.

14.8.14 Packaging

A final workstation has the task of testing the final product so that it meets all warranty requirements. The reliability of the product is the same and sometimes better than a new OEM part since all faults have been detected and corrected in the remanufacturing process. Upon meeting the test adequately, each unit is packaged individually with appropriate information on each label that includes: "Remanufactured Part", the part number, the remanufacture firm location, and the date when remanufactured.

14.9 Remanufactured Part Inventory

The packaged part is now a remanufactured part that is stored in the inventory awaiting a demand that eventually will consume and ship it away.

14.9.1 Production Schedule

A production scheduling system is run on a periodic basis of weeks or months. The system identifies the parts that require remanufacturing of cores for the next subsequent period, and the quantity of each, denoted here as Q. For convenience, the description to follow assumes the schedule is revised on a once-per-month basis. The input data needed for a given part is listed below:

$$\begin{aligned} F &= \text{average monthly forecast for the part} \\ P_{SS} &= \text{safety stock parameter(months)} \\ N &= \text{recommended lot size} \\ OH_p &= \text{current on-hand of the part} \\ OH_c &= \text{current on-hand of the core} \end{aligned}$$

With the above data, the safety stock is computed by: $SS = P_{SS} \times F$ and the order point is $OP = SS$. The order point is needed in the following computations. The algorithm assumes a part is not scheduled unless the core inventory is equal or above, N, the specified lot size. The logic to determine the schedule quantity for each part is listed below:

$$\begin{array}{lll} \text{If } [OH_c < N]: & & Q = 0 \\ \text{if } [OH_c \geq N] \text{ and } [OH_p - F] > OP: & & Q = 0 \\ \text{if } [OH_c \geq N] \text{ and } [OH_p - F] \leq OP: & & Q = N \end{array}$$

Example 14.1 A part has the following data: $F = 20$ is the forecast per month, $N = 50$ is the recommended lot size, $P_{SS} = 0.5$ (months), and thereby $SS = OP = (0.5 \times 20) = 10$ is the order point.

Suppose the on-hand data at a given month is $OH_p = 62$ and $OH_c = 44$. Because $[OH_c < N]$, $Q = 0$.

Assume at the next month, the monthly forecast remains at $F = 20$, while $OH_p = 41$ and $OH_c = 53$. Since $[OH_c \geq N]$ and $[OH_p - F] > OP$, $Q = 0$.

Assume at the next month, the forecast is still $F = 20$, while $OH_p = 23$ and $OH_c = 62$. Since $[OH_c \geq N]$ and $[OH_p - F] \leq OP$, $Q = 50$.

14.10 Remanufacturing in the Auto Industry

Remanufacturing is the process of taking a product that has been discarded, due to wear and deterioration, and rebuilding it to meet the specifications from the original equipment manufacturer (OEM). The product is rebuilt using a combination of reused, repaired and new components. Remanufacturing received a boost in 1995 when the Environmental Protection Agency (EPA) promoted the concept of reducing waste and promoting resource conservation by way of materials recovered from solid waste. The EPA provides a list of products that can be rebuilt with recovered material and encourages the state and federal agencies to purchase the rebuilt products, accordingly. The purchase price of a rebuilt product is often 50–70 % of the price from a new product.

Common examples occur in the vehicle industry where engines, transmissions, and brakes are rebuilt in this way. For convenience, the discussion below is from the truck industry where worn-out and discarded engines are removed from the vehicle and are remanufactured to like-new status. The removed products are called cores, and as such, the removed engines are core engines.

14.10.1 Rebuilt Engine Example

Consider a plant that specializes in renewing truck engines. The final product sells for 50–70 % of a new engine and the warranty is the same as a new engine. The plant has a large inventory of core engines of different models in the plant warehouse waiting for demands from customers. The core engines have been removed from trucks and replaced with newer engines. Each core engine includes the components that are associated with the engine, and thus, the core engine includes the set of core components that have been installed with the particular core engine. As the customer demands are received for the renewed engines by model, the demands are included on an order board with desired due dates.

When the demand builds up, a call is generated to build the engines on one of the plant's two assembly lines. The daily shift schedule typically is for $N = 10$ engines per line.

The operation of the plant is not easy; mainly because of the uncertainty in dealing with used engines and their components. Below describes how the planning takes place 1 week prior to the production phase. Planning's role begins by identifying an engine model to process in a forthcoming week. With the engine named, the next step is to check each component of the engine to determine if a purchase quantity is needed prior to the production process. The production process begins with disassembly of the components from the core engine unit, and ends with assembly of the components back into the unit.

14.10.2 Planning

A planning phase begins 1-week prior to production by running two computer systems: a master production schedule (MPS); and a material requirement planning (MRP) computer system. Subsequently, 1 week later, the production phase begins and is completed in 1 day. A description on the two computer systems is below.

14.10.3 Master Production Schedule (MPS)

The MPS system is run once a week and determines the daily schedule of core engine models for the next week on each of two lines. The number of engines for a model is denoted as N, and thereby, N is the number of engines that will be processed in disassembly and subsequently in assembly. MPS's role is to seek an engine model that has N or more core units in the warehouse and whose demand is large enough to warrant production. Upon selecting, the engine model is scheduled for production on one of the disassembly lines for a particular day of next week.

14.10.4 Material Requirement Planning (MRP)

After the MPS system is run, the MRP system begins. For every core engine model scheduled by MPS, the MRP system determines the stock status on each of K core components associated with the engine to determine how much, if any, new stock of the component needs to be purchased prior to the start of next week's assembly. The quantity to purchase for component k is denoted as Q_k for $k = 1–K$.

14.10.5 Production

A few days prior to the scheduled production day of the core engine, N core units of the engine model are taken from the warehouse and moved to the staging area of the disassembly area.

On the day the core engines are processed, the following functions take place in the plant: disassembly, cleaning, inspection, machining, assembly and testing. A description on each is below.

14.10.6 Disassembly (DSM)

In DSM, each of the major components: (crank shaft, crank case, cylinder heads, fuel pump, water pump, turbo, rods) are removed from the core engine. These units are called: core components.

Some core components are known in advance to be not worthy of salvaging and are 100 % scrapped upon disassembly.

14.10.7 Cleaning, Inspection and Machining

Upon disassembly of all core components, the following takes place:

1. The core engine is cleaned and inspected for any flaws.
2. The core components are individually cleaned and inspected for their capability on continuing use. The inspection yields three status outcomes, labeled as follows:

   A1 = acceptable as is
   A2 = acceptable with further machining and/or welding
   A3 = not acceptable

   After inspection, the A2 units are processed with further machining and/or welding as needed. The A3 units are scrapped.

14.10.8 Assembly (ASM)

Assembly begins after the core engine and all of the core components are cleaned, inspected, refurbished, and when all needed new component units are available.

14.10.9 Fallout

On some occasions, a shortage of one or more component units happens. This event is called a fallout, and requires buying more component units for later assembly, or disassembling another core engine to attain the component unit(s).

14.10.10 Testing

After assembly of all the refurbished cores on an engine, the engine is rigorously tested and upon acceptance, it is certified and is deemed ready for a customers use. The engine is now called a remanufactured (rebuilt or renewed) engine, with a warranty the same as a new engine.

14.10.11 MRP Computations

As stated earlier, the role of the MRP system is to compute the quantity of units to purchase (if any) for each core component of an engine 1-week prior to the production schedule date of the engine. The number of components is N_c and the schedule of N engines is planned for the production date. The description includes the following: status and probabilities, estimate of inspection outcomes, on-hand, on-order and safety stock, and the buy quantity.

14.10.12 Status and Probabilities

Based on previous history on core components, the probability on the outcome status from inspection is estimated for each of the N_c components of the engine model. These are labeled as:

p_1 = probability of A1
p_2 = probability of A2
p_3 = probability of A3

The sum of the three probabilities is one. Hence for an engine model, X, the following data is available for each core component:

Engine model = X

Component	p_1	p_2	p_3
1	–	–	–
. . .			
N_c	–	–	–

14.10.13 Estimate of Inspection Outcomes

For each core component, the MRP system estimates the number of units that will fall into each of the three categories A1, A2, A3, and these are denoted as x_1, x_2 and x_3, respectively. The method is shown below for a particular component where N is the shift assembly schedule quantity for the core engine.

status	Probability	Projection
A1	P_1	$x_1 = N \times p_1$
A2	P_2	$x_2 = N \times p_2$
A3	P_3	$x_3 = N \times p_3$

14.10.14 On-hand, On-order, and Safety Stock

More data is needed for the MRP computations. First is the inventory currently on-hand for each of the components of the engine model. The sum is called the effective on-hand inventory. Second is a projection of the on-order inventory; and third is a management parameter labeled as safety stock (SS). Below defines the data source.

$OH_0 =$ current new on-hand inventory for the component,
$OH_1 =$ current on-hand for units of A1 status for the component
$OH_2 =$ current on-hand for units of A2 status for the component
$OH = (OH_0 + OH_1 + OH_2) =$ effective on-hand inventory
$OO = (x_1 + x_2) =$ projected on-order inventory
$SS =$ safety stock $=$ a management parameter

The safety stock is an extra layer of component stock that might be needed in the assembly of N engines because a portion of the component inventory (on-order) is a projection and is not absolute. Safety stock lowers the probability of a fallout.

14.10.15 Buy Quantity

The quantity of units to buy for a component of the engine can now be computed, as described below. Recall, N is the number of engine units to assemble.

$$Q = \text{buy quantity for the component.}$$

The computations for Q is below:

$$\text{If } (N + SS) \le (OH + OO), \quad Q = 0$$
$$\text{If } (N + SS) > (OH + OO), \quad Q = (N + SS) - (OH + OO)$$

Example 14.2 Suppose a situation where the MPS calls for N = 10 core engines in disassembly and subsequently in assembly. Assume, one of the core components in the MRP computations has probabilities of: 0.2, 0.3 and 0.5, for status A1, A2 and A3, respectively. Since, N = 10, the projections for A1, A2 and A3 become, 2, 3, and 5, respectively. Assume further, the on-hand for the three categories are: 2, 1 and 3 for A1, A2 and new, respectively.

With the above information, the computations for the on-hand, OH, and on-order, OO, are as follows:

$$\mathrm{OH} = (2 + 1 + 3) = 6$$
$$\mathrm{OO} = (2 + 3) = 5$$

Recall, the buy quantity of new stock, Q, is obtained as below:

$$\mathrm{If}(\mathrm{N} + \mathrm{SS}) \leq (\mathrm{OH} + \mathrm{OO}): \quad \mathrm{Q} = 0$$
$$\mathrm{if}(\mathrm{N} + \mathrm{SS}) > (\mathrm{OH} + \mathrm{OO}): \quad \mathrm{Q} = (\mathrm{N} + \mathrm{SS}) - (\mathrm{OH} + \mathrm{OO})$$

Below shows how much to buy when the safety stock is set at SS = 0 or SS = 1 or SS = 2.

$$\text{If } \mathrm{SS} = 0: (10 + 0) \leq (6 + 5), \text{and thereby, } \mathrm{Q} = 0.$$
$$\text{If } \mathrm{SS} = 1: (10 + 1) \leq (6 + 5), \text{and thereby } \mathrm{Q} = 0.$$
$$\text{If } \mathrm{SS} = 2: (10 + 2) > (5 + 5), \text{and thereby } \mathrm{Q} = 1.$$

14.10.16 Probability of a Fallout

Recall, a fallout is the term used when the available inventory of a component part is short of the needs in the final assembly of the core-engine. The binomial probability distribution is used to compute the probability of a fallout. Assume the Example 14.2 where N = 10 core engines are disassembled and p = 0.5 is the probability of an acceptable unit in the disassembly process for one of the core components. The probability of x acceptable units using the binomial distribution is the following:

$$\mathrm{p(x)} = \mathrm{N!/[x!(N - x)!]p^x(1 - p)^{N-x}} \quad \text{for } \mathrm{x} = 0 \text{ to N}$$

The cumulative probability of x or less acceptable units is the following:

$$\mathrm{F(x)} = \sum\nolimits_{\mathrm{k}=0}^{\mathrm{x}} \mathrm{p(k)}$$

Table 14.1 lists the probabilities for k = 0–10 when p = 0.5.

Recall OH = on-hand, SS = safety stock, N = number of units, and OO = on-order. Note also where OO is uncertain and is a random variable the same as

Table 14.1 Probability p(x), cumulative probability F(x), when p = 0.5, and N = 10, for k = 0–10

x	p(x)	F(x)
0	0.001	0.001
1	0.010	0.011
2	0.044	0.055
3	0.117	0.172
4	0.205	0.377
5	0.246	0.623
6	0.205	0.828
7	0.117	0.945
8	0.044	0.989
9	0.010	0.999
10	0.001	1.000

x from the binomial distribution. Further, observe, the probability of a fallout is the following:

$$p(\text{fallout}) = p\left[OO \leq N - (OH + SS)\right]$$

When x = OO, the above becomes:

$$p(\text{fallout}) = p[x \leq N - (OH + SS)]$$

At N = 10, p = 0.5, OH = 6, the probability of a fallout for SS = 0, 1 or 2 are computed below.

$$\text{If SS} = 0\text{: } p(\text{fallout}) = p[x \leq 10 - (6 + 0)] = F(4) = 0.377$$
$$\text{If SS} = 1\text{: } p(\text{fallout}) = p[x \leq 10 - (6 + 1)] = F(3) = 0.172$$
$$\text{If SS} = 2\text{: } p(\text{fallout}) = p[x \leq 10 - (6 + 2)] = F(2) = 0.055$$

Note how the addition of safety stock reduces the probability of a fallout.

14.11 Summary

Products are purchased new by the customers and used till they don't function properly, whereby they are discarded and end in a return facility. The return items and their components are called cores. When possible, for some products, the unit in whole is refurbished to its original warranty status. For other products, the components are removed from the whole units, and the components are rebuilt like new. The rebuilt and refurbish processes requires a combination of disassembly, cleaning, inspection, machining, new parts, and assembly. The remanufactured items become available for sale at reduced costs, and mostly with new warranties to customers.

Recommended Reading

Altay, N., & Lewis, A. (2011). *Service parts management: Demand forecasting and inventory control*. New York: Springer.
Bedworth, D. D., & Bails, J. (1987). *Integrated production control systems, analysis design*. New York: Wiley.
Brown, R. G. (1959). *Statistical forecasting for inventory control*. New York: McGraw Hill.
Brown, R. G. (1962). *Smoothing, forecasting and prediction of discrete time series*. Englewood Cliffs, NJ: Prentice Hall.
Chase, C. W. (2013). *Demand driven forecasting*. New York: Wiley and SAS Business Series.
Chase, R. B., & Aquilan, N. J. (1989). *Production and operations management: A life cycle approach*. Homewood, IL: Irwin.
Chisholm, R. K., & Whitaker, G. R. (1971). *Forecasting methods*. Homewood, IL: Irwin.
Chow, W. (1990). *Assembly line design methodology and applications*. New York: CRC Press.
Doty, L. (1996). *Statistical process control*. South Norwalk, CT: Industrial Press.
Elsayed, E. A., & Boucher, T. O. (1985). *Analysis and control of production systems*. Englewood Cliffs, NJ: Prentice Hall.
Groover, M. P. (2001). *Automation, production systems and computer integrated manufacturing*. Englewood Cliffs, NJ: Prentice Hall.
Groover, M. P. (2010). *Fundamentals of modern manufacturing*. New York: Wiley.
Holt, C. C., Modigliani, F., Muth, J. F., & Simon, H. A. (1960). *Planning production inventories and work force*. Englewood Cliffs, NJ: Prentice Hall.
Hong, W. C. (2013). *Intelligent energy demand forecasting*. New York: Springer.
Lambert, A. J. D., & Gupta, S. M. (2005). *Disassembly modeling for assembly, maintenance, reuse, and recycling*. Boca Raton, FL: CRC Press.
Lewis, C. (1997). *Demand forecasting and inventory control*. New York: Wiley.
Makridakis, S., & Wheelwright, S. C. (1977). *Interactive forecasting*. San Francisco: Holden-Day.
Makridakis, S., & Wheelwright, S. (1978). *Time-series forecasting: Methods and applications*. New York: Wiley.
McGovern, S. M., & Gupta, S. M. (2011). *The disassembly line: Balancing and modeling*. New York: McGraw Hill.
Minner, S. (2000). *Strategic safety stocks in supply chains*. New York: Springer.
Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2008). *Introduction to time series analysis and forecasting*. New York: Wiley.
Montgomery, D. C., & Johnson, L. A. (1976). *Forecasting and time series analysis*. New York: McGraw Hill.
Oakland, J. S. (1996). *Statistical process control*. Oxford, England: Butterworth-Heinemann.

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0

Ortiz, C., & Ortiz, O. A. (2006). *Assembly, designing, constructing and managing a lean assembly line*. Boca Raton, FL: CRC Press.
Plosel, G., & Wight, O. (1967). *Production and inventory control*. Englewood Cliffs, NJ: Prentice Hall.
Prenting, T. O., & Thomopoulos, N. T. (1974). *Humanism and technology in assembly line systems*. Englewood Cliffs, NJ: Spartan–Hayden.
Rushton, A., Croucher, P., & Baker, P. (2006). *The handbook of logistics and distribution management*. London, England: Kogan Page.
Shewhart, W. A. (1939). *Statistical method from the viewpoint of quality control*. Washington, DC: U. S. Department of Agriculture.
Thomopoulos, N. T. (1980). *Applied forecasting methods*. Englewood Cliffs, NJ: Prentice Hall.
Thomopoulos, N. T. (1990). *Strategic inventory management and planning*. Carol Stream, IL: Hitchcock.
Thomopoulos, N. T. (2012). *Fundamentals of production, inventory and the supply chain*. New Delhi, India: Atlantic.
Thomopoulos, N. T. (2014). *Assembly line planning and control*. New York: Springer.
Thomopoulos, N. T. (2015). *Demand forecasting for inventory control*. New York: Springer.
Vollman, T. E., Berry, W. L., & Clay Whybark, D. (1997). *Manufacturing planning and control*. New York: McGraw Hill.
Weiss, H. J., & Gershon, M. E. (1989). *Production and operations management*. Boston: Allyn & Bacon.
Wild, R. (1975). *Work organization, a study of manual work and mass production*. New York: Wiley.

Index

N.T. Thomopoulos, *Elements of Manufacturing, Distribution and Logistics*,
DOI 10.1007/978-3-319-26862-0

P

Q

R