Process and Strategy (1)

Capacity (6)

Throughput (7)

Wailing Line Buffer (3)

Moving Average

Exponential Smoothing (2)

EOQ

ROP Part1

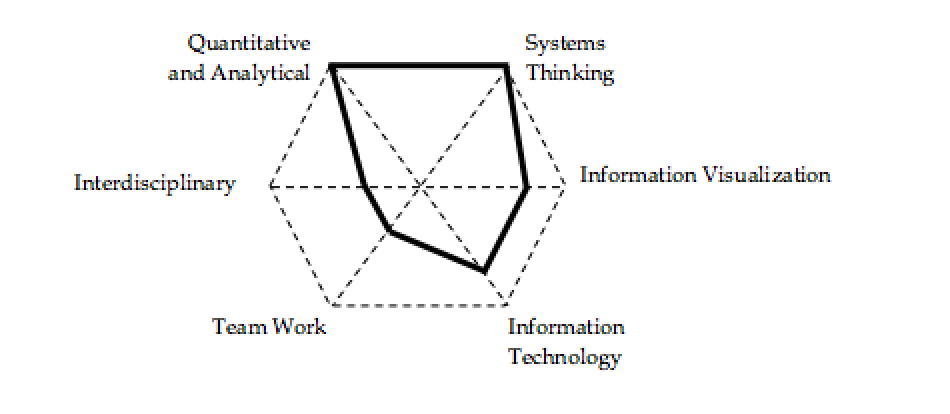
**Lecture I: Processes View & Strategy**

Customers have certain expectations about products and services that they buy. These expectations can be *physical* such as comfort, convenience, and safety; *psychological* such as relaxation and peace of mind; and *social/spiritual* such as feeding the poor. These expectations should be met within customers’ budgets. Business processes create manufactured products and deliver services. Some examples include the flow of cars in a General Motors assembly plant, flow of customers in a Wells Fargo branch, flow of patients at the UCLA Medical Center, flow of cash in Fidelity Investments, and flow of students during their two-to-five year program at CSUN. In all these systems, flow units (natural resources, semi-finished goods, products, customers, patients, students, and cash) flow through a set of processes (formed by a network of activities and buffers) using Human resources and Capital resources (such as equipment, buildings, tools) and an information infrastructure and value system to become a desired output. We mainly focus on the systems with discrete flow units – such as the systems stated above - as opposed to continuous flow, which is the domain of chemical engineering.

**Specific Features of the Course.** One of the most binding constraints of business school students – from the time they are admitted to college as “raw material” from high school to the time they graduate and leave college as “the final product” – is their low quantitative and analytical skills. According to the CEO of American Express in his 2011 interview with Fareed Zakaria on CNN, the low level of quantitative capabilities of our graduates has kept us from excelling beyond the graduates of rising countries such as China and India. Organization for Economic Cooperation and Development (OECD) Skills Outlook (2013) compares the literacy, mathematics, and computer skills of U.S. residents with people in other OECD countries. In mathematics, U.S. trailed 18 countries and beat Italy and Spain.

Believing that managers cannot go far if their quantitative and analytical capabilities are below a threshold, we have tried to improve these qualifications through our Operations Management (OM) classroom. In a typical traditional OM class, about 2/3 of the class time is spent on delivering the content. The rest is mainly spent on problem solving and case studies, term projects and simulation games. We have tried to improve these capabilities through flipping our Operations Management classroom. By delivering lectures using screen capture technology, students can learn the material at a time and location of their choice, which allows them to pause, rewind, or fast forward professor’s lectures when they need it. The class time is no longer spent on teaching basic concepts, but rather on more value-added activities such as problem solving, answering questions, creative-thinking, systems-thinking, as well as real world applications and discussions, potential collaborative exercises such as case studies, and virtual world applications such as web-based simulation games. A flipped classroom includes components of both an online and a traditional course. A flipped classroom is an online course because its online components must compete with the best of the online courses. A flipped classroom is also a traditional course because not even a single class session is cancelled while all the lectures are delivered online. This core concept is reinforced by a network of resources and learning processes, ensuring a smooth, lean, and synchronized course delivery system.

The specific features of the course and their relative importance are depicted below.



**Quantitative and Analytical.** We use Operations Management as a tool to improve the quantitative and analytical capabilities of our students. Students will learn to develop a structured, data-driven, analytical, and quantitative approach to discuss the core Operations Management concepts.

**Systems Thinking**. We try to improve systems thinking capabilities of our students by teaching the basic concepts of operations management not as isolated islands but as a total system designed towards improving process flow. Students will learn to implement the process view as the unifying paradigm to study the core concepts in the operations management (retrieved from Anupindi, et al., 2012).

**Visualization of Data and Information**. Besides quantitative representation (translating long writings into mathematical relationships), students will practice tabular representation (translating long writings into tables), and schematic representation (translating several pages of writing and tables into a graph, flow chart, or picture). Students also learn how to deal with large, unorganized, or erroneous big data sets.

**Information Technology.** We try to enhance students’ knowledge in spreadsheet modeling. We have learned that understanding the knowledge behind these models and developing small pilot spreadsheets leads to a better understanding of the course material. Through case studies, as well as web-based games, the stage is set to motivate the students to develop spreadsheet-based models.

**Teamwork.** We encourage collaborative learning and creative thinking. The first day of class is not spent on the syllabus, but rather on the importance of teamwork. Students are encouraged to have weekly team meetings to go over the already solved assignments and gain new insights in the web-based games and case studies. Academic integrity and ethics are also implicitly addressed in the course.

In this course we look at everything as a process. Process view: Input 🡪 Process🡪 Output. Here: Inputs can be tangible or intangible, natural or processed resources, parts and component**s**, energy, data, customers, cash, etc. Outputs can be tangible or intangible items such as products, byproducts, energy, information, served customers, cash, relief, etc., that flow from the system back into the environment.

Examples of Input 🡪 Process🡪 Output

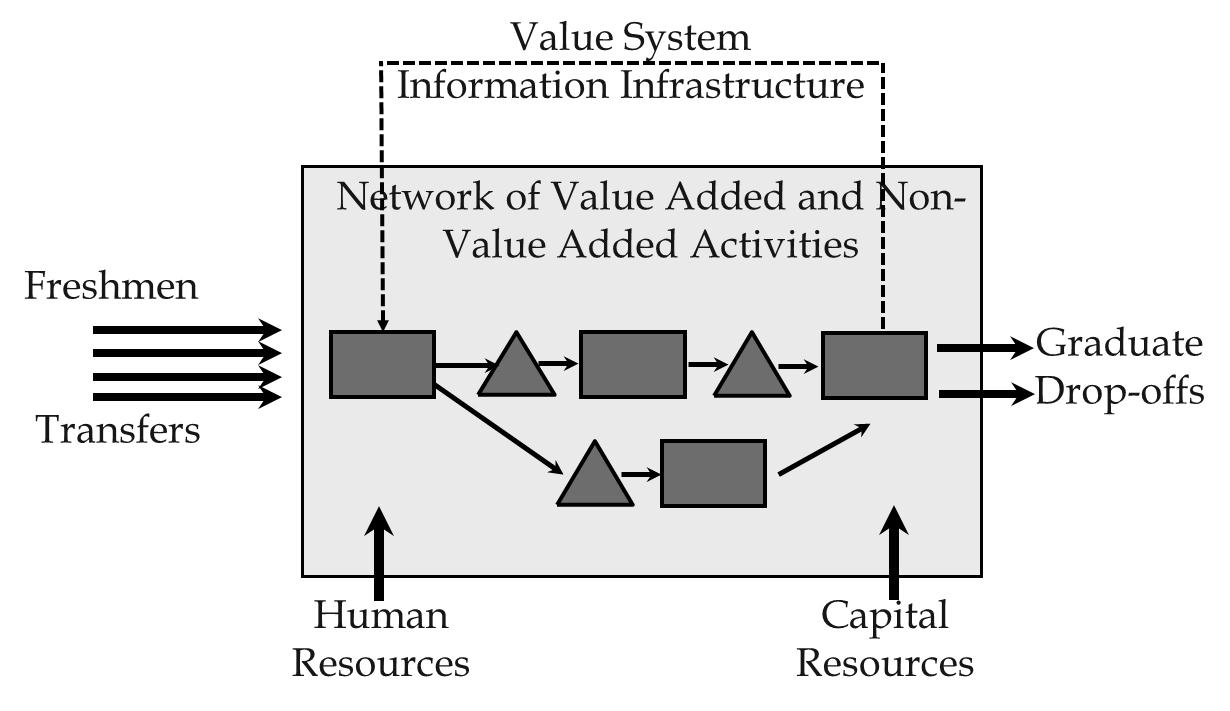
Raw material 🡪 Manufacturing Process 🡪 Finished goods

Data 🡪Accounting Process 🡪 Financial Statements

Accounts Receivable 🡪 Billing Process 🡪 Cash

Unsatisfied customer demand 🡪 Transformation Process 🡪 Satisfied customer demand

Five Components of Process View: (1) inputs, (2) outputs, (3) human resources and capital resources, a (4) network of value added/non-value added activities and buffers, and an (5) information structure and value system.



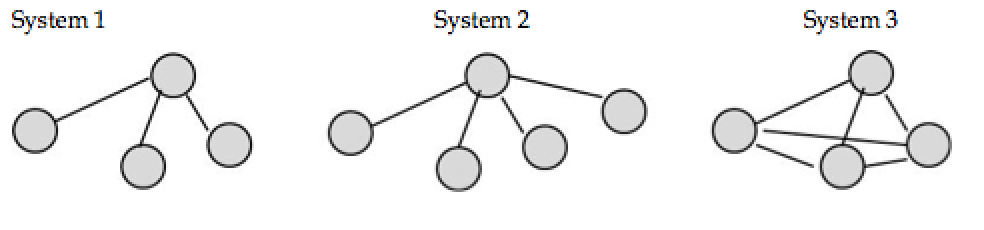
In process flow mapping (process blue printing), material flow is shown by solid lines, and information flow is shown by dashed lines. When inputs pass through the network, they are called flow units; when they leave the system they are output. A flow unit could be an item of inputs, outputs, or a combination of both-it depends on the reasons why we are looking at this process. Values added activities – activities making an input one step closer to its output form - are shown by rectangle; non-value added activities- buffers, storages, waiting lines- are shown by triangle. Not all writings are non-value added, for example, aging of cheese or hardening a concrete are value added activities.

The following are examples of different processes and their components. The fulfillment process starts from receipt of an order and ends at delivery of a product. Flow units of this process are orders. In an outbound logistics, process starts at the end of production and ends when the product is delivered to the customer; flow units are products. In a supply cycle, flow units are supplies; the system boundaries, the border limits of the system starts from issuing a purchase order and ends at receipt of the supplies. In a customer service process, customers are flow units. It starts from the point when an unsatisfied customer shows up until the point when the satisfied customer leaves the system. In a research and development process, flow units are projects. System starts from recognition of the need and ends at launching the project. In a cash cycle process, flow units are cash. The system boundary limits of the system start from the point when expenditure is accrued until the point when revenue is collected, regarding the product or services that this expenditure went through. So, this expenditure went to a product or a service; it went through a transformation process: it was sent it to a customer. Then, the revenue from the customer was collected, and that is from point when the cash was put into the system until the point that the cash was collected from the system.

Every component of a process is interconnected into a system. The relationship among those components and the objective is the goal of existence of this system. The battery limit of the system is the border between the environment and the system. Environment is everything outside the system. Usually, we do have control of variables and parameters inside the boundaries of the system, but we don't have much control over variables and parameters in the environment. Variables and parameters inside the system are called *endogenous* and outside the system-*exogenous.* A system is defined by its components, interrelationship between those components, and the objective of the system.

Systems can grow by increasing the number of their components. The second system in Figure 1 can be preferred to the first system because the second one has one additional component. Systems can also grow by increasing or enhancing their relationship between components. The third system can be preferred to the second system because the third one may have more integrated relationship between its components. Therefore, the third system can perform much better in a complex environment; it can also benefit from synergy between the components.

Figure 1.



However, the whole system performs better than its components. In a system view, the whole is greater than sum of its parts; 2>1+1!



Imagine desires of Sales, Purchasing, and Production departments in a seasonal industry. Imagine a company with two branches. Process flow in both branches is similar and contains two sequential operations. The first operation takes 5 minutes, the second 10 minutes. The first branch has produced an average of 11 units per hour in Operation 1 and 4 units per hours in operation 2. The second branch has produced an average of 5 units per hour in Operation 1 and 5 units per hours in operation 2. Who deserves appreciation?

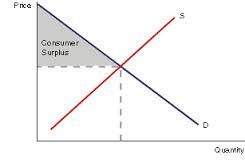
**Principle:** Performance measure of Sub-systems must be linked to the performance measure of the total system. Performance of a sub-system must be measured in terms of its impact on the performance of the total system.

Customers assign four attributes to a product; cost, quality, time, variety. These four attributes are often referred to as the four dimensional space. Companies develop a ***customer value proposition*** to fulfill customer expectations**.** Products have two classes of characteristics: order qualifiers and order winners. **Order qualifiers** are characteristics of a product that convince a customer to consider a product. **Order winners** are characteristics that convince the customer to buy the product.

Different market segments define order qualifiers and order winners differently. For example, order qualifiers in the eyes of a commercial airline flyer are entirely different from order qualifiers of a wealthy businessman who wants to buy a private jet.

Customers purchase products based on the value that they will derive from a product. That value is the greatest amount that the customer is willing to pay. Several companies might propose their products with different prices. If customers view different values in these different products, then they will buy the product with the largest gap between the value they derive compared to the product or process that the manufacturer or service provider offers. We refer to the difference between that value in the eye of the customer and the market price of a product/service as **consumer surplus.**

Figure 2. Consumer Surplus.



**Process Competencies Cost**

Customers are defined in four dimensional space:how much they are willing to pay, the quality they expect to get, the time that it takes to get the product and the variety of options they do have. Firms define their Customer Value Proposition to meet and exceed those expectations.

In order to deliver the necessary customer value proposition, firms create **Process Competencies** in four-dimensional space of **cost, quality, time and flexibility.** Producers need to have process competencies in these four dimensions to reflect a Customer Value Proposition in the product that meets and exceed customer expectations. For example, companies try to produce accurately priced products to satisfy the customer expectation in the price dimension. The cost dimension is the total cost of producing and delivering the products or outputs. Producers look at the process to discover what parts of the production process are value adding and what parts are non-value adding. Non-value adding processes do not play a role in the transformation process of producing a product or delivering a service. Therefore, non-value adding processes are removed to lower production cost.

To keep quality high and costs low, producers also need to allocate appropriate resources to each activity. If resource cost is higher than activity cost, activity cost goes up. If resource cost is lower than activity cost, quality of process or product goes down. Therefore, producers need to find out exactly the appropriate resource cost is.

Producers also need to have high standardization and low variations in arrival time and processing time along with high utilization. Producers need to fully utilize human and capital resources in order to breakdown costs on a large number of products. This helps producers create products and processes low in cost in a timely and flexible fashion.

For example, Zara is a well-known name in the apparel industry. Zara's business is design/manufacture/distribution/retailing. Zara differentiates itself from competitors by timely fashion for the masses. CVP of Zara-timely yet limited variety at modest cost and quality. It looks for a market segment that is willing to buy timely fashion and is not particularly anxious about the variety. The price should be average and buyers will expect average quality.

**The production line** is a common method for creating high utilization in processes. The key concept in production cost is to allocate appropriate resources to each operation. An appropriate resource cost is one that is not lower or higher than what is needed. If the resource cost is too high, that increases the cost unnecessarily. If the resource used is lower than what is needed, it will lower the quality.

 Once the appropriate resource is selected, producers need to reduce variability to help lower production costs. To do this, producers need to increase utilization of all human and capital resources to close to 100%. Reducing variability helps increase utilization. With high variability, it is impossible to reach even close to 100% utilization. Standardization, reduced variability, high utilization and appropriate

A production line at Ford Motor Companies, Highland Park in 1913

resources allocation are key components of cost reduction.

Process Quality and Quality at Source are both aspects of production lines. **Process quality** is the ability to deliver and produce quality products. **Quality at source**, is when products are produced and checked at the same minute. If there is a problem, the production line is stopped.

An example of the production line is Shouldice hospital in Canada. Shouldice focuses solely on hernia operations. They have created a production line where the hernia operation is done at a very high quality and very low price. They do this by performing standardized, repeatable outpatient procedures. They also minimize variability by rejecting patients with risk factors such as high blood pressure.

The second of the process competencies is flexibility. **Flexibility** is the ability to produce and deliver a variety of products at both high and low volumes. Key components for flexibility are **cross-trained workers,** **short setup time** **delayed differentiation.**

In order to create flexibility inside a production system producers, you need cross-trained workers. Cross-trained workers can shift from one operation to another. In addition to that, producers also need **general-purpose equipment.** General-purpose equipment is equipment that can produce many different types of products. Theoretically, all machines are general purpose, but in order to transfer them from producing one product to another producers may need to spend infinite financial resources. A flexible machine has a short setup time.

**Delayed differentiation** is when producers postpone the differences that they make in the product to the latest steps. An excellent example of delayed differentiation is Home Depot's paint station. Home Depot offers hundreds and hundreds of different colors. However, if Home Depot wanted to have all those colors on their shelves , all of Home Depot would need to be a painting department. Instead, they have a few base colors and mix them to create the needed colors. Home Depot has delayed differentiation to the last possible step.

For increased flexibility producers also need a small batch size. Small batch size is when you produce a small number of products each time. Producers do not generally produce a product for six months of demand. After six months, customers might change their preferences and might not want that product anymore. In addition to that, new technology may come, and if producers have already produced six months worth of products, they will need at least six months to implement the new technology. Therefore, flexible systems are more responsive both to changes in customer preferences and also to changes in technology.

**Flow Time.** The fourth dimension of process competencies is process flow time, which is the total time to transform a flow unit from input into output and delivery of the finished product or any services to the customer. Two main components of a short flow time are effective layout and smooth material. Other requirements of smooth flow time are including: less variability in arrival rate, processing rate, and quality.

In smooth flow time the activities must not stop because of starving. In starvation, one station is waiting for the output of the previous station, and therefore the station remains idle. Also, there is no blocking, which occurs when the activities have to stop due to the lack of space. Thus, in a smooth flow time is neither starvation nor blockage. In other words, smooth flow means no defect and no re-work.

Operation Management creates smooth flow. One aspect of the smooth flow is low production cost because the flow units should come into the process and leave quickly. Other characteristic of smooth flow is high quality, since as soon as a problem in quality appears the production line must stop the production, and a stop in production line doesn’t have smooth flow. High quality product is one requirement of smooth flow. Another feature of smooth flow is a flexible system as there is not too much inventory that can easily respond to technological advances and changes in customer preferences and switch from one product to another. All of these characteristics apply to production systems, service systems such as distribution systems, healthcare systems, and entertainment systems and so on. There are some examples for process competencies including: Corolla that has flow shop, decentralized assembly plants close to market, shop flow time and low cost. Ferrari has job shop that is only a single plant in Italy, long flow time, and high cost. To recognize which of above companies is better, we require sufficient information with regard to the fundamentals of these companies. However, it depends on the strategy and the market segment that they have focused on. If they are synchronized with those elements, they would be successful, otherwise they are not.

Another example is McMaster-Carr that is a material, repair, and operations and what they usually call it MRO. It is a product distributor, a process with high flexibility, high quality and short response time and at the high price. Wal-Mart is another instance for process competencies. Operational Strategy of Wal-Mart is short flow time and low inventory, while its Operations Structure is cross docking. Cross docking means when two trucks that one has red products and another one that carries blue products go into a warehouse with a simply conveyor system and carts. Then there will be two other trucks that both carry red and blue products to the Wal-Mart stores.

In Figure 3., trucks carry products from warehouse to the store. 

In summary, the process is done in the name of cross docking starts with two trucks— one with the red product and another with the blue product— reach suppliers in a place with a minimal storage using material handling systems. Then, in supply place, these products are put into two trucks, which now carry both the blue and red products, and they go to the corresponding Wal-Mart stores.

|  |  |
| --- | --- |
| Figure 4. Blue and red products in warehouse. | Figure 5. Products are available to customers. |

Cross docking is one stage of operations structure. Operations structure also have electronic data exchange, fast transportation system, focus locations which has enough market, and communication between the stores such that if inventorial product in one store is high and in another store is low, they can transfer products between these two stores.

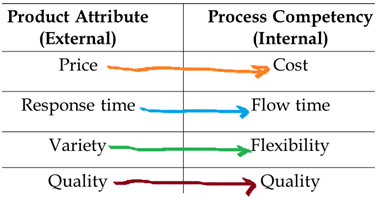
Compare inventory turns in Wal-Mart and in Target, which are retail stores. The times that inventory turns throughout the year in Wal-Mart is almost one and half times of that of Target’s. Sales per square foot in Wal-Mart is more than $400 per square feet, and in Target it is less than $300 per square feet.

**Operation Management and process competencies**. Operations management makes a smooth flow. It operates in hospital, university, bank, production system, assembly line, and in a distribution system as well. Consequently, operating management structures the process competencies in the direction of the customer value proposition. It develops measures to evaluate the effectiveness and efficiency of the processes. Thus, operations management develops process competencies to meet with customer value proposition. It develops measures to evaluate the effectiveness of these processes and efficiency of these processes. Operations management applies methods and techniques to improve process performance.

Process competencies are controllable whereas product attribute that are defined by customers are not controllable. Among the systems those ones are controllable, (such as process competencies, the environment and customer preferences) customer preferences that define the product attribute, are required to have a preparation of a customer value proposition, which meets and exceeds product attribute. Then, process competencies develop to be able to deliver customer value propositions and process competencies, which are controllable. There is no control on product attribute. There are three performance measures which help us understand if the process competencies are the best fit for the product attribute: Financial performance measures, External performance, and internal performance measure.

**Competitive Space and Strategy**. Second part of process view and strategy is related to the competitive space and strategy. Customers define the product attributes that they want in four-dimensional space of price, quality, variety and response time, whereas firms need to define process competencies in four-dimensional space of cost, quality, flexibility, and flow time. Therefore, to match these requirements, product attributes and process competencies are both defined in the four dimensional space.

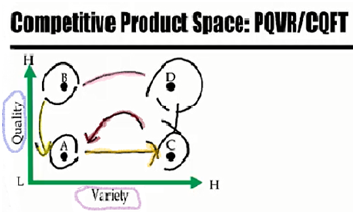
Figure 6. *Match of four dimensional Space for two Product Attribute and Process Competency*



It is not possible to visualize a four dimensional space. This means there is no way to represent them graphically, though they might be corresponded to mathematically by using matrix notation or vector. Also with regard to the three-dimensional space, even though the graphical representation of three-dimensional space is possible, the conception of such space is not as easy as that of a two-dimensional space. Therefore, the only way to demonstrate four-dimensional is using representation of two-dimensions.

The following is an example of the graphical illustration related to the two-dimensional space of variety and quality when two dimensions for the time are constant: For example, Figure 7 represents two dimensional space of variety and quality. It shows company A has low variety and low quality whereas Company C has high variety and low quality. Thus, company C has a better situation compared to that of company A in terms of variety dimension. Company B has the same variety as Company A, but quality of the product or quality of the process of company B is better than that of A’s. Hence, B dominates A and D dominates all of them.

Figure 7*. Products of companies in terms of two dimensions of variety and quality*



Consequently, in this 2 dimensional space, when move is from A to point B or from direction of A to C, a higher variety and higher quality will be created. As Figure 8 demonstrates, enhanced products or process require moving outside of the origin.

Figure 8. *Superior products require moving outside of origin*

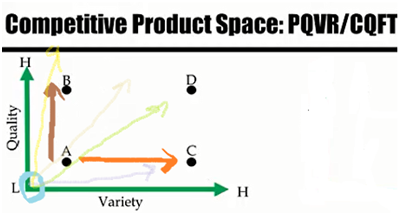
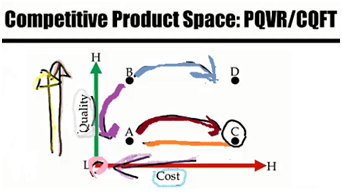


Figure 9 represents quality, the same as before, but in the horizontal direction with the variable cost instead of variety. The graph shows as long as quality is concerned, the company A and C are the same. Nevertheless, the difference is a higher cost compared to A and A dominates C. Both products of B and A have the same cost, but quality of B is much higher than that of A, and B dominates A. Also, products of D and B have the same quality, but cost of B is much lower than that of D, and B dominates D.

Figure 9 displays a quality-dimension, as moving is in direction to outwards, it will bring a higher quality. However, in the horizontal direction, direction towards origin (from right to the left) the situation would be better because costs will be dropped. In order to make these graphs consistent, cost would be replaced with the cost efficiency, which can be derived from the formula one divided by cost.

Figure 9. *The formula for efficiency*



**Efficiency = 1/ cost**

As a result, if product C or process C has a high cost, it will have a low cost efficiency since 1/cost becomes small. In contrast, a low cost in formula, creates a high cost efficiency, which shows a good point for products or process.

Figure 10. *The formula for Cost efficiency.*

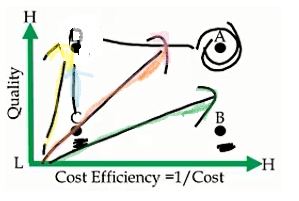
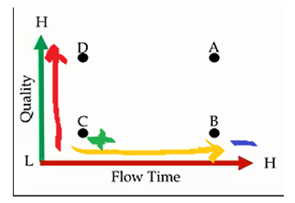


Figure 10 evidently demonstrates company B is more cost efficient compared to that of company C ; Product B has a higher cost efficient compared to product C; Process B is more cost efficient compared to that of C. C and D both have the same cost efficiencies, but D has higher quality. A and D both have the same quality, but A is more cost efficient and its cost is lower than that of D.

Basically, direction of outward moving makes a superior situation for the companies and products. This rule applies to the flow time variable. Figure 11 shows two variables of quality and flow time.

Figure 11. *Two quality and flow time dimensions*



For quality, direction toward up creates a high quality, and moving in horizontal direction from the left to the right side makes it worse. It means process requires more time or customer will get the product in the longer period of time, whereas moving from right to the left shows process takes less time. For this reason, on this dimension, flow time is replaced by responsiveness, which is derived from 1/Flow Time.

Figure 12. *The formula of Responsiveness*

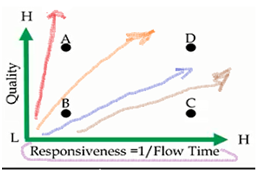
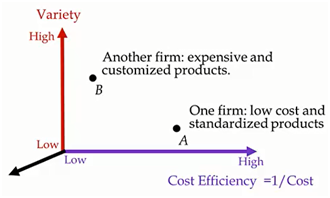


Figure 12 displays company C or product C or process C has a higher responsiveness compared to company B or process B or product B. Product D or process D has high responsiveness, which means production or process takes a short period of time with high quality. In this figure, moving outward makes an improved situation as well as the previous moving exhibited in prior figures.

In summary, one significant point in all figures is noticeable that when quality, variety, cost efficiency, and responsiveness move from the left to the right side or a direction to outward, an enhanced position will be made.

In Figure 13, product B has high variety, but it has high cost because its cost efficiency is low.

Figure 13. *Compare of two firms with regard to the variety and cost efficiency.* 

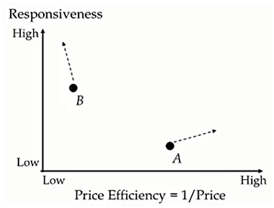
Product A has low variety, though at a high cost efficiency. One firm has a low cost and standardized product with a small variability; another firm has expensive product and customized product.

There is no possibility of determining which company would be more successful. It depends on the strategy and the market segment that these companies are looking at, and the customer value proposition they have prepared.

Company A might be Wal-Mart and Company B can be a jet manufacturer, which has very wealthy customers. Nonetheless, both companies may make high profits and both may go bankrupt in a couple of years.

**Strategy positioning** defines those positions that the firm wants to occupy in the competitive product space, such as the current position and the direction. There are two firms in a two-dimensional space of responsiveness and price or cost efficiency.

Figure 14. *Price efficiency formula and responsiveness*



Firm B has higher responsiveness compared to that of firm A, but its cost efficiency or price efficiency is much lower. The direction of the firm A is going to move toward the right side that causes higher price efficiency and increasing responsiveness at the same time. Now, Company A has low responsiveness. It means it takes more time to produce and deliver product or service to the customers, even though the price efficiency or cost efficiency is quite high. As direction of moving in Figure 14 shows, the strategy of this firm is to make a lower price and increase its responsiveness.

**Strategy should look like a sculpture.** A firm must ensure that its competitors are not capable to emulate its position. The strategy of a firm should be designed as a unique sculpture not as a block, which could easily be copied or imitated. It is difficult for competitors to imitate an array of interlocked activities, interlocks processes.

For instance, when Southwest Airline became successful many companies attempted to replicate it, but Southwest had created a resolute strategy similar to a single sculpture.

**Different Companies have Different Strategies:**

|  |  |
| --- | --- |
| Zara | Its strategy is timely, yet limited variety at modest cost and quality. |
| Aravind and Souldice | The strategy is low-cost, high quality, minimal variety, and average to long response time. |
| Corolla | Flow shop, decentralized assembly plants close to market, short flow time, low cost. |
| Ferrari | Job shop, a single plant, longer flow time, high cost. |
| McMaster-Carr | High flexibility, high quality, quick response time, and high price. |
| Walmart | Short flow time, low inventory, low cost, and average quality. |

**Some internal Measures and their relationship with process competencies.**

**Cost**

* Resource-Activity match
* High Utilization (Low Safety Capacity)
* Division of Labor (Job-Simplification)
* High Standardization and Modularization
* Effective Facility Layout
* Clear Material Flow Pattern
* Flow-Shop
* Value Analysis
* Training
* Method Improvement
* Technology

**Flexibility**

* Cross-trained Workers
* Short Set-up Time
* Delayed Differentiation (postponement)
* Small Batch Size
* Job-Shop
* U-Shaped Layout
* Internal Uniformity vs. External Variability

**Flow Time**

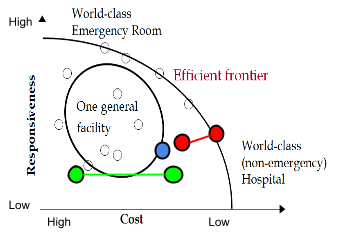
* Inventory Turnover
* Reliability in flow time
* No Starvation or Blockage
* Centralization
* Commonality
* Pooling
* Centralization
* Variance Reduction

**Quality**

* Conformance of Design and Manufacturing
* Quality at Source
* Reliability (quality over time)
* Service Level
* No Defect and Re-work
* Training
* Method Improvement
* Management

**Efficient Frontier**

Figure 15. Efficient Frontier.



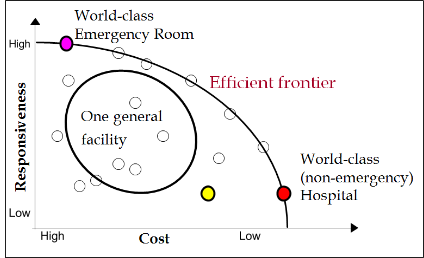
The small circles scattered around the graph represent different products or different processes or different firms. For example, these two processes, highlighted in green, have almost the same responsiveness, but the left process is an expensive process, and the right process is an inexpensive process. Since the responsiveness’ are the same, the one on the right is a better company.

These two companies, highlighted in red, demonstrate differences in responsiveness. The company on the right, along the efficient frontier, has higher responsiveness compared with the other and yet has lower cost. This process/product/company dominates this other process/product/company.

Efficient frontier is the minimal curve covering all the current positions in the industry. So if we want to find the minimal curve, it is on the efficient frontier. The processes/products/companies on the curve are world-class organizations that are trying to push the efficient frontier outward. The organizations inside the curve are not world-class organizations. However by improving themselves in both dimensions, these organizations, such as the one circled in blue, can push themselves onto the frontier and become world class, without or with little trade-off. If world-class organizations along the frontier want to become more responsive, then there is a trade-off, and they must increase their costs. Non-world class organizations inside the curve may improve their standing by improving responsiveness and costs without trade-off.

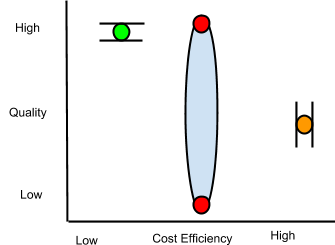
**Focused Strategy**

A focused process or a focused organization occupies a small portion of the four-dimensional space of competitiveness. For example, in a two-dimensional representation, the focused process highlighted in yellow below, has small cost variations, and small responsiveness variations. Thus, it can produce products at a given cost or lower cost products at another cost, fast operations in one area, and smaller operations in another.

Figure 16. Focused Strategy. 

The yellow highlighted organization is a focused organization, yet not a world class focused organization because it is not on the efficient frontier. Every organization is focused because everyone occupies some portion of the four-dimensional space (or in this example two dimensional space), yet some are world class while others are not. World-class organizations will fall on the efficient frontier. For example, as we move right along the graph we are demonstrating low cost and low responsiveness. A non-emergency hospital (highlighted in red) would be an example of a world-class organization falling on the efficient frontier, near the bottom right of the curve, due to its low responsiveness. You can also have the opposite. An emergency room or emergency hospital (highlighted in pink) is an example of a world-class organization that would fall on the upper left side of the efficient frontier, due to its high cost and high responsiveness. Both of these examples are focused, because as long as responsiveness is concerned they fall on the upper left and bottom right of the efficient frontier, and as far as cost is concerned they are also placed respectfully along the curve. They do have operations, which do have small variations in cost and small variations in responsiveness. A general, unfocused organization will fall within the center of the graph. This is because it has some operations, which are very inexpensive and require a long waiting time, and some operations which are very expensive that require fast responsiveness.

Figure 17. Cost Efficiency.



Within a focused strategy, if the graph above reflects cost efficiency and quality, the cost of operations will be in the mid level, far right range as pictured. Quality will also be in a similar small range. A quality focused organization will fall within a small range in the upper left portion of the curve with high cost. An organization within the center area of the graph will produce both high and low quality products. Companies that produce both high and low quality products at a similar cost cannot compete. A focused organization makes or has: all high, or average quality products; high cost, average cost, or low cost products; high, average, or low responsiveness products; high variety, low variety, or average variety in their products. It is impossible, however, to have a company that produces 1000 items ranging from high to low cost, or high to low quality all under the same management and operations.

A focused strategy is committed to a limited, congruent set of objectives, in terms of demand (product, market) and supply (inputs, technologies, and volumes). When we look at demand, this does not mean we produce 1000 types of products, for 100 different markets. We are committed to a limited number of products for a limited market(s). In terms of supply, we don’t use all types of input (low or high) or all types of technologies (manual, or automated).

A focused process is not limited to a few products, but all the products should fall within a small region of the four dimensional product space. If they don’t all fall within that space, then we require Plant-Within-Plant (PWP). This business strategy is diverse, but generally the entire business is divided into several mini plants, each with focused processes. One PWP may focus on low cost, while the other may focus on quick response. High and low volume products should be separated into different plants. High quality, high cost products should be produced in a separate plant than an average quality, low cost product. Different plants should also be under different management.

In a two dimensional space, with functions of cost efficiency horizontally and responsiveness vertically, unless world class organizations can push the boundary of the efficient frontier there is no way to increase or decrease cost without a trade off to responsiveness. Everything else within the curve is not world class unless it is along the frontier, yet it can move without a trade off by pushing simultaneously in multiple directions, on more than one dimension. One way to push the boundary of the efficient frontier is with new technologies.

Firms located on the same ray share strategic priorities. They all have the same cost efficiency, responsiveness, or tradeoffs. A trade off is simply the inability to increase one dimension or attribute, without decreasing, or without consequence to another. Firms on the frontier must trade off. Strategic positioning is the direction of the improvement from the previous position, or where the company wants to occupy along the efficient frontier. By not being able to move without tradeoffs, world-class companies try and push the boundaries of the efficient frontier. As technology and management technologies advance, they help to push the frontier outward, yet this is not the same across all industries.

Different companies intentionally choose different processes to achieve the same goal. For instance, McDonald’s vs. In-N-Out. have different processes that lead to different advantages and disadvantages, so we are always facing tradeoffs. Delivering books at a low cost can be easy. Delivering books fast can be easy. Delivering books fast and at a low cost, however, is not easy. You also cannot work and study for exams at the same time. The more you work, the less time you will have to study, and therefore the worse you will do on exams. The more you study, the better you will do on your exams, yet you will have less money. There is, therefore, a trade off between doing work and studying. We are always facing tradeoffs.

Operations management is a set of tools, techniques, and philosophies to create smooth flow. Operations management is also the knowledge necessary to understand tradeoffs, and come out with optimal tradeoffs. To create smooth flow, we are forced to have high quality products, with little inventory, because products are made and quickly leave. In such a system, if there is a change in customer preferences, or a change in technology, or a change in inputs, and requires variation, the system can immediately respond. Smooth flow means flexibility, short flow time, and high responsiveness. As soon as someone desires our product, we can quickly get it to them. Smooth flow also means low cost, because products have less time to absorb overhead costs. By creating smooth flow, we can determine the optimal tradeoffs. Operations management allows us to produce efficiently and determine the optimal levels of trade off.

**Operational Effectiveness**

Operational effectiveness is developing an operations strategy (encompassing the resources, processes values, and competencies within the four dimensional space of cost, quality, flexibility, and time) that supports the strategic positioning (customer value proposition), better than competitors.

In management, the general definition of effectiveness is doing the right things. If the thing you are doing is right, then you are effective. Efficiency is doing things right. You can be efficient, but not effective. You may be doing something wrong very well or quickly, but this does not make you effective. To be both effective and efficient, you should do the right things while doing things right.

In operations management, however, we define efficiency as cost efficiency. A process is efficient if we can produce output with minimal inputs and resources: low cost operations. An effective process is a process that supports the execution of a company’s strategy in the four dimensions of cost, quality, flexibility, and time. A synchronized process does well in all four dimensions, while supporting the customer value proposition. We are efficient if we do well in the cost dimension. We are effective if we are doing well in all four dimensions.

**Throughput**

**INTRODUCTION**

In any process, throughput can be defined as the number of elements passing through a given system in each unit of measurable time. That unit of time can vary in different cases, and can be anything from seconds to years. Throughput can be easily spotted inside any type process and with any type of unit. The process can be either a service or a manufacturing process such as: doing someone’s makeup, manufacturing a product, or producing a candy in a factory. In this picture, workers are putting the lids on the cars that are being assembled. Throughput in this particular case can be depicted as the work in process, or in this case, the number of cars being assembled that enter this process per second, minute, day, or year. Previously, throughput was touched upon in the lesson on Little’s Law. Throughput is an extremely important element of little’s law.

**Little’s Law:**

Inventory= Throughput ( R ) x Time

**Throughput= Inventory/Unit of time**

Inventory=(Inventory/Unit of time ) x Time

**Symbolic Representation of the Equation:**

I=R x T

It is imperative to realize that without throughput, there would be no way to calculate the amount of time it takes for a unit to go through a whole system or specific subsystems. In the same way, without throughput there would be no way to calculate inventory within a process, regardless if it’s in the manufacturing sector or in the service sector.

**HOW THROUGHPUT IS CALCULATED**

One may ask how is something like throughput calculated? There may be periods where there may be a lot of units coming in, such as Monday mornings in a Coffee Shop, or in the contrary dead hours, around 11 pm in a coffee shop, this concept is also known as variability. Throughput is calculated in any process in three steps:

1. Witnessing the process multiple times
2. Calculating the number of elements that go into any system or subsystem per unit of time
3. Calculating the average of those calculated rates

For example, let’s suppose Maria wants to know the number of customers that on average go in and out of her coffee shop. This particular experiment will be extremely helpful in deciding her staff during certain hours of the day. She collected the following data.

|  |  |
| --- | --- |
| 9 am | 5 customers/min (on average) |
| 10 am | 2 customers/min (on average) |
| 11 am | 2 customers/min (on average) |
| 12 pm | 4 customer/min (on average) |
| 1pm | 2 customer/min (on average) |
| 2 pm | 1 customer/min (on average) |

Based on these calculations, she can calculate the average number of customers per minute that come into her store, which is 2.67 people/min.

One can also describe throughput in the context of a flow system. In such cases, throughput can also be considered as the average rate in which an object flows.

Likewise, we have also come across the concept of capacity. Capacity is defined as the maximum rate at which a system can handle production. Average throughput can be anywhere below capacity value. However, the maximum flow at the period of the most congestion is maximum capacity. Throughput values for that specific process cannot exceed that capacity.

**CYCLE TIME AND TAKT TIME**

Cycle time measures how long an entire process, such as baking baklava, can take place at the maximum capacity of unit inflows. Cycle time relates to capacity by focusing on the internal capabilities of the process. However, throughput, the number of inflow of units per unit of time, focuses on the external demand for the specific process. Takt time is related to throughput. Therefore, Takt time, unlike cycle time, focuses on external demand, and measures how long an entire process takes at the average throughput or demand. The formulas are as follows:

Cycle Time= 1/Capacity

Takt time= 1/Throughput (demand)

Suppose Maria has a capacity of 50 people per hour. Her cycle time for the capacity would be:

Cycle Time= 1/50 hours x60 min/hour = 1.2 min

Suppose Maria has an average throughput of 28 customers per hour coming in to her coffee shop. Her Takt time would be:

1/28 hours x 60 min/hour= 2.14 min

It is extremely important to know that capacity should always be greater than throughput. This fact infers that the Cycle time will always be less than the takt time. Remember, throughput is the average number of inflow and outflow units per unit of time. When applying these concepts to make effective decisions, one needs to know the purpose of such equations. As managers, if we want a smooth flow, we want to make as much units as possible in small amount of time. Therefore, our goals would be to maximize capacity and throughput. In hand, our goals will also be to lessen takt time and cycle time.

**TERMS AND SYMBOLS RELATING TO THROUGHPUT AND CAPACITY**

In the future examples, we will be using the terms Ra, which refers to throughput, and Rp, which refer to capacity.

In problems we will be given resources, people or objects performing each activity or sub process of the entire process. We will see a value known as Tp, which is the unit load of a resource unit. That is, how long a resource needs to work for each flow unit produced. Using that one can, in turn, discover capacity!

**THE PROCESS OF EFFECTIVE CAPACITY, THROUGHPUT, AND UTILIZATION**

We will connect the concept of throughput, the equivalent of demand, after in depths look into the process of calculating effective capacity for individual resource units and, in turn, resource pools hourly and daily. We will then compare our demand, or throughput, with the maximum capacity our individual activities are able to withhold. That would determine our utilization. Based on our utilization, we can make effective decisions.

**CALCULATING EFFECTIVE CAPACITY FOR RESOURCE UNITS AND THEIR CORESPONDING POOL**

There are two types of capacity: Theoretical and Simple. Simple is also known as Effective Capacity. What makes them different from each other is that theoretical capacity does not take into account Capacity Waste factor, where as Simple, or effective capacity does consider Capacity Waste factor.

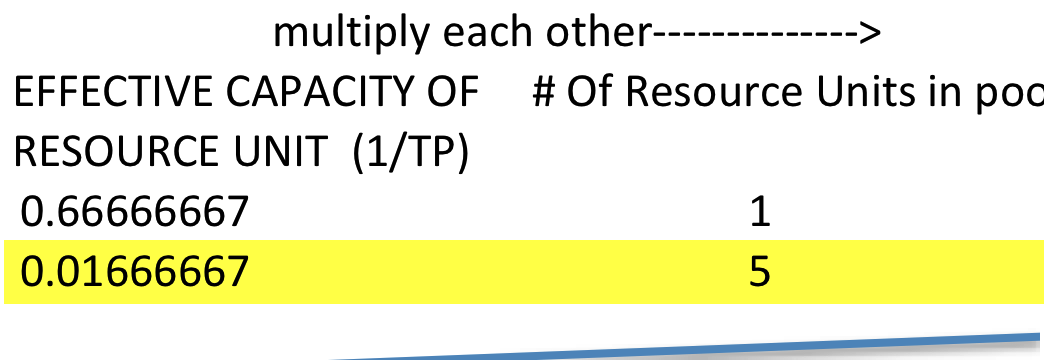
Continuing from the previous lesson of capacity, we can expand it to find out the effective capacity of each resource unit and resource pool. We, then, use the pool capacity (claims/min) to determine how many units are capable of going through a system each day or each month. This gives us great insight into the way the business is run, where it’s going, and where it could be.

**EXAMPLE**

You’ll go through this example to measure the effective capacity of each individual resource unit, such as workers, and the resource pool, which totals how many worker units the company has per resource position. We can then use these figures and the amount of hours that the workers work to determine utilization based on a throughput value.

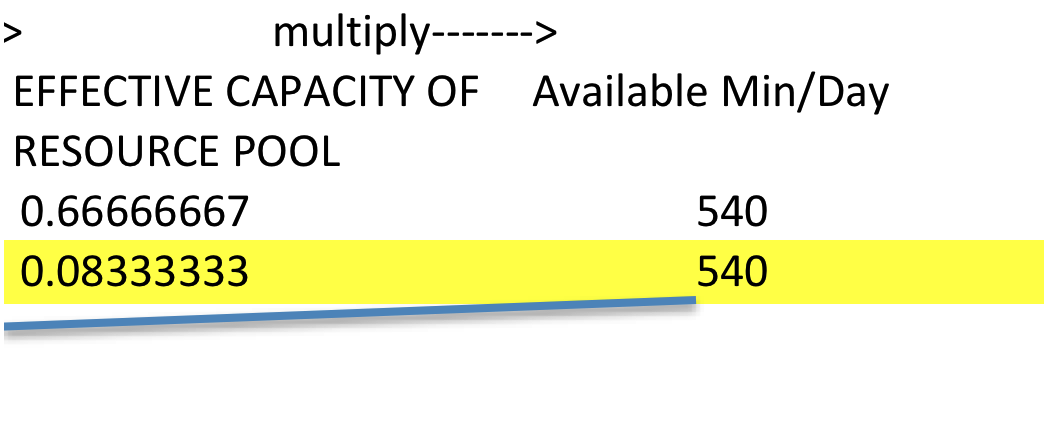
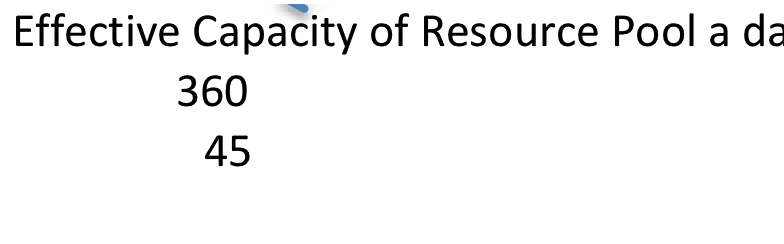
In this example, we will review the overall process of CSUN’s Counseling Service. We will observe the number of people that enter the office and utilize the resources daily. We will also see the demand in comparison to the capacity each resource (worker) is capable of doing. By doing this, we can calculate utilization. Through this number, we will see if the counseling service office needs to do more advertising in order to reach out to more students.

The counseling office consists of two kinds of workers: A receptionist and a Counselor. There is one receptionist checking in patients at the front desk. There are five counselors giving therapy to students who walk in the door. The receptionist spends a maximum of 1.5 minutes checking in each student. The counselor spends a maximum of 60 min with each student in therapy. Using this information, we want to find out the effective capacity of each resource unit. In order to do this, we need to inverse the fraction. 1.5 minutes becomes 1/1.5 minutes. And 60 minutes becomes 1/60. We are turning min/claim into claims/minute. This is to give us an easier time calculating effective capacity. We then multiply the claim per minute by the number of resources, or workers, on the scene. We only have one receptionist. However, we have 5 counselors.



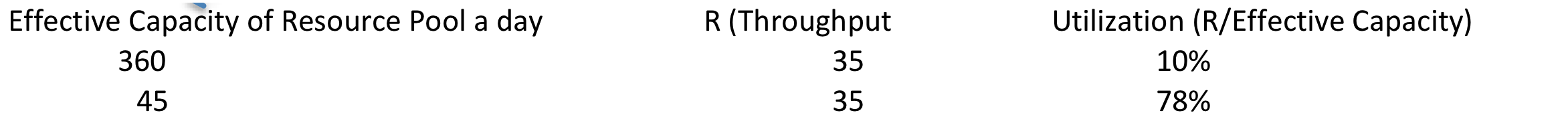
We do this because we need to find the capacity of the entire resource pool, so we can see what area of the process is the bottleneck of the entire process. The bottleneck of the process will be the resource with the minimum capacity. The number of people working on the job matters in how the job gets done. The first line contains the values of the receptionist; the second line contains the values of the counselors.

Now that we have the effective capacity of the resource pool, in other words, claims that they can handle per minute, we can calculate how many claims, or students, they can serve in a day. We do this by multiplying these values by the number of minutes they work in their shift. Both the receptionist and Counselors work 540 minutes in their daily shift. We multiply the effective capacity of the resource pool for each position by 540 to come to the number of people they are capable of serving a day



The receptionist can serve a maximum of 360 people during his shift, while the counselors can only serve a maximum of 45 people per shift. Throughput for the counseling services has been calculated. It is 35 students in one day.

Using the current daily demand, we can compare it to the maximum effective capacity that each resource pool serves per day. We do so by dividing throughput by the capacity of each resource pool. We turn this into a percentage which gives us the utilization value.



What does this tell us? Given the current demand, the receptionist only utilizes 10% of his capacity. Also, this table shows that given the current demand, the counselors utilize 78% of their capacity. Using these numbers, those who run the counseling services may decide to broadcast their services in order to reach out to the student population. As we all know, depression and suicide can be a major problem among young adults. Perhaps they may see a greater need to reach out to the student population more than ever before. They will have the quantitative capabilities to do so in addition to their qualitative motivations. Given the information from this specific case, we can also calculate values that we touched upon earlier.

**OTHER CALUCLUATIONS**

Rp (Capacity): 45

Cycle Time: 1/45\*5\*540=60

R (Throughput): 35

Takt Time: 1/35 \*5\*540=77

Flow Time: 60 + 1.5= 61.5

**Theoretical Capacity & Capacity waste Factor**

As we stated before, the effective capacity of a resource unit is defined as 1/Tp. Tp in this case is unit load, which is the time of production of a certain resource including the wasted time. In order to calculate Theoretical capacity, it is important to measure the Theoretical unit load, which is the time of production of a resource unit with no wasted time included.

Effective Capacity= c/Tp (Unit load)

Theoretical Unit load= Tp \* (1-CWF) (Capacity Waste factor)

Theoretical Capacity= c/Theoretical Unit load

In these types of cases, capacity waste factor represents a percentage amount of time that was wasted due to set- ups or perhaps break downs. In these cases, the capacity waste factors will be given in each problem.

For example, in car body shop, on average it takes a person 25 minutes to fully paint a car. And it has been determined that there is a Capacity Waste factor of 20% because it takes the mechanic 5 minutes to set up all of the tools needed. In this case these are the calculations.

Unit load= 25min

Effective capacity = c/Tp = 1/25\*60min per minute or 2.4 per hour

Theoretical Unit load= Tp \* (1-CWF) = 25\* (1- .20)= 20

Theoretical capacity= c/Theoretical Tp= 1/20 or 3 per hour

**CONCLUSION**

Mastering this process is crucial to understand how to compare process capacity to demand in order to come up with a utilization percentage. With utilization numbers, you have the numbers to back up making effective decisions in order to improve sales and customer turnabout. You can see how far the business would be comfortable in increasing customer numbers without reaching the capacity and stressing workers. Or, perhaps, you can see how many customers a business needs to hire in order to find a new bottleneck and improve capacity or lessen time. At the same time, as managers, it is imperative to eliminate as many non-value added activities that account for Capacity Waste factor. By eliminating these types of activities, we will see that our Effective capacity will increase, and our goals must be to maintain that effective capacity as close as possible to the theoretical capacity which would mean that there are not as many non value added activities and which infers that our process is running smoothly.

## *Problem 1*

On average it takes a barber 35 minutes to cut someone’s hair. During those 35 minutes, seven of those minutes is wasted by the barber in order to prepare his equipment and adjust any modifications to the machines he uses. If he work eight hours per day, what is his Simple and theoretical capacity?

CWF= 7/35= 0.2 or 20%

Unit load (Tp): 35 min

Effective Capacity= 1/Tp= 1/35\*60= 1.7/hr

Or 1.7/hr\* 8rhs/day= 13.6/day

Theoretical Unit Load (Ttp)= Tp-(1-.20)= 35min \*.80=28 min

Theretical Capacity: 1/TTP= 1/28\*60= 2.14/hr

Or 2.14/hr\*8hrs/day= 17.14

This means that if the barber did not waste any time in each haircut, he would be able to cut 3 people’s hair in addition to what he already had. Now with this data, he can work something out to decrease as much as possible his non value added activities and thus increase his capacity.

## *Problem 2*

ESP Guitar Company manufactures custom guitars from well known guitars players, Kirk Hammett, and James Hetfiled from legendary band Metallica. Both guitars required to spend some time in the design department, the painting department, and the detailing department. The ESP Explorer guitar requires four hours in the design department, two hours in the painting department, and three hours in the detailing department. On the other hand, the ESP Flying V guitar requires 3 hours in the design department, 2 hours in the panting department, and 2 hours in the detailing department. Each department works for 8 hours each day, and the Capacity waste factor are as followed: Design= 25%, painting= 0%, Detailing=50%. There are two people in the design department, three machines on the painting department, and five guitar engineers taking care of the detail department.

What is the effective capacity and theoretical capacity for each resource, and their corresponding pool?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unit load: Explorer | Unit load: Flying V | Unit Load: 60%Explorer 40% Flying V | No. of resources |
| design | 4 hours | 3 hours | 3.6 hours | 2 |
| Painting | 2 hours | 2 hours | 2 hours | 3 |
| Details | 3 hours | 2 hours | 2.6 hours | 5 |

Design:

Unit load (Tp)= 3.6 hours

Effective capacity= 1/Tp= 1/3.6hrs

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 3.6 hours\*(1-.25)= 2.7 hours

Theoretical capacity= 1/TTp= 1/2.7 hours

Painting:

Unit load (Tp)= 2 hours

Effective capacity= 1/Tp= 1/2hrs

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 2 hours\*(1-0)= 2hours

Theoretical capacity= 1/TTp= 1/2 hours

Details:

Unit load (Tp)= 2.6 hours

Effective capacity= 1/Tp= 1/2.6 hours

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 2.6 hours\*(1-.50)= 1.3hours

Theoretical capacity= 1/TTp=1/1.3 hours

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effective capacity per resource | Theoretical capacity per resource | Number of resources | Effective Capacity of pool | Theoretical capacity of pool | Effective Cap. of R-pool per day (x8hours) | Theoretical Cap. of R-pool per day  (x8hours) |
| Design | 1/3.6hr | 1/2.7hr | 2 | 0.56/hr | 0.74/hr | 4.44 | 5.93 |
| Painting | 1/2hr | 1/2hr | 3 | 1.5/hr | 1.5/hr | 12 | 12 |
| details | 1/2.6hr | 1/1.3hr | 5 | 1.92/hr | 3.8/hr | 15.4 | 30.8 |

In this case the bottleneck of this system is the design department because it is the one with the least capacity out of the entire pool

Cycle time= 8 hours /4.44= 1.8 hours

## *Problem 3*

Health blend produces many types of multivitamins which manufactures around the whole United States. All of their multivitamins have to go through a mixing department, and encapsulating department, and a bottling department. However, every single department can only handle a certain demand for each multivitamin. There is a demand 200,000 units for the month of April; half of the well known “Gold” multivitamin, and the other half for the “Silver” multivitamin. Health blend manufactures the multivitamins by orders of 1,000 units. In order to complete a Gold multivitamin order, it need to spend three hours in the mixing department, 4 hours in the encapsulating department, and 2 hours in the bottling department. On the other hand, the Silver multivitamin, on average, it spends four hours in the mixing department, five hours in the encapsulating department, and three hours in the bottling department. The capacity waste factors are as followed: mixing= 15%, encapsulating= 25%, and bottling=10%, and the product mix

What is the effective capacity and theoretical capacity for each resource, and their corresponding pool?

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | Unit load: Gold | Unit load: Silver | Unit load: 50%Gold 50%Silver | Number of Resources |
| Mixing | 3hrs/order | 4hrs/order | 3.5hrs/order | 3 machines |
| Encapsulating | 4hrs/order | 5hrs/order | 4.5hrs/order | 5 machines |
| Bottling | 2hrs/order | 3hrs/order | 2.5hrs/order | 2 machines |

Mixing:

Unit load (Tp)= 3.5hrs/order

Effective capacity= 1/Tp= 1/3.5hr

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 3.5hours\*(1-.15)= 2.98 hr

Theoretical capacity= 1/TTp= 1/2.98 hr

Encapsulating:

Unit load (Tp)= 4.5hrs/order

Effective capacity= 1/Tp= 1/4.5hr

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 4.5hours\*(1-.25)= 3.37hrs

Theoretical capacity= 1/TTp= 1/3.37hr

Bottling:

Unit load (Tp)= 2.5hrs/order

Effective capacity= 1/Tp= 1/2.5hr

Theoretical Unit load (TTp)= Tp\*(1 – CWF)= 2.5hours\*(1-.10)= 2.25hrs

Theoretical capacity= 1/TTp= 1/2.25hr

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | Effective capacity per resource | Theoretical capacity per resource | Number of resources | Effective Capacity of pool resource | Theoretical capacity of pool | Effective Cap. of R-pool per day (x8hours) | Theoretical Cap. of R-pool per day  (x8hours) |
| Mixing | 1/3.5hr | 1/2.98 hr | 3 | 0.86/hr | 1.01/hr | 6.9/day | 8.1/day |
| Encapsulating | 1/4.5hr | 1/3.37hr | 5 | 1.11/hr | 1.48/hr | 8.9/day | 11.8/day |
| Bottling | 1/2.5hr | 1/2.25hr | 2 | 0.8/hr | 0.89/hr | 6.4/day | 7.1/day |

Cycle time= 8 hours/ 6.9= 1.16 hours



**A. The Little's Law:**

The Fundamental long-term relationship between Work-In-Process, Throughput and Flow Time of a production system in a steady state.

The law explains the basic idea that you start a process with input (inventory), the inventory goes into a system (WIP) and then goes out as output (throughput). It helps us determine the time of the process from when an input goes into the system until it goes out. This can be explained by using the formula:

**1. The Application to Waiting Lines:**

In earlier chapters, we were introduced to formulas like:

1. Flow time

* Explains that flow time () is equal to buffer time () + processing time ()

1. Inventory

* Explains that Inventory () is equal to the number of flow units in the buffer () + the total number of flow units in all processors ()
* Explains that the number of flow units in the buffer () is equal to throughput (x buffer time ()
* Explains that the total number of flow units in all processors () is equal to throughput x processing time ()

**Hence we can conclude that:**

We also have to take into consideration the formulas that we learn in Capacity such as:

* Explains that utilization is equal to throughput () divided capacity ()



* Explains that Capacity () is equal to the number of processors () divided by processing time ()
* Explains that throughput () is equal to the total number of flow units in all processors () divided by processing time ()

**Hence we conclude that:**

**B. Characteristics of Waiting Lines**

1. Variability in arrival time and service time leads to
2. Idleness of resources
3. Waiting time of flow units

As utilization goes up, the number of people in the buffer goes up. The other one is variability; if every four minutes, exactly every four minutes, one customer arrives, and it takes exactly three minutes for that flow unit to pass the processor, then we never observe the waiting line.

In this case, utilization is one hundred percent, that is the maximum possible utilization, but there is absolutely no variability, neither in inter-arrival time nor flow time, then we never see flow units in the buffer.

1. We are interested in two measures
2. Average waiting time of flow units in the waiting line and in the system (Waiting line + Processor).
3. Average number of flow units waiting in the waiting line (to be then processed).

**C. Operational Performance Measures**

As explained earlier, if we know , and we know , we can easily computer . Therefore, all we need to compute is And if we know one of them because we know throughput (), then we can computer the other one.

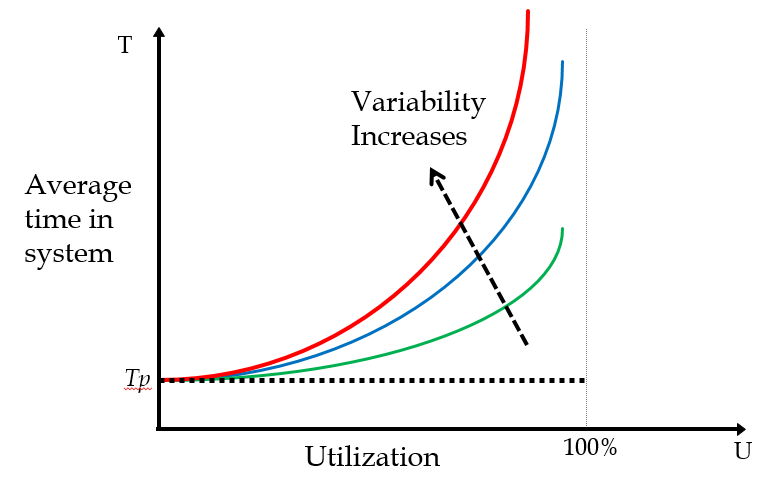
1. Flow time

1. Inventory

We have an approximation formula for Ii and we have exact formula for some specific cases of waiting line.

**D. Utilization – Variability - Delay Curve**

The graph below explains the relationship between utilization and variability. As utilization increases, the waiting time increases and as variability increases, the waiting time increases and vice versa.



* + - 1. **Utilization and Variability:**

1. Our two measures of effectiveness (average number of flow units waiting and their average waiting time) are driven by

**1. Utilization:** The higher the utilization the longer the waiting line/time.

**2. Variability:** The higher the variability, the longer the waiting line/time.

B. High utilization or low safety capacity, due to

**1. High inflow rate**

**2. Low** processing rate , which may be due to small-scale **c** and/or slow speed

**2. Drivers of Process Performance:**

Variability in the inter arrival time and processing time is measured using standard deviation (or Variance). Higher standard deviation (or Variance) means greater variability. The formula for variance is:

Standard deviation is computed by taking the square root of variance. The formula for standard deviation is:

Standard deviation is not enough to understand the extent of variability. Does a standard deviation of 20 *for an average of 80* represents more variability or a standard deviation of 150 *for an average of 1000?*

Coefficient of Variation: The ratios of the standard deviation of inter arrival time (or processing time) to the mean (average).

We refer to coefficient of variation of inter-arrival time as , and coefficient of variation of processing time as , capital . The small represents the number of servers coefficient of variation is shown by using capital . Capital small for coefficient of variation of inter-arrival time, and capital small for coefficient of variation of processing time.

**3. The Queue Length Approximation Formula:**

The approximation formula is used for the number of flow units in the buffer. Ii, the number of flow units in the buffer is equal to utilization to the power of two times one plus the number of servers divided by one minus the utilization, and we refer to it as utilization affect or u-part. Utilization affect is multiplied by variability affect. And that is squared coefficient of variability of inter-arrival time plus square of coefficient of variability of processing time divided by two. And we refer to it as variability affect or v-part.

We have to keep in mind that:

* and are the Coefficients of Variation
* Standard Deviation/Mean of the inter-arrival or processing times (assumed independent)

**4. Factors affecting Queue Length:**

1. Utilization effect; the queue length increases rapidly as approaches 1.
2. Variability effect; the queue length increases as the variability in inter arrival and processing times increases.
3. While the capacity is not fully utilized, if there is variability in arrival or in processing times, queues will build up and customers will have to wait.

**E. Coefficient of Variations for Alternative Distributions:**

* : average processing time:
* : average inter arrival time:
* : standard deviation of the processing time
* : standard deviation of the inter arrival time

If processing time does not have a specific distribution such as exponential, Poisson, or constant, and if its average is, and its standard deviation is , then its coefficient of variation, is divided by .

The same is applied for inter-arrival time. If inter-arrival time is a general distribution, and by general distribution we mean it is not Poisson, not exponential, and it is not constant, and it has an average of and standard deviation of , then for inter-arrival time, coefficient of variation is divided by

This is explained by the table below:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Inter arrival time or Processing Time distribution | General (G) | Poisson (M) | Exponential (M) | Constant (D) |
| Mean arrival time or Processing Time |  |  |  |  |
| Standard deviation ofInter arrival time or Processing Time |  |  |  | 0 |
| Coefficient of Variation of Inter arrival time or Processing Time |  | 1 | 1 | 0 |

Therefore, exponential distribution shows the time between two events and Poisson distribution shows the number of events over a specific time period.

**Example 1:**

A sample of 10 observations on inter arrival times in minutes are:

10, 10, 2, 10, 1, 3, 7, 9, 2, 6 minutes.

average of sample = 6 minutes

arrivals per minute

standard deviation, found through the use of excel: 3.94

= 3.94/6 = 0.66

**Example 1b:**

A sample of 10 observations on inter arrival times in minutes are:

7, 1, 7, 2, 8, 7, 4, 8, 5, 1 minutes.

average of sample = 5 minutes

arrivals per minute

standard deviation, found through the use of excel: 2.83

= 2.83 /5 = 0.57

Using examples 1 and 1b, we are able to figure out how many people are in the waiting line and what is the safety capacity with the use of the queue length approximation formula. We can also compute what is the waiting time in the line and system. (Example 1c)

**Example 1c:**

per minute or 10 per hour.

per minute or 12 per hour.

Since = (1/6)/(1/5) = 0.83

= 0.66 and = 0.57

By using the queue length approximation formula, we have the following:

1.56 passengers waiting in line.

To compute safety capacity, we use .

= (1/5) – (1/6) = 1/30 passengers per minute or (60)(1/30) = 2 passenger per hour.

We want to compute what is the waiting time for each person in the waiting line. The Little’s law states that *I = RT*, therefore . We can conclude that

1.56/(1/6) = 9.4 minutes

The approximate waiting time in the line is 9.4 minutes. We can also calculate the waiting time in the whole system.

Because we know that is 5 minute, the waiting time in the whole system is 14.4 minutes (5 + 9.4).

Using Little’s Law (*I = RT*) we are able to calculate the total number of people in the system.

2.4 people

We know that there are 1.56 people in the buffer. To calculate the total number of people in the processor, we add 1.56 to .83, the arrivals per minute to get 2.39.

Now let us suppose we have **two servers** instead of one, are we able to calculate the new waiting line and time? The following examples show how two servers change the waiting line and waiting time.

**Example 1c-a:**

per minute or 10 per hour.

per minute or 24 per hour.

Since = (1/6)/(2/5) = (5/12) = 0.417

= 0.66 and = 0.57

By using the queue length approximation formula, we have the following:

0.076 passengers waiting in line.

To compute safety capacity, we use .

= (2/5) – (1/6) = 7/30 passengers per minute or (60)(7/30) = 14 passenger per hour.

We want to compute what is the waiting time for each person in the waiting line. The Little’s law states that *I = RT*, therefore . We can conclude that

0.076/(1/6) = 0.46 minutes

The approximate waiting time in the line is 9.4 minutes. We can also calculate the waiting time in the whole system.

Because we know that is 5 minute, the waiting time in the whole system is 5.46 minutes (5 + 0.46).

Using Little’s Law (*I = RT*), we are able to calculate the total number of people in the system.

0.91 people

We know that there are 0.076 people in the buffer. To calculate the total number of people in the processor, we add 0.076 to 0.417, the arrivals per minute to get 0.91.

**Example 2:**

A call center has 11 operators. The arrival rate of calls is 200 calls per hour. Each of the operators can serve 20 customers per hour. Assume inter-arrival time and processing time follow Poisson and Exponential, respectively. What is the **average waiting time** (time before a customer’s call is answered?)

By using the queue length approximation formula, we have the following:

Therefore:

2.1 minutes

**Example 2a:**

The average waiting time before a call is answered is 2.1 minutes. Now let us suppose the service time is constant. Will the waiting time be more or less than 2.1 minutes?

Therefore:

1.03 minutes

**Example 3:**

Vons contains 7 checkout stands. The arrival rate of customers is 175 per hour. Each of the checkout stands can serve 15 customers per hour. What is the average waiting time (time before a customer gets to the cashier?)

By using the queue length approximation formula, we have the following:

minutes

**Example 4:**

A small room has been assigned on a University campus, for honor students to have a quiet place to study by themselves. Since the room is so small, the only other use for it would be storage. The room is open 16 hours every day. Only one student can use the room at a time. The dean wants to know if the room is being fully utilized by students, or if the room should be changed to a storage space. The best SOM professor at the University was assigned to solve this problem. In his research, he found that students arrive at the room at a rate of 3 per hour following Poisson distribution. The students stay in the room an average of 15 minutes and deviation of 5 minutes.

1. What percentage of time is the room idle (rounded)?
2. What is the utilization rate of the room?
3. What is the average number of students in the waiting line?



1. How much time, on average, does a student spend in the waiting line?

2.25

**Example 5:**

Yogurtland has one cashier. Every 5 minutes, a new customer comes in. It takes the cashier about 5 minutes to weigh the yogurt and ring up each customer. Yogurtland has been losing customers to the new frozen yogurt shop next door. In order to stop this, Yogurtland has implemented a new policy to pay customers waiting in line. The customer will get paid 1 dollar per minute of waiting time. A SOM professor has been recruited from a nearby university to analyze the cost of the new pay while you wait policy. Preliminary studies the SOM professor did, indicated that there are an average of .5 people waiting in line. Assume that arrivals follow Poisson and service time follows exponential distribution.

1. What is the capacity of the cashier per hour?

It takes a cashier 5 minutes per customer, therefore, 60 minutes / 5 minutes per customer = 12 customers.

1. What proportion of the time is the cashier busy?

6/12 = 50%

1. On average, how long does a customer wait in line?



.05 \* 60 minutes = 3 minutes

1. What is the hourly cost of the new policy?

The hourly cost of the new policy is 1 \* 60 = $60. Because there are .5 customers in the line at all ties, .5 \* 60 = $30.

1. What is the most Yogurtland would be willing to pay for another cashier?

Answer = $30.

Another cashier would increase Rp to 24 customers per hour. If Rp is 24 per hour, there would be no waiting line because there are only 12 customer per hour. Yogurtland would only be willing to pay the amount that would be saved by having a new cashier.

**FORECASTING**

This chapter focused on forecasting. Forecasting is the prediction of the future value of a variable of interest, such as demand. In addition, it is required in many business functions. For example, in accounting forecasting is used to provide cost and revenue estimates; in finance, forecasting is used to estimate cash flow, and provide sources and uses of funds for human resources needs, forecasting for hiring and training plans, marketing for pricing and promotion, and operations for production planning.

There are various forecasting techniques; qualitative techniques and quantitative techniques. An example of a qualitative technique is a *Delphi*. Some examples of quantitative forecasting techniques are time series analysis, and causal relationship forecasting. Time series analysis is used in analysis of data by time periods to determine if trends or patterns exist over time. Two examples of Time Series Analysis are moving average and exponential smoothing. Causal relationship forecasting relates the demand to an underlying factor other than time. Examples of causal forecasting are linear (single and multi-variables) and nonlinear (single and multi-variables). The measures of accuracy of forecasting are mean absolute deviation and tracking signal.

***All forecasts have four common characteristics.***

1. Forecasts are usually (always) inaccurate (wrong).
2. Forecasts should be accompanied by a measure of forecast error.
3. Forecasts for aggregate items are more accurate than individual forecasts
4. Long-range forecasts are less accurate than short-range forecasts.

The aggregate items are more accurate than individual forecasts because aggregate forecasts reduce the variability – actual demand for some items come out less than forecast, while the others come out greater than the forecasts, and they compensate each other’s. The standard deviation of the sum of two variables is less than the sum of the standard deviation of the two variables, because the variance of the sum of two variables is equal to the sum of the variances of the two variables. The reason long-range forecasts are less accurate than short-range forecasts is due to the fact that forecasts further into the future tend to be less accurate than those of more recent events. As time passes, we get better information, and make better predictions.

**Example:**

Here is an example of how businesses can gain from developed forecasts techniques. The San Pedro Bay Port is a combined port of Los Angeles and Long beach and, based on their container handling, are ranked 5th in the world, after the Port in Singapore and three ports in China. More than 50 percent of containers coming to the United States pass through San Pedro Bay ports and more than 1/3 of the other potential routes will take business from Southern California and its ports. What are the competing edges of Southern California ports? Deep water facilities for post Panama ships, which may contain more than 8,000 containers; state of the art on-dock facilities to transfer containers between ship and train; intermodal transfer between ship, truck, and train; consolidation and distribution facilities for trans-loading from 20-foot containers and 40-foot containers to 56-foot containers, which are allowed to move on California roads, but as important as this capability is and maybe more important than this capability, are the two last characteristics of all forecasting techniques.

If we want to transfer the load from Far East to East Coast, it will take 4 weeks. From Far East to West Coast, it takes 2 weeks, and from Far East to the mid-United States it takes something between 2 to 4 weeks. Now, if I am going to ship loads from Far East to East Coast, I should forecast the demand of the East Coast 4 weeks in advance. If I am going to ship from Far East to West Coast, I should estimate the demand for the West Coast 2 weeks in advance. Estimates of West Coast, which requires a forecast of 2 weeks, is more accurate than for the East Coasts, which is 4 weeks in advance.

Shorter time provides more accuracy. Look at the other property. The forecast for East Coast, West Coast, and mid-United States, are less accurate than the forecast for the total demand in the United States. So instead of forecasting for East Coast alone for 4 weeks and West Coast alone for 2 weeks and mid-United States for 3 weeks, I forecast the demand for all the United States for 14 days, 2 weeks in advance. Then when I send the container here, in one day I may transfer it to anywhere in California, in 2-3 days to somewhere in the mid-United States and 3 to 4 days somewhere in the East Coast. Now instead of estimating the demand of the East Coast alone, which is less accurate than the demand for the whole country, and instead of forecasting it for 4 weeks from now, I can forecast it for 14 days plus 3 days, which is 17 days from now. The forecast for the United States for the whole country between 14 days and 17 days in advance is much more accurate than the forecast for the East Coast, 4 weeks in advance and forecast for the mid-United States, which is 3 weeks in advance.

**Delphi**

Delphi is a forecasting technique, which uses the opinions of experts of a specific field to obtain a forecast of that respective field. For example, let’s say you own a business. You want to forecast the demand for your business in the coming year. As a result, you decide to use the Delphi forecasting technique. You ask experts in the field of your business for their opinion on the future demand of your business. You collect their responses, which are labeled anonymously and then return the first round of responses to the experts again. Then you ask them to recalculate their forecast after reviewing the first round of responses. After conducting several rounds of responses, the experts will find an average and that level will be assumed as the demand.

**Time Series**

Time series is a type of forecasting technique, which uses data from the past hoping to find a trend for the future. Although time series can measure any variable of interest, we will focus on measuring the variable of demand unless stated otherwise. Time series shows systematic and random components.

***There are three different types of systematic components.***

* Level: Where we think demand is.
* Trend: Growth or decline. This will change upward or downward over time.
* Seasonality: Predictable fluctuations.

We can identify and quantify systematic components.

The random component is the part that deviates from the systematic component and there is no control over this component. This is why forecasts will never be 100 percent accurate, because we are predicting them by assuming perfect conditions.

***Time Series Techniques***

* Naïve Forecast: The cheapest and easiest technique.
* Moving Average
* Exponential Smoothing

**Naïve Forecast**

Naïve forecast obtains its forecasts values for next period by matching it with the values for actual in the current period. Thus, the naïve forecasting method is very cheap, simple and easy. Its simple derivation makes it ideal to use as a base to compare it to the quality of other forecasting techniques. If the quality of other forecasting techniques is lower than naïve forecasting, then the latter is always preferred since it would be cheaper, easier and perform better, as well.

The variable F represents forecast and A represents actual. T represents this period, thus, T plus 1 is the next period. (If T is 1, T plus one is 2; If T is 5, T plus one is 6) At actual demand in period T, F(T+1) is the forecast of demand for period T+1.

**Moving Average**

Moving average is a quantitative forecasting technique that relies on historical data to predict the next period's forecast. It is called moving average because it uses the average of the most recent actual data in a certain period to forecast the next period. Furthermore, as time passes, there is more information about the actual demand. Moving average calculates the forecast using the newest pieces of data and drops the oldest data as time passes. Therefore, it is called moving average because the average moves as new data is available.

Now let’s suppose that we are manufacturing company that recently started six months ago and want to obtain the moving average to forecast the next month's level of production. We use the actual demand **A** of the passes period obtained the moving average. Lets note here that this period's (month) moving average becomes the next period’s forecast.

Two period moving average for period 7:

***MA72 = (A6+ A7)/2***

Three period moving average for period 4:

***MA73 = (A5+ 63 + A7 )/3***

There is no actual limitation on how many periods should be used when calculating moving average. Nevertheless, different period moving average will give us different results; we will discuss these issues later on in this chapter. For now, let’s focus on obtaining the moving average.

The general formula to calculate moving average for any ***n*** period moving average for period ***t*** is as follows:

**n** period moving average for period **t**

***MAtn = (At+ At-1+ At-2 +At-3+ ….+ At-n+1 )/n***

**MA** = Moving Average

**At** = Actual demand in period *t*

**Ft+1** = Forecast for a period t+1

**t** = period

**n** = Number of periods in a moving average

Every time as new data is available, the oldest piece of data is substituted to calculate the new moving average. As we have already explained, this period moving average becomes the next period's forecast, therefore can we assume the general assumption as follows:

The Forecast for period **t+1** is equal to the moving average for period **t**

***Ft+1 =MAtn***

**Exhibit 1 using an example with an Excel spreadsheet**

Suppose, you just got hired in the Hot Wireless store, which is the new a new wireless store that just opened a few months ago and is two blocks away from your house. During your first week there, your boss mentions to you that he is concerned about the amount of smartphones the store should keep in stock since he doesn't want to get overstocked with merchandise that will not be able to sell. It is important to your boss not to invest lot of money in inventory because as entrepreneur she is on a tight budget. Smartphones are supplied on order every two weeks by a local supplier. Having taken System Operations Management the past semester at CSUN, you think that you could calculate the sales of smartphones to obtain a forecast. After you explained this to your boss, she hands you with the bi-weekly sales since the store opened, which it only has been 6 months.

The biweekly sales are the following:

|  |  |
| --- | --- |
| Period (Biweekly) | Actual Demand |
| t | At |
| 1 | 43 |
| 2 | 56 |
| 3 | 79 |
| 4 | 75 |
| 5 | 84 |
| 6 | 87 |
| 7 | 91 |
| 8 | 99 |
| 9 | 96 |
| 10 | 108 |
| 11 | 106 |
| 12 | 113 |

You use the 3 period moving average of the 12th period to forecast the demand for smartphones for the coming two weeks. Therefore, you are able to place the order for inventory with more accurately.

***MA123 =*** (108+106+113)/3 = 109

Our Forecast for the smartphone sales for the next period (13th) will be a demand of 109 cellphones. Your boss usually places inventory orders based on the last's periods demand plus a markup of 20% to prevent being understock. She realized that she was holding thousands in inventory that you calculated being used for other important part of her business. Therefore, she places an order based on your forecast.

To her surprise, your forecast turns out really close to the actual demand of 109. Now, she is able to free some cash from inventory to reinvest in the business. Furthermore, you are now in charge to keep track on sales and placing orders.

**Exhibit 2.1 Continuous forecast**

As the time passes, hot wireless is attracting more customers but it still does not produce enough revenue to have a big inventory. Your boss actually likes to have low inventory because in that way Hot Wireless does not maintains the newest models without the risk of having outdated smartphones when new one come out. Therefore, you decide to analyze you data even further.

|  |  |  |
| --- | --- | --- |
| period | Actual Demand | 3 period Moving Average |
| t | At | MA |
| 1 | 43 |  |
| 2 | 56 |  |
| 3 | 79 | 59.33 |
| 4 | 75 | 70.00 |
| 5 | 84 | 79.33 |
| 6 | 70 | 76.33 |
| 7 | 91 | 81.67 |
| 8 | 99 | 86.67 |
| 9 | 90 | 93.33 |
| 10 | 108 | 99.00 |
| 11 | 106 | 101.33 |
| 12 | 113 | 109.00 |
| 13 | 110 | 109.67 |
| 14 | 116 | 113.00 |
| 15 | 120 | 115.33 |

***MA123 =*** (108+106+113)/3 = 109

The following period moving average are calculated using the most recent demand for smartphones and the oldest the piece of data is dropped.

***MA133  =*** ((-108) +106+113+110)/3 = 109.67

***MA143  =*** ((-106)+113+110 +116))/3 =113

It continuous in the same manner for the following periods.

***MAtn = (At+ At-1+ At-2 +At-3+ ….+ At-n+1 )/n***

To be able to have a better understanding of your demand, you can also use the moving average with different periods. For Example, you may use 4 period moving average, 5 period moving average, 6 period moving average, and so on; it is your choice as to how many to use.

Continuing with our Hot Wireless example, we have created the chart with two different moving average, which are 3 period moving average and 6 period moving average:

|  |  |  |  |
| --- | --- | --- | --- |
| period | Actual Demand | 3 period Moving Average | 4 period Moving Average |
| t | At | MA^3 | MA^6 |
| 1 | 43 |  |  |
| 2 | 56 |  |  |
| 3 | 79 | 59.33 |  |
| 4 | 75 | 70.00 |  |
| 5 | 84 | 79.33 |  |
| 6 | 70 | 76.33 | 67.83 |
| 7 | 91 | 81.67 | 75.83 |
| 8 | 99 | 86.67 | 83 |
| 9 | 90 | 93.33 | 84.83 |
| 10 | 108 | 99 | 90.33 |
| 11 | 106 | 101.33 | 94.00 |
| 12 | 113 | 109.00 | 101.17 |
| 13 | 110 | 109.67 | 104.33 |
| 14 | 116 | 113.00 | 108.17 |
| 15 | 120 | 115.33 | 112.17 |

Although, we are using the using the same method to forecast the demand for the next period, the results between moving average are different. This is due to the smaller moving average being more reactive to the most recent changes in data. As larger moving average takes into consideration older pieces of data and as result changes forecasts will be smoother and less reactive to the actual demand.



As you can see in the above scatter diagram, three period moving average is more reactive to the changes in demand, while the six period moving average has a smoother curve. From this comparison, we can conclude that smaller period moving average are better to forecast the demand in the short term. The bigger period moving average are better in the long run when a company wants to have an overall view of the demand or any other item of interest that the company wants to forecast, for example cost and profit. But which method is best forecast for our example? That is determined by the size of the error between actual and forecast. We need to find the Mean Absolute Deviation that we will use as a system to measure error between different periods moving average.

**Mean Absolute Deviation**

As we have already noted, different period moving average will give us different results. Although, the data still the same, a 3 period moving average will not be the same than that of a 6 period moving average. How can we know which system is the best forecast technique? The **Mean Absolute Deviation** (MAD) is used to determine the deviation of the forecasts compared with the actual numbers of a given period.

Mean Absolute Deviation (MAD) = *(Sum | At - Ft |) / Number of Periods*

One measure of effectiveness to find out whether our forecasting techniques are good for our data is summing out the difference between actual and Forecasts. Because the differences sometimes are positive and negative, the summation of the differences adds up to zero or close to zero. Therefore, to find the mean error of the whole data, we need to get rid of the negative signs.

The best way to get rid any number’s sign is to find its absolute value. For example, if we have a forecast for the sale of smartphones for this week of 140 and actuals sales were 135 smartphones, the difference, or error, between actual and forecast was 140 – 135 = -5. The absolute value of the error would be -5 = | 5 | and since we got rid of the negative sign, we can compute the mean absolute deviation MAD. MAD is the summation of the absolute value between actual demand and forecast divided by the number of observations.

|  |  |  |
| --- | --- | --- |
| **Exponential Smoothing** | May 5  2014 | |
|  | |  |

# What is Exponential Smoothing?

Exponential smoothing is another effective forecasting method that is similar to the weighted moving average forecasting technique. The moving average forecasting method functions differently from the weighted moving forecasting method. In the moving average method, the most recent end-pieces of data is added and divided by N. Alternatively, each piece of data can be divided by N and then added up, or you can multiply each piece of data by 1 over N and then add them up. Remember that At is the actual demand in period *t* and that F(t+1) is the forecast of demand for period *t+1*.

Example 1: In a 4-period moving average period take At, At-1, At-2, At-3 and divide them by 4. Or divide each one by 4, and dividing by 4 is the same as multiplying by .25; therefore, in a 4-period moving average, multiply At by .25, At-1 by .25, At-2 by .25, At-3 by .25. So the weights of all pieces of data are equal, in this example, the periods are multiplied equally by 25%. However, in a weighted moving average, the weights are not equal. The percentages (1/N) still sum up to 1, but it is not the same percentage multiplied to each period. For example, the most recent piece of data might get .4; the oldest piece of data might get a coefficient of .1. However, the summation of all the coefficients should come out equal to 1.

Exponential smoothing is similar to the weighted moving average method. The exponential smoothing formula is: forecast for the next period is equal to forecast for this period plus a fraction of the difference between actual demand and forecast in this period.

Ft+1 = Ft+α(At-Ft)

The forecast for this period and the actual demand for this period are used to solve for the forecasting of the next period. Another way to rearrange the exponential smoothing formula is by multiplying alpha by actual demand (At) and alpha minus 1 by Ft. Thus numerically, the exponential formula is:

Ft +1= (1-α) Ft+ αAt.

Example 2:

|  |  |  |  |
| --- | --- | --- | --- |
| t | 1 | 2 | 3 |
| At | 100 | 150 |  |
| Ft | 100 | 100 | 110 |

Suppose for period 1 the forecast is exactly the same as actual demand for period 1. Alternatively, add all pieces of data average them and assume that as the forecast for period 1. Therefore, forecast for period 2 is equal to the forecast for that period multiplied by (1 + α), plus the actual demand multiplied by alpha, α, in which α is .2. The actual demand, At, is equal to 100 and the forecast, Ft, is equal to 100. Now, Ft for period 2 is equal to 100.

Another approach is to find the difference between At and Ft, which is zero. Alpha is equal to .2. Therefore, F2 equals 100. In other words, F2 is equal to A1. Suppose a marketing department has informed us that actual demand for period two is 150 and we need to forecast for period three. Since we have no information for F1, we will enter A1,, which is 100. Alternatively, we may assume the average of all available data as our forecast for period 1.

Therefore, F3 = (1-α) F2 + αA2 or F3 = F2 + α. In which is .2 is multiplied by the difference between A2 and F2; where A2 = 150 and F2 = 100. The difference is 50. We then multiply 50 by .2, which equals 10. Therefore, since F2 = 100, we then add the forecast for period two, F2 plus 10. Which means that F3 equals 110. However, if we use the other formula, we would still get 110.

*Solution:*

*F2 🡪 A1, α = 0.2*

*F3 = (1-α)F2 + αA2*

*F3 = 0.8(100) + 0.2(150)*

*F3 = 80 + 30 = 110*

Example 2.1: The student store has been having a decline in demand for Universal Studios tickets during the spring semesters at California State University, Northridge. Actual demand in tickets for the spring semester of 2013 was 1500, and the forecast matched the same amount of tickets actually sold for last spring. The staff wants to know the forecast in Universal Studio tickets for the spring semester of 2014 to prevent any losses on unsold tickets. α = .3.

|  |  |  |
| --- | --- | --- |
| Spring Semester | Demand | Forecast |
| 2013 | 1500 | 1500 |
| 2014 |  | ? |

Solution:

F2014 = (1-α)F2013 + αA2013

F 2014 = (1-.3)1500 + (.3)1500 🡪 F 2014 = **1500**

Example 2.2: After the spring semester of 2014 ended, the actual demand for Universal Studios tickets ended up being 500. The staff of ticket office had over-estimated demand by 1000. Now they want to know what the forecast will be for spring semester of 2015. A2014 = 500; α = .3

|  |  |  |
| --- | --- | --- |
| Spring Semester | Demand | Forecast |
| 2013 | 1500 | 1500 |
| 2014 | 500 | 1500 |
| 2015 |  | ? |

Solution:

F2015 = (1-α)F2014 + αA2014

F2015 = (1-.3)1500 + (.3)500 🡪 F2015=**1200**

**OR**

F2015 🡪 α(A2014-F2014) = -300

F2015 = .3(500-1500) = -300 🡪 F2015 = 1500-300 = **1200**

Let’s look at an interesting observation and understand why Professor Asef tries to set the stage for the exponential smoothing while using the weighted moving average. F3 = (1-α)F2 + αA2; therefore, F3 is computed using F2 and A2. But how was F2 computed? F2 was computed using A1. Therefore, F3 was computed based on A2 and A2.

Example 3: At the end of period 3, you may realize that the demand for period 3 is 120 while the forecast is 110. Now let’s compute F4. Which means: F3 = 110, A3 = 120, α = .2.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| t | 1 | 2 | 3 | 4 |
| At | 100 | 150 | 120 |  |
| Ft | 100 | 100 | 110 | 112 |

Solution:

F4 = (1-α)F3 + αA3

F4 = 0.8(110) + 0.2(120)

F4 = 88 + 24 = **112**

Therefore, the forecast for period 4 is 112. To find F4, we used F3 and A3, since we know that F3 was computed using A2 and A1. Thus, F4 was computed based on A3, A2, and A1. Moreover, using the same logic, you will see that F5 is computed using A4, A3, A2, and A1; whereas, F10 is computed based on A9 to A1. Ft is computed based on At-1 and At-2 all the way to A1. Indeed, forecasting using exponential smoothing takes into account all pieces of data, but in different ways.

Example 3.1: When the spring semester of 2015 ended, the actual demand for Universal Studios tickets was again 500. This time the tickets were overestimated by 700. However, the staff of the ticket office is trying to keep a positive outlook since they had reduced the previous overestimated amount by 300 tickets. Help the staff forecast how many tickets they should purchase for the spring semester of 2016. A2015 = 500; α = .3

|  |  |  |
| --- | --- | --- |
| Spring Semester | Demand | Forecast |
| 2013 | 1500 | 1500 |
| 2014 | 500 | 1500 |
| 2015 | 500 | 1200 |
| 2016 |  | ? |

Solution:

F2016 = (1-α)F2015 + αA2015

F2016 = (1-.3)1200 + (.3)500 🡪 F2016 = **990**

Example 4: In this example, alpha is equal to .1. The data given is for the first 8 periods, and we need to determine the forecast for period 9. In period 1, the actual demand is 200 and the forecast is 200. However, since the actual demand for period 1 is equal to the forecast for period 1, the forecast for the following period is 200. For period 2, actual demand is 250, and the forecast for this period is 200; with alpha equal to .1. Apply the formula to solve for the next period’s forecast using F3 = (1-α)F2+αA2; the answer being 205.

|  |  |  |
| --- | --- | --- |
| Week | Demand | Forecast |
| 1 | 200 |  |
| 2 | 250 | 200 |
| 3 | 175 | 205 |
| 4 | 186 | 202 |
| 5 | 225 | 200 |
| 6 | 285 | 203 |
| 7 | 305 | 211 |
| 8 | 190 | 220 |

Now the forecast for period 3 is 205, and actual demand is 175. Therefore, using the same formula, we discover that F4 = 202. The forecasts for the rest of the periods can be found using the same procedure. Thus, the forecast for period 9, F9, is equal to 217.

What is the relationship between alpha and N in N-period moving average? What does large alpha mean? What does small alpha mean?

Using the formula, Ft + 1 = Ft, our previous forecast plus alpha times At-Ft. When alpha goes up, more weight is given to the recent deviation. When alpha goes down, less weight is given to the recent deviations. In moving average, when N is small it is more reactive. When N is larger, the curve is smoother. The larger the N, the smaller the alpha becomes. This is the relationship between exponential smoothing and moving average. The problem below serves as another example to observe the differences in alpha between exponential smoothing and moving average while demand remains constant.

Example 4.1: There are two small stores that sell the same brand of pepper spray. One store is called Reseda Life and the other store is called Nordoff Market. α = .1

Reseda Life’s historical demand of pepper spray is shown as follows:

|  |  |  |
| --- | --- | --- |
| Year | Demand | Forecast |
| 1 | 1000 | 1000 |
| 2 | 1500 | 1000 |
| 3 | 1700 | 1050 |
| 4 | 2000 | 1115 |
| 5 | 1400 | 1204 |
| 6 | 1000 | 1224 |
| 7 | 3000 | 1202 |
| 8 | 2500 | 1382 |

What is the forecast demand of pepper spray for year 9 at Reseda Life?

Solution:

F9 = (1-α)F8 + αA8

F9 = (1-.1)1382 + (.1)2500 🡪 F9 = **1494**

Example 4.2: Nordoff Market is the rival of Reseda Life, and had some of its staff do some undercover work on Reseda Life. Nordhoff Market wanted its staff to inspect how much pepper spray its customers purchase every year. Nordhoff Market offers discounts to increase demand. Thus, demand until year 8 matches Reseda Life’s demand of pepper spray. α = .5

|  |  |  |
| --- | --- | --- |
| Year | Demand | Forecast |
| 1 | 1000 | 1000 |
| 2 | 1500 | 1000 |
| 3 | 1700 | 1250 |
| 4 | 2000 | 1475 |
| 5 | 1400 | 1738 |
| 6 | 1000 | 1569 |
| 7 | 3000 | 1285 |
| 8 | 2500 | 2143 |

What is the forecast demand for year 9 of pepper spray?

Solution:

F9 = (1-α)F8 + αA8

F9 = (1-.5)2143 + (.5)2500 🡪 F9 = 2322

Thus, you would use the same computations as you would when alpha equals .1 as, let’s say compared to alpha being .4. As alpha decreases, the curve becomes smoother and gets closer to a horizontal line. However, when alpha increases, it becomes more reactive to the recent changes. As shown in Exhibit 1.1 below. Conversely, in moving average, when periods are small, it is more reactive in exponential smoothing. When alpha is large, it is more reactive. In moving average, when N is large it has a smoother curve versus exponential smoothing when alpha is small, it has a smoother curve.

*Exhibit 1.1:*

|  |  |  |
| --- | --- | --- |
| Week | Demand | Forecast |
| 1 | 200 |  |
| 2 | 250 | 200 |
| 3 | 175 | 220 |
| 4 | 186 | 202 |
| 5 | 225 | 196 |
| 6 | 285 | 207 |
| 7 | 305 | 238 |
| 8 | 190 | 265 |

But which is better: an exponential smoothing with a larger value of alpha, or one with a smaller value of alpha? We cannot answer that question, but what we do know is that it is the same as moving average. However, we can use mean absolute deviation (MAD) to determine which one is better. The forecasts using alpha equal to .1, will be compared to forecasts using alpha equal to .4.

Example 6: Look at period 2 to period 8.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Week** | **Demand** | **Forecast for 0.1 alpha** | **AD** | **Forecast for 0.4 alpha** | **AD** |
| **1** | **200** |  |  |  |  |
| **2** | **250** | **200.00** | **50.00** | **200.00** | **50.00** |
| **3** | **175** | **205.00** | **30.00** | **220.00** | **45.00** |
| **4** | **186** | **202.00** | **16.00** | **202.00** | **16.00** |
| **5** | **225** | **200.40** | **24.60** | **195.60** | **29.40** |
| **6** | **285** | **202.86** | **82.14** | **207.36** | **77.64** |
| **7** | **305** | **211.07** | **93.93** | **238.42** | **66.58** |
| **8** | **190** | **220.47** | **30.47** | **265.05** | **75.05** |
|  |  |  | **46.73** |  | **51.38** |

Solution:

Ft = αAt-1 + (1-α)Ft-1

Ft-1 = αAt-2 + (1-α)Ft-2

Ft = αAt-1 + (1-α)αAt-2 + (1-α)Ft-2

Ft-2 = αAt-3 + (1-α)Ft-3

Ft = αAt-1 + (1-α)αAt-2 + (1-α)2αAt-3 + (1-α)3Ft-3

= αAt-1 + (1-α)αAt-2 + (1-α)2αAt-3 + (1-α)3αAt-4 + (1-α)4αAt-5 + (1-α)5αAt-6 + (1-α)6αAt-7 + …

Actual demand is 250, Forecast is 200, and absolute deviation is 50. Demand is 175, forecast is 205, absolute deviation 30. Add up all absolute deviations and divide it by 7, which results to 46.73. For alpha equal to .4, add up all absolute deviations and divide it by 7, and this result to 51.38. For this specific set of data, alpha equal to .1 is better coefficients because MAD is lower than alpha being equal to .4.

Example 6.1: Analyze the relationship between Reseda Life and Nordoff Market. Which alpha is better .1 or .5 for these two stores?

Solution:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Week** | **Demand** | **Forecast for 0.1 alpha** | **AD** | **Forecast for 0.5 alpha** | **AD** |
| **1** | **1000** | 1000 | - | 1000 | - |
| **2** | **1500** | 1000 | 500 | 1000 | 500 |
| **3** | **1700** | 1050 | 650 | 1250 | 450 |
| **4** | **2000** | 1115 | 885 | 1475 | 525 |
| **5** | **1400** | 1204 | 196 | 1738 | 338 |
| **6** | **1000** | 1224 | 224 | 1569 | 569 |
| **7** | **3000** | 1202 | 1798 | 1285 | 1715 |
| **8** | **2500** | 1382 | 118 | 2143 | 357 |
|  |  |  | MAD:4371/7=**624.43** |  | MAD:4454/7=**636.29** |

The MAD that is lower is the better coefficient: 624.43<636.29, thus, .1 alpha is better than .5 alpha.

When graphing the analysis of both stores, Nordoff Market seems to be more responsive to demand because alpha = .5. Whereas, Reseda Life has alpha = .1; therefore, data is less reactive to demand and the data looks smoother.

Here is an example to show how the forecasting changes when alpha goes up or comes down. If alpha is equal to 1, Ft + 1 = (1-α)Ft + αAt. Alpha is 1, and 1-α, which equals 0. So 0 is multiplied by Ft, and Ft + 1 is equal to At; which is exactly like the naïve technique. The forecast is the same as actual demand for the previous period.

# Exponential Smoothing and Excel

Review: Let us examine what happens when the value of alpha goes down. When alpha decreases we become less reactive. When alpha is 0, there is a straight line. In other words, there is no reaction to what is happening in reality. Conversely, when the value of alpha increases we become more reactive. In exponential smoothing, a forecast is calculated using the forecast from the previous period, and the actual data from the previous period. This means that we only use two values, actual and forecast. It is important to realize, however, that the forecast from the previous period includes the history from all previous periods, and when we add it to the actual from the previous period, the whole history of data can be calculated. Now, let us take another look at the exponential smoothing formula.

Forecast for period T is a function of actual in the previous period and forecast in the previous period, while actual is multiplied by alpha and forecast is multiplied by 1-α. This is expressed in the following equation: Ft = α At-1 + (1 – α) Ft-1. Each time we receive a new piece of data, we multiply it by alpha; however, Ft-1 is not just Ft-1. Ft-1 is a history of data from previous periods. If we look one period earlier, Ft-1 is a function of At-2 and Ft-2, where At-2 is multiplied by α and Ft-2 is multiplied by 1-α. If we replace Ft-1 with the equation for Ft-1, we will get a new equation in which each actual piece of data is multiplied by alpha. This is expressed in the following equation: Ft = α At–1+(1– α) α At–2+(1– α)2Ft–2. Now, if we replace Ft-2 with the equation for Ft-2, the new equation for Ft is expressed as follows:

Ft = αAt–1 + (1– α)αAt–2 + (1– α)2αAt–3 + (1 – α) 3Ft–3

As you can see, all pieces of data are still multiplied by alpha. A piece of data which is one period old is multiplied by alpha, a piece of data which is 2 periods old is multiplied by 1-α, and a piece of data which is 3 periods old is multiplied by 1-α to the power of 2. If we continue extending this general formula for Ft, we will see that all actual pieces of data are multiplied by alpha, but also by 1-α to the power of something. If data is *t* periods old, it is multiplied by 1-α to the power of *t*-1.

For example, if data is 4 periods old, it should be multiplied by 1-α to the power of *4-1*; if data is 5 periods old, it should be multiplied by 1-α to the power of *5-1*; if data is 6 periods old, is should be multiplied by 1-αto the power of *6-1*, and so forth. If alpha is equal to 0.5, then 1-α to the power of 0 is 1, 1-α to the power of 1 is 0.5, 1-α to the power of 2 is 0.25, and 1-α to the power of 5 would be around .01 or .02. Therefore, exponential smoothing takes into account all pieces of data, but as the data gets older, its coefficient gets smaller.

This is also evident in another example. If alpha is equal to 0.9, then 1-α to the power of 0 is 1, 1-α to the power of 1 is 0.1, 1-α to the power of 2 is 0.01, and 1-α to the power of 3 is 0.001. Because alpha will always be between 1 and 0, increasing the power to which it is raised will make alpha smaller and smaller. Therefore, exponential smoothing is a weighted moving average, which uses all actual pieces of data, but the coefficient of each piece of data gets smaller as that piece of data gets older.

Example 7: We will now solve a problem involving the coefficients of data. Assume α = 0.6.

A) Find the coefficient of data when data is 10 periods old.

Solution: (1 – α) t – 1

(1 – 0.6) 10 – 1 = 0.4 9 = 0.00026

B) Find the coefficient of data when data is 2 periods old.

Solution: (1 – α) t – 1

(1 – 0.6) 2 – 1 = 0.4 1 = 0.4

C) Find the coefficient of data when data is 20 periods old.

Solution: (1 – α) t – 1

(1 – 0.6) 20 – 1 = 0.4 19 = 0.000000027

When we talk about an N-period moving average, the newest piece of data is 1 period old and the oldest piece of data is N-periods old. On average, data is 1 + N divided by 2 periods old, or 1+N/2. In exponential smoothing, the age of data is equal to 1 over alpha, or 1/α. Although these techniques are not exactly the same, they are similar enough that we can set them equal to each other.

Therefore:

(1+N)/2 = 1/α.

For example, in a 6-period moving average, the newest piece of data is 1 year old and the oldest piece of data is 6 years old. The data on average is 3.5 periods old. Since the exponential smoothing equation is almost equivalent to the moving average equation, we can set 3.5 equal to 1 over alpha, and we will find that alpha equals 0.28.

Let us look at another example. In a 10-period moving average, the newest piece of data is 1 period old and the oldest piece of data is 10 years old. We will first find the average age of data, and then we will find the value of alpha: (1+10)/2 = 5.5. The average age of data is 5.5 periods. We will now find the value of alpha by setting 5.5 equal to 1 over alpha: 5.5 = 1/α. After solving the problem, we find that alpha is equal to 0.18.

Example 8: We will now solve a problem involving the average age of data.

Find the value of alpha and the average age of data in a 15 period moving average.

Solution: (1+N)/2

(1+15)/2 = 8. The average age of data is 8 periods. Now we will solve for α.

1/α = 8, α = 1/8 = 0.125. The value of α is 0.125.

Now we will describe how to solve exponential smoothing problems using Excel, and introduce some Excel features. Suppose there is demand in period 0, and the demand from period 0 will be used as my forecast for period 1. Next, we will find the forecast for period 2. The forecast for period 2 is equal to 1-α, and we must make it absolute by pressing F4 times forecast + alpha. This will make it absolute times actual, which is our forecast for period 2. We can now use the forecast and actual for period 2 to calculate the forecast for period 3. We find the forecast for period 3, by multiplying the forecast of period 2 by 1-α + actual of period 2 times alpha, we can then copy the formula down for all periods.

Once we have the actual and forecast data, we can then calculate actual minus forecast, and copy the formula down for all periods. This will give us the deviation between At and Ft. In the next column, we will calculate the absolute deviation and copy the formula down for all periods.

Now we will compute MAD. We can compute MAD as the summation of all appropriate cells divided by the period number. If we copy this formula all the way down, however, it will give us incorrect values for MAD. This is because we must make the first element in the range absolute. Now when we copy the formula all the way down, the values for MAD are correct.

After computing MAD, we will compute the summation of deviations. We will do this by including all positive and negative signs, and finding the summation of all deviations in the range. We must make sure that the first element in the range is absolute. Next we will compute the tracking signal.

The tracking signal is computed by dividing MAD by the summation of deviations, and then copying the formula all the way down. Lastly, we will perform “what-if” analysis to see what happens when alpha changes. Alpha is currently equal to 0.5.

When performing “what-if” analysis, it is important that the MAD column is one column to the right, and one row north. After performing “what-if” analysis in a one dimensional table, we received different MAD values for every value of alpha between 0.1 and 1. We will then reduce the number of decimal points, and find the lowest MAD value by typing “= MIN” and selecting the range of MAD values. This will give us the smallest MAD value in the range. Since finding a single value in a large range of numbers can be difficult, we will use Excel’s conditional formatting function to highlight the smallest MAD value that was given to us by the MIN function.

Inventory Models

## Importance of Inventory

Today, we will talk about inventory models.  **Inventory models** are perfect examples of applying mathematical models to real world problems.  In general, an organization’s inventory is a very prominent cost, for example, 20 percent of the budgets of hospitals are spent on medical, surgical, and pharmaceutical supplies. For all hospitals in the United States, it adds up to $150 billion annually.  The average inventory in the United States economy is about $1.13 trillion, and that is for $9.66 trillion of sales per year.

However, there are times where inventory may be at the detriment of a company. For instance, a company with a large work in process and finished goods inventory may discover that the market is shifting from one product to another product. In this case, the company will have a large amount of work in process and finished goods inventory of a product that customers have already shifted to another product.  Thus the company will have *two choices*.

One choice is **to fire-sell** all inventories and finished goods what they have, which involves the selling of goods at extremely discounted prices. A drawback to fire-selling is that it may turn into a significant loss because the inventory is potentially sold at 20 percent, 10 percent, or 30 percent of their actual value. The second way is that they could sell their finished goods and also at the same time, turn their work in process into sellable finished goods. Unfortunately, this means that there would be a lot of delay in entering the product into the market, and they a company could lose a substantial portion of market share. Therefore, in both of these alternatives, they lead to loss. Thus, what is the message? We need to reduce our inventory as much as possible; we need to have minimal inventory.

## Inventory Classified

There are three types of inventory.  **Input inventory** is composed of raw materials, parts, components, and sub-assemblies that we buy from outside. **In-process inventory** are parts and products, sub-assemblies and components that are being processed; part, products, sub-assemblies, and components that are there to decouple operations.  For example, if operation B follows operation A, in order to not completely have operation B dependent on operation A, we may put a little bit of inventory between those two.  And the third type of in-process inventory is when we realize that if we buy at large volumes, we get lower expense due to economies of scale.

Then we have **output inventory**.  We need to have some inventory “over there,” because when customers come, we cannot tell them “wait, I will make your product.”  The best strategy is if I can have such a low flow time in which I can deliver manufactured product and give it to the customer, but we are not there yet.  Therefore, I should have inventory on shelf when customers come to satisfy demand.  Also, sometimes, the demand in one season is high and in another season is low; therefore, I should produce in low season and put it in inventory to satisfy demand in high season. Another type of output inventory is our products that are in a pipeline from manufacturing plants to warehouses or to distribution or to retailers. If you look at the huge volumes of inventory on our highways, those are **pipeline** or **in transit inventories.**

## Inventory

Poor inventory management *hampers operations*, *diminishes customer satisfaction*, and *increases operating costs*.  A typical firm probably has about 30 percent of its *current assets* in inventories or about 90 percent of its *working capital* (and working capital as you may remember from accounting and from finance is the difference between current asset and current liabilities). Using 90 percent of working capital, they may switch to another company forever or another vendor.

**Understocking** is not good because customers come and when we tell them that we don’t have the product, we will then have dissatisfied customers.  This will lead to a loss of sales, and the customer may go to another vendor forever.

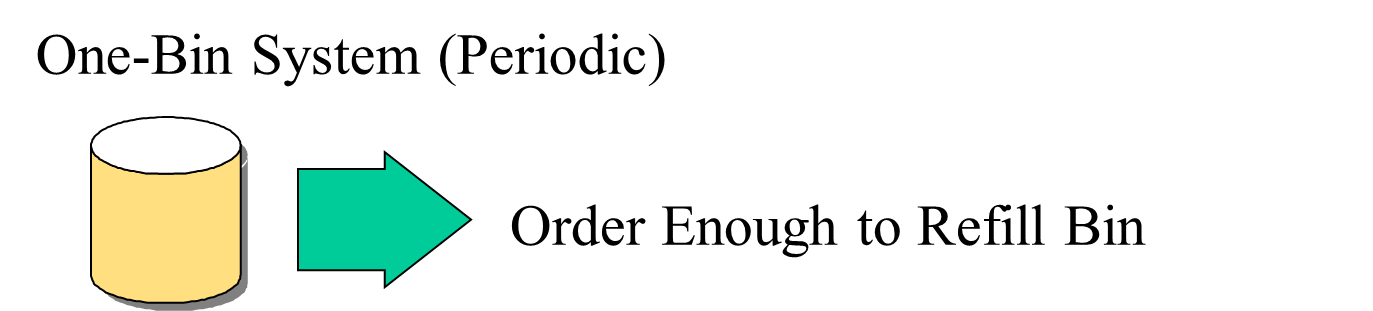
**Overstocking** is not good either because it has three types of costs: financial costs, physical costs, and obsolescence costs. *Financial costs*, instead of having our money in a city or in a profitable business, we put it in inventory.  *Physical cost,* our inventory should be put in safe keeping somewhere.  Thus, we either lease a warehouse or allocate a portion of our shop to a physical location of these products.  And finally we may have *obsolescence cost* , if we purchase for a large amount of inventory for a product that eventually gets low consumer demand, we may never be able to sell them; thus rendering the product obsolete.

## Periodic Inventory [Counting] Systems

We have *two types* of inventory systems—inventory counting systems.  One of them is called periodical.  The other is called perpetual.  In **periodical inventory system,** at the beginning of each period, the existing inventory level is identified and the additional required volume to satisfy the demand during the period is ordered.  So the quantity of order is variable, but the timing of order is fixed.  It is the beginning of each period.  **Re-order point (ROP)** – when we reorder, is defined in terms of time.  It is the beginning of the period.

How do we order?  At the beginning of the order – at the beginning of the period, we go and look at our inventory.  You may imagine it like a one-bin system, that is, there is one bin and our products are in there.  We can look and see that it is filled up to this point and the remainder is empty, so I order that much.   Each time, we order enough to refill the single bin.  The quantity that I order each time depends on how much I need to fill the bin, but the timing is exact.  It is at the beginning of the period.  Reorder point is defined in terms of time.

**Figure 1. Periodic Inventory: One-Bin System**

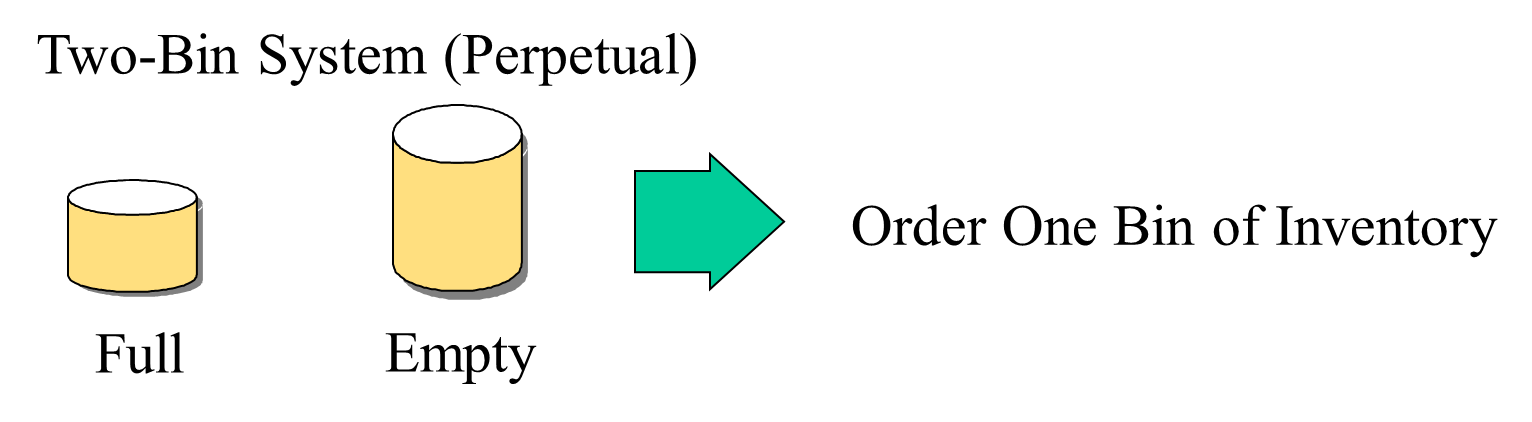


The *disadvantage* of this system is that during this period, we don’t have information about our inventory.  I don’t look at that virtual bin during the period.  I only look at it at the end of each period, which is the beginning of the next period.  The *advantage* is because our inventory counting system counts for a lot of items at one day, then at one day I could order for a lot of items. So then WE make orders for many items at the same time.  My ordering cost will perhaps go down because of it.

## Perpetual Inventory Systems

**Perpetual inventory system** is entirely different.  When inventory reaches reorder point, we order a specific quantity, and usually we order **economic order quantity (EOQ),** which we will discuss later.  The quantity of order, unlike periodic inventory system is fixed, but the timing of the order is variable.  Whenever our inventory has reached a specific level, we will order.  Reorder point is defined in terms of quantity, or inventory on hand.  So you may think of it as a virtual two-bin system.  Whenever the first bin gets empty, we order enough products to fill this first order.  While we are waiting to get this product, we are using the inventory of this bin. A perpetual inventory system is like a two-bin system in which when one bin gets empty, we order enough to fill it up.  The other one was a single bin at the end of each period where we look to see how much we need and we order that much.  The benefit of this system is that it keeps track of removal from inventory continuously.

**Figure 2. Perpetual Inventory: Two-Bin System**



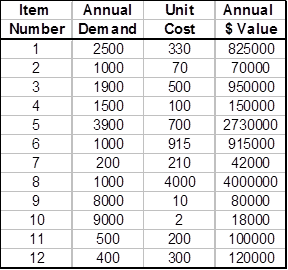
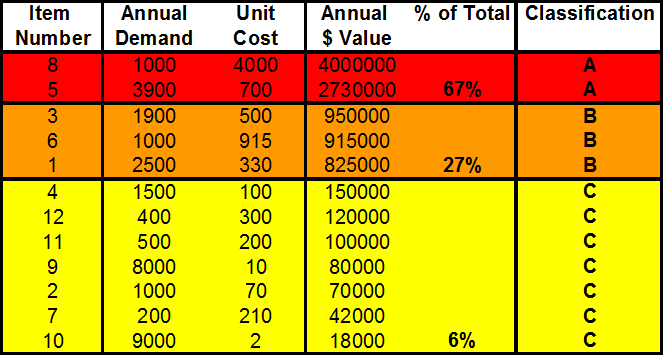
## A Classification Approach: ABC Analysis

In **ABC analysis**, the question is which type of inventory counting system is preferred?  Is it periodical or perpetual?  Perpetual is always better but more expensive because we need an automated system to continuously count our inventory.  Therefore, we may conduct an ABC analysis.

**Example 1.**

Here are our 12 parts (see Figure 3).  Here is the annual demand of each part.  Here is the unit cost of each part.  If we multiply them, we will get annual value of all items in our inventory system in our warehouse.  If we sort them in non-increasing order, we will see that two items, which is 2 divided by 12, which is something between 15 to 20 percent of items, form 67 percent of the annual value. And here, 7 items divided by 12 is a little more than 50 percent, say 55 percent of items, form 6 percent of the value.  These are group C.  These are group A and obviously these are group B.  For group A, we use perpetual, and group C, we may use periodical, and for group B, one of the two options.

**Figure 3. ABC Analysis**

**Example 1a.**

Here we have 12 different video games along with their list of annual demand and unit cost. Using these values, we can find the annual dollar value, the percentage of their value, and then classify them as A, B, or C to see which type of inventory counting system is preferred.



## The Basic Inventory Model: Economic Order Quantity

Now we reach to our basic inventory model, which we try to develop a mathematical formula for a portion of real world, and this model is called **economic order quantity**, or **EOQ.**  In this model, we assume that we only have one product; only a single product.  Demand is known, and demand is constant throughout the year.

So for example, we know that we need 5,000 units of product per year, and if a year is 50 weeks, then 1/50 of this number, we need every week.  If a week is 5 days, 1/5 of whatever we need per week we need per day, so demand is known and it is constant.  Every day, every minute, every hour, we have the same demand as any other minute, hour, or day.

Each order is received in a single delivery.  When we order, we should wait for a lead time, 1 day, 2 days, 3 days, it is known, and it is fixed.  After lead time, we receive the inventory that we have ordered. Therefore, as soon as inventory is on hand, reach to demand during lead time as soon as our inventory is equal to what we need for lead time. Iif lead time is 3 days, as soon as our inventory reaches a level that we need for 3 days, then we order.  During those 3 days, because lead time is fixed, after 3 days, we will get the product.  And because demand is fixed and constant, at the second we get the product, our inventory reaches 0, and then we replenish.

There are only two costs involved in this model: **ordering cost,** cost of ordering and receiving the order; and **holding** or **carrying costs,** costs to carry an item in inventory for one year; cost to count one item in inventory for one year.  Unit cost of product does not play any role in this model because we do not get a quantity discount.  It does not matter if we order one unit or one million units, the price is the same.

## The Basic Inventory Model

Allow us to give you an example and this will clarify the topic we are going to discuss.

**Example 2.**

Annual demand for a product is 9600 units.

D = 9600

Annual carrying cost per unit of product is $16.

H = 16

Ordering cost per order is $75.

S = 75

Annual demand for a product is 9,600 units (D), so we need 9,600 units per year.  In every minute of a year, we need the same number of units as another minute.  So D, which is demand, is equal to 9,600.  Annual carrying cost per unit of a product is $16 per unit per year (H). What does it mean?  If we have one unit of inventory in our warehouse and if we keep it for one year, it costs us $16.  That includes, for example, financial cost, physical cost of holding this inventory, and obsolescence cost.  That is, $16 per unit per year, or inventory carrying cost.  And we show it by H, and that is equal to 16.  Ordering cost per order is $75 (S).  Each time we place an order, it costs us $75.  S is for ordering cost, and it is equal to 75.

**Example 2a.**

Apple’s MacBook has an annual demand of 15,000 units per year. Therefore, D = 15,000 units. They have an annual carrying cost per unit of $22 per unit per year. This means that for every year that they keep the product; it costs them $22, which is their H. Every time an order is placed, it will cost them $80, and so S will be $80.

Annual demand for the MacBook is 15,000 units.

D = 15,000

Annual carrying cost per MacBook is $22

H = 22

Ordering cost per order is $80

S = 80

How much should we order each time to minimize our total cost?

## Economic Order Quantity

Now we need to answer this question, how much should we order each time to minimize our total cost?  That is one question.  Should we order 9,600 at the beginning of the year or should we order twice a year each time 4,800 units, or say 96 times a year each time 100 units?  What is the economic order quantity?  So like many other business problems here, we have two different cost structures.  One goes up if ordering cost goes up and the other goes down as ordering quantity goes down.

Then the other questions we should answer are how many times should we order?  What is the length of order cycle if we have 288 working days per year?  What is the total cost of this system?  As we said, we don’t consider purchasing cost because it is constant, and it does not depend on our strategy.  In EOQ—in mathematical problems—if you get fractional numbers, don’t worry about it.  Just put it as is.

## Ordering Cost

So let’s look at **ordering cost**.  If we order one time a year (9,600 units at the beginning of the year) we have one order and therefore, one ordering cost.  But if we put an order for 100 units each time, then we have 96 orders and 96 ordering costs.  In this case, it is 96 times 75 because each order costs us $75.  That is our ordering cost, which is very high.  Alternatively, if we order at the beginning of the year, then ordering cost is 1 times 9,600.  So if D is demand in units per year, Q is what we order each time per order.  Therefore, the number of orders per year is equal to what?  This is all we need.  This is what we order each time.  Therefore, to find out how many times we order:, we need divided by what we order each time.  So that would be D divided by Q.  Q is we order each time.  D is all we need per year.  We said we don’t have quantities gone.  It doesn’t matter what Q is.  Whatever we order, the price is still the same. If we order 100 units, each time the price is the same; the same as if we order 9,600 units at the beginning.  S is order cost per order.  This is the number of orders.  Order cost per order is S; therefore, annual ordering cost is S multiplied by the number of orders.

**Example 3.**

D = Demand in units / year

Q = Order Quantity in units / order

Number of orders / year =

S = Order Cost / Order

Annual Order Cost = S

**Example 3a.**

If demand for MacBooks are 15,000 units and the order quantity is 100 units, then the number of orders per year would be 150 units. We would multiply that by our ordering cost of $80 to get an annual cost of $12,000 per year.

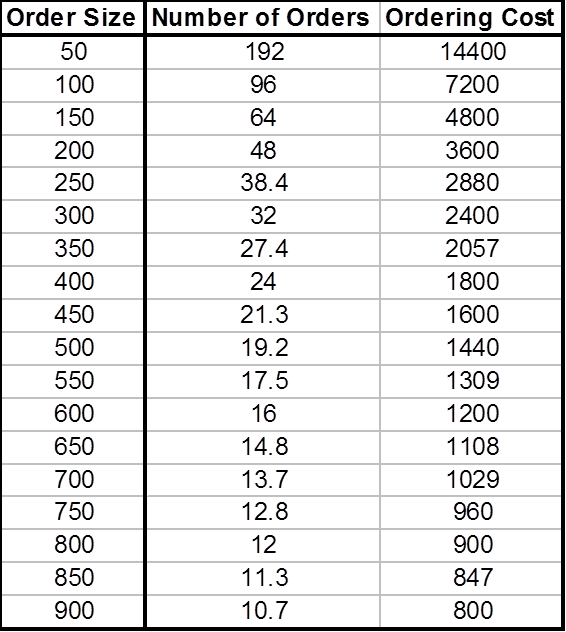
## Annual Ordering Cost

**Example 4.**

Therefore, if we order 50 units each time because we need 9,600 units per year, we will have 192 orders per year.  If we multiply 192 by 75, we get, 14,400, which is a large number.  If each time we order 100 units, we will have 9,600 divided by 100, which is 96 orders.  Then we multiply by 75 and we get 7,200.  So if we increase our ordering size and order 500 units each time, we will order 19.2 times.  19.2 times in this period means we order 20 times in this year, but 19.2 of it is for this year and 0.8 will go toward next year.  So if you don’t get integer numbers here, don’t worry.  In the first year, you have 27 orders and one other order, which 0.4 of it is for this year and 0.6 of it is for next year.

Don’t forget all these assumptions are valid for the next year and the next year after that.  This is the most basic inventory model.  We can develop all types of inventory models, but they are more difficult to develop, and we are sure you don’t want us to go through very difficult models.  So let’s first try to understand this basic model.  Then the pathways to others are not really difficult.  You only need interest and time.  So here if we order 800 units, we should order 12 times, and our cost is 900.

**Figure 4. Annual Ordering Cost Table**



So in summary, as order size goes up, number of orders goes down.  Because we have fixed cost per order, total ordering cost per year goes down.  This is number of orders per year.  This is ordering cost per year.

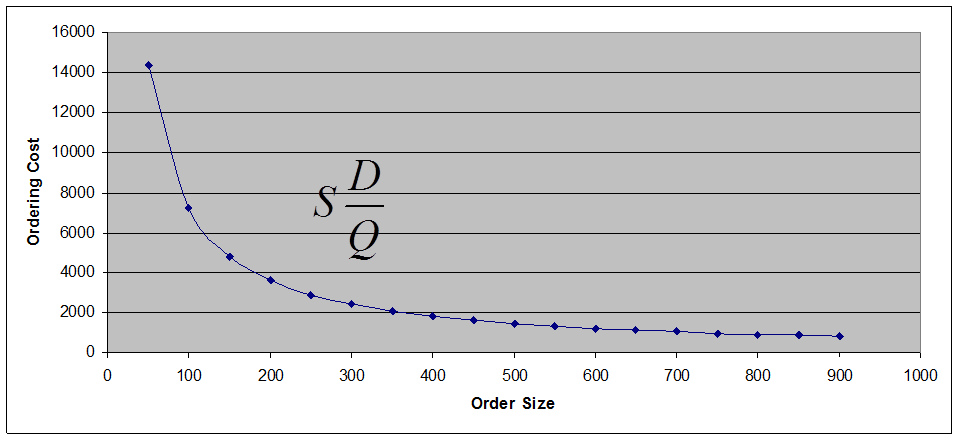
**Example 4a.**



**Example 5.**

The curve looks like this—S times D divided by Q looks like this:

**Figure 5: Ordering Cost per year Curve**



It decreases as ordering size increases.  So it is a beneficial to order as much as we can, and the best thing is to order all we need at the beginning of the year.  One order a year for 9,600 units at the cost of $75, but unfortunately the story has another side and that is inventory carrying cost.

**Example 5a.**

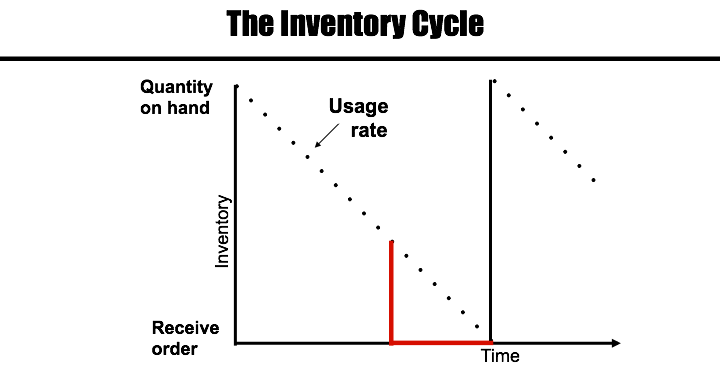
**Figure 5a: Annual Ordering Cost Chart**

## The Inventory Cycle

**Example 6.**

When we order, when we get this Q, whatever it is, if it is 100 you need or 9,600 at the beginning of the year, when we receive the order, this will be our quantity in hand.  Then we start using it with a constant rate.  So that is why it is a straight line coming down.  Usage rate is constant.  It comes down, down, down.  We are consuming it, consuming it, consuming it.  Then inventory reaches 0.  Before inventory reaches 0, suppose this is our lead time.  Suppose this is 4 days.  If this is our model, this is 1 day, 2 day, 3 day, 4 days.  We will go to see how much demand is in 4 days.  As so as soon as our inventory in hand reaches this much, we place an order.

**Figure 6. The Inventory Cycle**



As soon as our inventory on hand reaches demand in lead time, will see what we need in lead time, and then we place an order.  Because demand rate is fixed, usage rate is fixed.  And because lead time is fixed, exactly at the same time our inventory reaches 0, we will get the next order so we will not be out of stock.  Then we consume it.

Therefore, this is the story.  We get the inventory, and we consume it.  As soon as it gets to 0 we get the next batch.  We will then consume it again.  As soon as we get 0 inventory, we get the next batch.  This cycle will constantly repeat.

**Example 6a.**

For the Macbook example, we know that demand is 15,000 units annually. Therefore D = 15,000. Assuming that we are seeking to reduce ordering cost, the optimal order size will be 15,000 units bought at the beginning of the year to give us 15,000 units on hand when the order is received. Apple would use these products at a constant rate (shown as a straight line with a downward slope), which can be calculated as Quantity on Hand/ Working Days.

Usage Rate: 15,000/ 300 = 50 Macbooks per day

Let’s say we have an order size of 500 units. To satisfy demand, we order this quantity of Macbooks 30 times per year. In addition, we can figure out the amount of time 500 units will last during the period.

**Example 7.**

**Figure 7. The Inventory Cycle: Order Quantity**



 Suppose this is one year (see Figure 7), we have 4.2 orders per year.   Q, order quantity at the beginning of the period we get Q.  No matter what Q is 100, 9,600, or whatever, at the beginning of the period we get Q units.  At the end of the period we have 0 units; therefore, Q units in one course, 0 units in another course.  Your GPA is Q + 0 divided by 2, which equals to Q divided by 2. Therefore, if each time you get Q, and if that Q goes to 0 at the end of the period, this is like this throughout the period and your average inventory was Q divided by 2.  One day, you have a salary of Q dollars.  The other day, you have a salary of 0.  Your average salary during these two days is Q + 0 divided by 2, which is Q divided by 2.  Therefore, if each time we ordered Q and we don’t have any safety stock and our inventory reaches 0, we get the next order.

**Example 7a.**

At the beginning of a period we start with Q units. At the end of a period we end with 0 units. Therefore we can compute the average.

Q= Order Quantity

Average: (Q + 0)/2

0

Q

If our order size was 15000 units, we’d have 15,000 as our beginning inventory Q. To calculate the average we would use the formula Average= (Q+0)/2 or Q/2.

(15,000+ 0)/2 = 7500 units

If our order size is 500 units, our beginning inventory is 500. Average would be equal to 250.

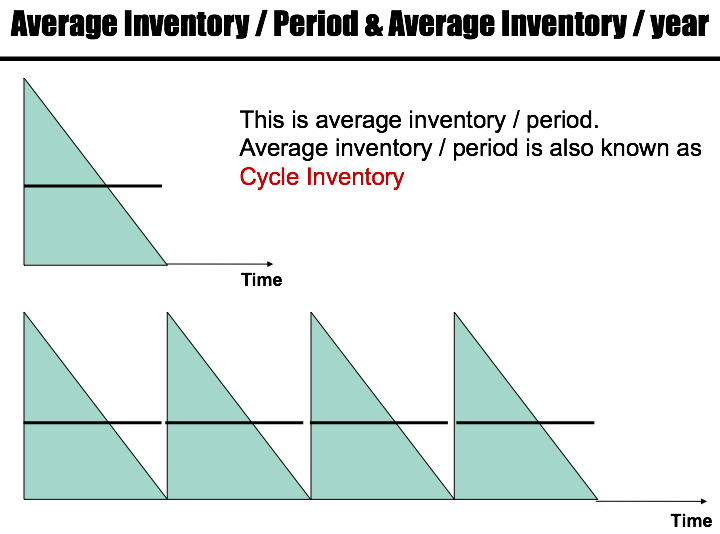
(500+0)/2 = 250 units.

## Average Inventory per Period and Average Inventory per Year

**Example 8.**

**Average inventory per period** and **average inventory per year**.  ,At the beginning of the period we have Q units, and at the end of the period we have 0 units.  Average inventory is Q + 0 divided by 2, Q divided by 2.  This is our average inventory.  Average inventory per periodis whatever we ordered divided by 2.  That is what we do have throughout the period.  Average inventory per period is also known as **cycle inventory**, but what is average inventory per year?

**Figure 8. Cycle Inventory**



Suppose this is a year (see Figure 8).  In this year, we have 4 periods.  Average inventory per period is Q divided by 2.  What is average inventory per year?  Let’s say we have 9,600 total demand.  We have ordered 4 times, that is 9,600 divided by 4, which is 2,400.  Each time we ordered 2,400.  It goes to 0, 2,400, 0, 2,400, 0, 2,400, and 0.  In general, Q, 0, Q, 0, Q, 0, Q, 0.  Average inventory per period is Q + 0 divided by 2, which is Q divided by 2.  In this period is Q divided by 2.  For each period of the four periods, Q is divided by 2; therefore, average inventory per year is multiplied by 4.  That is Q divided by 2, Q divided by 2, Q divided by 2, Q divided by 2; therefore, throughout the period it is Q divided by 2 throughout the year.  Each time we order Q throughout the year, we have Q divided by 2 average inventories.

**Example 8a.**

Cycle Inventory

Q

Time (Year)

Time: 1 Year

Number of Periods: 5

Demand: 15,000 units

Inventory per period: 15,000 units/ 5 periods = 3000 units

Cycle Inventory = Average per period

= (3000 + 0) / 2

= 1500 units

Average inventory per year: Cycle inventory x # of periods

= 1500 units x 5

= 7500 units

## Inventory Carrying Cost

**Example 9.**

So Q is equal to order quantity in units per order.  Average inventory per year, each time we order Q is Q divided by 2.  H is inventory **carrying cost** for one unit per year; therefore, for one unit per year we have H cost if we are carrying 2 units per year, then our carrying cost is H times Q divided by 2.

Q = Order quantity in units / order

Average inventory / year = Q/2

H = Inventory carrying cost / unit / year

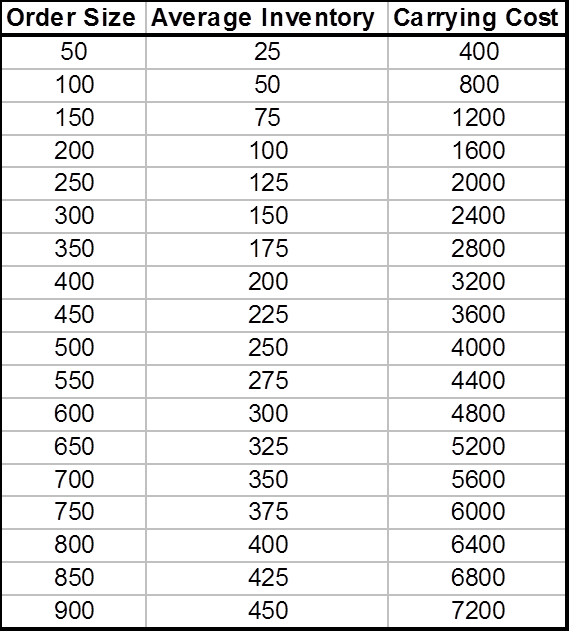
Annual Carrying Cost = H

If Q goes up, annual carrying cost goes up.  In the other scenario, cost structure we had, in ordering cost, when Q goes up, ordering cost comes down, but in this one, when Q goes up, cost also goes up.  We said we are system analysts.  We try to make the optimal solution for everybody and not for only inventory counting cost.  If we are talking about inventory counting costs, it tells us order as little as you can.  If we are talking only about ordering cost, it tells us order as much as you can.  These two solutions contradict each other, thus we should see what the benefit for the whole system is.

## Annual Carrying Cost

**Example 10.**

**Figure 9. Annual Carrying Cost Table**



In this carrying cost, if order size goes up, then average inventory is here.  This order size—half of it is average inventory. Order size 600, average inventory 300.  Order size 800, average inventory 400.  Order size 900, average inventory 450.  So this is average inventory.  We should multiply by 16, and we will get this number.  As we go up, inventory carrying cost goes up.   H time Q divided by 2 increases as Q increases.

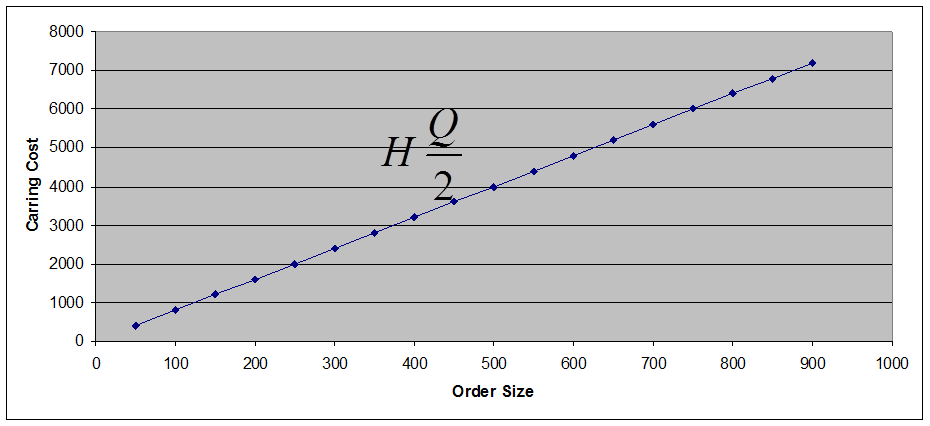
**Example 10a.**

**Figure 9a. Annual Carrying Cost Table for Macbooks**



**Example 11.**

**Figure 10. Annual Carrying Cost Curve**



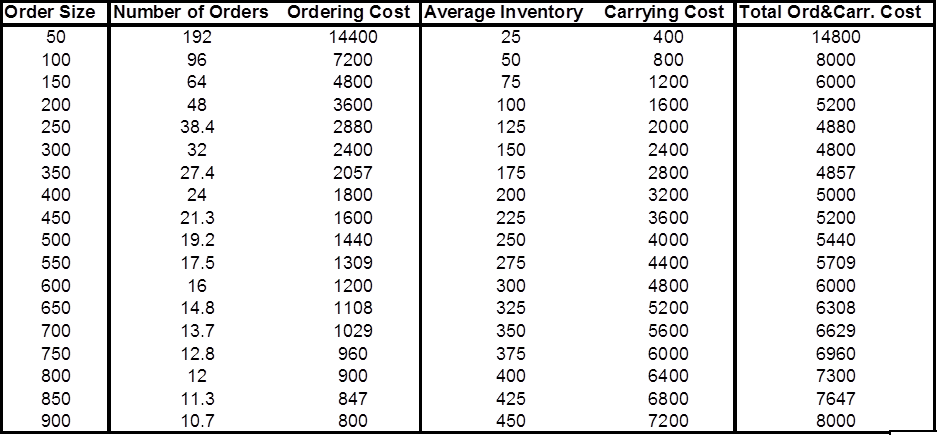
**Example 11a.**

**Figure 10a. Annual Carrying Cost Curve for Macbook**

## Total Cost

**Example 12.**

**Figure 11. Total Cost Table**



So we have order size, number of orders goes down as order size goes up, ordering cost goes down.  Average inventory goes up as order size goes up, carrying cost goes up, and total cost goes down and then goes up.  It’s like this.

**Example 12a.**

So let’s assume here that we have an annual demand for textbooks of 17,000 at the CSUN bookstore every year. D = 17,000. Then carrying cost of each textbook is $25 per unit, H = $25/per unit. The ordering cost per order is $60, S = $60/per order. Going off of this, CSUN has an order quantity of 2000 textbooks, Q = 1550. So with that, we can find out that they will need to put in 11 orders per year because Annual Orders Per Year = D/Q, which is 11 = 17,000/1550.

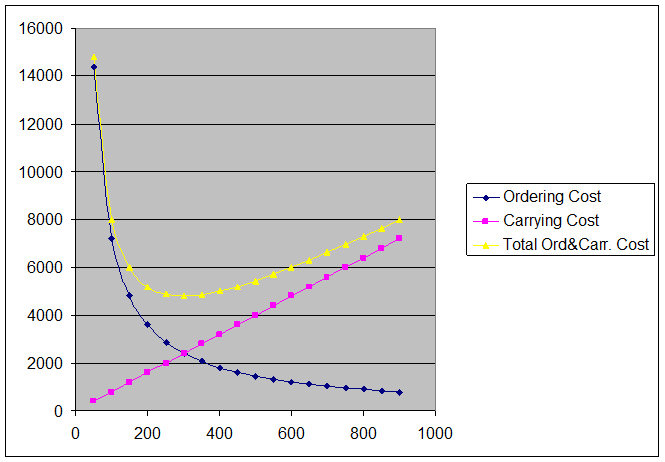
We then find our average inventory per year by dividing our quantity order by 2, Q/2 = 1550/2 = 775. We then we find the annual carrying cost by multiplying H, inventory carrying cost, with average inventory per year, H(Q/2) = $25 x 775 = $18,125. So with this information we can graph out the relationship between Ordering Cost and Carrying Cost to choose the right amount of Order Size for minimal total cost.

**Figure 11a. Total Cost Table**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Order Size** | **Number of Orders** | **Ordering Cost** | **Average Inventory** | **Carrying Cost** | **Total Ord. & Carr. Cost** |
| 100 | 170.0 | 10,200 | 50 | 1250 | 11,450 |
| 150 | 113.3 | 6,800 | 75 | 1875 | 8,675 |
| 250 | 68.0 | 4,080 | 125 | 3125 | 7,205 |
| 350 | 48.6 | 2,914 | 175 | 4375 | 7,289 |
| 450 | 37.8 | 2,267 | 225 | 5625 | 7,892 |
| 550 | 30.9 | 1,855 | 275 | 6875 | 8,730 |
| 650 | 26.2 | 1,569 | 325 | 8125 | 9,694 |
| 750 | 22.7 | 1,360 | 375 | 9375 | 10,735 |
| 850 | 20.0 | 1,200 | 425 | 10625 | 11,825 |
| 950 | 17.9 | 1,074 | 475 | 11875 | 12,949 |
| 1050 | 16.2 | 971 | 525 | 13125 | 14,096 |
| 1150 | 14.8 | 887 | 575 | 14375 | 15,262 |
| 1250 | 13.6 | 816 | 625 | 15625 | 16,441 |
| 1350 | 12.6 | 756 | 675 | 16875 | 17,631 |
| 1450 | 11.7 | 703 | 725 | 18125 | 18,828 |
| 1550 | 11.0 | 658 | 775 | 19375 | 20,033 |

**Example 13.**

**Figure 12. Total Cost Curve**



The pink line represents in Figure 12 carrying cost., the blue line represents ordering cost, and the yellow line represents total cost.  It reaches a minimum and then goes up.  The best strategy for us is to order this much.  This is what minimizes our total cost.  That is what we call economic order quantity (EOQ).

**Example 13a.**

**Figure 12a. Total Cost Curve**

Here we have ordering cost, carrying cost, and total cost.

## EOQ (Economic Order Quantity)

TC = (Q/2)H + (D/Q)S

EOQ is at the intersection of the two costs.

(Q/2)H = (D/Q)S

Q is the only unknown. If we solve it

**Total cost** is inventory carrying cost plus ordering cost.  You can use calculus computed derivative of TC with respect to Q and compute EOQ.  If you do that, you will find that EOQ is at the intersection of both costs.  EOQ is where ordering cost intersects with carrying cost.  Therefore, these two are equal at EOQ. We can solve this equation, and therefore we can find EOQ is equal to square root to 2D times S divided by H:

**Figure 13. EOQ Formula**



This is 2; this is demand per year, order cost per order, carrying cost per unit per year.  

## Back to the Original Questions

Now we go back to our previous example, as we are in a good position to solve this problem.  So we know that annual demand is 9,600.  Ordering cost is 75, carrying cost is 16. How much should we order each time to minimize total cost, which is to order EOQ?

## What is the Optimal Order Quantity?

**Example 14.**

EOQ 2DS divided by H.  we put the numbers over there.  D is 9,600, H is 16, S is 75.  EOQ is 300.

D = 9600, H = 16, S = 75

The best strategy is to order 300, not 100, not 9,600, 300.  That is what minimizes our overall cost: both carrying cost and ordering cost.  And as we said we don’t include purchasing cost. Purchasing cost plays no role because no matter if we order 100 units, 200 units, 300 units, or 9,600 units each time, we don’t get quantity discount.  We will buy it for the same price.

**Example 14a.**

So now we want to find out what our optimal order quantity would be with our given data using Economic Order Quantity.

D = 17000, H = 25, S = 60

So we find out that the best amount to be ordered is 286 books, which would result in the minimal overall cost between carrying and ordering cost.

## How Many Times Should We Order?

**Example 15.**

How many times should we order?  Annual demand is 9,600 units. How much do we order each time?  300.  Each time we order EOQ, therefore, how many times should we order?  We need 9,600.  Each time we order 300.  That is D divided by EOQ, which is 9,600 divided by 300, which is 32.

Annual demand for a product is 9600 units.

D = 9600

Economic Order Quantity is 300 units.

EOQ = 300

Each time we order EOQ.

How many times should we order per year?

D/EOQ

9600/300 = 32

**Example 15a.**

So now it is up to finding out what our amount of orders per year will be which will be easy:

Annual demand for books is 17000 units.

D = 17000

Economic Order Quantity is 286 units.

EOQ = 286

How many times should we order per year?

D/EOQ

17000/286 = 59.4, so 60 times.

## What is the Length of an Order Cycle?

**Example 16.**

 9,600 units are required for 288 days.  300 units are enough for how many days?  Each day, we need 9,600 divided by 288. .  Then if we divide 300 by that number, it gets us the number of days.  Alternatively, we can say 9,600 are for 288 days, and 300, we divided by 9,600, which gives us a fraction of a year.  Then we multiply by 288 and get the number.  Just think about it.  You don’t need to memorize any formula. You have 288 days.  Each time you order 300 units, all you need during that 288 days is 9,600.  How much is the length of each cycle?  It is 9 days.

Working Days = 288/year

9600 units are required for 288 days.

300 units are enough for how many days?

(300/9600) × (288) = 9 days

**Example 16a.**

So now we want to find out how much we a single order will last, or the length of an order cycle. This is relatively easy as well and can be found out with our given data. Assuming that an annual amount of work days is 300, we know that we need 17,000 books during those 300 days due to demand. So in order to find out how long an order lasts, you can divide the EOQ (which is 286) by the demand of books (17,000) and then multiply it by the amount of days in the year.

Working Days = 300/year

17000 units are required for 300 days.

286 units are enough for how many days?

(286/17000) × (300) = 5 days

## What is the Optimal Total Cost?

**Example 17.**

You can go through other different ways to come up with the same 9 days.  What is the optimal total cost?  Total cost formula is Q divided by 2 times H + D divided by Q times S.  The economic quantity is Q.  All other things unknown.  So, 300 divided by 2 times 16 plus 9,600 divided by 300, times 75 gives us 2,400 plus 2,400.  Remember one test to know that your computations are correct is these two costs should come out equal at EOQ. Before EOQ, ordering cost is higher.  After EOQ, carrying cost is higher.  At EOQ, they are equal. Total cost is 4,800.  This is the optimal policy that minimizes the total cost.

The total cost of any policy is computed as:

TC = (Q/2)H + (D/Q)S

The economic quantity is 300.

TC = (300/2)16 + (9600/300)75

TC = 2400 + 2400

TC = 4800

**Example 17a.**

So now we want to find the Optimal Total Cost that we derive from using EOQ. To find total cost, we use the following formula:

TC = (Q/2)H + (D/Q)S

We know that our EOQ was 286 units, so that is Q, and we know that our carrying cost is 25 per unit, demand is 17,000 units, and ordering cost is 60. Knowing all of this, we can apply it to the Total Cost equation as follows:

TC = (286/2)25 + (17000/286)60

TC = 3575 + 3566.4

TC = 7142

## Centura Health Hospital

**Example 18.**

Now, we solve another problem. You have Centura Hospital.  Demand in this hospital is 31,200 units per year.  In our current strategy, we order 6,000 units each time.  Ordering cost is $130, carrying cost is $0.90.  A year is 52 weeks.  What is the average inventory or cycle inventory?  Here, Q is 6,000.  It reaches 0 and goes up to 6,000. Therefore, average inventory is Q divided by 2, which is 3,000.

**Average inventory** = Q/2 = 6000/2 = 3000

What is total annual carrying cost?  This is average inventory per period and average inventory per year.  Therefore, on average we have 3,000 units in inventory and each one costing $0.90; therefore, carrying cost per units times average inventory is 2,700.  

**Carrying cost** = H(Q/2) = 0.9×3000=2700

How many times do we order?  Each time we order 6,000.  All we need is 31,200.  31,200 divided by 6 means 5.2.  We place 6 orders.  5 orders are completely utilized in this period and the 6th one, 20% of it is utilized this year and the remainder is utilized next year.  But we simply stay with this 5.2.  The number of order per year is 5.2

31200/6000 = 5.2

What is the total annual ordering cost?  You order 5.2 times, ordering cost per each time you order is $130.  It is S times D divided by Q, which is 130 times 5.2, which equals $676.

**Total ordering cost** = S (D/Q)

**Ordering cost** = 130 (5.2) = $676

**Example 18a.**

Assume you are running a restaurant that serves jumbo shrimp to its customers; they are essentially world renown for how great they taste. But the secret lies in the special spice that is used to marinate the shrimp in. Because of this essential ingredient the demand that the restaurant has for the spice is 72,000 crates (units) per year. The restaurant orders 9000 units each time at an ordering cost of $900 and carrying cost of $25. The year is 52 weeks long. So now you are asked to find the **Average Inventory**, **Annual Carrying Cost**, **Total Ordering Cost**, and the **EOQ** with this given data.

Q = 9000 and Avg. Inventory is Q/2, so…

**Average Inventory** =Q/2 = 9000/2 = 4500

So now that we have the average inventory, we can find the total annual carrying cost; which is the **Average Inventory** multiplied by the **Carrying Cost**. Our data states that H = 25 and Avg. Inventory is 4500, so…

**Annual Carrying Cost** = **H (**Q/2) = 25(4500) = 112,500

Now we want to find out the total annual ordering cost, which consists of using the demand, amount being ordered, and of course, the ordering cost of each unit. First we need to find the total amount of times we will be ordering given our **Quantity Ordered** (Q) and **Demand** (D).

Q = 9000, D = 72,000 so… D/Q = (72,000/9000) = 8 orders per year

Now we can find the Annual Total Ordering cost since we know that our ordering cost per order is $900 and we order 8 times per year:

**Total Annual Ordering Cost** = **S (**D/Q) = 900(8) = 7200

So now we see what our cost is when we order 9000 units, but we don’t know if this is the optimum amount we can order in order to have the lowest carrying and ordering cost! So in order to do this we use the EOQ formula to see if 9000 is too much or too little.

**Economic Order Quantity** =

D = 72,000, H = 25, S = 900

Given this new ordering quantity, we can test out and see the new carrying and ordering costs the restaurant would have:

**Annual Carrying Cost =** H (Q/2) = 25(1138.4) = **28,460**

**Annual Ordering Cost =** S (D/Q) = 900(72,000/2276.8) = **28,460.9**

Now we can see how much more money is saved by comparing the differences in annual carrying and ordering costs between the Q of 9000 and 2276.8.

**Re-Order Point**

**Introduction**

In the previous chapter, we discussed how much to order. For instance, when there is no price discount, we order the economic order quantity (EOQ). Similarly, when there is a price discount we may order EOQ, or we may order more than EOQ. As long as cost is our main concern, we never order less than EOQ; however, if we are concerned with flow time rather than production cost, we may order less than EOQ. Why? In the previous chapter, we addressed how much to order, and now we are interested in determining when to order. When there is a periodic inventory system we always order at the start of the period, e.g. start (or end) of every week, start (or end) of every month. However, in perpetual (as opposed to periodic) inventory system, reorder point is defined in terms of quantity, not in terms of time. Instead of ordering whenever the end of the period is reached, we order when inventory on hand reaches a specific level. This defines the Re-Order-Point (ROP). In the previous chapter, we discussed how much to order (EOQ), in this chapter we discuss when to order (ROP).

In a periodic inventory system, ROP is the start (or end) of the period. However, in a perpetual inventory system or in continuous reviewing inventory system, reorder point is when inventory on hand drops to a predetermined quantity. In the EOQ model, we assume that there is no variation in demand, i.e. demand is known, constant, and it will remain constant. If there was no variation in demand and demand was solely constant, then ROP is when inventory on hand is equal to average demand during lead-time. If demand is constant, then demand will always be a specific number. Therefore, average demand during lead-time is equal to actual demand during lead-time. In contrast, if demand is a variable, average demand during lead-time differs from actual demand during the lead time. But if the demand is constant with no variation, demand during lead-time is average demand during the lead-time. Reorder point is a point when inventory on hand is equal to demand during lead-time.

**Lead-time** is the time interval from the moment an order is placed until it is received. If demand is fixed, we know what actual demand is during lead-time. But if demand is variable, we know average demand during lead-time and its variance. The greater the variability of demand during lead-time, the greater the need for additional inventory. These additional inventories help reduce the risk of shortages. The additional inventory is known as **safety stock**. Therefore, the more variability in the demand during lead time, the more safety stock that is needed. When variability exists in the demand or lead-time, actual demand during lead time will be different than average demand during lead time. It may be greater or less than the average demand during lead time. However, if there is no variation in demand during the lead-time, then demand during the lead-time is equal to average demand during the lead-time.

**Understanding ROP**

In the case of variability, we assume demand during lead-time follows a normal distribution. If I order at a point when inventory on hand is equal to average demand during the lead time, because normal distribution is a symmetric bell-shaped curve, the probability of being greater than average is 50% and probability of being less than average is 50%. Therefore, if my reorder point is at the point when inventory on hand is equal to average demand during the lead time, there is 50% probability that demand during lead time exceeds the average demand during the lead time. In a practical environment, if there is a 50% probability, it is likely that a store will not have a product requested by a customer. Hence, you do not have the product to satisfy the product demand. You will lose the profit that you could have made through the sale. However, if the client comes to the store once or twice after this occurrence and is told that the product is not available, this client will most likely shop at another store. Now the client’s store loyalty is lost. For this reason, usually retailers, manufacturers, and distributors don’t want a 50% probability of shortage. They want 1%, 5%, or 10%. Therefore, we would like to have a probability of 99%, 95%, or 90% to be able to fulfill your request. Therefore, you should order when inventory on hand is equal to average demand during lead-time. Safety stock is what you have to add to the average demand during the lead time. Thus, you order EOQ (or any other order quantity) when inventory on the hand is equal to average demand during lead time plus safety stock. When you order at this level of inventory on hand, it will reduce the probability of stock out during the lead-time. Reorder point in perpetual inventory control is the inventory level equal to the average demand during lead time plus a safety stock.

**Solving ROP- Example 1**

Suppose average demand for an inventory item is 200 units per day with a lead time of 3 days. There is no variation in demand, and no variation in lead time. Therefore, demand is constant and is completely known in advance. Hence, since there is no need for any safety stock, we set it equal to 0. We order at the point when inventory on hand is equal to the demand during the lead-time and when demand during the lead-time is known. Therefore, reorder point is equal to 3 x 200. Hence, whenever inventory drops to 600, we place an order. The demand during those 3 days is exactly 600 units. We will receive the order exactly in 3 days.

**Solving ROP- Example 2**

Average demand for an inventory item is 200 units per day. Lead time is 3 days, and safety stock is 100 units. What is the reorder point? In a perpetual inventory system the reorder point is the point at which inventory on hand is equal to the average demand during lead time plus safety stock. So lead time is 3 days, average demand during lead time is 200. 3 x 200=600 +100 = 700. When inventory level drops to 700 units, we place an order. Since there is variation in demand during the next three days, our demand is not 600. It may be 500, 400, 429, 678, 750 and so forth. Now that we have gone over two simple examples of ROP, the rest of the chapter will go further into depth.

**No Variation in Demand: Linear Relationship**

Figure 1 shows the inventory on hand. The y-axis represents inventory and the x-axis represents time. As we start consuming that inventory, it will gradually go down. Because demand is constant, it goes down in a linear fashion. If lead-time is three days, it will be 3 days until we expect inventory to reach 0. If the relationship between inventory and time is linear, we can assume that there is no variation in demand. Because we assume that lead time is 3 days, we can then draw a perpendicular line and find out whenever inventory reached this level or when an order should be placed. When we place an order at this level of inventory, after this many days we will receive the next order. No variation in demand, no variation in lead time.

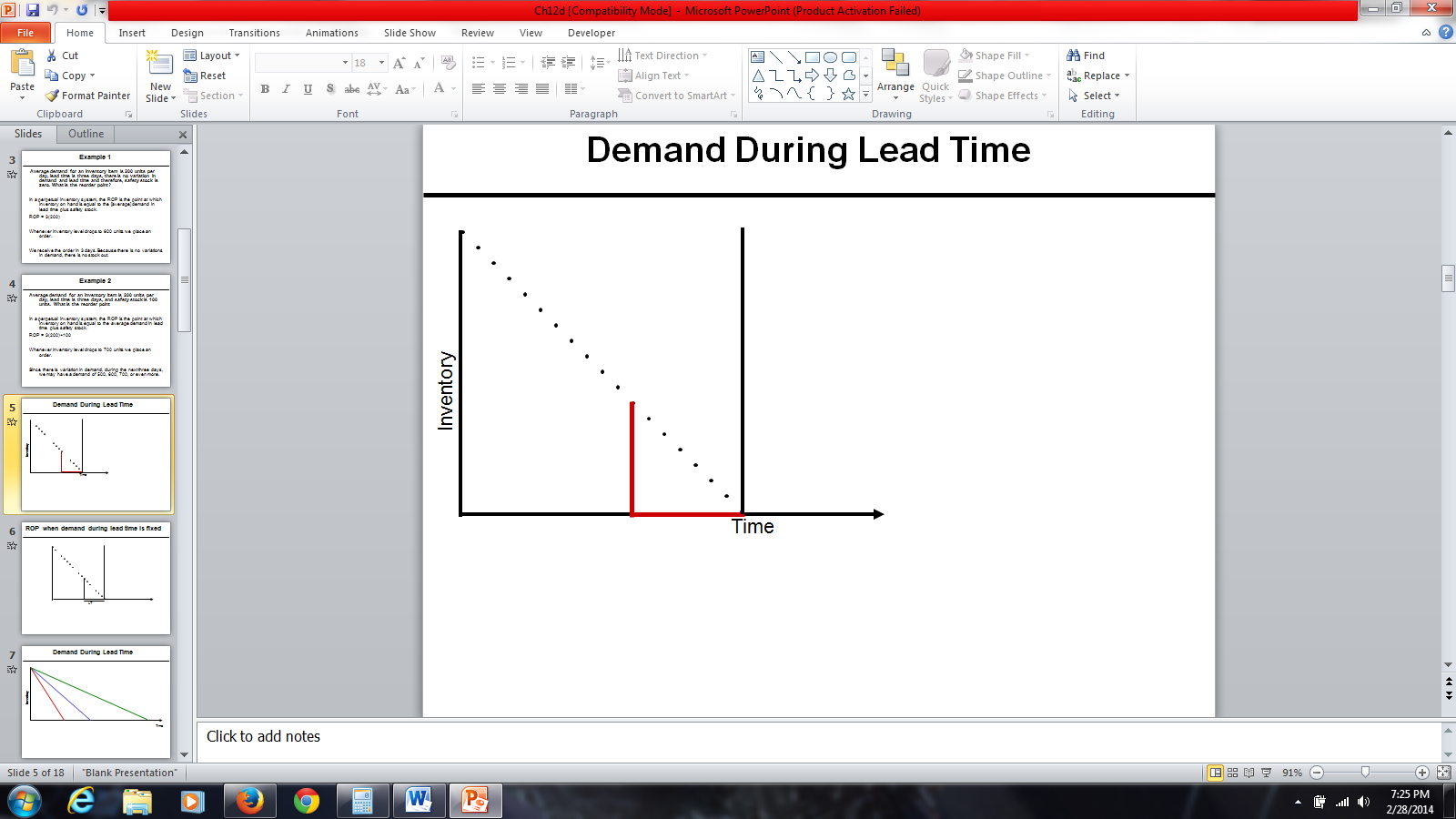


Figure 1: Demand During Lead Time

**Demand During Lead Time is Fixed**

Figure 2 represents lead time. This is the point that I expected the demand to become equal to 0. When inventory on hand is 0, we place an order. Inventory goes down at a fixed pace, and is represented by the linear downward line on the graph in Figure 2.

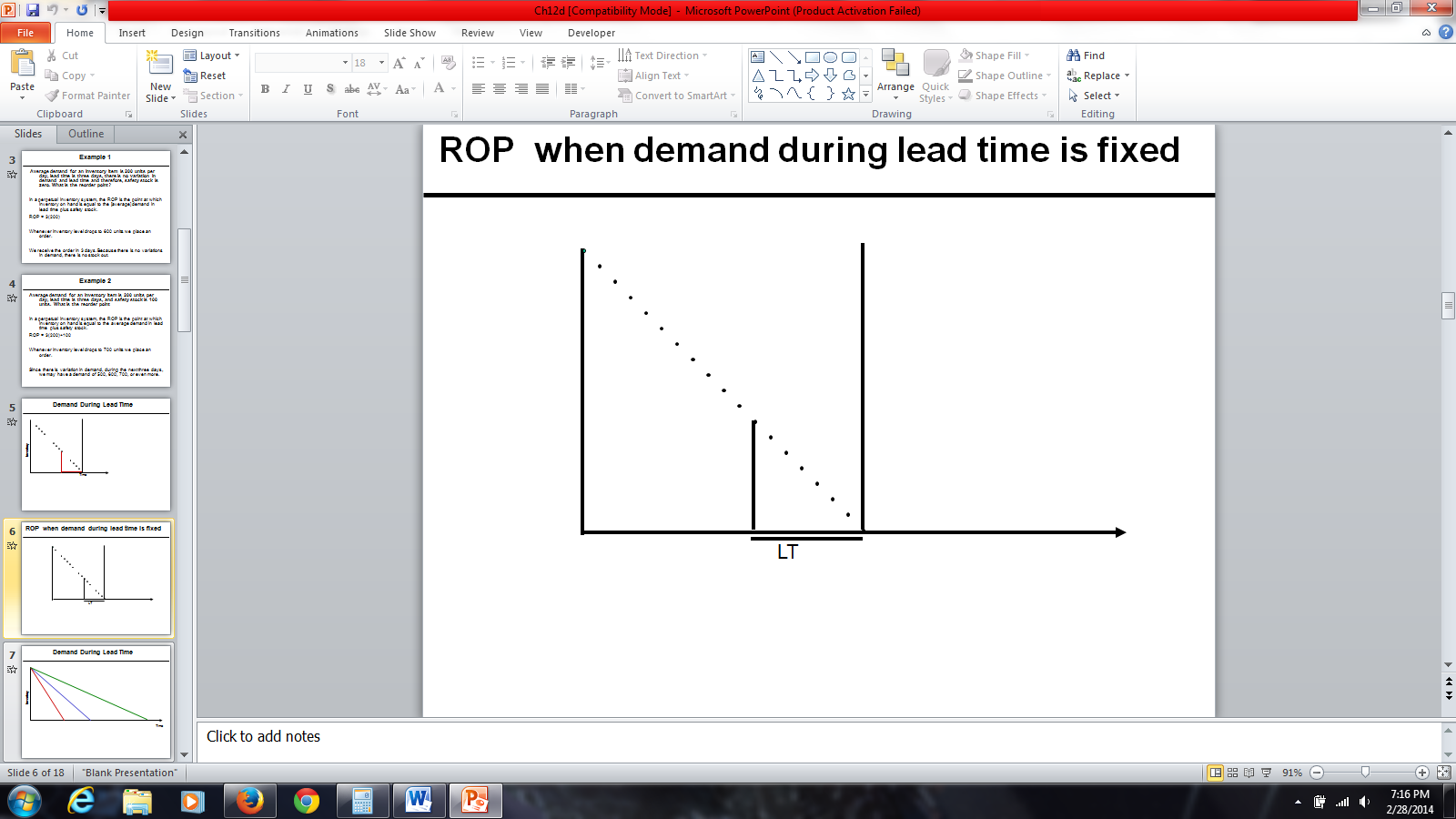


Figure 2: ROP When Demand During Lead Time is Fixed

Figure 3 represents demand during lead time. This graph shows three different lead times for the same amount of inventory. It is possible to consume the inventory at a faster rate. Suppose rather than consuming it at the speed represented by the red line, perhaps we can consume it at the speed of the green line. Thus, we would be reaching the reorder point earlier than the previous one. Alternatively, we may consume it with a slower rate. In all three cases, whenever inventory reaches this level we will put an order. Suppose rather than consuming inventory at the speed of the green line, we want to consume it at a speed of the blue line. Reorder point is determined based on inventory on hand. Inventory on hand is determined based on consumption rate. At a faster rate, inventory on hand reaches a specific level sooner. At a slower rate, inventory on hand reaches a specific level deferred in to the future.

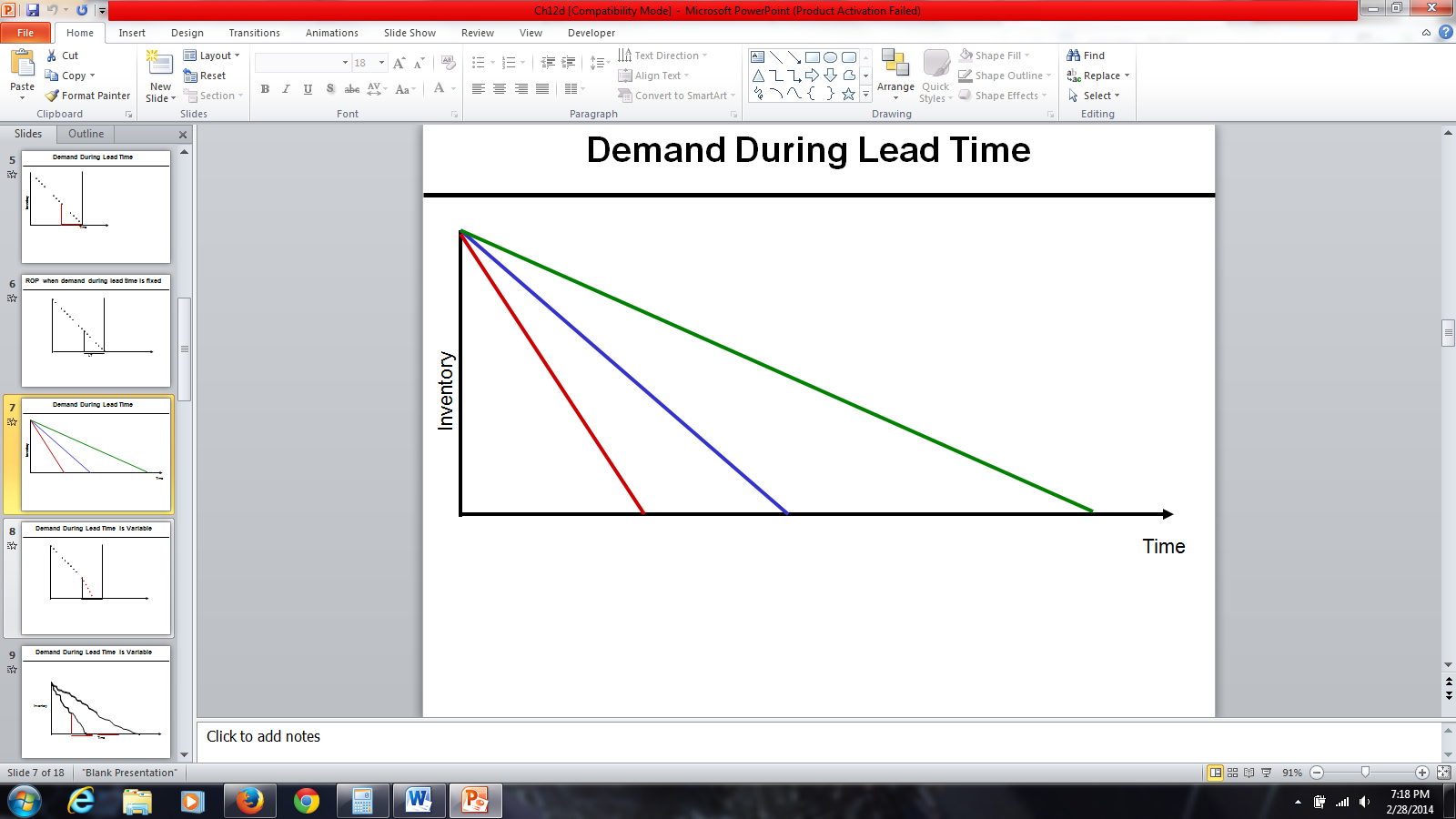


Figure 3: Demand During Lead Time

Figure 4 represents when demand during lead time is variable. Suppose we have a certain quantity of inventory and have reached reorder point. We expect to consume the inventory at the same linear rate; therefore, exactly at the point that inventory on hand reaches 0, we will get the next order. But who can guarantee that customers will come to the store based on the rate we have assumed? No one can guarantee that. Therefore, it is possible that the product is consumed at a faster rate. The red line represents the consumption at the faster rate. In that case, inventory on hand reached 0 at this point and we are left without inventory for the remaining days. Any customer who comes to the store asking for the product will not be able to purchase it because there will be no inventory on hand for the product in question. Thus, the customer will have to wait until the inventory arrives.

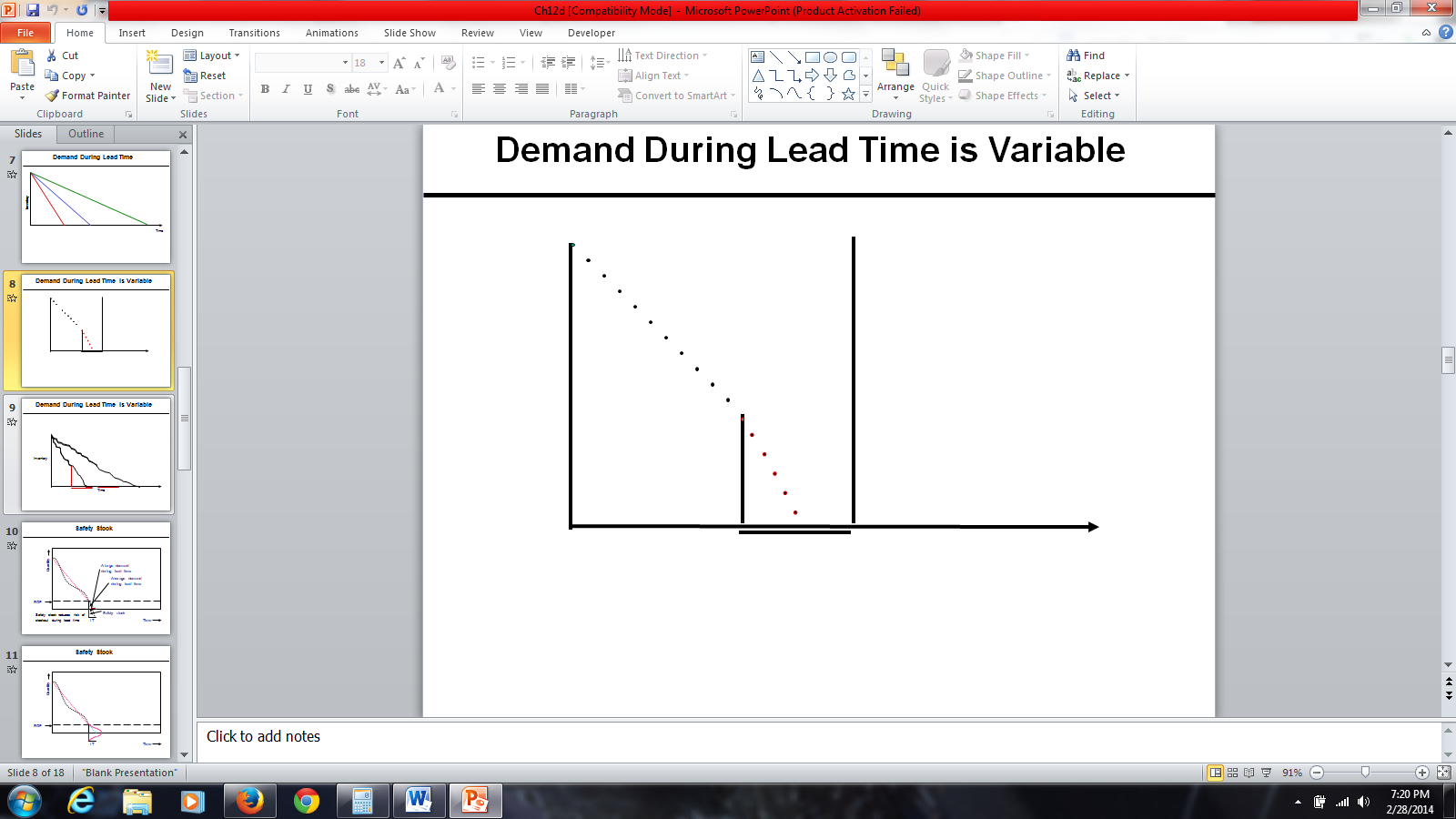


Figure 4: Demand During Lead Time is Variable

In general, customers do not come to the store to ask for a product in a linear fashion. There might be more or less customers, depending on the day. In contrast to the basic assumptions of the inventory model that demand is known and constant, suppose demand is unknown. If only average demand is known, we will need a computation for EOQ based on average demand. However, some days we may have more orders and other days we may not. In Figure 5, the first linear has a steeper slope, thus inventory drops quickly. In the second linear, the slope is less steep than the first, thus inventory drops slowly. As soon as inventory on hand reaches to the level in which we think inventory on hand is enough for average inventory during lead time, we will order. We expect to receive the order at this point, and consume this inventory at this rate. It is possible that we consume it at a faster rate. In that case, we will be out of stock for a certain amount of time. Suppose we reach the reorder point, but the demand after that was quite high. Therefore, before we get the next order, the entire inventory on hand was consumed. Alternatively, the demand may be less than the average demand that we have assumed. We expect to consume the inventory in a linear fashion, and we expect that by the next order, inventory on hand is equal to 0. But if we consume the inventory at the lower rate, we would still have inventory on hand by the arrival of the next order.

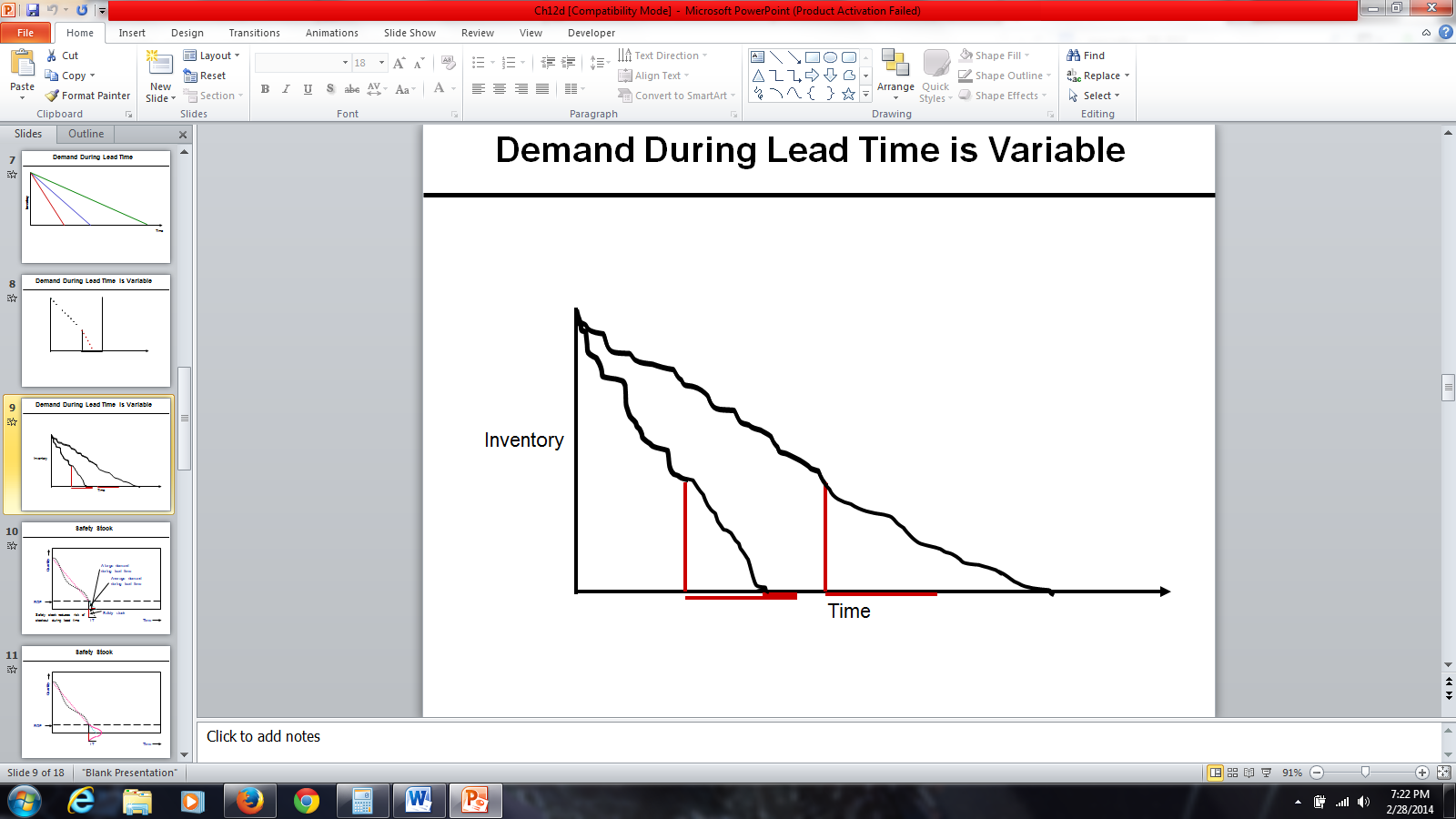


Figure 5: Demand During Lead Time is Variable

**Safety Stock**

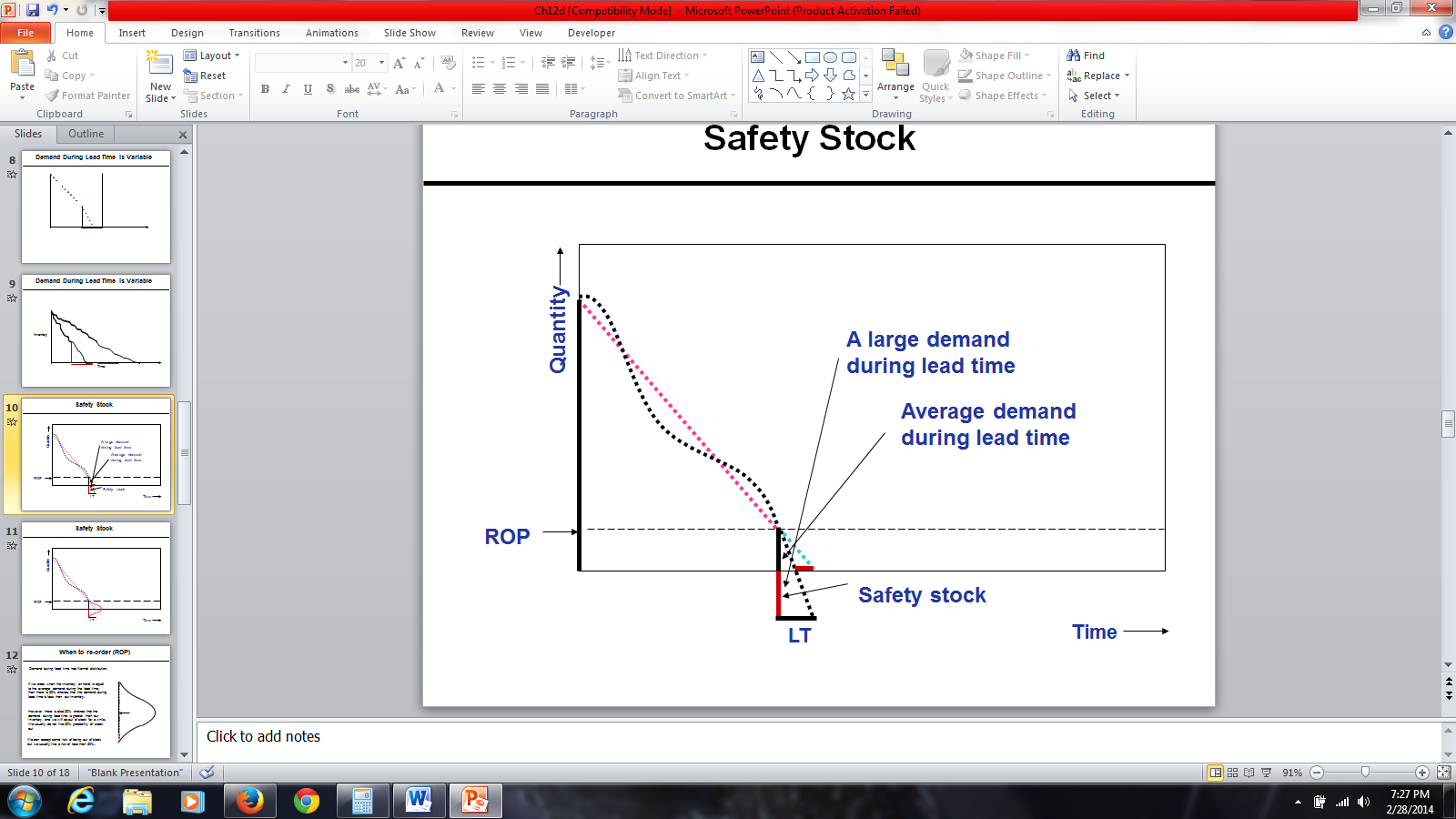


Figure 6: Safety Stock

Why do we need safety stock? We need safety stock because we want to reduce the probability of being short of stock. The y-axis shows quantity (inventory on hand). The x-axis represents time. This is the order that I have received; it is either EOQ or greater than EOQ. It doesn’t contain anything about when to order. It is either EOQ, or if we have a price discount, it is possible that it is greater than EOQ to take advantage of price discount. How much to order and when to order are completely different concepts.

We consume inventory in a non-linear fashion (a curve), which sometimes is steep or mild. To find the slope of this line, we will use average inventory. The vertical axis displays quantity and the horizontal axis displays time. In this pictorial representation, we want to show how inventory is consumed over time. It is either EOQ or if there is a price discount, it may be something more than EOQ to take advantage of price discount. In Figure 6, we assume that it is consumed in a linear fashion. It may be consumed either at a rapid rate or at a gradual rate. The slope of the line is equal to average demand. So we assume that it is consumed in a linear way with this slope, but in reality it may be consumed in a rate lower than that rate or in a rate higher than that rate. So while actual demand is non-linear, we assume it is linear because we want to model it and get some conclusions, some guidelines from that model. Don’t forget that whenever we want to model a portion, we need to simplify it. Then based on that simplifications, we would be able to develop a mathematical formulation and collect information about the simplified version of the real world. However, later we need to incorporate at least some parts of complexities of real world into our simple model.

Whenever inventory on hand reaches average demand during lead time, we place an order. The reorder point is a point in which inventory on hand is equal to average demand during the lead time. It is possible that during lead time the demand is quite higher than what is expected, which depends on the slope of the line. Because demand during lead time was quite higher than the average demand, at this point inventory reaches 0. Therefore, for this much of time, which is almost half of the lead time, we don’t have any product to give to customers. Therefore, in order to reduce this probability, we need a safety stock. Safety stock reduces risk of stock out during the lead time. That is why we add to average inventory during lead time.

Figure 7 represents an illustration of safety stock with inventory on hand when we received the order along with the reorder point when we reorder. The reorder point is equal to average demand during lead time. We expect to consume this inventory at the average rate, and therefore we may go back this many days. Demand during the lead time is not constant. Demand during lead time has normal distribution. That amount is only average of the demand during lead time. Actual demand during lead time could be any of these numbers. If it is more than average, probability of stock out is 50%. If we want probability of stock out not to be 50%, but to be a smaller amount, then we will add safety stock. We order not at the point when inventory on hand is equal to 0 but at the point when inventory on hand is equal to average demand during lead time.

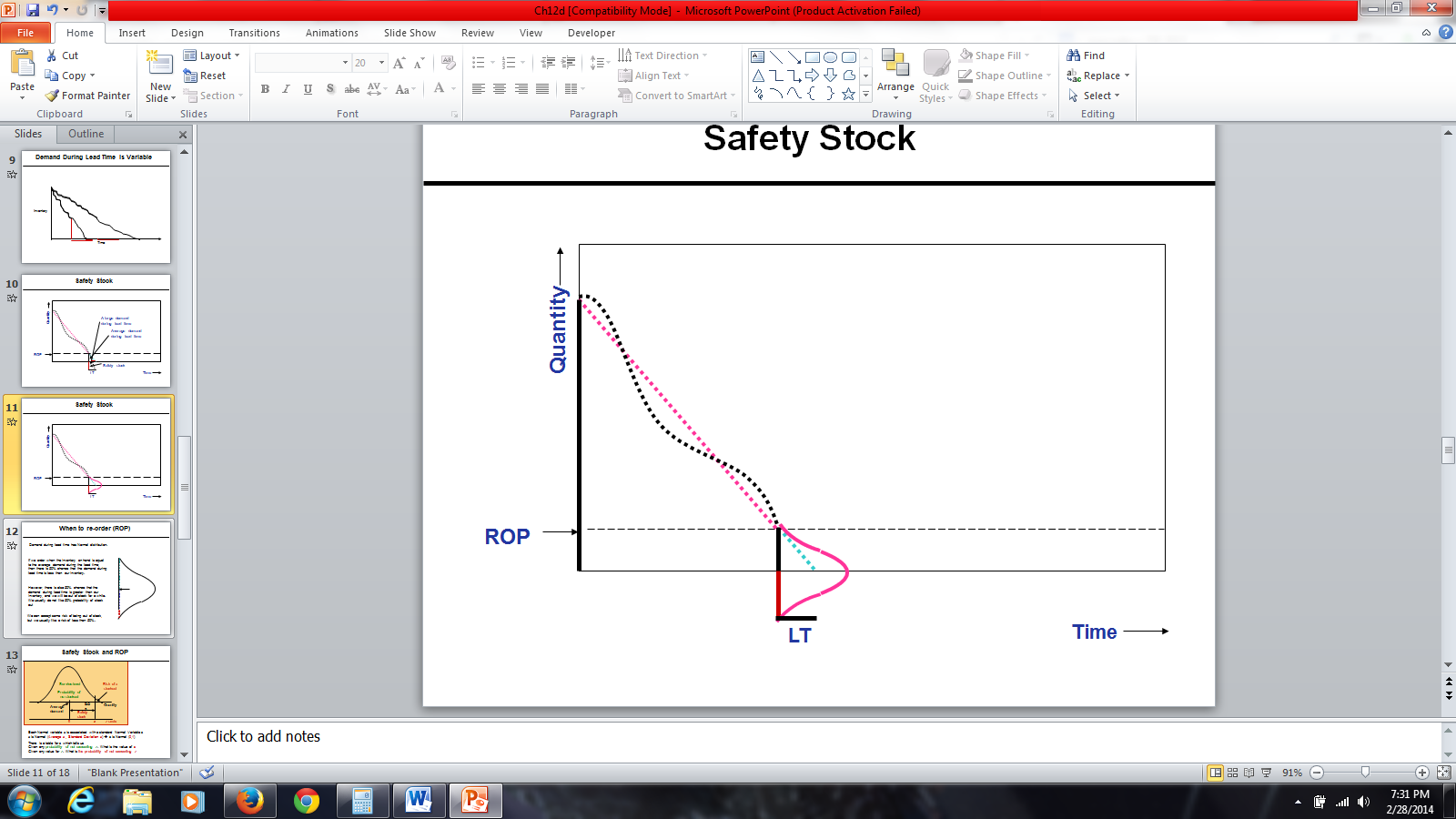


Figure 7: Safety Stock

Figure 8 illustrates a simple graph of reorder point. We assume demand during lead time has a normal distribution. This middle arrow is the average. There is 50% probability that demand is less than this average, and there is 50% probability that demand is greater than this average. Therefore, if we order at the point when inventory on hand is equal to average demand during the lead time, there is 50% probability that the demand during lead time is less than this average. It is also possible that demand during lead time is greater than average demand during lead time. There is 50% possibility that demand is greater than the average demand. Demand could be greater than the average demand. In all of these cases, we want to be able to satisfy the demand. We don’t want the probability of facing shortage to be 50%. We want it to be less than 50 percent. With 100% probability, we can satisfy all the demand. Therefore, we are ready to accept some risk to be out of stock. We want that risk probability that we will not be able to satisfy demand to be 1%, 5 %, 10%, but not 50%.

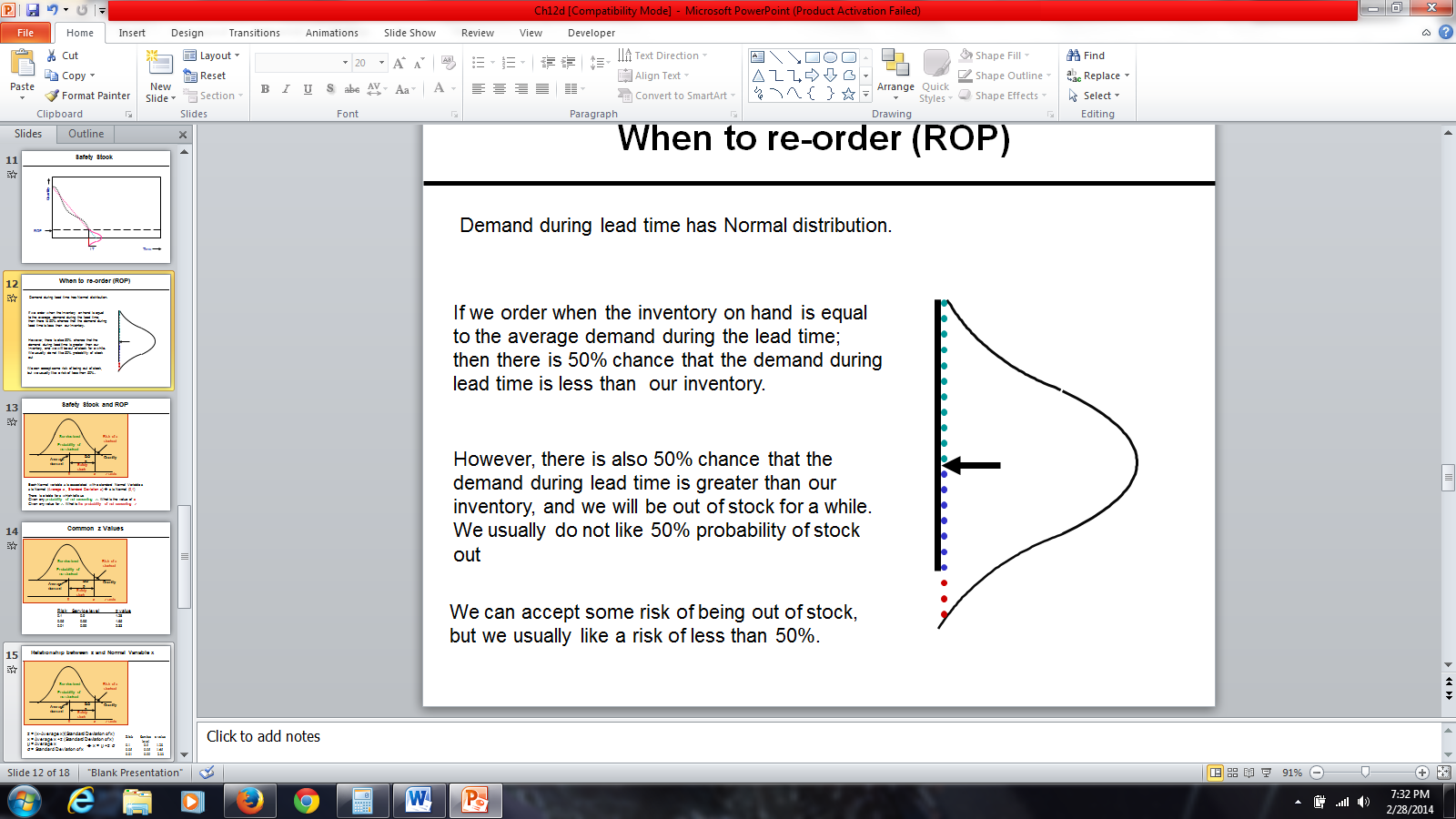


Figure 8: Re-Order Point

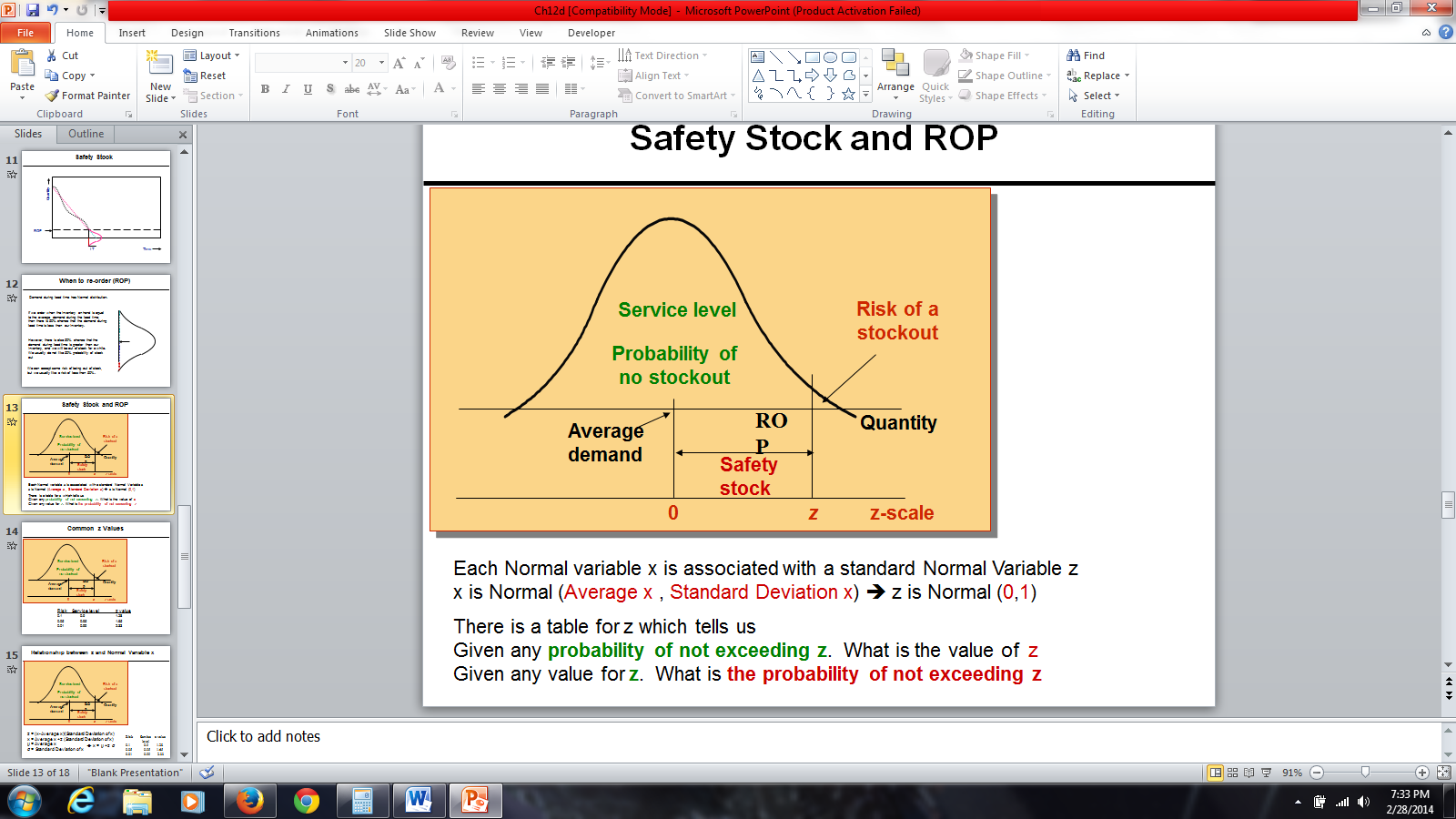


Figure 9: Safety Stock and ROP

Figure 9 represents a normal distribution. This is average demand during lead time. We are ready to accept this much risk to be out of stock. This risk could be 1%, 5%, 10%, depending on managerial policies. Probability of no stock out is quite high. Risk or probability of stock out is quite low compared to this probability. This is what we call a service level. If probability of risk is 5 %, service level is 95%. If probability of risk is 1%, service level is 99%. If probability of risk is 10%, service level is 90%. If management says with 90% probability we want to be able to satisfy the demand during lead time, we are ready to accept the risk of 10%. This is the safety stock, which is added to average inventory to ensure that this probability is less than 50 %.

Demand during lead time is a normal variable X. What do you know about this variable? It’s normal, so its distribution is bell-shaped. When we say it’s a normal variable X, we should also know its average, which we could say is a hundred units during lead time, and we should have its standard deviation. When we talk about a normal variable, we know it’s normal and has a specific mean, which could be anything, and has a specific standard deviation. This normal variable, this random normal variable has two attributes, mean, and a standard deviation, which could be any number. A normal variable X is associated with a standard normal variable Z. What is the attributes of a normal variable X? It has a mean, an average, and a standard deviation. But those two attributes in standard normal variable Z are 0 and 1, therefore, standard normal variable Z has mean of 0, and a standard deviation of 1. Given any probability of not exceeding a specific Z, that table will give us the value of Z. But this Z is a standard normal variable. It is not the variable X that we are looking for and it’s specific average and specific standard deviation.

So, if we know the service level or risk, which is one service level, then we can compute a specific Z, which is related to that risk and that service level. If we know a specific Z, we can find out what is the service level for that specific Z and what is the risk for that specific Z. The table in Figure 10 gives us service level and risk. While we are talking about a general normal variable with a specific average and a specific standard deviation, there is a relationship which connects that general normal variable X with a standard normal variable Z. We need this mapping because we don’t have a table of probabilities for all types of normal variables, such as different averages and different standard deviations. However, we have that table for standard normal variable which has mean of 0 and a standard deviation of 1. Also, we have a relationship which transforms any normal variable with any mean and any standard deviation into standard normal variable.

So if risk is 10%, service level is 1 minus 10%, which is 90%, and if we go to normal table, it will tell us Z value is 1.28. The Z value for the probability of service level is 90% and probability of risk is 10%, the specific Z is 1.28. If the average demand during lead time was 0, standard deviation of lead time was 1, inventory on hand is 1.28, then there is 90% probability that we could satisfy all the demand. There is 10% probability that we cannot satisfy all the demand. We usually don’t have Z distribution for my demand during the time. We have normal distribution, which can easily transform into Z distribution or standard normal distribution. The standard normal distribution table tells us if risk is 5% and service level is 95%, then your Z value is 1.65. If risk is 1% and service level is 99%, then the Z value is 2.33. So it simply tells us that the Z value for which the probability of being greater than Z is 1%, and the probability of being less than or equal to Z is 99%.

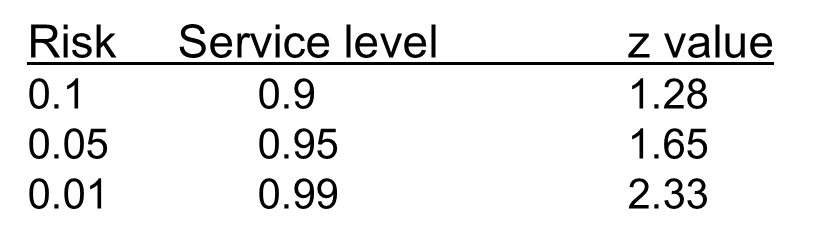


Figure 10: Common z Values

If we take any normal variable X, and subtract it from its average, and divide it by its standard deviation, we get Z value. We can find what Z is for each of the different percentages of risk. Therefore, we can find average demand during lead time, the standard deviation of demand during lead time. Hence we can find the reorder point. X is equal to average X + Z times the standard deviation of X. Mu is average of X. Sigma is the standard deviation of X. X is equal to Mu + Z times sigma. So if we know Z, we can transform it into X. If we have X, we can transform it into Z. Given any service level or any risk, we can easily compute Z value.

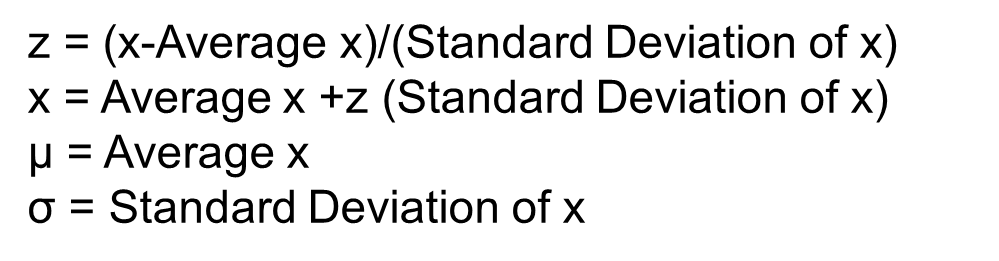


Figure 11: Relationship between z and Normal Variable x

LTD is lead time demand. Reorder point is average lead time demand + Z times the standard deviation of lead time demand. Reorder point is equal to average lead time demand + Z times the standard deviation of lead time demand, which is what we call safety stock. If we only reorder at a point when inventory on hand is equal to average lead time demand, then probability of stock out is 50%. But if we add safety stock to it and reorder at the point where the inventory on hand is equal to average lead time demand + safety stock, then the probability of stockout would be quite low. It could be 10 %, 5 %, or 1%. Therefore, if you want probability of stockout to be 10%, you will find 1.28, which is Z. Then you would take that Z and you multiply it by the standard deviation of lead time demand, and that would be safety stock. Then you add it to average lead time demand, and that would give you the reorder point. If you order at that reorder point, probability of stockout would be 10%.

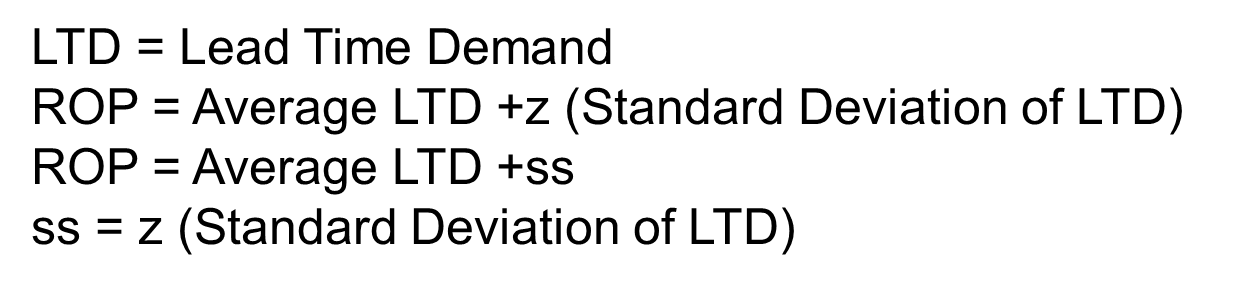


Figure 12: Relationship between z and Normal Variable ROP

**Practice Set #1**

1. **Demand for Honda cars in one week consist of 1400 units. (7 days). What is the average demand? Lead time is 2 days; there is no variation in demand and lead time. Safety stock is 0. What is the reorder point?**

**Solution: demand/n (1400/7)=200.**

**ROP= 2(200)=400.**

**Whenever inventory level drops to 200 units, we place an order. We should receive the order in 2 days.**

1. **Average demand for an inventory car part (tires) is 400 units per day, lead time is 2 days and safety stock is 50 units. What is the reorder point?**

**Solution: ROP=lead time (average demand)+safety stock.**

**Therefore, 2(400)+50 = 850.**

**Whenever inventory level drops to 850 units, we place an order. Since there is variation in demand, during the next two days we may have a demand of 650, 750, 850, or even more.**

**True or False?**

1. **Safety stock reduces risk of stock out during lead time. (True)**
2. **If we order when inventory on hand is less than the average demand during the lead time, then there is 50% chance that the demand during time is less than out inventory (False)**

**Practice Set #2**

1. **At Bed Bath & Beyond, customers purchase on average 21,000 Magic Bullets machines per week. The lead time is 20 days. Assuming zero safety stock. Compute ROP.**

**ISafety=0**

**L=10**

**R= (21,000U units/week) X (1week/7days)=3000 units/day**

**Solution: ROP= LTD+Isafety**

**ROP=LTD+0**

**ROP=L x R**

**ROP= 20 x 3000=60,000**

1. **The following week, the Magic Bullet gained media popularity. Customers at Bed Bath & Beyond demanded on average 105,000 Magic Bullets per week. The lead time increased to a month during the month of February. Assuming 50,000 safety stock. Compute ROP.**

**ISafety=50,000**

**L= 28 days (during the month of February, there are always 28 calendar days)**

**R=(105,000 units/week) x (1 week/7days)=15,000 units per day**

**Solution: ROP= LTD+Isafety**

**ROP= LTD+50,000**

**ROP= (28 x 15,000) +50,000=470,000**

1. **After a great success with the Magic Bullet, Bed Bath & Beyond decided to introduce a new product called the Pocket Hose. After a successful introduction, customer began to demand on average 1,000 units per day. The standard deviation of demand is 5 per day, and the lead time is 10 days. Compute ROP at 90% service level. Compute safety stock (round to nearest dollar) Assume demand is variable and lead time is fixed.**

**Questions to think about:**

**What is the average demand during the lead time?**

**What is standard deviation of demand during lead time?**

**L: Lead Time**

**R: Demand per period (per day, week, month)**

**R: Average Demand per period (day, week, month)**

**σR: Standard deviation of demand (per period)**

**LTD: Average Demand During Lead Time**

**LTD = L × R**

**σLTD: Standard deviation of demand during lead time**

σ\_LTD=√L σ\_R

**L: 10 days**

**R: 1,000 units/day**

**σR: Standard deviation of daily demand =5**

**LTD: Average Demand During Lead Time**

**LTD = L × R = 10 × 1,000 = 10,000**

**σLTD: Standard deviation of demand during lead time**

σ\_LTD=√10 (5)=15.811

**Solution:**

**SL = 90% 🡺 z = 1.28**

**x = μ + z σ**

**ROP= LTD +z σLTD**

**LTD = 10,000**

**σLTD = 15.811**

**Thus,**

**ROP= 10,000+ 1.28 × 15.811**

**ROP= 10,000 + 20**

**ROP = 10,020.**

**Isafety = 20**

1. **Use the same information provided on problem 3. However, assume that demand is fixed and lead time is variable.**

**If Lead time is variable and Demand is fixed**

**L: Lead Time**

**L: Average Lead Time = 10 days**

**σL: Standard deviation of Lead time = 5 days**

**R: Demand per period = 1,000 per day**

**LTD: Average Demand During Lead Time**

**LTD = 10 × 1,000 = 10,000**

**σLTD: Standard deviation of demand during lead time**

σ\_LTD=10,000(5) =50,000

σ\_LTD=Rσ\_L

**LDT = 10,000**

**σLTD = 50,000**

**SL = 90%**

**z =1.28**

**Thus,**

**ROP = LTD +zσLTD**

**ROP = 10,000 +1.28(50,000)**

**ROP = 320 + 64,000**

**Isafety = 64,000**

**5.) At the Arbor Grill, average lead time demand for potatoes (to make French fries) is 18,000 units. Standard deviation of lead time demand is estimated to be 9,000 units. The store can only order a 7-day supply, 21,000 units, each time the inventory level drops to 26,000 units. Due to its limited space, suppose holding cost is $3 per unit per year.**

**LTD = 18,000, *σLTD* = 9,000, ROP = 26,000, H=$3, R = 3,000 / day, Q or EOQ = 21,000.**

**a) Compute the service level.**

**ROP = LTD + Isafety 🡺 Isafety = 26,000 – 18,000 = 8,000**

**ROP = LTD +z *σLTD* 🡺 Isafety = z *σLTD* =8000**

**z(9,000) = 8,000 🡺 z = 8,000/9,000 = 0.89**

**z =0.89 🡺 P(z ≤ Z) = 0.8133 = 81.33%**

**In 81.33 % of the order cycles, Arbor Grill will not have a stock-out. Risk = 18.67%.**

**b) Compute the cycle inventory, and average inventory.**

**At reorder point we order 21,000 units.**

**Icycle = Q/2 = 21,000/2 = 10,500**

**Isafety = 8,000**

**Average Inventory = I = Icycle + Isafety =10,500 + 8,000 = 18,500.**

**c) Compute the total holding costs per year.**

**H(Average Inventory) = H(I) = 3 × 18,500 = 55,300/year**

**d) Compute the average flow time.**

**R = 3,000 / day, I = 18,500**

**RT = I 🡺3000T = 18,500**

**T = 6.167**

**6.167 days**

**6.)Wilson Inc. produces a 3-week supply of its FIFA World Cup Soccer Ball Model when stock on hand drops to 400 units. It takes 1 week to produce a batch. Orders average 350 units per week, and standard deviation of forecast errors is estimated at 175 units.**

**ROP=400, L=1 week, LTD 350/week, σ\_LTD=175, Q=3(350)**

**LTD = N(350, 175).**

**ROP = LTD + Isafety**

**ROP = LTD +z *σLTD***

**400= 350+z(175)**

**50 = 175z**

**z = 50/175 = 0.29, Thus, risk is 61.41%**



|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| SL | % change | Z | σLTD | Is =zσLTD | % change | ROP = LTD +Is |
| 0.8 | 100 | 0.84 | 175 | 147 | 100 | 497 |
| 0.9 | 113 | 1.28 | 175 | 224 | 152 | 574 |
| 0.95 | 119 | 1.64 | 175 | 288 | 195 | 638 |
| 0.99 | 124 | 2.33 | 175 | 407 | 276 | 757 |