Summaries pre-master

Industrial Engineering and Management

Study Association Stress
RAVELIJN BUILDING | RA1336

Table of contents

Stress

Stress is the study association for Business Administration (BA), International Business Administration (IBA) and Industrial Engineering and Management (IEM) of the University of Twente. Stress was founded on May 21st, 1974. Currently, Stress has over 2100 members and is the largest UT (study) association. Stress organizes various activities to support, expand and complement all of its studies. Stress has five principles, which will greatly enhance your time as a student: Study, Meet, Practice, Develop and International. Following these, several activities are organized by the roughly 120 active members, and are partially made possible thanks to the sponsorship and participation of various companies. Moreover, we have regular contact with other business and management study associations across the Netherlands.

Education

As a study association, Stress represents its members towards the faculty. Therefore we have a Commissioner of Educational Affairs who, with help of the Education Committee, deals with everything concerning your education. From collecting summaries to handling complaints to organizing educational events. If you have any questions concerning your study, the teachers, the faculty or anything else, feel free to ask the Commissioner of Educational affairs!

Education Committee

The committee consists of representatives of every cohort. The representative of the freshmen will be introduced during the first module. This committee is aimed at forming the bridge between teachers and students and therefore, improving the quality of the study. They keep the summary database on the Stress site up to date, organise study evenings and try to keep an eye out for the quality of our education. When there are complaints from the students concerning education, about one of the courses of one of our studies for example, they will be handled by this committee.

Summaries

Next to handling complaints, we also collect and check summaries of all the courses you follow! So it does not matter if the course is in the first, fifth or eighth module, you can send your summary in and we will check if your summary will make a good addition to our collection. To hand in your summary, simply send the file to ec@stress.utwente.nl. The Education Committee will then check the summary and if it is found to be sufficient, you will be compensated for your efforts. If your summary is the first one of a course, you will receive €15. If it is the second one, €10, and for the third one you will get €5. If you think you have made a better summary than the ones online, you can also send yours in and earn €5,-. Our summary collection can be found at the bottom of the 'study' page on www.stress.utwente.nl.

Panel Meetings

The panel meetings are organised in every module to improve the education. Here, the teachers of the module together with some students discuss the module. The students are able to give their opinion about the module and what they would like to see improved. The teachers can also ask questions about the opinions of the students. This way teachers know what went right in a module and what went wrong so they can improve the module for next year. The panel meetings are for all the students which are taking the module. You can also join the feedback panel. This means that you join a group of student who attends the panel meeting each module and gives valuable feedback to the programme.

Study sessions

For some courses, Stress organises study sessions. During these afternoons or evenings one or two student assistants of the course will be present. The study sessions are free to attend and coffee, tea and snacks are provided for you by Stress. If you think a study session will be valuable for a course you are following, please contact the Commissioner of Educational Affairs or the Education Committee. They will check if there is more demand for a study evening for this course and act accordingly.

Complaints

If you have a complaint, you can submit it at the 'Study' page on www.stress.utwente.nl or talk to someone of the Education Committee. However, if you feel it is a really important complaint or you want to explain it personally, you can come to the Stress room and talk to the Commissioner of Educational Affairs or send an email. We will then contact the programme management team and discuss what actions can be taken. They value bundled complaints greatly because it tells them a lot more when multiple people have the same complaint, this is the most important reason to always voice your opinion.

Ordering Books

For the first module, you can order your books during the Kick-In. For the following modules, you will have to order them by yourself. You can do this online, at our website. The only requirement to order the books is that you are a member of Stress.

To order books online you have to go to the 'Study' page on www.stress.utwente.nl. On the left of the screen you find a header: 'BOOKSALE', and below the option: 'Order your books'; select this option. Next, you can use the dropdown menus to select your study and module. Once you have chosen the correct options, press 'To booklist'. After this, you can select all the books you would like to order, and then proceed to 'Checkout. After paying, the books will be shipped to the address you enter.

For any questions about the books you need, the ordering of the books or anything else book-related, you can send an email to books@stress.utwente.nl.

Tutor platform

If you are having trouble studying for a course, we have the tutor platform to provide you with the right student for your struggles. We have a wide variety of students who have gone before you and who are willing to help you out for a small compensation. Send an email to tutor@stress.utwente.nl and mention your study, study-year, course you need help with, how many hours you need and any requirements you might have for the tutor. The payment of the tutor can be negotiated but keep in mind that students get paid €10 to €15 by the university when working as a teaching assistant, and you have to pay for it yourself.

The other way around, we are always looking for new tutors. If you are interested in joining our tutor pool let us know. We will add you to the WhatsApp group and you can reply to students asking for tutoring.

HELP!

Often, students do not know where to go with any problems, either study-related or personal. Here you find some information about the most common places to find help.

Study advisor

The study advisor is not only there to answer all your questions about your study, but also there to help you with any personal conditions or other issues that might affect you or your study progress. If you have any problem at all, go see your study advisor. Even if they are not the person who can help you, they can send you to someone who can. Every talk with study advisors is confidential and they will always do their best to help you. You can make an appointment with the study advisors on www.bms.planner.utwente.nl. The study advisors for IEM are Cornelis ten Napel and Ellen van Zeijts. The office of Cornelis is RA3246 and the office of Ellen is RA3256. Their emails are c.tennapel@utwente.nl and e.w.g.vnzeijts@utwente.nl.

Red desk / Student Affairs Coaching & Counselling

If your study and personal life are all on track, this bit of information might not be really relevant for you. But if it is not the case, when your study is completely going the wrong way, or you find it hard to adapt to living away from your parents or you have a difficult situation back at home, the Student Affairs Coaching & Counselling, also called the 'Red Desk', is the place where they can help you. Every possible question about study or personal issues will be answered here, or you will be forwarded to a trained professional. The Red Desk can be contacted at sacc@utwente.nl and is located in the Vrijhof (building 47), third floor, room 311.

Become active at Stress!

Next to your study, you can become an active member of our association! Stress offers many different committees which have organisational tasks or supporting tasks. On our website, you can check out all the committees from Stress. To find out which committee suits you best, email the Commissioner of Internal Affairs at internal@stress.utwente.nl.

Member Initiative

Have you always wanted to organize something big, but never had the resources? We appreciate initiatives from our members! So, if you have a clever idea for something within Stress or the committees, please contact us and we can see what is possible.

More information about Stress

Do you want to know more about Stress? Or do you want to check out our website and social media? Make sure to scan the QR code:

Part one

! Disclaimer: always check what you need to study corresponds with the content of the summaries, courses can be changed which could cause changes in study material for your exams
If you made a summary for a course this module you can sent them to education@stress.utwente.nl and depending on how many summaries we have for this course you will receive compensation for your work.

Each specialisation has a specific pre-master programme. If you are a student with a technical programme from a Research University and you are admitted to one of the pre-master programmes, you have to take up to 15 EC of courses. (one part)

Courses PLM and HCTM

- Statistics & Probability for premaster IEM	202001176	5 EC
- OR models for premaster IEM	202000450	10 EC

Courses FEM

- Statistics & Probability for premaster IEM	202001176	5 EC
- Financial Engineering for premaster IEM	201500020	10 EC

If you are a student with a technical programme from a University of Applied Science or a student with a social science programme from a Research University and you are admitted to one of the pre-master programmes, you have to take up to 30 EC of courses. (two parts)

Courses PLM and HCTM

- Calculus A for premasters	202001172	4 EC
- Academic Skills for premaster IEM	202000451	1 EC
- OR models for premaster IEM	202000450	10 EC

Courses FEM

- Calculus A for premasters	202001172	4 EC
- Academic Skills for premaster IEM	202000451	1 EC
- Engineering for premaster IEM	202000454	10 EC

Summary 1

Course: Calculus A for premasters
Book: Thomas' Calculus
Chapters: Functions, Limits & Continuity, Integration, Techniques of Integration, First-Order Differential Equations, Parametric Equations and Polar Coordinates, Vectors and the Geometry of Space, Vector-Valued Functions and Motion in Space, Partial Derivatives
Year the summary: 2016/2017

Summary 2

Course: OR models for premaster IEM
Book: Winston, W. L. (2004). *Operations* Research. Thomson Brooks/Cole.
Chapters: 1, 3, 9, 15, 16, 18, 20
Year the summary: 2016/2017

Summary 3

Course: OR models for premaster IEM
Book: slides
Chapters: -
Year the summary: 2016/2017

Summary 1: Calculus A for premasters

WEEK 1: FUNCTIONS, LIMITS, CONTINUITY, ASYMPTOTES

Chapter 1: Functions

1.1 Functions and Their Graphs

Domain: set D of all possible input values

Range: set of all values of f(x) as x varies throughout D

Natural domain: the largest set of real x-values for which the formula gives real yvalues

Function	Domain (x)	Range (y)
$y=x^2$	(-∞,∞)	[0, ∞)
y = 1/x	(-∞,0) ∪ (0, ∞)	(-∞,0) ∪ (0, ∞)
y = √x	[0, ∞)	[0, ∞)
y = √(4-x)	(-∞,4]	[0, ∞)
$y = \sqrt{(1-x^2)}$	[-1,1]	[0,1]

Vertical line test: a function f can have only one value f(x) for each x in its domain, so no vertical line can intersect the graph of a function more than once.

Piecewise-defined functions: sometimes a function is described by using different formulas on different parts of its domain, i.e. **the absolute value function** *(figure 1.8)*.

Greatest integer function/integer floor function: the function whose value at any number x is the greatest integer less than or equal to x.

Least integer function/integer ceiling function: the function whose value at any number x is the smallest integer greater than or equal to x.

Even function	Odd function
f(-x) = f(x)	f(-x) = -f(x)
Symmetric about the y-axis	Symmetric about the origin
+c doesn't change symmetry	+c symmetry is lost

Common functions

- **Linear functions** y = mx +b. If b=0 **proportional variables:** one is always a constant multiple of the other (y = mx). **Inversely proportional:** if the variable y is proportional to the reciprocal 1/x. *Figures 1.15 and 1.16.*
- **Power functions**
  $y = x^a$. *Figure 1.17.*
- **Polynomials**

$y = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x_0$ where n is a nonnegative integer and the numbers a_n are real constants (the coefficients of the polynomial). All polynomials have the domain $(-\infty, \infty)$. *Figure 1.18.*

- **Rational functions** $y = p(x)/q(x)$, where p and q are polynomials. *Figure 1.19.*
- **Algebraic functions**
  Any function constructed from polynomials using algebraic operations (addition, subtraction, multiplication, division, taking roots)
- **Trigometric functions**
  Sine, cosine, tangent, cosecant, secant, cotangent *(section 1.3)*
- **Exponential functions** $y = a^x$. All exponential functions have domain $(-\infty,\infty)$ and range $(0, \infty)$, so an exponential function never assumes the value 0. *Section 1.5. Figure 1.22.*
- **Logerithmic functions** $y = \log_a x$ Inverse functions of exponential functions. *Figure 1.23.* Domain $(0, \infty)$ and range $(-\infty,\infty)$.

1.2 Combining functions; Shifting and Scaling Graphs

$(f \circ g)(x) = f(g(x))$

First find g(x), second find f(g(x))

Shift formulas

- Vertical shifts
  $y = (x) + k$

  Shifts the graph of f up k units if $k > 0$ or down if $k < 0$

- Horizontal shifts
  $y = f(x+h)$

  Shifts the graph of f left h units if $h > 0$ or down if $h < 0$

Scaling and reflecting formulas For c > 1

$y = cf(x)$	stretches the graph of f vertically by a factor of c
$y = 1/c\ f(x)$	compresses the graph of f vertically by a factor of c
$y = f(cx)$	compresses the graph of f horizontally by a factor of c
$y = f(x/c)$	stretches the graph of f horizontally by a factor of c

For c=-1
$y = -f(x)$ reflects the graph of f across the x-axis
$y = f(-x)$ reflects the graph of f across the y-axis

Ellipses
$(c^2)x^2 + y^2 = r^2$, where r is the radius

Center $(h, k) = \frac{(x-h)^2}{a^2} + \frac{(y-k)^2}{b^2} = 1$

1.6 Inverse functions and logarithms

One-to-one function: a function that has distinct values at distinct elements in its domain; they take on any one value in their range exactly once, i.e. square roots of different numbers.

Horizontal line test: a function is one-to-one if and only if its graph intersects each horizontal line at

Inverse functions: since each output of a one-to-one function comes from just one input, the effect of the function can be inverted to send an output back to the input from which it came: $f^{-1}(b) = a$ if $f(a) = b$. The domain of f^{-1} is R and the range is D.

From f to f^{-1}

Example 3 and 4 on page 43

1. Solve the equation y = f(x) for x. This gives a formula $x = f^{-1}(y)$ where x is expressed as a function of y
2. Interchange x and y, obtaining a formula $y = f^{-1}(x)$ where f^{-1} is expressed in the conventional format with x as the independent and y as the dependent variable

Logarithmic functions

If a is any positive real number other than 1, the base a exponential functions $f(x) = a^x$ is one-to-one. It therefore has an inverse. Its inverse is called the logarithm function with base a: $\log_a x$.

$\text{Log}_e x = \ln x$

$\text{Log}_{10} x = \log x$

Ln x = y and $e^y = x$

Ln e = 1

Algebraic properties of the natural logarithm

Example 5 on page 45

1. Product rule: ln bx = ln b + ln x
2. Quotient rule: ln b/x = ln b – ln x
3. Reciprocal rule: ln 1/x = -ln x
4. Power rule: $\ln x^r = r \ln x$

Inverse properties for a^x and $\log_a x$

1. Base a: $a^{\log x} = x$, $\log_a a^x = x$
2. Base e: $e^{\ln x} = x$, $\ln e^x = x$

a^x
$= e^{x \ln (a)}$

Change of base formula

$\text{Log}_a x = \ln x / \ln a$

Chapter 2: Limits and Continuity

2.2 Limit of a Function and Limit Laws

Identity function: $f(x) = x$

Constant function: $f(x) = k$

Non-existing limits

Figure 2.10 on page 67

- Jumps (the left and the right side do not approach the same value): "a function f(x) has a limit as x approaches c if and only if it has left-hand and right-hand limits there and these one-sided limits are equal".
- Arbitrarily large values for $x \to 0$
- Oscillating

Limit laws

1. **Sum rule:** lim (f(x) + g(x)) = L + M
2. **Difference rule:** lim (f(x) - g(x)) = L - M
3. **Constant multiple rule:** lim (k * f(x)) = k * L
4. **Product rule:** lim (f(x) * g(x)) = L * M
5. **Quotient rule:** lim f(x)/g(x) = L / M
6. **Power rule:** lim [f(x)]^n = L^n
7. **Root rule:** lim ^n sqrt (f(x)) = ^n sqrt L

Limits of polynomials: if $f(x) = a_n x^n + a_{n-1} x^{n-1} + \ldots + a_1 x_0$, then $\lim_{x \to c} = P(x) = P(c) = a_n c^n$ + an-1cn-1 + … + a0.

Limits of rational functions: If P(x) and Q(x) are polynomials, then $\lim_{x \to c} = P(x) / Q(x) = P(c) / Q(c)$

Calculating limits

- Try to replace x with the value that the function is approaching
- Find or create common factors in order to eliminate them / rewrite the function, *example 9 on page 71.*

The sandwich theorem

Suppose that $g(x) \leq f(x) \leq h(x)$ for all x in some open interval containing c, except possibly at x = c itself and that lim g(x) = lim h(x) = L. Then lim f(x) = L.

Limit rules

1. $\lim_{\theta \to 0} \sin \theta = 0$
2. $\lim_{\theta \to 0} \cos \theta = 1$
3. For any function f, $\lim_{x \to c} |f(x)| = 0$ implies $\lim_{x \to c} f(x) = 0$

2.4 One-Sided Limits

One sided limits: limits as x approaches the number c from the left-hand side (x < c) or the right-hand side (x > c) only

- Right-handed: $x \to c+$
- Left-handed: $x \to c-$

$\lim_{\theta \to 0} \sin\theta / \theta = 1$ with θ in radians *Example 5b on page 89/90*

2.5 Continuity

A function $y = f(x)$ is continuous at an **interior point** c of its domain if $\lim_{x \to c} f(x) = f(c)$

A function $y = f(x)$ is continuous at a left **endpoint** a or is continuous at a right endpoint b of its domain if $\lim_{x \to a+} f(x) = f(a)$ or $\lim_{x \to b-} f(x) = f(b)$

Examples 1 – 3 on pages 93-94

Continuity test

A function f(x) is continuous at an interior point x = c of its domain if and only if it meets the following three conditions:

1. f(c) exists — c lies in the domain of f
2. $\lim_{x \to c} f(x)$ exists — f has a limit as $x \to c$
3. $\lim_{x \to c} f(x) = f(c)$ — the limit equals the function value

   Example 4 on page 94

Discontinuities

Figure 2.40 on page 95

- Removable
- Jump
- Infinite
- Oscillating

Properties of continuous functions

Examples 6 and 6 on page 96

1. **Sums** $f + g$
2. **Differences** $f-g$
3. **Constant multiples** kf
4. **Products** fg
5. **Quotients** f/g **6. Powers** f^n
7. **Roots** $\sqrt[n]{f}$

If f is continuous at c and g is continuous at f(c), then the composite $g \circ f$ is continuous at c. *Example 8 on page 97*

If g is continuous at the point b and $\lim_{x \to c} f(x) = b$ then

$\lim_{x \to c} g(f(x)) = g(b) = g(\lim_{x \to c} f(x))$

Example 9 on page 98

Example 10 on page 99

The intermediate value theorem for continuous functions

If f is a continuous function of a closed interval [a,b] and if y_0 is any value between f(a) and f(b), then y_0 = f(c) for some c in [a,b]

Examples 11-12 on page 100

2.6 Limits Involving Infinity; Asymptotes of Graphs

WEEK 2: DIFFERENTIATION AND APPLICATIONS

Chapter 3: Differentiation

3.1 Tangents and the Derivative at a Point

3.2 The Derivative as a Function

3.3 Differentiation Rules

3.5 Derivatives of Trigonometric Functions

3.6 The Chain Rule

3.8 Derivatives of Inverse Functions and Logarithms

Chapter 4: Applications of Derivatives

4.1 Extreme Values of Functions

4.3 Monotonic Functions and the First Derivative Test

4.5 Indeterminate Forms and L'Hôpital's Rule

4.8 Antiderivatives

WEEK 3: INTEGRATION AND APPLICATIONS

Chapter 5: Integration

5.1 Area and Estimating with Finite Sums About finite sums

Definite integral: if the number of terms contributing to the sum approaches infinity and we take the limit of these sums.

5.2 Sigma Notation and Limits of Finite Sums

Algebra rules for finite sums

1. **Sum rule** $\sum(a + b) = \sum a + \sum b$
2. **Difference rule** $\sum(a - b) = \sum a - \sum b$
3. **Constant multiple rule** $\sum ca = c \sum a$
4. **Constant value rule** $\sum c = nc$

First n squares: $\sum k^2 = n(n+1)(2n+1) / 6$

First n cubes: $\sum k^3 = ((n(n+1)/2)^2$

5.3 The Definite Integral

Rules satisfied by definite integrals

Figure 5.11 on page 317

1. **Order of integration** ${}_b\int^a f(x)\,dx = -\,{}_a\int^b f(x)\,dx$
2. **Zero width interval** ${}_a\int^a f(x)\,dx = 0$
3. **Constant multiple** ${}_a\int^b kf(x)\,dx = k\,{}_a\int^b f(x)\,dx$
4. **Sum and difference** ${}_a\int^b f(x) + g(x) = {}_a\int^b f(x) + {}_a\int^b g(x)$ / ${}_a\int^b f(x) - g(x) = {}_a\int^b f(x) - {}_a\int^b g(x)$
5. **Additivity** ${}_a\int^b f(x)\,dx + {}_b\int^c f(x)\,dx = {}_a\int^c f(x)\,dx$
6. **Max-min inequality** If f has a maximum value max f and minimum value min f on [a,b] then $\min f * (b - a) \le \max f * (b - a)$
7. **Domination** $f(x) \ge g(x)$ on [a,b] $\rightarrow {}_a\int^b f(x)\,dx \ge {}_a\int^b g(x)\,dx$ $f(x) \ge 0$ on [a,b] $\rightarrow {}_a\int^b f(x)\,dx \ge 0$

If f is integrable on [a,b] then its average value on [a,b], also called its mean, is $av(f) = 1/(b-a)\,{}_a\int^b f(x)\,dx$

5.4 The Fundamental Theorem of Calculus

Net Change: the net change in a function F(x) over an interval $a \le x \le b$ is the integral of its rate of change: $F(b) - F(a) = {}_a\int^b F'(x)\,dx$

To find the area between the graph of y = f(x) and the x-axis over the interval [a,b]:

1. Subdivide [a,b] at the zeros of f
2. Integrate f over each subinterval
3. Add the absolute values of the integrals

5.5 Indefinite Integrals and the Substitution Method

Indefinite integral of the function f with respect to x as the set of all antiderivatives of f: $\int f(x)\,dx = F(x) + C$

Definite integral $_a\int^b f(x)\,dx$ is a number. An indefinite integral is a function plus an arbitrary constant C.

Substitution

Examples 1 – 10 on pages 337 – 342.
If u = g(x) is a differentiable function whose range is an interval I and f is continuous on I, then $\int f(g(x))g'(x)\,dx = \int f(u)\,du$

5.6 Substitution and Area Between Curves

Substitution in definite integrals: If g' is continuous on the interval [a,b] and f is continuous on the range of g(x) = u then $_a\int^b f(g(x)) * g'(x)\,dx = {}_{g(a)}\int^{g(b)} f(u)\,du$. To use the formula, make the same usubstitution u = g(x) and du = g'(x) dx you would use to evaluate the corresponding indefinite integral. Then integrate the transformed integral with respect to u from the value g(a) (the value of u at x = a) to the value g(b) (the value of u at x = b). *See example 1 and 2 on page 345.*

Let f be continuous n the symmetric interval [-a,a]:

Figure 5.24 and example 3 on page 346

- If f is even, then $_{-a}\int^a f(x)\,dx = 2\ {}_0\int^a f(x)\,dx$
- If f is odd, then $_{-a}\int^a f(x)\,dx = 0$

Area between curves: if f and g are continuous with f(x) ≥ g(x) throughout [a,b], then the area of the region between the curves y = f(x) and y = g(x) from a to b is the integral of (f-g) from a to b: $A = {}_a\int^b [f(x) – g(x)]\,dx$. *Example 4-6 on page 348/349*

Integration with respect to y: $A = {}_c\int^d [f(y) – g(y)]\,dy$. In this equation f always denotes the right-hand curve and g the left-hand curve, so that f(y) – g(y) is nonnegative. *Example 7 on page 350.*

Chapter 8: Techniques of Integration

8.1 Integration by Parts

Integration by parts is a technique for simplifying integrals of the form ∫ f(x)g(x) dx. This is useful when **f can be differentiated repeatedly and g can be integrated repeatedly without difficulty**.

E.g. ∫ x cos x dx and ∫ x^2e^x because x or x^2 can be differentiated repeatedly to become zero and cos x or e^x can be integrated repeatedly without difficulty.

Integration by parts formula

∫ u dv = uv - ∫ v du   with u = f(x) | v = g(x) | du = f'(x) dx | dv = g'(x) dx

Examples 1 – 6 on pages 455 – 458

Tabular integration

Examples 7 and 8 on page 458 - 459

8.2 Trigonometric Integrals

Products with powers of sines and cosines

Examples 1-3 on page 463 ∫
$\sin^m x \cos^n x$ dx

- **m is odd:** write m as 2k + 1 and use the identity $\sin^2 x = 1 - \cos^2 x$ to obtain $\sin^m x = \sin^{2k+1} x = (\sin^2 x)^k \sin x = (1 - \cos^2 x)^k \sin x$

  Then combine the single sin x with dx in the integral and set x dx equal to –d(cos x)
- **m is even and n is odd:** in ∫ $\sin^m x \cos^n x$ dx, we write n as 2k + 1 and use the identity $\cos^2 x = 1 - \sin^2 x$ to obtain

  $\cos^n x = \cos^{2k+1} x = (\cos^2 x)^k \cos x = (1 - \sin^2 x)^k \cos x$

  Then combine the single cos x with dx and set cos x dx equal to d(sin x)
- **Both m and n are even:** in ∫ $\sin^m x \cos^n x$ dx, we substitute $\sin^2 x = (1 - \cos 2x) / 2$, $\cos^2 x = (1 + \cos 2x) / 2$

Eliminating square roots

Example 4 on page 464

Integrals of powers of tan x and sec x

Example 5 and 6 on page 464-465

Products of sines and cosines

Example 7 on page 466

- sin mx sin nx = ½ [cos (m – n)x – cos (m + n)x]
- sin mx cos nx = ½ [sin (m – n)x – sin (m + n)x]
- cos mx cos nx = ½ [cos (m – n)x – cos (m + n)x]

8.4 Integration of Rational Functions by Partial Fractions

Method of partial fractions: rewriting rational functions as a sum of simpler fractions.

- The degree of f(x) must be less than the degree of g(x). If it isn't, divide f(x) by g(x) and work with the remainder term (*see example 3 on page 474*).
- We must know the factors of g(x).

See method of partial fractions on pages 472-473 and examples 1 – 5 on pages 473-476.

Heaviside 'cover-up' method for linear factors: *see Heaviside Method on page 477 and examples 6 and 7 on pages 476-478.*

Other ways to determine coefficients: *example 8 on page 478.*

8.7 Improper Integrals

Improper integrals of type 1: integrals with infinite limits of integration

1. If f(x) is continuous on [a, ∞), then $\int_a^\infty f(x)\,dx = \lim_{b\to\infty} \int_a^b f(x)\,dx$
2. If f(x) is continuous on (-∞, b], then $\int_{-\infty}^b f(x)\,dx = \lim_{a\to-\infty} \int_a^b f(x)\,dx$
3. If f(x) is continuous on (-∞,∞), then $\int_{-\infty}^\infty f(x)\,dx = \int_{-\infty}^c f(x)\,dx + \int_c^\infty f(x)\,dx$

If the limit is finite, we say that the improper integral **converges** and that the limit is the value of the improper integral. If the limit fails to exist, the improper integral **diverges.** *See examples 1 and 2 on page 497-498.*

The integral $\int_1^\infty dx/x^p$: *example 3 on page 498.*

Improper integrals of type 2: integrals of functions that become infinite at a point within the interval of integration.

Example 4 and 5 on page 500.

1. If f(x) is continuous on (a,b] and discontinuous at a, then $\int_a^b f(x)\,dx = \lim_{c\to a^+} \int_c^b f(x)\,dx$
2. If f(x) is continuous on [a,b) and discontinuous at b, then $\int_a^b f(x)\,dx = \lim_{c\to b^-} \int_a^c f(x)\,dx$
3. If f(x) is discontinuous at c, where a < c < b and continuous on [a, c) ∪ (c,b] then $\int_a^b f(x)\,dx = \int_a^c f(x)\,dx + \int_c^b f(x)\,dx$.

Direct comparison test

See example 7 on page 502
Let f and g be continuous on a [a, ∞) with 0 ≤ f(x) ≤ g(x) for all x ≥ a, then

1. $\int_a^\infty f(x)\,dx$ converges if $\int_a^\infty g(x)\,dx$ converges
2. $\int_a^\infty g(x)\,dx$ diverges if $\int_a^\infty f(x)\,dx$ diverges

Limit comparison test

See examples 8 and 9 on page 503

If the positive functions f and g are continuous on [a, ∞), and if $\lim_{x\to\infty} f(x)/g(x) = L$ with $0 < L < \infty$, then $_a\int^\infty f(x)\,dx$ and $_a\int^\infty f(x)\,dx$ both converge or both diverge.

Summary of types of improper integrals can be found on page 504

WEEK 4: DIFFERENTIAL EQUATION

Chapter 7: Integrals and Transcendental Functions

7.2 Exponential Change and Separable Differential Equations

See examples 1 and 2 on pages 429-430.

Chapter 9: First-Order Differential Equations

9.1 Solutions, Slope Fields and Euler's Method

A first-order differential equation is an equation y' = dy/dx = f(x,y) in which f(x,y) is a function of two variables defined on a region in the xy-plane. *See example 1 on page 515.*

First-order initial value problem: a differential equation y' = f(x,y) whose solution must satisfy an initial condition $y(x_0) = y_0$. *See example 2 on page 515.*

Euler's Method: given a differential equation dy/dx = f(x,y) and an initial condition $y(x_0) = y_0$, we can approximate the solution y = y(x) by its linearization: $L(x) = y_0 + f(x_0, y_0)(x-x_0)$. *See example 3 and 4 on pages 518-519.*

9.2 First-Order Linear Equations

A first-order linear equation is one that can be written in the form dy/dx + P(x)y = Q(x). Rewriting in standard form, *see example 1 on page 522.*

Integrating factor: by multiplying both sides by a positive function v(x), which is called the integrating factor, that transforms the left-hand side into the derivative of the product v(x)*y, we can solve the linear equation. Choose v(x) so that v(dy/dx) + Pvy = d/dx (v * y):

1. dy/dx + P(x)y = Q(x)	set equation in standard form
2. v(x) dy/dx + P(x)v(x)y = v(x)Q(x)	multiply by positive v(x)
3. d/dx (v(x)y) = v(x)Q(x)	
4. v(x)y = ∫v(x)Q(x) dx	integrate with respect to x
5. y = 1/v(x) ∫v(x)Q(x) dx	

The terms v (dy/dx) should cancel (*see page 523*), so choose v so that $v = e^{\int P dx}$, in the following way:

1. dv/dx = Pv	
2. dv/v = P dx	variables separated, v > 0
3. ∫dv/v = ∫P dx	integrate both sides
4. ln v = ∫P dx	
5. $e^{\ln v} = e^{\int P\,dx}$	exponentiate both sides to solve for v
6. $v = e^{\int P\,dx}$	

See examples 2 and 3 on pages 524 – 525.

WEEK 7: VECTOR SPACES, PLANAR AND SPACE CURVES

Chapter 11: Parametric Equations and Polar Coordinates

11.1 Parametrizations of Plane Curves

If x and y are given as functions $x = f(t)$ and $y = g(t)$ over an interval I of t-values, then the set of points $(x,y) = (f(t), g(t))$ defined by these equations is the **parametric curve.** The equations are **parametric equations.** The variable t is a **parameter** for the curve and its domain I is the **parameter interval.** If I is a closed interval, $a \le t \le b$, the point $(f(a), g(a))$ is the **initial point** of the curve and $(f(b), g(b))$ is the **terminal point.** *Examples 1 – 7 on pages 629-631.*

Brachistochrones and tautochrones

11.2 Calculus with Parametric Curves

A parametrized curve $x = f(t)$ and $y = g(t)$ is differentiable at t if f and g are differentiable at t.

Parametric Formula for dy/dx: $dy/dx = (dy/dt) / (dx/dt)$

Parametric Formula for d^2y / dx^2: $d^2y / dx^2 = (dy'/dt) / (dx/dt)$ with $y' = dy/dx$

Examples 1 and 2 on page 637.

If a curve C is defined parametrically by $x = f(t)$ and $y = g(t)$, $a \le t \le b$, where f' and g' are continuous and not simultaneously zero on [a,b], and C is traversed exactly once as t increases from t = a to t = b, then the **length of curve C** is the definite integral

$L = {}_a\int^b \sqrt{\,[f'(t)]^2 + [g'(t)]^2\,}\ dt$

$f'(t) = dx/dt$ $g'(t) =$
dy/dt

Example 4 and 5 on page 640

Length of a Curve y = f(x)

Differential arc length: $ds = \sqrt{((dx/dt^2) + (dy/dt)^2)}$ *See example 6 on pages 641 – 642*

If a smooth curve $x = f(t)$, $y = g(t)$, $a \le t \le b$, is traversed exactly once as t increases from a to b, then the areas of the surfaces generated by revolving the curve about the coordinate axes are as follows:

See example 7 on page 643

- **Revolution about the x-axis**
  $S = {}_a\int^b 2\pi y$ *differential arc length formula* dt
- **Revolution about the y-axis**
  $S = {}_a\int^b 2\pi x$ *differential arc length formula* dt

11.3 Polar Coordinates

To define polar coordinates, we first fix an **origin** O (called the **pole**) and an **initial ray** from O (*see figure 11.18 on page 645*). Then each point P can be located by assigning to it a **polar coordinate pair** (r, Θ) in which r gives the directed distance from O to P and Θ gives the directed angle from the initial ray to ray OP. *Example 1 on pages 645 – 646.*

If we hold r fixed at a constant value of r = a, the point P(r, Θ) will lie |a| units from the origin O. As Θ varies over any interval of length 2π, P then traces a circle of radius |a| centered at O. If we hold Θ fixed at a constant value $\Theta = \Theta_0$ and let r vary between -∞ and ∞, the point P(r, Θ) traces the line through O that makes an angle of measure Θ_0 with the initial ray.

See examples 2 and 3 on page 646.

- r=a circle of radius |a| centered at O
- $\Theta = \Theta_0$ line through O making an angle Θ_0 with the initial ray

Equations relating polar and Cartesian coordinates

See example 4 – 6 on pages 647 - 648

- $x = r \cos \Theta$
- $y = r \sin \Theta$
- $r^2 = x^2 + y^2$
- $\tan \Theta = y/x$

11.4 Graphing in Polar Coordinates

Symmetry

- **Symmetry about the x-axis:** if the point (r, Θ) lies on the graph, then the point (r, - Θ) or (-r, π-Θ) lies on the graph. *See figure 11.26a on page 650.*
- **Symmetry about the y-axis:** if the point (r, Θ) lies on the graph, then the point (r, π- Θ) or (-r, -Θ) lies on the graph. *See figure 11.26b on page 650.*
- **Symmetry about the origin:** if the point (r, Θ) lies on the graph, then the point (-r, Θ) or (r, Θ+π) lies on the graph. *See figure 11.26c on page 650.*

Slope of the Curve r = f(Θ): $dy/dx|_{(r,\Theta)} = f'(\Theta) \sin \Theta + f(\Theta) \cos \Theta \;\;/\;\; f'(\Theta) \cos \Theta - f(\Theta) \sin \Theta$

Examples 1 and 2 on pages 650 – 651.

A technique for graphing

See example 3 on page 652

1. First graph r=f(Θ) in the Cartesian rΘ-plane
2. Use the Cartesian graph as a "table" and guide to sketch the polar coordinate graph

Chapter 12: Vectors and the Geometry of Space

12.1 Three-Dimensional Coordinate Systems

- **Points on the x-axis:** (x, 0, 0) □ **Points on the y-axis**: (0, y, 0)
- **Points on the z-axis**: (0, 0, z)

E.g. the plane y = 3 is the plane perpendicular (loodrecht) to the y-axis at y=3. *See figure 12.3 and example 1 on page 697.*

Distance between $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$: $|P_1P_2| = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$ *See example 3 on page 680*

The standard equation for the sphere of radius a and center (x_0, y_0, z_0):
$(x - x_0)^2 + (y - y_0)^2 + (z - z_0)^2 = a^2$

See examples 4 and on pages 680-681

12.2 Vectors

A **vector** is represented by a **directed line segment.**

- If **v is a two-dimensional vector** in the plane equal to the vector with initial point at the origin and terminal point (v_1, v_2), then the component form of v is $v = \langle v_1, v_2 \rangle$
- If **v is a three-dimensional vector** in the plane equal to the vector with initial point at the origin and terminal point (v_1, v_2, v_3), then the component form of v is $v = \langle v_1, v_2, v_3 \rangle$

The **magnitude or length** of the vector v = PQ is the nonnegative number

$|v| = \sqrt{v_1^2 + v_2^2 + v_3^2} = \sqrt{(x_2-x_1)^2 + (y_2-y_1)^2 + (z_2-z_1)^2}$

See example 1 on page 684

Vector algebra operations

See figures 12.12 – 12.14 on pages 867 – 868
Example 3 on page 868

- **Addition:** $u + v = \langle u_1 + v_1, u_2 + v_2, u_3 + v_3 \rangle$
- **Scalar multiplication:** $ku = \langle ku_1, ku_2, ku_3 \rangle$

Properties of vector operations

Let u, v, w be vectors and a, b be scalars

1. u + v = v + u
2. (u + v) + w = u + (v + w)
3. u + 0 = u
4. u + (-u) = 0
5. 0u = 0
6. 1u = u
7. a(bu) = (ab)u

8. $a(u+v) = au + av$
9. $(a+b)u = au + bu$

Unit vector: a vector f of length 1

- $i = <1, 0, 0>$
- $j = <0, 1, 0>$
- $k = <0, 0, 1>$

So, the vector from $P_1(x_1, y_1, z_1)$ to $P_2(x_2, y_2, z_2)$ is $P_1P_2 = (x_2-x_1)i + (y_2-y_1)j + (z_2-z_1)k$

- **v/|v|** is a unit vector in the direction of v
- **v = v/|v|** expresses v as its length times its direction *See example 6 on page 688*

The **midpoint** M of the line segment joining points $P_1(x_1, y_1, z_1)$ and $P_2(x_2, y_2, z_2)$ it the point

$(x_1 + x_2)/2$, $(z_1 + z_2)/2$, $(z_1 + z_2)/2$.

See examples 8 an 9 on pages 688 – 689.

12.3 The Dot Product

Angle between vectors: the angle Θ between two nonzero vectors $u = <u_1, u_2, u_3>$ and $v = <v_1, v_2, v_3>$ is given by

$\Theta = \cos^{-1} ((u_1v_1 + u_2v_2 + u_3v_3) / |u||v|) \quad = \quad \cos^{-1} ((u \cdot v) / |u||v|)$

The dot product u · v of vectors $u = <u_1, u_2, u_3>$ and $v = <v_1, v_2, v_3>$ **=** $u_1v_1 + u_2v_2 + u_3v_3$ *Examples 1-3 on pages 693-694*

Vectors u and v are **orthogonal** (or perpendicular) if and only if $u \cdot v = 0$. *Example 4 on page 694.*

Dot product properties and vector projections

1. $u \cdot v = v \cdot u$
2. $(cu) \cdot v = u \cdot v = u \cdot (cv) = c(u \cdot v)$
3. $u \cdot (v + w) = u \cdot v + u \cdot w$
4. $u \cdot u = |u|^2$
5. $0 \cdot u = 0$

The **vector projection** of u onto v is the vector

$\text{proj}_v u = ((u \cdot v) / |v|^2)\, v$

The **scalar component** of u in the direction of v is the scalar |u|
$\cos \Theta = (u \cdot v) / |v|$

Examples 5 and 6 on pages 696 – 697

The **work** done by a constant force F acting through a displacement D = PQ is W = F*D *Example 7 on page 697.*

12.4 The Cross Product

$u \times v = (|u||v|\sin\Theta)\, n$

Parallel vectors: nonzero vectors u and v are parallel if and only if u x v = 0

Properties of the cross product

If u, v and w are any vectors and r, s are scalars, then

1. (ru) x (sv) = (rs)(u x v)
2. u x (v + w) = u x v + u x w
3. v x u = -(u x v)
4. (v + w) x u = v x u + w x u
5. 0 x u = 0
6. u x (v x w) = (u·w)v – (u·v)w

- i x j = -(j x i) = k
- j x k = -(k x j) = i
- k x i = -(i x k) = j
- i x i = j x j = k x k = 0

|u x v| is the area of a parallelogram: $|u \times v| = |u||v||\sin\Theta||n| = |u||v|\sin\Theta$. Here |u| is the base of the parallelogram and $|v||\sin\Theta|$ the height. *See figure 12.30 on page 701.*

Calculating the cross product as a determinant

Examples 1-4 on page 702

If $u = u_1 i + u_2 j + u_3 k$ and $v = v_1 i + v_2 j + v_3 k$, then

$$u \times v = \begin{matrix} i & j & k \\ u_1 & u_2 & u_3 \\ v_1 & v_2 & v_3 \end{matrix}$$

Torque

Triple scalar or box product

Example 6 on page 704

$$(u \times v) \cdot w = \begin{matrix} u1 & u2 & u3 \\ v_1 & v_2 & v_3 \\ w1 & w2 & w3 \end{matrix}$$

12.5 Lines and Planes in Space

A vector equation for the line L through $P_0(x_0, y_0, z_0)$ parallel to v is

$r(t) = r_0 + tv$, where r is the position vector of a point $P(x,y,z)$ on L and r_0 is the position vector of $P_0(x_0,y_0z_0)$.

The standard parametrization of the line through $P_0(x_0,y_0z_0)$ parallel to $v = v_1i + v_2j + v_3k$ is $x = x_0 + tv_1$ $y = y_0 + tv_2$ $z = z_0 + tv_3$ $-\infty \le t \le \infty$

Examples 1-4 on pages 707-708

Distance from a point S to a line through P parallel to v $d = |PS \times v| / |v|$

See example 5 on page 708

Equation for a plane

Example 6 – 7 on pages 709 - 710

The plane through $P_0(x_0,y_0,z_0)$ normal to $n = Ai + Bj + Ck$ has

- **Vector equation:** $n \cdot P_0P = 0$
- **Component equation:** $A(x-x_0) + B(y-y_0) + C(z-z_0) = 0$
- **Component equation simplified:** $Ax + By + Dz = D$
  where $D = Ax_0 + By_0 + Cz_0$

Two planes are parallel if and only if their normal are parallel, or $n_1 = kn_2$

Examples 8-10 on pages 710-711

Distance from a point to a plane

$d = |PS \cdot n/|n| \,|$

Example 11 on page 711

Angles between planes

Example 12 on page 712

Chapter 13: Vector-Valued Functions and Motion in Space

13.1 Curves in space and their tangents

Helices: figure 13.3 and 13.4 on page 726

For r(t) = OP = f(t)i + g(t)j + h(t)k r has a limit L = $\lim_{t\to t0}$ r(t).

$\lim_{t\to to}$ = ($\lim_{t\to t0}$ f(t)) i + ($\lim_{t\to t0}$ g(t)) j + ($\lim_{t\to t0}$ h(t)) k

r'(t) = dr/dt = $\lim_{t\to 0}$ (r(t + dt) – rt) / dt = (df/dt) i + (dg/dt) j + (dh/dt) k

If r is he position vector of a particle moving along a smooth curve in space, then

v(t) = dr/dt.

Example 4 on page 729

- Velocity is the derivative of position: v = dr/dt
- Speed is the magnitude of velocity: speed = |v|
- Acceleration is the derivative of velocity: a = dv/dt = d^2r/dt^2
- The unit vector v/|v| is the direction of motion at time t

Differentiation rules for vector functions

1. **Constant function rule:** d/dt C = 0
2. **Scalar multiple rules:** d/dt [cu(t)] = cu'(t)
   dt [f(t)u(t)] = f'(t)u(t) + f(t)u'(t)
3. **Sum rule:** d/dt [u(t) + v(t)] = u'(t) + v'(t)
4. **Difference rule:** d/dt [u(t) - v(t)] = u'(t) - v'(t)
5. **Dot product rule:** d/dt [u(t) · v(t)] = u'(t) · v(t) + u(t) · v'(t)
6. **Cross product rule:** d/dr [u(t) x v(t)] = u'(t) x v(t) + u(t) x v'(t)
7. **Chain rule:** d/dt [u(f(t))] = f'(t)u'(f(t))

If r is a differentiable vector function t of constant length, then r ·
dr/dt = 0

13.2 Integrals of vector functions; projectile motion

The indefinite integral of r with respect to t is the set of all antiderivatives of r, denoted by ∫ r(t) dt = R(t) + C. *Example 1 on page 734.*

If the components of r(t) = f(t)I + g(t)j + h(t)k are integrable over [a,b], then so is r and the definite integral of r from a to b is

$_a\int^b$ r(t) dt = ($_a\int^b$ f(t) dt) i + ($_a\int^b$ g(t) dt) j + ($_a\int^b$ h(t) dt) k

See example 2 and 3 on page 734

Ideal projectile motion equation

$r = (v_0 \cos a)\, t\, i + ((v_0 \sin a)\, t - ½ gt^2)\, j$

Example for on page 736

Maximum height: $y_{max} = (v_0 \sin a)^2 / 2g$

Flight time: $t = (2v_0 \sin a) / g$

Range: $R = v_0^2/g \sin 2a$

Example 5 on page 737

13.3 Arc length in space

The length of a smooth curve $r(t) = x(t)i + y(t)j + z(t)k$, $a \le t \le b$, that is traced exactly once as t increases from $t = a$ to $t = b$, is

$L = \int_a^b \sqrt{ (dx/dt)^2 + (dy/dt)^2 + (dz/dt)^2 }\, dt$

Arc length formula: $L = \int_a^b |v|\, dt$

Arc length parameter with base point $P(t_0)$

$s(t) = = \int_{t_0}^t \sqrt{ [x'(\tau)]^2 + [y'(\tau)]^2 + [z'(\tau)]^2 }\, dt = \int_{t_0}^t |v(\tau)|\, dt$

Example 2 on page 743

Speed on a smooth curve

$ds/dt = |v(t)|$

Unit tangent vector: $T = v/|v|$ *Example 3 on page 744*

WEEK 8: PARTIAL DERIVATIVES

Chapter 14: Partial derivatives

14.1 Functions of several variables

Domain and ranges: *example 1 on page 766*

A region in the plane is **bounded** if it lies inside a disk of fixed radius.

The set of points in the plane where a function f(x,y) has a constant value f(x,y) = c is called a level **curve** of f. The set of all points (x,y,f(x,y)) in space for (x,y) in the domain of f, is called the **graph** of f. The graph of f is also called the **surface** z = f(x,y). *Example 3 on page 768*

The set of points (x,y,z) in space where a function of three independent variables has a constant value f(x,y,z) = c is called a **level surface** of f. *Example 4 on page 768.*

14.2 Limits and continuity in higher dimensions

We say that a function f(x,y) approaches the **limit** L as (x,y) approaches (x_0, y_0) and write
$\lim_{(x,y)\to(x_0,y_0)} f(x,y) = L$

if, for every number $\varepsilon > 0$ there exists a corresponding number $\delta > 0$ such that for all (x,y) in the domain of f, $|f(x,y) - L| < \varepsilon$ whenever $0 < \sqrt{(x-x_0)^2 + (y-y_0)^2} < \delta$.

Properties of limits of functions of two variables *Example 1-4 on pages 775-777*

1. **Sum rule** $\lim_{(x,y)\to(x_0,y_0)} f(x,y) + g(x,y) = L + M$
2. **Difference rule** $\lim_{(x,y)\to(x_0,y_0)} f(x,y) - g(x,y) = L - M$
3. **Constant multiple rule** $\lim_{(x,y)\to(x_0,y_0)} k\, f(x,y) = k\, L$
4. **Product rule** $\lim_{(x,y)\to(x_0,y_0)} f(x,y) \cdot g(x,y) = L \cdot M$
5. **Quotient rule** $\lim_{(x,y)\to(x_0,y_0)} f(x,y) / g(x,y) = L / M$
6. **Power rule** $\lim_{(x,y)\to(x_0,y_0)} f(x,y)^n = L^n$
7. **Root rule** $\lim_{(x,y)\to(x_0,y_0)} \sqrt[n]{f(x,y)} = \sqrt[n]{\mathbf{L}}$

A function f(x,y) in **continuous** at the point (x_0, y_0) if

Example 5 on page 777

1. F is defined at (x_0, y_0)
2. $\lim_{(x,y)\to(x_0,y_0)} f(x,y)$ exists
3. $\lim_{(x,y)\to(x_0,y_0)} f(x,y) = f(x_0,y_0)$

Two-path test for nonexistence of a limit

if a function f(x,y) has different limits along two different paths in the domain of f as (x,y) approaches (x_0, y_0) then $\lim_{(x,y)\to(x_0,y_0)} f(x,y)$ does not exist. *Example 6 on page 778.*

If f is continuous at (x_0, y_0) and g is a single-variable function continuous at $f(x_0,y_0)$, then the composite function $h = g \circ f$ defined by $h(x,y) = g(f(x,y))$ is continuous at (x_0, y_0).

14.3 Partial Derivatives

The partial derivative of f(x,y) with respect to x at the point (x_0,y_0) is $\delta f/\delta x =$
$\lim_{h\to 0} f(x_0 + h, y_0) – f(x_0, y_0) / h$

The partial derivative of f(x,y) with respect to x at the point (x_0,y_0) is $\delta f/\delta y =$
$\lim_{h\to 0} f(x_0, y_0 + h) – f(x_0, y_0) / h$

Examples 1 – 5 on pages 785 – 786

Functions of more than two variables *Examples 6 and 7 on page 786*

Partial derivatives and continuity *Example 8 on page 787*

Second order partial derivatives *Example 9 on page 788*

The mixed derivative theorem

If f(x,y) and its partial derivatives f_x, f_y, f_{xy} and f_{yx} are defined throughout an open region containing a point (a,b) and are all continuous at (a,b) then $f_{xy}(a,b) = f_{yx}(a,b)$.

Partial derivatives of still higher order

Example 11 on page 789

The increment theorem for functions of two variables

Suppose that the first partial derivatives of f(x,y) are defined throughout an open region R containing the point (x_0, y_0) and that f_x and f_y are continuous at (x_0, y_0). Then the change

$\Delta z = f(x_0 + \Delta x, y_0 + \Delta y) – f(x_0, y_0)$

in the value of f that results from moving from (x_0, y_0) to another point $(x_0 + \Delta x, y_0 + \Delta y)$ in R satisfies an equation of the form

$\Delta z = f_x(x_0, y_0)\, \Delta x + f_y(x_0, y_0)\, \Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$ in which each of $\varepsilon_1, \varepsilon_2 \rightarrow 0$ and both $\Delta x, \Delta y \rightarrow 0$.

A function z = f(x,y) is differentiable at (x_0, y_0) if $f_x(x_0,y_0)$ and $f_y(x_0, y_0)$ exist and Δz satisfies an equation of the form

$\Delta z = f_x(x_0,y_0)\Delta x + f_y(x_0,y_0)\, \Delta y + \varepsilon_1 \Delta x + \varepsilon_2 \Delta y$

in which each of $\varepsilon_1, \varepsilon_2 \rightarrow 0$ and both $\Delta x, \Delta y \rightarrow 0$. We call f differentiable if it is differentiable at every point in its domain and say that its graph is a smooth surface.

14.4 The Chain Rule

The chain rule for functions of one independent variable and two(/three) intermediate variables
dw/dt = df/dx * dx/dt + df/dy * dy/dt (+ df/dz * dz/dt)

See branch diagram + example 1 and 2 on pages 794-795

Two independent variables: *see page 796 and example 3-4*

A formula for implicit differentiation

Suppose that F(x,y) is differentiable and that the equation F(x,y) = 0 defines y as a differentiable function of x. Then at any point where $F_y \neq 0$, $dy/dx = - F_x/F_y$

See examples 5 and 6 on pages 798-799.

14.5 Directional derivatives and gradient vectors

The **derivative of f at $P_0(x_0, y_0)$ in the direction of the unit vector $u = u_1 i + u_2 j$** is the number $(df/ds)_{u, P0}$ $= \lim_{s\to 0} [f(x_0 + su_1, y_0 + su_2) - f(x_0, y_0)] / s$

Example 1 on page 803

The **gradient vector** of f(x,y) at a point $P_0(x_0,y_0)$ is the vector $\nabla f =$
$(\delta f/\delta x)\, i + (\delta f/\delta y)\, j$

The directional derivative is a dot product

If f(x,y) is differentiable in an open region containing $P_0(x_0,y_0)$ then

$(df/ds)_{u,P0} = (\nabla f)P_0 \cdot u$

Example 2 on page 804

Properties of the directional derivative $D_u f = \nabla f \cdot u = |\nabla f| \cos\Theta$

1. The function f increases most rapidly when cosΘ=1 or when Θ=0 and u is the direction of ∇f. That is, at each point P in its domain, f increases most rapidly in the direction of the gradient vector ∇f at P. The derivative in this direction is
   $D_u f = |\nabla f|\cos(0) = |\nabla f|$
2. Similarly, f decreases most rapidly in the direction of -∇f. The derivative in this direction id $D_u f =$ $|\nabla f|\cos(\pi) = -|\nabla f|$
3. Any direction u orthogonal to a gradient $\nabla f \neq 0$ is a direction of zero change in f because Θ then equals π/2 and
   $D_u f = |\nabla f|\cos(\pi/2) = |\nabla f| \cdot 0 = 0$

Example 3 on page 805

At every point (x_0,y_0) in the domain of a differentiable function f(x,y), the gradient of f is normal to the level curve through (x_0, y_0). *See figure 14.30 on page 806. See example 4 on page 806.*

Algebra rules for gradients

Examples 5 and 6 on page 807

1. **Sum rule** $\nabla(f+g) = \nabla f + \nabla g$
2. **Difference rule** $\nabla(f-g) = \nabla f - \nabla g$
3. **Constant multiple rule** $\nabla(kf) = k\nabla f$
4. **Product rule** $\nabla(fg) = f\nabla g + g\nabla f$

5. **Quotient rule** $\nabla(f/g) = (g\nabla f - f\nabla g) / g^2$

14.6 Tangent planes and differentials

The **tangent plane** at the point $P_0(x_0,y_0,z_0)$ on the level surface $f(x,y,z) = c$ of a differentiable function f is the plane through P_0 normal to $\nabla f|P_0$. The **normal line** of the surface at P_0 is the line through P_0 parallel to $\nabla f|P_0$.

Tangent plane to $f(x,y,z) = c$ at $P_0(x_0, y_0, z_0)$ $f_x(P_0)(x-x_0) +$
$f_y(P_0)(y-y_0) + f_z(P_0)(z-z_0) = 0$

Normal line to $f(x,y,z) = c$ at $P_0(x_0, y_0, z_0)$ x
$= x_0 + f_x(P_0)t$ $y = y_0 + f_y(P_0)t$

$z = z_0 + f_z(P_0)t$

Example 1 on page 810

Plane tangent to a surface $z = f(x,y,z)$ at $(x_0, y_0, f(x_0, y_0))$

The plane tangent to the surface $z = f(x,y)$ of a differentiable function f at the point $P_0(x_0, y_0, z_0) = (x_0, y_0, f(x_0, y_0))$ is

$f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0) – (z-z_0) = 0$

Examples 2 and 3 on page 811

Estimating the change in f in a direction u

To estimate the change in the value of a differentiable function f when we move a small distance ds from a point P_0 in a particular direction u, use the formula $df = (\nabla f|P_0 \cdot u)\, ds$

Example 4 on page 812

The **linearization** of a function $f(x,y)$ at a point (x_0, y_0) where f is differentiable is the function

$L(x,y) = f(x_0, y_0) + f_x(x_0,y_0)(x-x_0) + f_y(x_0,y_0)(y-y_0)$ $f(x,y) \approx L(x,y)$ is the **standard linear approximation** of f at (x_0,y_0)

Example 5 on page 813

The error in the standard linear approximation

$|E(x,y)| \leq ½ M (|x-x_0|+|y-y_0|)^2$

Example 6 on page 814

Total differential of f:

$df = f_x(x_0,y_0)\, dx + f_y(x_0,y_0)\, dx$

Example 7-9 on pages 815-816

Functions of more than two variables *Example 10 on page 816*

14.7 Extreme values and saddle points

Let f(x,y) be defined on a region R containing the point (a,b). Then

1. f(a,b) is a **local maximum** value of f if $f(a,b) \geq f(x,y)$ for all domain points (x,y) in an open disk centered at (a,b)
2. f(a,b) is a **local minimum** value of f if $f(a,b) \leq f(x,y)$ for all domain points (x,y) in an open disk centered at (a,b)

First derivative test for local extreme values

If f(x,y) has a local maximum or minimum value at an interior point (a,b) of its domain and if the first partial derivatives exist there, then $f_x(a,b) = 0$ and $f_y(a,b) = 0$.

Critical point: an interior point of the domain of a function f(x,y) where both f_x and f_y are zero or where on or both of f_x and f_y do not exist.

A differentiable function f(x,y) has a **saddle point** at a critical point if in every open disk centered at (a,b) there are domain points (x,y) where $f(x,y) > f(a,b)$ and domain points (x,y) where $f(x,y) < f(a,b)$. The corresponding point (a,b, f(a,b)) on the surface z=f(x,y) is called a saddle point of the surface.

Examples 1 and 2 on page 822

Suppose that f(x,y) and its first and second partial derivatives are continuous throughout a disk centered at (a,b) and that $f_x(a,b) = f_y(a,b) = 0$. Then

1. f has a **local maximum** at (a,b) if $f_{xx} < 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a,b)
2. f has a **local minimum** at (a,b) if $f_{xx} > 0$ and $f_{xx}f_{yy} - f_{xy}^2 > 0$ at (a,b)
3. f has a **saddle point** at (a,b) if $f_{xx}f_{yy} - f_{xy}^2 < 0$ at (a,b)
4. the test is **inconclusive** at (a,b) if $f_{xx}f_{yy} - f_{xy}^2 = 0$ at (a,b)

Example 3 and 4 on page 823

We organize the search for the absolute extrema of a continuous function f(x,y) on a closed and bounded region R into three steps:

Example 5 and 6 on page 824-825

1. **List the interior points** of R where f may have local maxima and minima and evaluate f at these points
2. **List the boundary points** of R where f has local maxima and minima and evaluate f at these points.
3. **Look through the lists for the maximum and minimum values of f.** These will be the absolute maximum and minimum values of f on R. Since absolute maxima and minima are also local maxima and minima, the absolute maximum and minimum values of f appear somewhere in the lists made in steps 1 and 2.

For the exam, you have to memorize the standard integrals 1-7 and 19 on page 453 and the trigonometry identities

Summary 2: OR models for premaster IEM

Chapter 1: An Introduction to Model Building

1.1 An Introduction to Modelling

- **Static model:** a model in which the decision variables do not involve sequences of decisions over multiple periods
- **Dynamic model:** a model in which the decision variables do involve sequences of decisions over multiple periods

- **Linear models:** decision variables are multiplied by constants and added together
- **Nonlinear models**: e.g. exponential

- **Integer:** only integer values allowed
- **Non integer:** also fractions allowed

- **Deterministic:** for any value of the decision variables, the value of the objective function and whether or not the constraints are satisfied is known with certainty
- **Stochastic:** uncertainty

1.2 The Seven-Step Model-Building Process

1. Formulate the problem
2. Observe the system
3. Formulate a mathematical model of the problem
4. Verify the model and use the model for prediction
5. Select a suitable alternative
6. Present the results and conclusion of the study to the organization
7. Implement and evaluate recommendations

1.3 – 1.5 Examples cases

Chapter 3: Introduction to Linear Programming

3.1 What Is a Linear Programming Problem?

Linear programming problem (LP): an optimization problem for which we do the following:

1. We attempt to maximize (or minimize) a linear function of the *decision variables*. The function that is to be maximized or minimized is called the *objective function*.
2. The values of the *decision variables* must satisfy a set of *constraints.* Each constraint must be a linear equation or linear inequality.
3. A *sign restriction* is associated with each variable. For any x_i the sign restriction specifies that x_i must be either nonnegative ($x_i \geq 0$) or *unrestricted in sign (urs).*

Decision variables: describe the decision to be made

Objective function: the function to be maximized (revenue/profit) or minimized (costs) e.g.
maximize $z = \underline{3}x_1 + \underline{2}x_2$ → 3 and 2 are **price coefficients**

Constraints: restrictions

e.g. $\underline{2}x_1 + \underline{3}x_2 \leq \underline{100}$ → 2 and 3 are **technological coefficients**, 100 is the **right hand side (rhs):** often represents the quantity of a resource that is available

Sign restriction: the decision variables can only have positive values → $x_1 \geq 0$

Unrestricted in sign (urs): both positive and negative values.

Assumptions:

- **Proportionality assumption of linear programming:** It takes exactly three times as many finishing hours to manufacture three soldiers as it takes to manufacture one soldier.
- **Additivity assumption of linear programming:** The contribution of a variable to the left-hand side of each constraint is independent of the values of the variable. No matter what the value of x_1, the manufacture of x_2 trains uses x_2 finishing hours and x_2 carpentry hours.
- **Divisibility assumption:** each decision variable be allowed to assume fractional values. A linear programming problem in which some or all of the variables must be nonnegative integers is called an **integer programming problem.** It depends on the situation if rounding off leads to a reasonable solution.
- **The certainty assumption:** each parameter (objective function coefficient, right hand side and technological constraint) is known with certainty.

Optimal solution to an LP: a point in the *feasible region* with the largest/smallest (maximize/minimize) objective function value. Most LPs have only one optimal solution, however some LPs have none, others have an infinite number of solutions.

The feasible region: for an LP is the set of all points that satisfies all the LPs constraints and sign restrictions. **Infeasible point:** any point that is not in the feasible region.

3.2 The Graphical Solution of Two-Variable Linear Programming Problems

Isoprofit/isocost line: a line on which all points have the same z-value. If one isoprofit line is found, all can be found by moving parallel to the isoprofit line. The last isoprofit line intersecting (touching) the feasible region defines the largest z-value of any point in the feasible region and indicates the optimal solution.

A constraint is **binding** if the left-hand side and the right-hand side of the constraint are equal when the optimal values of the decision variables are substituted into the constraint. A constraint is **nonbinding** if the left-hand side and the right-hand side of the constraint are unequal when the optimal values of the decision variables are substituted into the constraint.

A set of points S is a **convex set** if the line segment joining any pair of points in S is wholly contained in S. For any convex set S, a point P in S is an **extreme point** if each line segment that lies completely in S and contains the point P has P as an endpoint of the line segment.

[Example 2: Dorian Auto on page 60-62]

3.3 Special cases

1. Some LPs have an infinite number of optimal solutions (alternative or multiple optimal solutions): [example 3 on page 63-64]. Goal programming is often used to choose between alternative optimal solutions
2. Some LPs have no feasible solutions (infeasible LPs): [example 4 on page 65-66].
3. Some LPs are unbounded: there are points in the feasible region with arbitrarily large (or small) z-values. Probably caused by an error made when formulating the LP: [example 5 on page 67].

3.4 – 3.12: Different applications of linear programming in different areas

Diet problems, work-scheduling problems, capital budgeting problems, short-term financial planning, blending problems, production process models, multiperiod decision problems, multiperiod financial models, multiperiod work-scheduling.

Chapter 9: Integer Programming

9.1 Introduction to Integer Programming

- **Pure integer programming:** all variables are required to be integers
- **Mixed integer programming:** some of the variables are required to be integer
- **0-1 integer programming:** all variables must equal 0 or 1
  The LP obtained by omitting all integer or 0-1 constraints on variables is called the LP relaxation of the IP → the feasible region for any IP must be contained in the feasible region for the corresponding LP relaxation: *optimal z-value for LP relaxation ≥ optimal z-value for IP*

9.2 Formulating Integer Programming Problems

Fixed charge problem: a cost is associated with performing an activity at a nonzero level that does not depend on the level of the activity. Gandhi example: if we make any shirts at all (no matter how many we make), we must pay the fixed charge of €200 to rent a shirt machine.

[Example 3 on page 480-482]

Set-covering problem: each member of a given set 1 must be "covered" by an acceptable member of some set 2. The objective is to minimize the number of elements in set 2 that are required to cover all elements in set 1, e.g. cities to be reached within a certain time from fire stations or hospitals.

[Example 5 on page 486-487]

Either-or constraints

When at least one of the formulas below should be satisfied.

a. $f(x_1, x_2, \ldots, x_n) \leq 0$
b. $g(x_1, x_2, \ldots, x_n) \leq 0$

This can be ensured by adding the two constraints below:

c. $f(x_1, x_2, \ldots, x_n) \leq My$
d. $g(x_1, x_2, \ldots, x_n) \leq M(1-y)$

Here, y is a 0-1 variable and M is a number chosen large enough to ensure that the constraints are satisfied for all values $x_1, x_2, \ldots, x_n$ (that satisfy the other constraints in the problem).

- ➔ If y = 0, then f ≤ 0 and g ≤ M. So, if y = 0, then a (and possibly b) must be satisfied
- ➔ If y = 1, then f ≤ M and g ≤ 0. So, if y = 1, then b (and possibly a) must be satisfied ➔ So: whether y = 0 or y = 1, c and d ensure that at least one of a and b is satisfied [Example 6 on page 488-489]

If-then constraints

If a is satisfied, than b should be satisfied too. (If a is not satisfied b may or may not be satisfied.) a. $f(x_1, x_2, \ldots, x_n) > 0$

b. $g(x_1, x_2, \ldots, x_n) \geq 0$

This can be ensured by adding the two constraints below:

c. $-g(x_1, x_2, \ldots, x_n) \leq My$

d. $f(x_1, x_2, \ldots, x_n) \leq M(1-y)$

Here, y is a 0-1 variable and M is a number chosen large enough to ensure that the constraints are satisfied for all values $x_1, x_2, \ldots, x_n$ (that satisfy the other constraints in the problem).

- ➔ If $f > 0$, then d can only be satisfied if $y = 0$. Then c implies $-g \leq 0$ or $g \geq 0$, which is the desired result. Thus, if $f > 0$, then c and d ensure that $g \geq 0$.
- ➔ If $f > 0$ is <u>not satisfied</u>, then d allows $y = 0$ or $y = 1$. For $y = 1$, c is automatically satisfied. Thus if $f > 0$ is not satisfied, then the values of $x_1, x_2, \ldots, x_n$ are unrestricted and $g < 0$ or $g \geq 0$ are both possible.

  [Example on page 490 (under if-then constraints)]

Chapter 15: Deterministic EOQ Inventory Models

15.1 Introduction to Basic Inventory Models

Inventory models answer the following questions:

- When should an order be placed for a product?
- How large should each order be?

Costs involved in inventory models

- Ordering and setup costs
- Unit purchasing costs
- Holding or carrying costs
- Stockout or shortage costs
    - Back-ordering
    - Lost sales

The Economic Order Quantitiy (EOQ) Model Assumptions:

- Repetitive ordering
- Constant demand
- Constant lead time
- Continuous ordering

15.2 The Basic Economic Order Quantity Model

Assumptions:

1. Demand is deterministic and occurs at a constant rate
2. If an order of any size (say, q units) is placed, an ordering and setup cost K is incurred
3. The lead time for each order is zero
4. No shortages are allowed
5. The cost per unit-year of holding inventory is h

Assumption 2 implies that the unit purchasing cost p does not depend on the size of the order. In section 15.3, quantity discounts are discussed.

Assumption 3 implies that each order arrives as soon as it is placed. This is relaxed later in this section 15.2.

Assumption 4 implies that all demands must be met on time: negative inventory is not allowed. In section 15.5 this assumption is relaxed.

Assumption 5 implies that a carrying cost of h dollars is incurred if 1 unit is held for one year, 2 units for half a year or ¼ units for four years. In short, if I units are held for T years, a holding cost of ITh is incurred.

Derivation of Basic EOQ Model

Since orders arrive instantaneously, we should never place an order when *I* (inventory level) is greater than zero. If we place an order when *I* > 0, we are incurring an unnecessary holding cost. On the other hand, if *I* = 0, we must place an order to prevent a shortage from occurring. So, every time an order is placed, we are facing the same situation, i.e. *I* = 0. This means, each time we place an order, we should order the same quantity *q.*

We now determine the value of *q* that minimizes annual cost (*q*).*

$$TC(q) = annual\ cost\ of\ placing\ orders + annual\ purchasing\ cost + annual\ holding\ cost$$

Since each order is for *q* units, D/q orders per year have to be placed so that annual demand *D* is met. Hence

$$\frac{ordering\ cost}{year} = \left(\frac{ordering\ cost}{order}\right)\left(\frac{orders}{year}\right) = \frac{KD}{q}$$

$$\frac{purchasing\ cost}{year} = \left(\frac{purchasing\ cost}{unit}\right)\left(\frac{units\ purchased}{year}\right) = pD$$

To compute the annual holding cost, not that if we hold *I* units for a period of one year, we incur a holding cost of (*I* units)(1 year)(*h* dollars/unit/year) = *hI* dollars. Suppose the inventory level is not constant and varies over time. If the average inventory level during a length of time *T* is *I,* the holding cost for the time period will be *hTI.* See figure 1 on page 850.

If we define *I(t)* to be the inventory level at time *t*, then during the interval [0, *T*], the total inventory cost is given by:

$$h(area\ from\ 0\ to\ T\ under\ the\ I(t) curve) = hTI$$

More formally, the inventory level from time 0 to time *T* is given by:

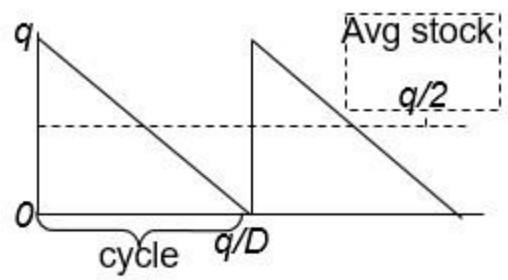
$$I(t) = \frac{\int_0^T I(t)dt}{T}$$

and the total holding cost incurred between time 0 and time *T* is:

$$\int_0^T hI(t)dt = hTI(T)$$

To determine the annual holding cost, we need to examine the behaviour of *I* over time. Assume that an order of size *q* has just arrived at time 0. Since demand occurs at a rate of D/year, it will take q/D years for inventory to reach zero again. Since demand during any period of length *t* is *Dt*, the inventory level over any time interval will decline along a straight line of slope *–D.* When inventory reaches zero, an order of size *q* is placed and arrives instantaneously, raising the inventory level back to *q*.

Cycle: any interval of time that begins with the arrival of an order and ends the instant before the next order is received



The figure consists of repeated cycles of length q/D. Hence, each year will contain 1/(1/D) = D/q cycles. The average inventory level is simply half of the maximum inventory level attained during the cycle. This result will hold in any model for which demand occurs at a constant rate and no shortages are allowed. Thus, for our model, the average inventory level during a cycle will be q/2 units.

We are now ready to determine the annual holding cost:

$$\frac{holding\ cost}{year} = \left(\frac{holding\ cost}{cycle}\right)\left(\frac{cycles}{year}\right)$$

Since the average inventory level during each cycle is q/2 and each cycle is of length q/D,

Then,

$$\frac{holding\ cost}{year} = \frac{q^2h}{2D}\left(\frac{D}{q}\right) = \frac{hq}{2}$$

Combining ordering cost, purchasing cost and holding cost, we obtain

$$\boldsymbol{TC(q) = \frac{KD}{q} + pD + \frac{hq}{2}}$$

To find the value of *q* that minimizes TC(q), we set TC’(q) equal to zero:

$$TC'^{(q)} = -\frac{KD}{q^2} + \frac{h}{2} = 0$$

This is satisfied for *q* = $(2KD/h)^{1/2}$. Therefore:

$$\boldsymbol{q^* = EOQ = \sqrt{\frac{2KD}{h}}}$$

Proof that annual holding cost = annual ordering cost, when EOQ (q*) is ordered:
This is illustrated in figure 3 on page 852.

$$\frac{holding\ cost}{year} = \frac{hq^*}{2} = \frac{h}{2}\left(\frac{2KD}{h}\right)^{1/2} = \left(\frac{KDh}{2}\right)^{1/2}$$

$$\frac{ordering\ cost}{year} = \frac{KD}{q^*} = \frac{KD}{\left(\frac{2KD}{h}\right)^{1/2}} = \left(\frac{KDh}{2}\right)^{1/2}$$

[Example 1 on EOQ on page 852]

Sensitivity of total cost to small variations in the order quantity

HC(q) = annual holding cost if the order quantity is q

OC(q) = annual ordering cost if the order quantity is q

Holding costs *(h)* and ordering costs *(K)* are often hard to estimate. However, the curve of HC(q) + OC(q) is very flat near q* (see figure 4 on page 853). This means that even if *h* and *K* are inaccurate, costs will only increase a little. In the example of page 853, ordering 20% more than the EOQ raises HC(q) + OC(q) only from 20 to 20.33 (which is an increase of under 2%). Purchasing costs are not taken into account here, because these are unaffected by the order quantity *q.*

Determination of EOQ when holding cost is expressed in terms of dollar value of inventory Sometimes the annual holding cost is expressed in terms of the cost of holding one dollar's worth of inventory for one year (h_d). Then the cost of holding one unit of inventory for one year will be ph_d instead of *h*. [Example 1 on EOQ on page 852]

The effect of a nonzero lead time

A nonzero lead time leaves the annual holding and ordering costs unchanged. Therefore, the EOQ still minimizes total costs. To prevent shortages from occurring and to minimize holding costs, each order must be placed at an inventory level that ensures that when each order arrives, the inventory level will be zero. So, the **reorder point** (the inventory level at which an order should be placed) must be determined. In order to do so, two cases should be considered:

1. *Demand during lead time does not exceed the EOQ (LD ≤ EOQ)*
   In this case, the reorder point occurs when the inventory level equals LD. Then the order will arrive L time units later, and upon arrival of the order, the inventory level will equal LD – LD = 0. E.g. in example 1, suppose lead time is one month (1/12 year) and demand is 500, an order should be placed at LD = (1/12)*500 = 41,67 units.
2. *Demand during lead time exceeds the EOQ (LD > EOQ)*
   The reorder point does not equal LD. Suppose that in example 1, L = 15 months = 15/12 year. Then LD = (15/12)500 = 625 units. Why can't we place an order each time the inventory level

reaches 625 units? Because the EOQ = 250 and the inventory level will never reach 625. To determine the correct reorder point, observe that orders are placed every six months. Suppose that an order has just arrived at time 0. Then an order must have been placed L = 15 months ago (at T = -15 months). Since orders arrive every six months, orders must be placed at T = -9 months, T = -3 months, T = 3 months, … Since at T = 0 an order has just arrived, the inventory at T = 0 is 250. Then at T = 3 (or any other point when an order is placed), the inventory level will equal 250 – (3/12)(500) = 125. So, the reorder point is at 125 units. In general, the reorder point equals the remainder when LD is divided by the EOQ: 625/250. 2*250 = 500. The remainder is 625 – 500 = 125.

[Example 3, noninventory problem on page 855]

Power-of-two ordering policies: page 857

15.3 Computing the Optimal Order Quantity When Quantity Discounts Are Allowed

Till now on, we have assumed that the annual purchasing cost does not depend on the order size. This allowed us (in section 15.2) to ignore the annual purchasing cost when calculating the EOQ. In real life, suppliers often reduce the unit purchasing price for large orders, these are called quantity discounts. If a supplier gives quantity discounts, the annual purchasing cost will depend on the order size. Sometimes, holding costs are expressed as a percentage of an item's purchasing cost. In that case, also the annual holding cost will depend on the order size. Because purchasing costs now depend on the order size, we cannot ignore these costs anymore and a new approach is needed.

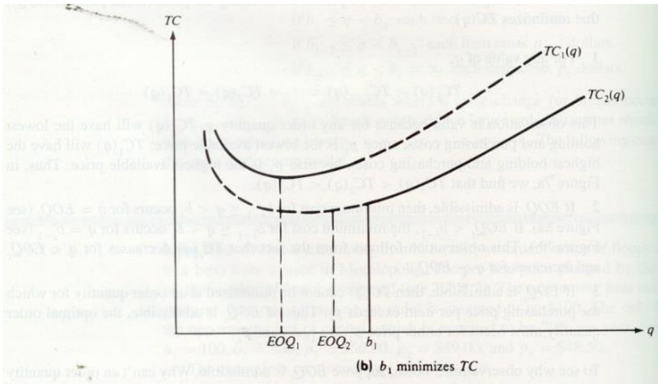
If we let *q* be the quantity ordered each time an order is placed, the general quantity discount model analysed in this section, may be described as follows:

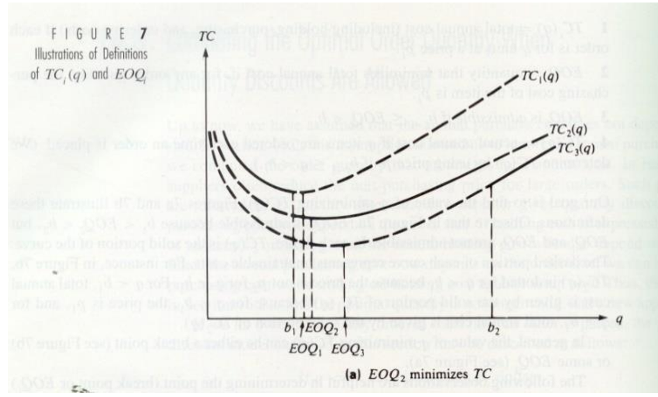
- If $q < b_1$, each item costs p_1 dollars
- If $b_1 \le q < b_2$, each item costs p_2 dollars
- If $b_{k-2} \le q < b_{k-1}$, each item costs p_{k-1} dollars
- If $b_{k-1} \le q < b_k = \infty$, each item costs p_k dollars

Price break points: the points where a price change (or break) occurs, i.e. $b_1, b_2, \ldots, b_{k-1}$. Since larger quantities should be associated with lower prices, $p_k < p_{k-1} < \ldots < p_2 < p_1$.

Definitions:

- $TC_i(q)$ = total annual cost (including holding, purchasing and ordering costs) if each order is for *q* units at a price p_i
- EOQ_i = quantity that minimizes total annual cost if, for any order quantity, the purchasing cost of the item is p_i
- EOQ_i is admissible if $b_{i-1} \le EOQ_i < b_i$
- TC(q) = actual annual cost if *q* items are ordered each time an order is placed (we determine TC(q) by using price p*i* if *bi-1* $\le q < b_i$)





Important observations:

- For any value of q,
  $TC_k(q) < TC_{k-1}(q) < \ldots < TC_2(q) < TC_1(q)$

  Meaning that for any order quantity, $TC_k(q)$ will have the lowest holding and purchasing costs, since p_k is the lowest available price. $TC_1(q)$ will have the highest holding and purchasing costs, because p_1 is the highest available price.
- If EOQ_i is admissible, then minimum cost for $b_{i-1} \leq q < b_i$ occurs for $q = EOQ_i$. If $EOQ_i < b_{i-1}$, the minimum cost for $b_{i-1} \leq q < b_i$ occurs for $q = b_{i-1}$. This observation follows from the fact that $TC_i(q)$ decreases for $q < EOQ_i$ and increases for $q > EOQ_i$.
- If EOQ_i is admissible, then TC(q) cannot be minimized at an order quantity for which the purchasing price per item exceeds p_i. Thus, if EOQ_i is admissible, the optimal order quantity must occur for either price p_i, p_{i+1}, … or p_k.

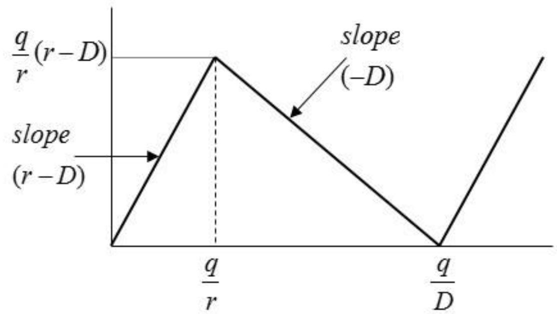
[Example 4, EOQ with price discounts on page 863]

Determining EOQ with price discounts

1. *Calculate EOQ for the lowest price and see if EOQ is admissible*
   E.g. for p = 48,50 and 300 ≤ q. This yields EOQ = 143,59. This is not admissible, because 143,59 < 300. Therefore, for q ≥ 300, TC(q) is minimized by q* = 300.
2. *If not, continue with second lowest price*
   E.g. for p = 49,00 and 100 ≤ q < 300. This yields EOQ = 142,86. This is admissible, because 100 ≤ EOQ < 300. Therefore, higher prices cannot yield the order quantity minimizing TC(q). So, either q* = 300 or q* = 142,86.
3. Calculate TC(q)'s
   In this case TC(300) and TC(142,86). For TC(300) take p = 48,50. For TC(142,86) take p = 49. TC(300) < TC(142,86). Therefore q* = 300 will minimize TC(q).

15.4 The Continuous Rate EOQ Model

If a company makes its own products, the continuous rate EOQ model will be more realistic than the traditional EOQ model. This model assumes that a firm can produce a good at a rate of *r* units per time period. This means that during any time period of length *t*, the firm can produce *rt* units. Again, we assume that demand is deterministic and occurs at a constant rate; we also assume that shortages are not allowed.



q = number of units produced during each production run

K = cost of setting up production run

h = cost of holding one unit in inventory for one year

D = annual demand for the product

Assuming that per-unit production costs are independent of run size, we must determine the value of q that minimizes annual holding costs + annual setup costs. Since demand occurs at a constant rate, we know that the average inventory level = ½ * maximum inventory level. The maximum inventory level occurs at time q/r. Since between zero and q/r, the inventory level is increasing at a rate of r-D units per year, the inventory level at time q/r will be (q/r) (r-D). So, the average inventory level will be (1/2) (q/r) (r-D), or

$$\frac{Holding\ cost}{year} = h(average\ inventory) = \frac{h(r-D)q}{2r}$$

Since the ordering costs are the same, i.e. KD/q, this yields

$$\text{q} * \text{(optimal run size)} = \sqrt{\frac{2r}{h(r-D)}KD} = \sqrt{\frac{2KD}{h}} * \sqrt{\frac{r}{r-D}} = EOQ * \sqrt{\frac{r}{r-D}}$$

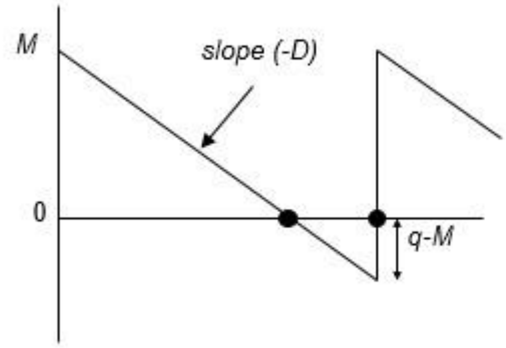
For large r, the rate model EOQ should approach the instantaneous delivery situation of the EOQ model. This is true because for a large *r*, r/(r-D) approaches 1, so that q* approaches the EOQ.

[Example 5 on page 867]

15.5 The EOQ Model with Back Orders Allowed

We assume that all demand is backlogged and no sales are lost. Let *s* be the cost of being short one unit for one year. Let q-M be the maximum shortage that occurs under an ordering policy. Assuming zero lead time, the firm will be q-M units short each time an order is placed.

Since purchasing cost do not depend on *q* and *M*, we can minimize annual cost by determining the values of *q* and *M* that minimize annual holding costs + annual shortage costs + annual order costs.



$$A = \frac{M}{D}$$

$$B - A = \frac{q-M}{D}$$

$$\frac{D}{q}\ cycles/year$$

See summary slides for derivation

15.6 When to Use EOQ Models

One assumption for using EOQ models is that demand should be constant. However, in practice, demand is often irregular. To determine if demand is sufficiently regular to justify the use of EOQ models, calculate the variability coefficient (VC):

1. Determine the estimate d of the average demand per period by

$$d = \frac{1}{n}\sum_{i=1}^{n} d_i$$

2. Determine an estimate of the variance of the per-period demand D by

$$Est.\ var\ D = \frac{1}{n}\sum_{i=1}^{i=n} d^2{}_i - d^2$$

3. Determine an estimate of the relative variability of demand (VC) by

$$VC = \frac{est.\ var\ D}{d^{-2}}$$

Demand for the next 6 weeks is given in the following table:

Week	1	2	3	4	5	6
Demand	20	50	40	30	40	30

1. d = (20 + 50 + 40 + 30 + 40 + 30) / 6 = 35
2. Est. var D = (1/6) (20^2 + 50^2 + 40^2 + 30^2 + 40^2 + 30^2) - 35^2 = 91,667
3. VC = 91,667 / 35^2 = 0,075

VC < 0.20 so an EOQ model can be used.

Chapter 16: Probabilistic Inventory Models

All inventory models discussed in chapter 15 require that demand is known with certainty. In this chapter, inventory with uncertain demand are considered.

16.1 Single-Period Decision Models

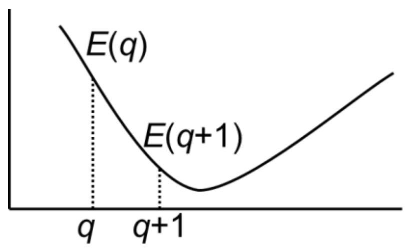
After *q* has been determined, the value d assumed by a random variable *D* is observed. Depending on the values *d* and *q*, the decision maker incurs a cost c(d, q). We assume that the decision maker wants to choose *q* so that costs are minimized. Since the decision is made only once, we call this a singleperiod decision model.

16.2 The Concept of Marginal Analysis

We now assume that *D* is an integer-valued discrete random variable with P(D = d) = p(d). Let E(q) be the expected cost if q is chosen. Then

$$E(q) = \sum_{d} p(d)c(d, q)$$

E(q) is a convex function of q. Then, the graph looks like this:



Let q* be the value of q that minimizes E(q). We see that q* is the smallest value of q for which E(q*+1) – E(q*) ≥ 0. This is the change in expected cost that occurs if we increase the decision variable *q* to *q + 1*.

Marginal analysis: determining q* by repeatedly computing the effect of adding a marginal unit to the value of q.

Approach: begin with q = 0. If E(1) – E(0) ≤ 0, we can benefit by increasing 1 from 0 to 1. Now check to see whether E(2) – E(1) ≤ 0. If this is true,, also increasing q from 1 to 2 will reduce expected costs. Continuing will show that increasing q by 1 will reduce expected costs up to the point where we try to increase q from q* to q*+1. In this case, increasing q by 1 will increase expected costs.

16.3 The News Vendor Problem: Discrete Demand

Organizations often face inventory problems where the following sequence of events occur:

1. The organization decides how many units to order. We let q be the number of units ordered
2. With probability p(d), a demand of d units occurs. In this section, we assume that d must be a nonnegative integer. We let D be the random variable representing demand
3. Depending on d and q, a cost c(d,q) is incurred

News vendor problem: a vendor of newspapers will have costs if demand is higher than the amount of newspapers on hand, but also is demand is lower than the amount of newspapers on hand. In the first case because he could have sold more if he had more newspapers (because there is still demand), in the second case because he has too many newspapers and they are worthless at the end of the day. The news vendor must order the number of papers that properly balances these two costs.

Overstocking cost: $c(d, q) = c_0 q \qquad for\ (d \leq q)$

Understocking cost: $c(d, q) = -c_u q \quad for\ (d \geq q + 1)$

To derive the optimal order quantity via marginal analysis, let E(q) be the expected cost if an order is placed for q units. Find the value q* that minimizes E(q). So, we must determine the smallest value of q for which E(q + 1) – E(q) ≥ 0. There are two possibilities:

1. $d \leq q$: in this case, ordering q+1 units instead of q units causes us to be overstocked by one more unit. This increases costs by c_o. The probability that $d \leq q$ will occur is $P(D \leq q)$.
2. $d \geq q + 1$: in this case, ordering q+1 units instead of q units enables us to be short one less unit. This will decrease cost by c_u. The probability that $d \geq q + 1$ = 1 - $P(D \leq q)$.

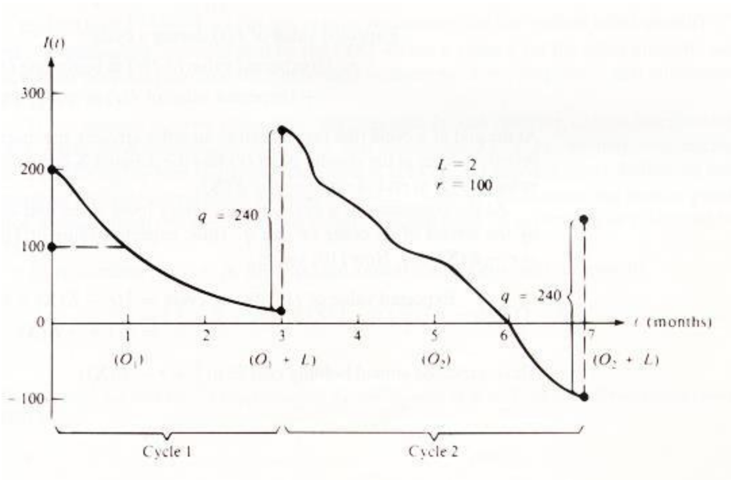
$$E(q+1) - E(q) \geq 0 \ \ will\ hold\ if \ \ P(D \leq q) \geq \frac{c_u}{c_o + c_u}$$

[Example 1 on page 883]

16.6 The EOQ with Uncertain Demand: The (r,q) and (s, S) Models

In this section we discuss a modification of the EOQ that is used when **lead time is nonzero** and the **demand during each lead time is random**. We begin by assuming that all demand can be **backlogged**. As in chapter 15, we assume a continuous review model, so that orders may be placed at any time and we define:

- **K** = ordering cost
- **h** = holding cost/unit/year
- **L** = lead time for each order (assumed to be known with certainty)
- **q** = quantity ordered each time an order takes place
- **D** = random variable (assumed continuous) representing annual demand, with mean E(D), variance var D and standard deviation σ_D
- $\mathbf{C_B}$ = cost incurred for each unit short, which does not depend on how long it takes to make up stockout
- **OHI(t)** = on-hand inventory (amount of stock on hand) at time t
- **B(t)** = number of outstanding back orders at time t
- **I(t)** = net inventory level at time t = OHI(t) – B(t)
- **r** = inventory level at which order is placed (reorder point)
- **X** = random variable representing demand during lead time



X is an continuous random variable having density function f(x) and mean, variance and standard deviation of E(X), var X and σ_X respectively

If we assume that the demands at different points in time are independent, it can be shown that the random lead time demand satisfies:

E(X) = LE(D), var X = L(var D), $\sigma_X = \sigma_D\sqrt{L}$

If D is normally distributed, X will also be normally distributed. Suppose we allow the lead time L to be a random variable L, with mean E(L), variance var L and standard deviation σ_L. If the length of the lead time is independent of the demand per unit time during the lead time, then

E(X) = E(L)E(D) and var X = E(L)(var D) + E(D)^2(var L)

Determination of Reorder Point: The Back-Ordered Case

All demand must eventually be met and no sales are lost. We assume fixed purchasing costs (each unit is purchased for the same price).TC(q,r) = expected annual cost incurred if each order is for q units and is placed when the reorder point is r. Then TC(q,r) = (expected annual holding cost) + (expected annual ordering cost) + (expected annual cost due to shortages).

$$Expected\ value\ of\ I(t) during\ a\ cycle = (\ expected\ value\ at\ beginning\ of\ cycle + expected\ value\ at\ end\ of\ cycle\ ) / 2$$

At the end of a cycle (the instant before an order arrives), the inventory level will equal the inventory level at the reorder point (r) less the demand X during lead time. Thus, **the expected value of I(t) at the end of cycle = r – E(X).** At the beginning of a cycle, the inventory level at the end of the cycle is augmented by the arrival of an order of size q. Thus, **the expected value of I(t) at the beginning of cycle = r – E(X) + q.** So:

$$Expected\ value\ of\ I(t) during\ a\ cycle = \frac{r - E(X) + r - E(X) + q}{2} = \frac{q}{2} + r - E(X)$$

$$Expected\ annual\ holding\ cost = h\left(\frac{q}{2} + r - E(X)\right)$$

B_r = random variable representing the number of stockouts or back orders during a cycle if the reorder point is r

$$\frac{\textit{Expected shortage cost}}{\textit{cycle}} = C_B E(B_r)$$

$$\textit{Expected annual shortage cost} = \frac{C_B E(B_r) E(D)}{q}$$

$$\textit{Expected annual order cost} = K\left(\frac{\textit{expected orders}}{\textit{year}}\right) = \frac{KE(D)}{q}$$

$$TC(q,r) = h\left(\frac{q}{2} + r - E(X)\right) + \frac{C_B E(B_r) E(D)}{q} + \frac{KE(D)}{q}$$

q* may be adequately approximated by the EOQ. Given a value q, marginal analysis can be used to determine a reorder point r* that minimizes TC(q,r). If we assume that q is given, the expected annual order costs are independent of r. Therefore, when minimizing TC(q,r) these can be excluded. Increase the reorder point from r to r+Δ, will TC increase or decrease?

$$P(X \geq r^*) = \frac{hq^*}{C_B E(D)}$$

If the right hand side is larger than 1, the equation will have no solution and holding cost is prohibitively high relative to the stockout cost. Management should set the reorder point at the smallest acceptable level. The same if the equation yields a negative value for r*.

[Example 5 on page 894]

Determination of Reorder Point: The Lost Sales Case

Now we assume that all stockouts result in lost sales and that a cost of C_{LS} dollars is incurred for each lost sale. Realize that expected inventory in lost sales case = (expected inventory in back-ordered case) + (expected number of shortages per cycle). This equation follows because in the lost sales case, we find that during each cycle, an average of (expected shortages per cycle) fewer orders will be filled from inventory, thereby raising the average inventory level by an amount equal to expected shortages per cycle. In the lost sales case equation, the right hand side is smaller than in the back-ordered case equation. Thus, the lost sales assumption will yield a lower stockout probability (and a larger reorder point and safety stock level) than the back-ordered assumption.

$$P(X \geq r^*) = \frac{hq^*}{hq^* + C_{LS} E(D)}$$

Continuous Review (r,q) Policies

(r,q) policy: a continuous review inventory policy, in which we order a quantity q whenever our inventory level reaches a reorder level r. Also called an two-bin policy.

Continuous Review (s,S) Policies

Suppose that a demand for more than one unit can arrive at a particular time. Then an order may be triggered when the inventory level is less than r, and our computation of expected inventory level at the end and beginning of a cycle is incorrect. We see that if demands are larger than 1 unit at the same time, then the (r,q) model may not yield a policy that minimizes expected annual cost. Here, a (s,S) policy is optimal: an order is placed when whenever the inventory level is less than or equal to s. The size of the order is sufficient to raise the inventory level to S.

16.7 The EOQ with Uncertain Demand: The Service Level Approach to Determining Safety Stock Level

Managers often decide to control shortages by meeting a specified service level. We assume that all shortages are backlogged.

- Service Level Measure 1: **SLM_1**, the expected fraction of all demand that is met on time
- Service Level Measure 2: **SLM_2**, the expected number of cycles per year during which a shortage occurs

[Example 6 on page 898]

Determination of Reorder Point and Safety Stock Level for SLM_1

$$1 - SLM_1 = \frac{expected\ shortages\ per\ year}{expected\ demand\ per\ year} = \frac{E(B_r)}{q}$$

Can be used to determine the reorder point that yields a desired service level. We now assume that lead time demand is normally distributed, with mean E(X) and standard deviation σ_X. If X is normally distributed, the determination of $E(B_r)$ requires knowledge of the normal loss function.

Normal loss function: NL(y) is defined by the fact that $\sigma_X NL(y)$ is the expected number of shortages that will occur during a lead time if (1) lead time demand is normally distributed with mean E(X) and standard deviation σ_X and (2) the reorder point is E(X) + y σ_X.

In short, if we hold y standard deviations (in terms of lead time demand) of safety stock, then $NL(y)\sigma_X$ is the expected number of shortages occurring during a lead time. For example NL(2) = 0.0085 means that if the reorder point exceeds the mean lead time demand by $2\sigma_X$, then an average of $0.0085\sigma_X$ shortages will occur during a given lead time.

$$E(B_r) = \sigma_X NL\left(\frac{r - E(X)}{\sigma_X}\right)$$

So:

$$1 - SLM_1 = \frac{\sigma_X NL\left(\frac{r - E(X)}{\sigma_X}\right)}{q}$$

With:

$$NL\left(\frac{r - E(X)}{\sigma_X}\right) = \frac{q(1 - SLM_1)}{\sigma_X}$$

[Example 7 on page 903]

Determination of Reorder Point and Safety Stock Level for SLM_2

Suppose that a manager wants to hold sufficient safety stock to ensure that an average of s_0 cycles per year will result in a stockout. Given a reorder point of r, a fraction P(X>r) of all cycles will lead to a stockout. With an average of E(D)/q cycles per year, an average of [P(X>r)E(D)] / q cycles per year will result in a stockout. Thus, given s_0, the reorder point is the smallest value of r satisfying:

$$P(X \geq r) = \frac{s_0 q}{E(D)} \qquad \textit{for continuous lead time demand}$$

$$P(X > r) \leq \frac{s_0 q}{E(D)} \qquad \textit{for discrete lead time demand}$$

16.8 (R,S) Periodic Review Policy

On-order inventory level: the sum of on-hand inventory and inventory on order.

Every R units of time, we review the on-hand inventory level and place an order to bring the on-order inventory level up to S. In general, an (R,S) policy will incur higher holding costs than a cost-minimizing (r,q) policy, but an (R,S) policy is usually easier to administer than a continuous review policy. With an (R,S) policy (unlike a continuous review policy), we can predict with certainty the times when an order will be placed. We assume that R already has been determined, that all shortages are backlogged, that demand is a continuous random variable and that the per-unit purchase price is constant (so that annual purchasing costs do not depend on our choice or R and S). We define:

- **R** = time (in years) between reviews
- **D =** demand (random) during a one-year period
- **E(D) =** expected demand during a one-year period
- **K =** cost of placing an order
- **J =** cost of reviewing inventory level
- **h =** cost of holding one item in inventory for one year
- **C_B =** cost per-unit short in the backlogged case
- **L** = lead time for each order (assumed constant)
- **D_{L+R} =** demand (random) during a time interval of length L + R
- **$E(D_{L+R})$** = mean of D_{L+R}
- **σ_{DL+R} =** standard deviation of D_{L+R}

The expected costs are given by:

(Annual expected purchase costs) + (annual review costs) + (annual ordering costs) + (annual expected holding costs) + (annual expected shortage costs)

$$Expected\ annual\ holding\ cost = h\left[S - E(D_{L+R}) + \frac{E(D)R}{2}\right]$$

The value of S minimizing the sum of annual expected holding and shortage costs will occur for the value of S satisfying

$$P(D_{L+R} \geq S) = \frac{Rh}{C_B}$$

[Example 8 on page 910]

Suppose that all shortages result in lost sales, and a cost of C_{LS} is incurred for each lost sale. Then the value of S minimizing the sum of annual expected holding and shortage costs is given by

$$P(D_{L+R} \geq S) = \frac{Rh}{Rh + C_{LS}}$$

Determination of R

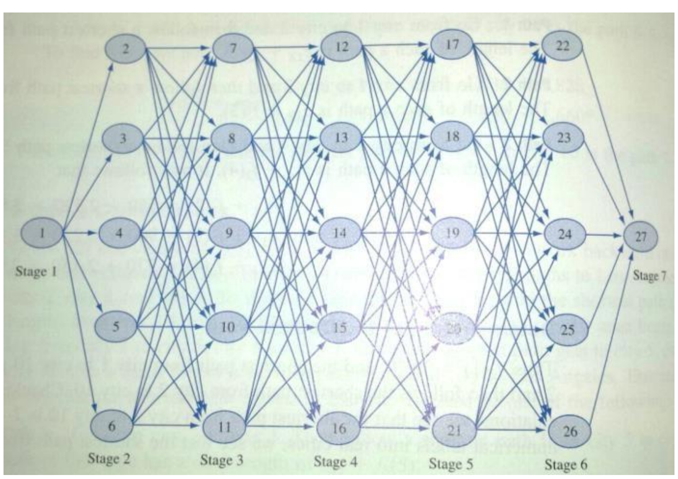
Often, the review interval R is set equal to EOQ/E(D). This makes the number of orders placed per year equal the number recommended if a simple EOQ model were used to determine the size of orders. Since each order is accompanied by a review, however, we must set the cost per order to K+J (instead of K). This yields

$$EOQ = \sqrt{\frac{2(K+J)E(D)}{h}}$$

Chapter 18: Deterministic EOQ Inventory Models

18.1 Two Puzzles Introduction 18.2 A Network Problem

[Example 3 on page 963 – 965]



In the example 3, enumerating all possible paths is easy because there are only 18. However, sometimes the problem is bigger (see figure to the left). Then, for example $f_5(21)$ can be found by the following equation:

$f_5(21)$ = min {c_{21},j + f_6(j)} *with (j = 22, 23, 24, 25, 26)*

Characteristics of dynamic programming applications

1. *The problem can be divided into stages with a decision required at each stage*
   In many problems,, the stage is the amount of time that has elapsed since the beginning of the problem. However, in some situations decisions are not required at every stage.
2. *Each stage has a number of states associated with it*
   State: the information that is needed at any stage to make an optimal decision.
3. *The decision chosen at any stage describes how the state at the current stage is transformed into the state at the next stage*
   However, in many problems a decision does not determine the next stage's state with certainty: then the current decision only determines the probability of the state at the next stage.
4. *Given the current state, the optimal decision for each of the remaining stages must not depend on previously reached states or previously chosen decisions*
   The principle of optimality: e.g. if the shortest path from city 1 to city 10 is known to pass through city 2, then the shortest path from city 1 to city 10 must include a shortest path from city 2 to city 10 (2-5-8-10). This follows because any path from city 1 to 10 that passes through city 2 and does not contain a shortest path from city 2 to city 10 will have a length of c_{12} + [something bigger than $f_2(2)$]. Of course, such a path cannot be a shortest path from city 1 to city 10.
5. *If the states for the problem have been classified into one of T stages, there must be a recursion that relates the cost or reward earned during stages t, t + 1, … T to the cost or reward earned from stages t + 1, t + 2, …, T.*
   In example 3, the recursion could have been written as $\mathbf{f_t(i) = \min \{ c_{ij} + f_{t+1}(j) \}}$ where j must be a stage t+1 city and $f_5(10) = 0$.

First backward, then making choices (as in example 3)

The use the recursion, we begin by finding the optimal decision for each state associated with the last stage. Then we use the recursion described in characteristic 5 to determine $f_{T-1}(\cdot)$ (along with the optimal decision) for every stage T-1 state. Then we use the recursion to determine $f_{T-2}(\cdot)$ (along with the optimal decision) for every stage T-2 state. We continue in this fashion until we have computed $f_1(i_1)$ and the optimal decision when we are in stage 1 and state i_1. Then our optimal decision in stage 1 is chosen from the set of decisions attaining $f_1(i_1)$. Choosing this decision at stage 1 will lead us to some stage 2 state (call it state i_2) at stage 2.

Then at stage 2, we choose any decision attaining $f_2(i_2)$. We continue in this fashion until a decision has been chosen for each state.

18.3 An Inventory Problem

An inventory problem with the following characteristics can be solved by dynamic programming:

1. Time is broken up into periods, the present period being period 1, the next period 2, and the final period T. At the beginning of period 1, the demand during each period is known.
2. At the beginning of each period, the firm must determine how many units should be produced. Production capacity during each period is limited.

3. Each period's demand must be met on time from inventory or current production. During any period in which production takes place, a fixed cost of production as well as a variable per-unit cost is incurred.
4. The firm has limited storage capacity. This is reflected by a limit on end-of-period inventory. A per-unit holding cost is incurred on each period's ending inventory.
5. The firm's goal is to minimize the total cost of meeting on time the demands for periods 1, 2, ..., T.

In this model, the inventory position is reviewed at the end of each period and then the production decision is made (**periodic review model**). If setup costs are exclude, the inventory model in [Example 4 on page 970] is similar to the Sailco inventory problem in section 3.10 (compare).

18.4 Resource-Allocation Problems

Limited resources must be allocated among several activities (similar to e.g. Giapetto problem). However, linear programming can only be used if the following three assumptions are fulfilled:

1. The amount of a resource assigned to an activity may be any nonnegative number
2. The benefit obtained from each activity is proportional to the amount of the resource assigned to the activity
3. The benefit obtained from more than one activity is the sum of the benefits obtained from the individual activities..

If assumptions 1 and 2 do not hold, dynamic programming can still be used if assumption 3 is valid and when the amount of the resource allocated to each activity is a member of a finite set.

[Example 5 on page 975]

Generalized resource allocation problem

Suppose we have w units of a resource available and T activities to which the resource can be allocated. If activity t is implemented at a level x_t (we assume x_t must be a nonnegative integer), then $g_t(x_t)$ units of the resource are used by activity t and benefit $r_t(x_t)$ is obtained. The problem of determining the allocation of resources that maximizes total benefit subject to the limited resource availability ay be written as:

$$\max \sum_{t=1}^{t=T} r_t(x_t) \qquad s.t. \sum_{t=1}^{t=T} g_t(x_t) \leq w$$

Where x_t must be a member of {0, 1, 2, …}.

To solve the function above by dynamic programming, define $f_t(d)$ to be the maximum benefit that can be obtained from activities t, t+1, …, T if d units of the resource may be allocated to activities t, t_1, …, T:

$F_{T+1}(d) = 0$ for all d

$F_t(d) = \max\{r_t(x_t) + f_{t+1}[d - g_t(x_t)]\}$

where xt must be a nonnegative integer satisfying $g_t(x_t) \leq d$. Let $x_t(d)$ be any value of x_t that attains $f_t(d)$. To use the formulas above to determine an optimal allocation of resources to activities 1, 2, …, T, we begin by determining all $f_T(\cdot)$ and $X_T(\cdot)$. We then use the last two formulas to determine all $f_{T-1}(\cdot)$ and $X_{T-1}(\cdot)$, continuing to work backward in this fashion until all $f_2(\cdot)$ and $X_2(\cdot)$ have been determined. We now calculate $f_1(w)$ and $x_1(w)$. Then we implement activity 1 at a level $x_1(w)$. At this point, we have $w - g_1[x_1(w)]$ units of the resource available for activities 2, 3, …, T. Then activity 2 should be implemented at a level of $x_2\{w - g_1[x_1(w)]$. We continue in this fashion until we have determined the level at which all activities should be implemented.

Solution of knapsack problems by dynamic programming

[Example 6 on page 979] + alternative recursion

Turnpike theorem
c_j = benefit obtained from each type j item w_j =
weight of each type j item

In terms of benefit per unit weight, the best item is the item with the largest value of c_j/w_j. Assume there are n types of items that have been ordered, so that

$$\frac{c_1}{w_1} \geq \frac{c_2}{w_2} \geq \cdots \geq \frac{c_n}{w_n}$$

Thus, type 1 items are the best, type 2 items are the second best and so on. Recall that it is possible for the optimal solution to a knapsack problem to use none of the best item. Assume that

$$\frac{c1}{w1} > \frac{c2}{w2}$$

So, there is a unique best item type. It can be shown that for some number w*, it is optimal to use at least one type 1 item if the knapsack is allowed to hold w pounds, where w ≥ w*. This result holds for

$$w^* = \frac{c_1 w_1}{c_1 - w_1(\frac{c_2}{w_2})}$$

18.5 Equipment-Replacement Problems

[Example 7 on page 985] + alternative recursion

18.6 Formulating Dynamic Programming Recursions

$f_t(i) = \min\{(\text{cost during stage } t) + f_{t+1}\ (\text{new state at stage } t+1)\}$

$f_t(i)$ is the minimum cost incurred from stage t to the end of the problem given that at stage t the state is i. The minimum cost incurred from stage t to the end of the problem must be attained by choosing at stage t an allowable decision that minimizes the sum of the costs incurred during the current stage (stage t) plus the minimum cost that can be incurred form stage t+1 to the end of the problem. A correct formulation of a recursion of this form requires that we identify three important aspects of the problem:

1. **The set of decisions that is allowable, or feasible, for the given state and stage.**
   This often depends on both t and i. E.g. example 18.3: d_t *= demand during month t* i_t *= inventory at beginning of month t*
   In this case, the set of allowable month t decisions consists of the members of {0, 1, 2, 3, 4, 5} that satisfy $0 \le (i_t + x_t - d_t) \le 4$.
2. **We must specify how the cost during the current time period (stage t) depends on the value of t, the current state, and the decision chosen at stage t.** E.g. example 18.3:
   Suppose a production level x_t *is chosen during month t. Then the cost during month t is given by* $c(x_t) = (1/2)(i_t+x_t-d_t)$.
3. **We must specify how the state at stage t+1 depends on the value of t, the state at stage t, and the decision chosen at stage t**

   Example 18.3: *month t+1 state is* $i_t+x_t-d_t$

[Example 8 on page 990 (fishery)]

Incorporating the time value of money into dynamic programming formulations

The formulation above says that profits received during later years are weighted the same as profits received during earlier years. However, later profits should be weighted less. E.g. $\beta < 1$, \$1 received at the beginning of year t+1 is equivalent to β dollars received at the beginning of year t. See page 991.

[Example 9 on page 991 (power plant)]

[Example 10 on page 992 (wheat sale)]

[Example 11 on page 993 (refinery capacity)]

[Example 12 on page 994 (traveling salesperson)]

Nonadditive recursions

[Example 13 on page 997 (minimax shortest route)]

[Example 14 on page 994 (sales allocation)]

18.7 The Wagner-Whitin Algorithm

Dynamic lot-size model:

1. Demand d_t during a period $t(t = 1,2,\ldots,T)$ is known at the beginning of period 1
2. Demand of period t must be met on time from inventory or from period t production. The cost c(x) of producing x units during any period is given by c(0)=0 and for x>0, c(x)= K + cx, where K is

a fixed cost for setting up production during a period and c tis the variable per-unit cost of production

3. At the end of period t, the inventory level i_t is observed, and a holding cost hi_t is incurred. We let i_0 denote the inventory level before period 1 production occurs
4. The goal is to determine a production level x_i for each period t that minimizes the total cost of meeting (on time) the demands of period 1, 2, …, T.
5. There is a limit c_t placed on period t's ending inventory
6. There is a limit r_t placed on period t's production

Discussion of the Wagner Within algorithm on page 1002

[Example 15 on page 1003]

Chapter 20: Queuing Theory

20.1 Some queuing terminology

Queue discipline:

- FCFS (first come, first served)
- LCFS (last come, first served)
- SIRO (service in random order
- Priority queuing disciplines

20.2 Modelling arrival and service processes

We define t_i to be the time at which the ith customer arrives. For $i \geq 1$, we define $T_i = t_{i+1} - t_i$ to be the ith interarrival time. Thus in figure 1 (page 1053), $T_1 = 8 - 3 = 5$ and $T_2 = 15 - 8 = 7$. In modelling the arrival process, we assume that the T_i's are independent, continuous random variables described by the random variable A. We assume that A has a density function a(t):

$$P(A \leq c) = \int_0^c a(t)dt$$

and

$$P(A > c) = \int_c^\infty a(t)dt$$

$$Average\ interarrival\ time = \frac{1}{\lambda} = \int_0^\infty ta(t)dt$$

$$Arrival\ rate = \lambda$$

An exponential distribution is the most common choice for A. An exponential distribution with parameter λ has a density $a(t) = \lambda e^{-\lambda t}$.

The average interarrival time E(A) is given by $E(A) = 1/\lambda$.

Using the fact that A = $E(A^2) - E(A)^2$, we can show that var A = $1/\lambda^2$.

No-memory property of the exponential distribution

If we know that at least t time units have elapsed since the last arrival occurred, then the distribution of the remaining time until the next arrival (h) does not depend on t.

[See Lemma 1 on page 1054]

Relation between Poission distribution and exponential distribution

Interarrival times are exponential with parameter λ if and only if the number of arrivals to occur in an interval of length t follows a Poisson distribution with parameter λt. Assumptions:

1. Arrivals defined on nonoverlapping time intervals are independent (for example, the number of arrivals occurring between times 1 and 10 does not give us any information about the number of arrivals occurring between times 30 and 50).
2. For small Δt (and any value of t), the probability of one arrival occurring between times t and t+Δt is λΔt + o(Δt), where o(Δt) refers to any quantity satisfying $\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0.$

If assumptions 1 and 2 hold, then N_t follows a Poisson distribution with parameter λt and interarrival times are exponential with parameter λ; that is $a(t) = \lambda e^{-\lambda t}$. This theorem states that if the arrival rate is stationary, if bulk arrivals cannot occur, and if past arrivals do not affect future arrivals, then interarrival times will follow an exponential distribution with parameter λ and the number of arrivals in any interval of length t is Poisson with parameter λt.

The Erlang distribution

If interarrival times do not appear to be exponential, they are often modelled by an Erlang distribution. An Erlang distribution is a continuous random variable (T) whose density function f(t) is specified by two parameters: a rate parameter R and a shape parameter k (k must be a positive integer):

$$f(t) = \frac{R(Rt)^{k-1}e^{-Rt}}{(k-1)!}$$

Using integration by parts, we can show that if T is an Erlang distribution with rate parameter R and shape parameter k, then

$$E(T) = \frac{k}{R} \qquad \text{and} \qquad var\ T = \frac{k}{R^2}$$

See figure 4: for a large k, the Erlang distribution behaves like a normal distribution. For extremely large values of k, the ERlang distribution approaches a random variable with zero variance (that is, a constant interarrival time). If we model interarrival times as an Erlang distribution with shape parameter k, we are really saying that he interarrival process is equivalent to a customer going through k phases (each of which has the no-memory property) before arriving. Therefore, the shape parameter is often referred to as the number of phases of the Erlang distribution.

Modelling the service process

We assume that the service times of different customers are independent random variables and that each customer's service time is governed by a random variable S having a density function s(t). We let 1/μ be the mean service time for a customer.

$$\frac{1}{\mu} = \int_0^\infty ts(t)\,dt$$

The variable $\frac{1}{\mu}$ will have units of hours per customer, so μ has units of customers per hour. For this reason, we call μ the service rate. For example, $\mu = 5$ means that if customers were always present, the server could serve an average of 5 customers per hour and the average service time would be 1/5 hour.
No memory property: Consider a three-server system in which each customer's service time is governed by an exponential distribution s(t)= $\mu e^{-\mu t}$. Suppose all three servers are busy, ,and a customer is waiting. What is the probability that the customer who is waiting will be the last of the four customers to complete service? One of customers 1-3 will be the first to complete service. Then customer 4 will enter service. By the no-memory property, customer 4's service time has the same distribution as the remaining service ties of customers 1 and 2. Thus, by symmetry, customers 4, 1 and 2 will have the same chance of being the last customer to complete service. This implies that customer 4 has a 1/3 chance of being the last customer to complete service.

Unfortunately, actual service times may not be consistent with the no-memory property. For this reason, we often assume that s(t) is an Erlang distribution with shape parameter k and rate parameter $k\mu$. This yields a main service time of $1/\mu$. Modelling service times as an Erlang distribution with shape parameter k also implies that a customer's service time may be considered to consist of a passage through k phases of service, in which the time to complete each phase has the no-memory property and a mean of $1/k\mu$.

If interarrival and service times are deterministic, each interarrival time will be exactly 1/ λ and customer's each service time will be exactly 1/μ.

The Kendall-Lee Notation for Queuing Systems Assumptions:

- One single line
- s identical parallel servers
- FCFS discipline

Six characteristics: 1/2/3/4/5/6 1.
Nature of the arrival process:

- M = interarrival times are independent, identically distributed (iid) random variables having an exponential distribution
- D = interarrival times are iid and deterministic
- E_k = interarrival times are iid Erlangs with shape parameter k
- GI = interarrival times are iid and governed by some general distribution

2. Nature of the service times
   - M = service times are iid and exponentially distributed

- D = service times are iid and deterministic
- E_k = service times are iid Erlangs with shape parameter k
- G = service times are iid and follows some general distribution

3. The number of parallel servers
4. The queue discipline
   - FCFC = first come, first served
   - LCFS = last come, first served
   - SIRO = service in random order
   - GD = general queue discipline
5. The maximum allowable number of customers in the system
6. The size of the population from which customers are drawn

In many models 4/5/6 = GD/∞/∞, then 4/5/6 is often omitted.

E.g. $M/E_2/8/FCFS/10/\infty$ might represent a health clinic with 8 doctors, exponential interarrival times, two-phase Erlang service times, an FCFS queue discipline and a total capacity of 10 patients.

The waiting time paradox

If A is the random variable for the time between buses, then the average time until the next bus (as seen by an arrival who is equally likely to come at any time) is given by

$$\textit{Expected waiting time} = \frac{1}{2}\left(E(A) + \frac{var\ A}{E(A)}\right)$$

20.3 Birth-death processes

State: the number of people present in any queuing system at time t.

$P_{ij}(t)$: the probability that j people will be present in the system at time t, given that at time 0, i people are present.

Steady state or equilibrium probability: π_j, the fraction of time that j customers are present.

Birth-death process: a continuous-time stochastic process for which the system's state at any time is a nonnvegative integer.

1. With probability $\lambda_j\Delta t + o(\Delta t)$, a birth occurs between time t and time t + Δt. A birth increases the system state by 1, to j+1. The variable λ_j is called the **birth rate** in state j. In most queuing systems a birth is simply an arrival.
2. With probability $\mu_j\Delta t + o(\Delta t)$, a death occurs between time t + Δt. A death decreases the system state by 1, to j-1. The variable μ_j is the **death rate** in state j. $\mu_0 = 0$ must hold, so that no negative state could occur. In most queuing systems a death is a service completion.
3. Births and deaths are independent of each other.

Relation of exponential distribution to birth-death processes: page 1064 **Derivation of steady-state probabilities for birth-death processes:** page 1066

Flow balance equations

j=0 $p_0\lambda_0 = \mu_1 p_1$

j > 0 $(\lambda_{j-1})(p_{j-1}) + (\mu_{j+1})(p_{j+1}) = p_j(\lambda_j+\mu_j)$

Solution of flow balance equations

See page 1068 – 1069 p

$p_1 \quad \frac{{}_0\lambda_0}{\mu_1} =$

$$c_j = \frac{\lambda_0\lambda_1\ldots\lambda_{j-1}}{\mu_1\mu_2\ldots\mu_j}$$

$p_j = p_0 c_j$

Because the system has to be in some state at any given time, the steady-state probabilities must sum to 1:

$$\sum_{j=0}^{j=\infty} p_j = 1 \qquad so \qquad p_0\left(1+\sum_{j=1}^{j=\infty} c_j\right) = 1 \qquad and \qquad p_0 = \frac{1}{1+\sum_{j=1}^{j=\infty} c_j}$$

20.4 The M/M/1/GD/∞/∞ queuing system and the queuing formula L = λW

$$p_1 = \frac{p_0\lambda}{\mu}, \qquad p_2 = \frac{p_0\lambda^2}{\mu^2}, \qquad p_j = \frac{p_0\lambda^j}{\mu^j}$$

$$\rho = \frac{\lambda}{\mu}$$

$$p_0(1+\rho+\rho^2+\cdots) = 1$$

$p_0 = 1 - \rho \qquad and \qquad p_j = \rho^j(1-\rho)$

Derivation of L, L_q and L_s: page 1073 – 1074

$$L = \frac{\lambda}{\mu-\lambda}$$

$$L_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$$

$L_s = L - L_s$

Queuing formula/Little's queuing formula λ = average number of arrivals *entering* the system per unit time

L = average number of customers present in the queuing system

L_q = average number of customers waiting in line

L_s = average number of customers in service

W = average time a customer spends in the system

W_q = average time a customer spends in line

W_s = average time a customer spends in service

- $L = \lambda W = \frac{\rho}{1-\rho}$
- $W = \frac{L}{\lambda} = \frac{1}{\mu-\lambda}$
- $L_q = \lambda W_q = \frac{\lambda^2}{\mu(\mu-\lambda)}$
- $L_s = \lambda W_s$

[Example 3 on page 1076]
[Example 4 on page 1076]
[Example 5 on page 1078]
[Example 6 on page 1079]
[Example 7 on page 1079]

20.5 The M/M/1/GD/c/∞ queuing system

When c customers are present, new arrivals are turned away and are forever lost to the system.
$\lambda_j = \lambda$ $\lambda_c =$
0 $\mu_0 = 0$
$\mu_j = \mu$

$$p_0 = \frac{1-\rho}{1-\rho^{c+1}}$$
$$p_j = \rho^j p_0$$

$$W = \frac{L}{\lambda(1-p_c)}$$

[Example 8 on page 1084]

20.6 The M/M/s/GD/∞/∞ queuing system

If j ≤ s customers are present, then all j customers are in service. If j > s customers are present, then all s servers are occupied and j-s customers are waiting in line.

$$\rho = \frac{\lambda}{s\mu}$$

$$P(j \geq s) = \frac{(s\rho)^s \pi_0}{s!\,(1-\rho)}$$

$$L_q = \frac{\mathrm{P(j \geq s)}\rho}{1-\rho}$$

$$W_q = \frac{\mathrm{P(j \geq s)}}{s\mu - \lambda}$$

$$\mathrm{L} = \mathrm{L_q} + \frac{\lambda}{\mu}$$

$$W = \frac{\mathrm{P(j \geq s)}}{s\mu - \lambda} + \frac{1}{\mu}$$

[Example 9 on page 1089]

[Example 10 on page 1089]

20.7 The M/G/∞/GD/∞/∞ and GI/G/∞/GD/∞/∞ Models

Infinite service (or self service)

$$L = \frac{\lambda}{\mu}$$

$$\pi_j = \frac{\left(\frac{\lambda}{\mu}\right)^j e^{-\frac{\lambda}{\mu}}}{j!}$$

[Example 11 on page 1096]

20.8 The M/G/1/GD/∞/∞ queuing system

Here, service time distribution is not exponential and the service times do not longer have the momemory property.

$$L_q = \frac{\lambda^2\sigma^2 + \rho^2}{2(1-\rho)}$$

$$L = L_1 + \rho$$

$$W_q = \frac{L_q}{\lambda}$$

$$W = W_q + \frac{1}{\mu}$$

$$L = \frac{\lambda}{\mu - \lambda}$$

20.9 Finite Source Models: The Machine Repair Model

There are two situations where the assumption of the state-independent arrivals rates may be invalid:

1. If customers do not want to buck long lines, the arrival rate may be a decreasing function of the number of people present in the queuing system
2. If arrivals to a system are drawn from a small population the arrival rate may greatly depend on the state of the system.

Finite source models: models in which arrivals are drawn from a small population. M/M/R/GD/K/K

See table 8 and figure 22 on page 1100
$The\ total\ rate\ at\ which\ a\ breakdown\ occurs$ = $\lambda_j = \lambda + \lambda + \cdots + \lambda = (K - j)\lambda$
Formulas see page 1100 - 1101

20.11 The M/G/s/GD/s/∞ system (blocked customers cleared)

Blocked customers cleared: arrivals when no service is available are turned away and forever lost to the system.

$$L = L_s = \frac{\lambda(1 - p_s)}{\mu}$$

Erlang's loss formula: p_s depends on the service time distribution only through its mean ($1/\mu$).

[Example 15 on page 1112]

Summary 3: OR models for premaster IEM

(Integer) Linear Programming

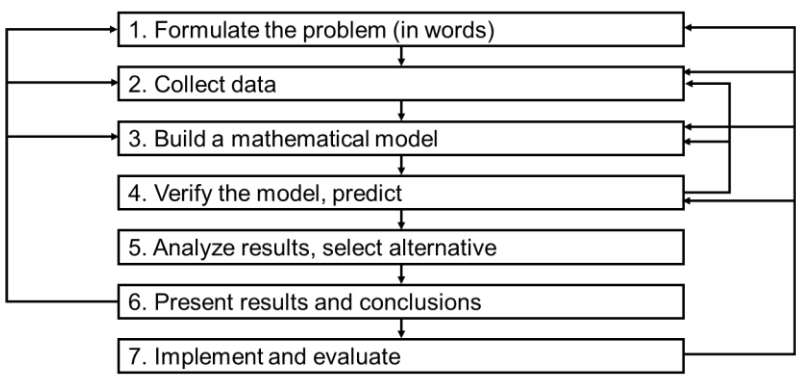
Advantages using quantitative models

- Focus on most important factors in decision making (skip details)
- Obligation to structure the problem and collect data
- Obligation to quantify information
- Experiment outside reality

Disadvantages using quantitative models

- Model might be inadequate:
- Important factors forgotten
- Problem forced in a known model
- Wrong relations specified

Problem solving approach



Summation signs

$$\sum_{t=1}^{n} x_t = \sum_{k=1}^{n} x_k = x_1 + x_2 + \dots + x_n$$

$$\sum_{t=m}^{n} x_t = x_m + x_{m+1} + \dots + x_n$$

$$\sum_{i=1}^{m}\sum_{j=1}^{n} x_{ij} = (x_{11} + x_{12} + \dots + x_{1n}) + (x_{21}+x_{22} + \dots + x_{2n}) + \dots + (x_{m1}+x_{m2} + \dots + x_{mn})$$

[Example productmix problem (Giapetto)]

- Manufacturing and sales of soldiers and trains
- Requirements on carpentry and finishing
- Constraints on available capacities
- Problem: What production plan maximizes weekly prc

Production	Carpentry	Finishing
Costs per hour	$ 6	$ 4
Capacity (hours/week)	80	100

Products	Soldier	Train
Sales price per unit	$ 27	$ 21
Material costs per unit	$ 10	$ 9
Carpentry hours per unit	1	1
Finishing hours per unit	2	1
Maximum sales per week	40	∞

MODEL

Decision variables

x_1 = number of soldiers produced each week
x_2 = number of trains produced each week

Objective function

MAX $z = \underline{3}x_1 + \underline{2}x_2$ — price coefficients

Restrictions

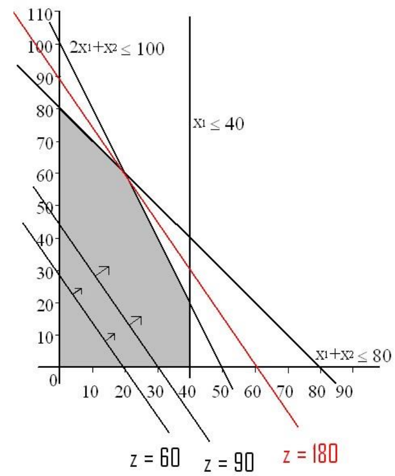
$\mathbf{2}x_1 + x_2 \le \mathit{100}$	(finishing)	**technological coefficient**
$x_1 + x_2 \le \mathit{80}$	(carpentry)	*right hand side*
$x_1 \le \boldsymbol{\mathit{40}}$	(demand)	
$x_1 \ge 0, x_2 \ge 0$	(sign restrictions)	

Terminology

- Linear programming problem: optimize a linear function under linear constraints and sign restrictions
- Optimize: minimize or maximize
- Constraints: equations or inequalities (=, ≤, ≥)
- Sign restrictions: variables are nonnegative or unrestricted in sign (urs)

How to solve an LP?

1. <u>Graphically</u>
   - For 2 decision variables only
2. Apply algorithm
3. Use computer program (LINGO)



The graph shows the following solution is optimal:

z = 180, x_1 = 20 and x_2 = 60

- The feasible region S is a convex set (i.e. the line segment joining two points in S is wholly contained in S)
- The optimal solution is a corner point (also called an extreme point)
- The optimal solution can be found by moving an isoprofit line in a direction that increases the *z*-value (or an isocost line in a direction that decreases the *z*value)

Special cases

- LP has an alternative optimal solution (isoprofit line parallel to a restriction)
- LP in infeasible (constraints are too restrictive)
- LP is unbounded (constraints are missing)

Problem formulation process

1. Detailed verbal description of the problem
2. Determine the overall objective
3. Determine the constraints
4. Define the decision variables. (Explicitly state units of measurement!)
5. Formulate the objective function. (Use all decision variables!)
6. Formulate the constraints
7. Check the entire formulation (especially for linearity)

[Example work scheduling at post office]

- Number of full-time employees needed on day j: c_j
- Work 5 consecutive days, then receive 2 days off
- Problem: minimize the number of full-time employees needed

MODEL Decision variables

x_j = number of full-time employees working on day j

Objective function

$z = \Sigma_j x_j$

Restrictions

s.t. $x_1 + \quad x_4+x_5+x_6+x_7 \geq c_1$ (Mo)

$x_1+x_2+ \quad\quad x_5+x_6+x_7 \geq c_2$ (Tu)

... ...

$x_3+x_4+x_5+x_6+x_7 \geq c_7$ (Su)

$x_j \geq 0$, j=1..

[Example production of sailboats]

- Demand for the next four quarters: $d_1 = 40$, $d_2 = 60$, $d_3 = 75$, $d_4 = 25$ units
- Production capacity with regular-time labor: 40 boats
- Production cost in regular-time: $400 per sailboat
- Production cost in overtime: $450 per sailboat
- Starting inventory: $i_0 = 10$ sailboats
- Production during a quarter can be used to meet demand in that quarter
- Inventory cost: $20 per quarter per sailboat

Problem: determine a production schedule that minimizes the sum of production and inventory costs

MODEL

Decision variables

x_t = number of sailboats produced by regular-time labor during quarter t (t = 1, 2, 3, 4)

y_t = number of sailboats produced by non regular-time labor during quarter t (t = 1, 2, 3, 4) i_t = number of sailboats on hand at end of quarter t (t = 1, 2, 3, 4)

Objective function

Min $400(x_1 + x_2 + x_3 + x_4) + 450(y_1 + y_2 + y_3 + y_4) + 20(i_1 + i_2 + i_3 + i_4)$

Constraints

$i_t = i_{t-1} + x_t + y_t - d_t$ (t = 1, 2, 3, 4)

inventory balance at the end of quarter t = inventory at the end of quarter t-1 – demand quarter t
$x_t \leq 40$ (t = 1, 2, 3, 4)
regular-time production should not exceed 40
$x_t, y_t, i_t \geq 0$, t = 1, 2, 3, 4
demand should be met in time

[Example marketing]
MSI is specialized in evaluating consumer reaction to new products, services and advertising campaigns. For a large client MSI has to conduct 1000 personal interviews with a number of households across the country. Some interviews will take place during daytime and some during

the evenings, households with and without children should be interviewed. The interview costs are as follows.

	Day	Evening
Children	€20	€25
No children	€18	€20

The following constraints have to be fulfilled:

1. In total 1000 interviews
2. At least 400 households with children
3. At least 400 households without children
4. Total number of evening interviews at least as great as total number of daytime interviews
5. At least 40 per cent of interviews for households with children must be conducted during the evening
6. At least 60 per cent of interviews for households without children must be conducted during the evening

MODEL Decision variables

DC = number of daytime interviews of households with children

EC = number of evening interviews of households with children

DNC = number of daytime interviews of households without children

ENC = number of evening interviews of households without children

Objective function

MIN $20DC + 25EC + 18DNC + 20ENC$

Restrictions

- *1000 interviews*
  $DC + EC + DNC + ENC = 1000$
- *At least 400 households with children*
  $DC + EC \geq 400$
- *At least 400 households without children*
  $DNC + ENC \geq 400$
- *Total number of evening interviews at least as great as total number of daytime interviews*
  $EC + ENC \geq DC + DNC$ or $EC + ENC - DC - DNC \geq 0$
- *At least 40 per cent of interviews for households with children must be conducted during the evening*
  $EC \geq 0.4(DC + EC)$ or $-0.4DC + 0.6EC \geq 0$
- *At least 60 per cent of interviews for households without children must be conducted during the evening*
  $ENC \geq 0.6(DNC + ENC)$ or $-0.6DNC + 0.4ENC \geq 0$

Assumptions

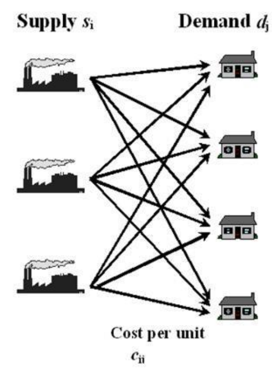
1. Proportionality
2. Additivity
3. Divisibility *(rounding off gives good solutions in general, otherwise ILP)*
4. Certainty *(sensitivity analysis)*

An assignment problem is a balanced transportation problem for which $s_i = d_j = 1$ for all *i* and *j,* e.g.

assigning machines to jobs, assigning individuals to tasks

[Example assignment problem]

- Powerco has 3 electric power plants that supply 4 cities
- Problem: minimize the cost of meeting each city's power demand



m plants, n cities

MODEL Decision variables

x_{ij} = number of kwh produced at plant *i* and sent to city *j*

Objective function

$$\sum_{i=1}^{m}\sum_{j=1}^{n} c_{ij}x_{ij}$$

Constraints

$\sum_{j=1}^{n} x_{ij} \leq s_i,\ i = 1, \dots, m$ *supply*

$\sum_{i=1}^{m} x_{ij} \geq d_j,\ j = 1, \dots, n$ *demand*

$x_{ij} \geq 0,\ i = 1, \dots, m,\ j = 1, \dots, n$ *sign restrictions*

[Example PerfumeCo]

PerfumeCo manufactures Brute and Chanelle perfumes. The raw material needed to manufacture each type of perfume can be purchased for \$3 per pound. Processing 1 lb of raw material requires 1 hour of laboratory time. Each pound of processed raw material yields 3 oz of Regular Brute Perfume and 4 oz of Regular Chanelle Perfume. These perfumes can be sold for \$7/oz and \$6/oz, respectively. Each year PerfumeCo has 6000 hours of laboratory time available and can purchase up to 4000 lb of raw material. Assume that the cost of laboratory hours is a fixed cost. [Remark: 8 oz. = 0.5 lb.]

How can PerfumeCo maximize yearly profit?

MODEL

Decision variables x_1 = number of ounces of Regular Brute sold annually x_3 = number of ounces of Regular Chanelle sold annually x_5 = number of pounds of raw material purchased annually

Optimization function

MAX $7x_1 + 6x_3 - 3x_5$

Restrictions $x_5 \leq 4000$ $x_5 \leq 6000$ $x_1 - 3x_5 = 0$ $x_3 - 4x_5 = 0$

$x_i \geq 0$, i = 1,3,5

New situation

Consider the PerfumeCo case described on slide 43. Assume that PerfumeCo also has the option of further processing Regular Brute and Regular Chanelle to produce Luxury Brute, sold at \$18/oz, and Luxury Chanelle, sold at \$14/oz. Each ounce of Regular Brute processed further requires an additional 3 hours of laboratory time and \$4 of processing cost and yields 1 oz of Luxury Brute. Each ounce of Regular Chanelle processed further requires an additional 2 hours of laboratory time and \$4 of processing cost and yields 1 oz of Luxury Chanelle.

How can PerfumeCo maximize yearly profit in this case?

Differences with the situation described earlier?

- Regular Brute/ Chanelle can be sold or processed further
- Laboratory time can be used only once, i.e. it can be used to produce regular or it can be used to produce luxury perfume

MODEL Decision variables

x_1 = number of ounces of Regular Brute sold annually

x_2 = number of ounces of Luxury Brute sold annually

x_3 = number of ounces of Regular Chanelle sold annually
x_4 = number of ounces of Luxury Chanelle sold annually

x_5 = number of pounds of raw material purchased annually

Objective function

Max $7x_1 + 14x_2 + 6x_3 + 10x_4 - 3x_5$

(Contribution to profit = revenues from perfume sales – processing costs – costs of processing raw materials = $7x_1 + 18x_2 + 6x_3 + 14x_4 - (4x_2 + 4x_4) - 3x_5$)

*Ounces of Regular Brute sold + ounces of Luxury Brute sold = ounces of Brute produced = (ounces of Brute produced/pound of raw material)*pounds of raw material purchased = $3x_5$*
$x_1 + x_2 = 3x_5$ or $x_1 + x_2 - 3x_5 = 0$

Ounces of Regular Chanelle sold + ounces of Luxury Chanelle sold = ounces of Chanelle produced
$x_3 + x_4 = 4x_5$ or $x_3 + x_4 - 4x_5 = 0$

Restrictions

$x_5 \leq 4000$

$3x_2 + 2x_4 + x_5 \leq 6000$

$x_1 + x_2 - 3x_5 = 0$ *ounces regular brute sold + luxury brute sold = brute produced*
$x_3 + x_4 - 4x_5 = 0$ *ounces of Regular sold + ounces of Luxury sold = total ounces produced* $x_i \geq 0$, $i = 1,\ldots,5$

[Example blending problem]

- Sunco manufactures 3 types of gasoline by blending 3 types of crude oil.
- Production cost: \$4 per barrel of gasoline
- Production capacity: 14000 barrels a day
- Max. 5000 barrels of each type of crude oil daily can be purchased
- \$1 advertising → 10 barrels extra demand
- Requirements on octane rating and sulfur content
- Known: sales and purchase prices
- Problem: maximize daily profits (= revenues – costs)

Gasoline	gas 1	gas 2	gas 3
sales price per barrel	\$ 70	\$ 60	\$ 50
min. octane rating	10	8	6
max. sulfur content	1%	2%	1%
daily demand in barrels	3000	2000	1000

Crude oil	crude 1	crude 2	crude 3
purchase price per barrel	\$ 45	\$ 35	\$ 25
octane rating	12	6	8
sulfur content	0.5%	2%	3%

MODEL Decision variables

- x_{ij} = number of barrels of crude oil i used daily to produce gas j (i = 1,2,3, j = 1,2,3)

- a_j = number of dollars spent daily on advertising gas *j* (*j* = 1,2,3)

Most difficult constraints: what makes a blend of crude oils gas 1 (or 2 or 3)?

A mixture may be sold as gas 1 if it has an octane rating ≥ 10 and a sulfur content ≤ 1%

- Number of barrels of gas *j* produced and sold $= x_{1j} + x_{2j} + x_{3j}$
- Number of barrels of crude oil *i* purchased $= x_{i1} + x_{i2} + x_{i3}$

Optimization function

Maximize: revenues gas– purchasing costs crude oil– production costs – advertising costs

Restrictions

- Revenues: $70*(x_{11} + x_{21} + x_{31}) + 60*(x_{12} + x_{22} + x_{32}) + 50*(x_{13} + x_{23} + x_{33})$
- Purchasing costs: $45*(x_{11} + x_{12} + x_{13}) + 35*(x_{21} + x_{22} + x_{23}) + 25*(x_{31} + x_{32} + x_{33})$
- Production costs: $4*(x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} + x_{31} + x_{32} + x_{33})$
- Advertising costs: $a_1 + a_2 + a_3$

- Production for each type of gas meets demand:
  $x_{11} + x_{21} + x_{31} = 3000 + 10*a_1$ (advertising effect!)

  $x_{12} + x_{22} + x_{32} = 2000 + 10*a_2$
  $x_{13} + x_{23} + x_{33} = 1000 + 10*a_3$

- Max. of 5000 barrels crude oil per type available for purchase:
  $x_{11} + x_{12} + x_{13} \leq 5000$
  $x_{21} + x_{22} + x_{23} \leq 5000$
  $x_{31} + x_{32} + x_{33} \leq 5000$

- Production capacity: no more than 14000 barrels of gas are to be produced:
  $x_{11} + x_{12} + x_{13} + x_{21} + x_{22} + x_{23} + x_{31} + x_{32} + x_{33} \leq 14000$ (Is this constraint redundant?)

A mixture may be sold as gas 1 if it has an octane rating ≥ 10 and a sulfur content ≤ 1%

- Octane:

  $(12x_{11} + 6x_{21} + 8x_{31})/(x_{11} + x_{21} + x_{31}) \geq 10$

  <u>This constraint is not allowed in an LP model, since it is nonlinear; rewrite it as:</u>

  $12x_{11} + 6x_{21} + 8x_{31} \geq 10(x_{11} + x_{21} + x_{31})$ or

  $2x_{11} - 4x_{21} - 2x_{31} \geq 0$

- Sulfur:

$(0.5x_{11} + 2x_{21} + 3x_{31})/(x_{11} + x_{21} + x_{31}) \le 1$
<u>Again nonlinear, rewrite as:</u>
$-0.5x_{11} + x_{21} + 2x_{31} \le 0$

- Restrictions on octane level:
  $(12x_{11} + 6x_{21} + 8x_{31})/(x_{11} + x_{21} + x_{31}) \ge 10 \rightarrow 2x_{11} - 4x_{21} - 2x_{31} \ge 0$

  $(12x_{12} + 6x_{22} + 8x_{32})/(x_{12} + x_{22} + x_{32}) \ge 8 \rightarrow 4x_{12} - 2x_{22} \ge 0$

  $(12x_{13} + 6x_{23} + 8x_{33})/(x_{13} + x_{23} + x_{33}) \ge 6 \rightarrow 6x_{13} + 2x_{33} \ge 0$

(Is this last constraint redundant?)

- Restrictions on sulfur level:

  $(0.5x_{11} + 2x_{21} + 3x_{31})/(x_{11} + x_{21} + x_{31}) \le 1 \rightarrow -0.5x_{11} + x_{21} + 2x_{31} \le 0$

  $(0.5x_{12} + 2x_{22} + 3x_{32})/(x_{12} + x_{22} + x_{32}) \le 2 \rightarrow -1.5x_{12} + x_{32} \le 0$

$(0.5x_{13} + 2x_{23} + 3x_{33})/(x_{13} + x_{23} + x_{33}) \le 1 \rightarrow -0.5x_{13} + x_{23} + 2x_{33} \le 0$ - Sign restrictions:

$x_{ij}, a_j \ge 0, i = 1,2,3, j = 1,2,3$

[Example capital budgeting]
Star Oil Company is considering five different investment opportunities. The cash outflows and net present values (in millions of dollars) are given in the table on the next slide. Star Oil has $40 million available for investment now (time 0); it estimates that one year from now (time 1) another $ 20 million will be available. Star Oil may purchase any fraction of each investment. In this case, the cash outflows and NPVs are adjusted accordingly. For example, if Star Oil purchases one-fifth of investment 3, then a cash outflow of 1/5*(5) = $1 million dollars would be required at time 1 and the NPV equals 1/5*16 = $3.2 million. Star Oil wants to maximize the NPV that can be obtained by investing in investments 1-5.

Formulate an LP that will help achieve this goal. Assume that any funds left over at time 0 cannot be used at time 1.

Investment →	1	2	3	4	5
Cash outflow ↓					
Time 0	11	53	5	5	29
Time 1	3	6	5	1	34
NPV	13	16	16	14	39

MODEL Decision variables

xi = fraction of investment i purchased by Star Oil (i = 1, 2, 3, 4, 5)

Objective function

MAX $13x_1 + 16x_2 + 16x_3 + 14x_4 + 39x_5$

Constraints

- *Star Oil cannot invest more than $40 million at time 0* $11x_1 + 53x_2 + 5x_3 + 5x_4 + 29x_5 \leq 40$
- *Star Oil cannot invest more than $20 million at time 1*
  $3x_1 + 6x_2 + 5x_3 + x_4 + 34x_5 \leq 20$
- *Star Oil cannot purchase more than 100% of investment i* *(i = 1, 2, 3, 4, 5)* $x_i \leq 1$ (i = 1, 2, 3, 4, 5)
- *Sign restrictions*
- $x_i \geq 0$ (*i* = 1, 2, 3, 4, 5)

Integer Linear Programming

Consider an IP problem. The LP obtained by omitting all integer or 0-1 constraints on variables is called the LP-relaxation of the IP. Rounding off the optimal solution of the LP-relaxation of an IP does not yield the optimal IP solution in general. It might even yield a solution that is not feasible for the IP!

MAX-problem:

Optimal z-value for LP-relaxation ≥ optimal z-value for IP

MIN-problem:

Optimal z-value for LP-relaxation ≤ optimal z-value for IP

[Example investment]

Stockco is considering 4 investment options. The company has $14,000 available at the moment. For each of the investments the cash outflow at this moment and the net present value are known. How can Stockco maximize the net present value obtained from investments 1-4?

INVESTMENT	CASH OUTFLOW (in $000)	NET PRESENT VALUE (in $000)
1	5	16
2	7	22
3	4	12
4	3	8

MODEL Decision variables

x_j = 1 if investment *j* is made, 0 otherwise, *j* = 1,2,3,4

Objective function

MAX $z = 16x_1 + 22x_2 + 12x_3 + 8x_4$

Restrictions

$5x_1 + 7x_2 + 4x_3 + 3x_4 \leq 14$

$x_j = 0$ or 1, $j = 1,\ldots,4$

[Example fixed-charge]

- Gandhi manufactures shirts, shorts and pants
- Each type of manufacturing requires an appropriate type of machinery
- Machinery is rented per week at a fixed rate
- Resource constraints (labor and cloth)
- Variable unit cost and selling price for each type of clothing

Problem: Maximize Gandhi's weekly profits

	LABOR (hrs)	CLOTH (sq yd)	SALES PRICE	VAR. COST	FIXED COST
SHIRTS	3	4	$12	$6	$200
SHORTS	2	3	$8	$4	$150
PANTS	6	4	$15	$8	$100
MAX	150	160			

MODEL

Decision variables

x_1 = number of shirts produced each week
x_2 = number of shorts produced each week
x_3 = number of pants produced each week
y_1 = 1 if any shirts are manufactured, 0
otherwise y_2 = 1 if any shorts are manufactured, 0
otherwis y_3 = 1 if any pants are manufactured, 0
otherwise

$x_i > 0 \rightarrow y_i = 1$ and $x_i = 0 \rightarrow y_i = 0$

Optimization function

MAX $z = 6x_1 + 4x_2 + 7x_3 - 200y_1 - 150y_2 - 100y_3$

Restrictions

$3x_1 + 2x_2 + 6x_3 \leq 150$

$4x_1 + 3x_2 + 4x_3 \leq 160$

$x_1, x_2, x_3 \geq 0$

y_1, y_2, y_3 = 0 or 1

The link between x_i and y_i is missing in the constraints

Solution: add big M constraints, where M_i is so large that it does not restrict the values of x_i

$\boldsymbol{x}_i \leq \boldsymbol{M}_i \boldsymbol{y}_i$, i = 1, 2, 3 (*), so

$M_1 \geq \min\{160/4, 150/3\} = 40$, $M_2 \geq \min\{160/3, 150/2\}=53.3$ and

$M_3 \geq \min\{160/4, 150/6\} = 25$

Remark: (*) can be replaced by $x_i \leq My_i$, i = 1, 2, 3 where $M \geq \max\{M_1, M_2, M_3\}$

Set covering problem

Set covering problem: each member of a given set 1 must be "covered" by an acceptable member of some set 2. The objective is to minimize the number of elements in set 2 that are required to cover all elements in set 1, e.g. cities to be reached within a certain time from fire stations or hospitals.

Special constraints

- **Either-or constraint:**

  At least one of the constraints $f(x_1,\ldots,x_n) \leq 0$ and $g(x_1,\ldots,x_n) \leq 0$ holds.

  1. Choose a large M s.t. $f(x_1,\ldots,x_n) \leq M$ and $g(x_1,\ldots,x_n) \leq M$. (This can always be done, because there are other constraints in the problem)
  2. Let y be a 0-1 variable.
  3. Add constraints: $f(x_1,\ldots,x_n) \leq My$ $g(x_1,\ldots,x_n) \leq M(1-y)$

  [Example either-or constraint]
  Model the following constraint:
  $x \leq 0$ or $x \geq 250$

  Define f=x, g = 250-x

  Then $f(x_1,\ldots,x_n) \leq My$, $g(x_1,\ldots,x_n) \leq M(1-y)$ results
  in $x \leq My$ $250-x \leq M(1-y)$

- If-then constraint

  If the constraint $f(x_1,\ldots,x_n) > 0$ holds, then the constraint $g(x_1,\ldots,x_n) \geq 0$ must be satisfied, while if $f(x_1,\ldots,x_n) > 0$ is not satisfied, then $g(x_1,\ldots,x_n) \geq 0$ may or may not be satisfied.

  In short: $f(x_1,\ldots,x_n) > 0$ implies $g(x_1,\ldots,x_n) \geq 0$.

  1. Choose a large M s.t. $f(x_1,\ldots,x_n) \leq M$ and $-g(x_1,\ldots,x_n) \leq M$ 2.
  Let y be a 0-1 variable
  3. Add constraints:

$-g(x_1,\ldots,x_n) \le My$

$f(x_1,\ldots,x_n) \le M(1-y)$ [Example if-then constraint]

Suppose x_1 and x_2 are 0-1 variables. Model the following constraint: If $x_1 = 1$, then $x_2 = 0$

Define f = x_1, g = $-x_2$

Then $-g(x_1,\ldots,x_n) \le My$, $f(x_1,\ldots,x_n) \le M(1-y)$ results in

$x_2 \le My$

$x_1 \le M(1-y)$

Inventory Management

ABC classification

- Class A items: 20% of high value items which account for 80% of total stock value
- Class B items: the next 30% of medium-value items which account for around 10% of the total stock value
- Class C items: the remaining 50% of low-value items which account for around the last 10% of the total stock value

- Annual usage, value → usage value ("omzet")
- Consequences of stockout
- Uncertainty of supply
- High obsolescence of deterioration risk

Measuring inventory

Item	***Average number in stock***	***Cost per item***	***Annual demand***
A	*400*	*3*	*2000*
B	*600*	*5*	*1000*

- Investment in inventory relative to total throughput
- Stock cover ("How long would inventory last without replenishment")

[Example]

$$stock\ cover(A) = \frac{stock}{2000} = \frac{400}{} * 50(weeks\ per\ year) = \mathbf{10\ weeks}\ demand$$

$$stock\ cover(B) = \frac{stock}{demand} = \frac{600}{1000} * 50(weeks\ per\ year) = \mathbf{30\ weeks}$$

$$2000\ (demand\ A) \qquad 1000\ (demand\ B)$$

$$average\ stock\ cover(A,B) = \mathbf{10} * \frac{}{3000\ (total\ demand)} + \mathbf{30} * \frac{}{3000\ (total\ demand)}$$

- Stock turn (“How often is the stock used up in a period”)

$$stock\ turn(A) = \frac{demand}{stock} = \frac{2000}{400} = \mathbf{5\ times/year}$$

$$stock\ turn(B) = \frac{demand}{stock} = \frac{1000}{600} = \mathbf{1.67\ times/year}$$

$$average\ stock\ turn(A,B) = \mathbf{5} * \frac{400\ (stock\ A)}{1000\ (A,B)} + \mathbf{1.67} * \frac{600\ (stock\ B)}{1000\ (A,B)} = 3\ times/year$$

NVENTORY MODELS

INDEPENDENT VS. DEPENDENT DEMAND

When demand for components depend on demand for end-products, taking this into account can significantly reduce stock levels (and thus costs)

[Example dependent demand]*:* a PC requires 2 disk drives.

- Annual demand PCs = 2400, EOQ = 400
- Production in months 1, 3, 5, 7, 9 and 11
- EOQ for disk drives = 1600
- Disk drive inventory at the end of month 1 is zero
- Production of disk drives takes one month, so production that starts in month *i* enters inventory in month (*i+1*)

When should the company produce disk drives?

EOQ for disk drives = 1600, demand is 2*2400=4800. So 3 times per year disk drives need to be produced.

DETERMINISTIC VS. STOCHASTIC DEMAND (AND LEAD TIMES)

DETERMINISTIC DEMAND (15):

Continuous demand:

- Economic Order Quantity (15.2)
- EOQ with quantity discounts (15.3)
- Continuous rate EOQ/EBQ (15.4)
- EOQ with backorders allowed (15.5)

Periodic demand:

- Wagner-Within, Silver-Meal (18.7)

STOCHASTIC DEMAND (16):

- Single-period inventory models
    - Newsboy problem (16.1-16.3)

- Continuous review models
  - Cost based (r,q) and (s,S) models (16.6)
  - Service level based (16.7)
- Periodic review models
  - (R,S) models (16.8)

Backlogging vs. lost sales

DETERMINISTIC DEMAND MODELS

Continuous demand:

- Economic Order Quantity (15.2)
- EOQ with quantity discounts (15.3)
- Continuous rate EOQ/EBQ (15.4)
- EOQ with backorders allowed (15.5)

Periodic demand:

- Wagner-Within, Silver-Meal (18.7)

CONTINUOUS DEMAND

Economic Order Quantity

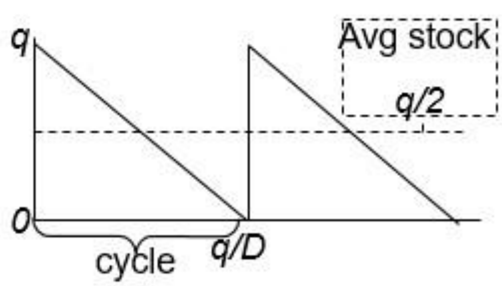
q is order quantity

Costs:

- Ordering and setup costs (K)
- Unit purchasing costs (p)
- Holding costs (h)

Assumptions:

- Repetitive ordering
- Constant demand
- Continuous ordering



Total costs (q) = annual ordering cost (K) + annual purchasing costs (p) + annual holding costs (h)

$$TC(q) = \frac{KD}{q} + pD + \frac{hq}{2}$$

$$TC'(q) = \frac{h}{2} - \frac{KD}{q^2} = 0 \ if$$

$$q = q* = \sqrt{\frac{2KD}{h}} = EOQ$$

[Example (p. 854)]

Demand D = 10,000 cameras per year

Ordering cost K = €5

Cost price p = €100

Capital opportunity cost h_d = 20%

$$h = p * h_d = 100 * 0.20 = €20$$

$$EOQ = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 * 5 * 10,000}{20}} = 70.71$$

So: 70 or 71

Remarks:

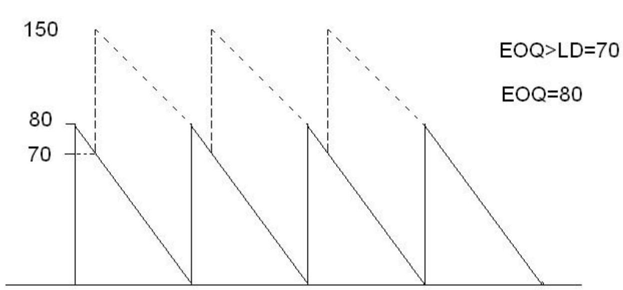
- TC(q) is reasonable flat around the EOQ
- Small variations in K, D and h have little influence on the EOQ
- Therefore it does not really matter if you choose to round to 70 or 71

Economic Order Quantity with nonzero lead time (L>0)

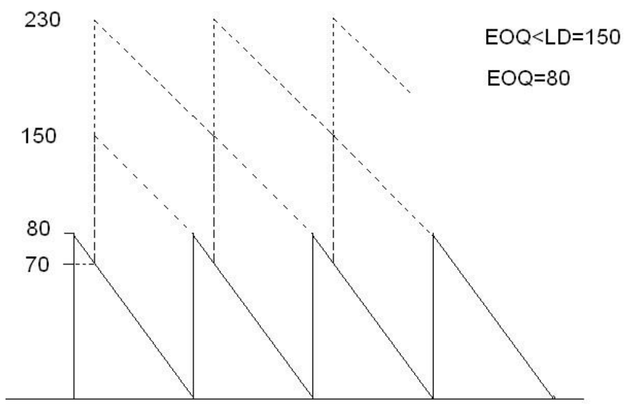
If the lead time L>0, each order must be placed at a level that ensures that when the order arrives the inventory equals zero

Reorder point: inventory level at which an order should be placed

- EOQ ≥ LD: reorder point equals LD, then upon arrival time the physical stock level equals LD-LD = 0



- EOQ < LD: reorder point equals LD, but the physical stock level then equals LD *mod* EOQ e.g. if LD = 150, EOQ = 80 then LD = EOQ + 70, the reorder point = 150 and the order is placed when the physical stock is 70. This brings the economic inventory to 230 (80 + 150).



[Exercise 1 (EOQ)]
A computer company needs to train 27 service representatives per year. A training program takes a month. It costs $12,000 to run a training program no matter how many people are trained. A service representative earns $1,500 per month.

D = 27

K = 12,000

h = 18,000 (12*1500)

$$EOQ = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 * 12{,}000 * 27}{18{,}000}} = 6$$

27 / 6 = 4,5 trainings per year

Economic Order Quantity with quantity discounts
[Exercise 2 (EOQ with quantity discounts)]
A hospital buys drip needles at a wholesaler. The price per needle is €0.60 in case the order size is less than 1000 units and €0.50 in case the order size is at least 1000 units. The annual holding cost are 20% of the purchasing price. Yearly demand is 5,000 drip needles. The cost per order is €5.

P =€o,60 for D < 1000, €0,50 for D ≥
1000

h = 0.20p

D = 5,000

K = €5

First calculate EOQ for the lowest price:

$$EOQ = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 * €5 * 5{,}000}{0.20 * €0.50}} = 707.106 \ldots$$

707 is less than 1000 and you cannot buy needles at a price of €0.50 for order sizes lower than 1000. So again calculate EOQ for the higher price:

$$EOQ = \sqrt{\frac{2KD}{h}} = \sqrt{\frac{2 * €5 * 5{,}000}{0.20 * €0.60}} = 645.497 \ldots$$

$$TC(1000) = \frac{KD}{q} + \frac{hq}{2} = \frac{5 * 5000}{1000} + \frac{(0.20 * €0.50) * 5000}{2} = 325$$

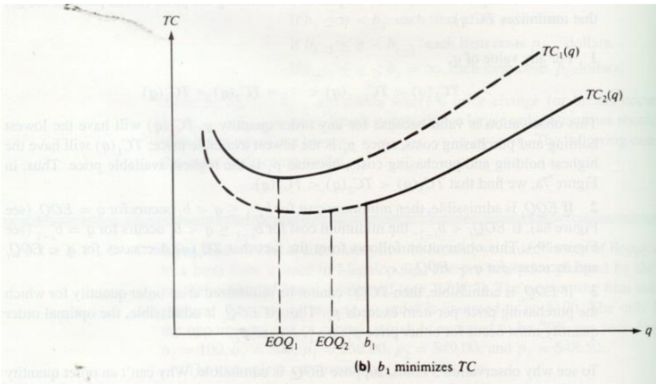
$$TC(645) = \frac{KD}{q} + \frac{hq}{2} = \frac{5 * 5000}{645} + \frac{(0.20 * €0.60) * 5000}{2} = 338{,}759$$

$q < b_1$ price per unit	$\$p_1$
$b_1 \leq q < b_2$	$\$p_2$
...	
...	
$b_{k-2} \leq q < b_{k-1}$	$\$p_{k-1}$
$b_{k-1} \leq q < b_k = \infty$	$\$p_k$

Larger quantities →lower prices, so: $p_1 > p_2 > \ldots > p_k$

Find q that minimizes total annual cost TC(q), call it q*

1. $TC_i(q)$ = total annual cost if each order is for q units at a price p_i
2. EOQ_i = min $TC_i(q)$
3. EOQ_i is admissable if $b_{i-1} \leq EOQ_i < b_i$
4. $TC(q)$ = actual annual cost if q items are ordered each time an order is placed (use p_i if $b_{i-1} \leq q < b_i$)



(b) b_1 minimizes TC

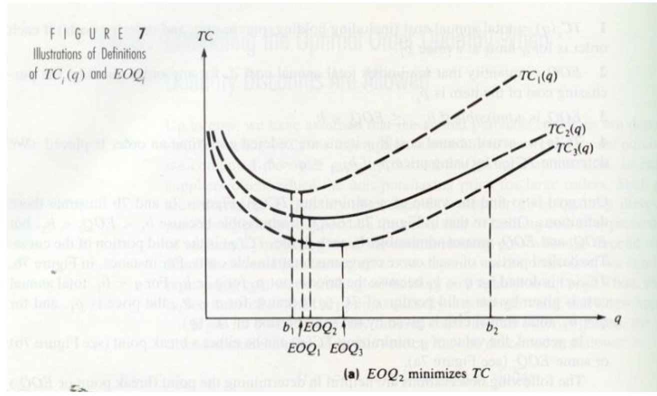


FIGURE 7
Illustrations of Definitions of $TC_i(q)$ and EOQ_i

(a) EOQ_2 minimizes TC

Note that:

- $\forall q$: $TC_k(q) < \ldots < TC_2(q) < TC_1(q)$
- If EOQ_i is admissable, then minimum cost for $b_{i-1} \leq q < b_i$ occurs for $q = EOQ_i$
- If EOQ_i is not admissable, then TC(q) cannot be minimized at an order quantity for which the purchasing price per item > p_i

Economic Order Quantity with continuous rate
Also called the Economic Batch Quantity model

Internal production, so constant arrival rate

- Production at rate r per time-unit ($r \geq D$)
  - Per unit production cost independent of run size

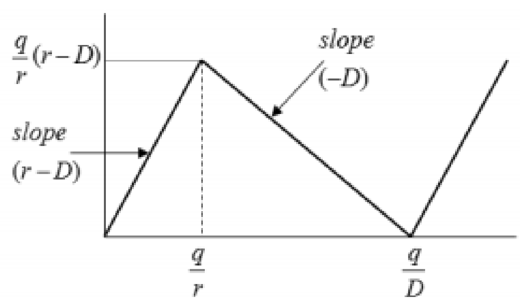
Determine the value of q that minimizes: $F(q) = \frac{\text{holding costs}}{year} + \frac{\text{setup costs}}{year}$

$$F(q) = \frac{\text{holding costs}}{year} + \frac{\text{setup costs}}{year}$$

$$F(q) = \frac{1}{2} * \frac{q}{r}(r - D) * h + \frac{KD}{q}$$

$$F'(q) = \frac{h(r-D)}{2r} - \frac{KD}{q^2} = 0 \; if$$

$$q* = \sqrt{\frac{2r}{h(r-D)}KD} = \sqrt{\frac{2KD}{h}} * \sqrt{\frac{r}{r-D}} = EOQ * \sqrt{\frac{r}{r-D}}$$



$If \; r \to \infty: \; q^* = EOQ$

[Exercise 3 (EOQ with continuous rate)]
The production process at Father Dominic's pizza can produce 400 pizza pies per day; the firm operates 250 days per year. Father Dominic's has a cost of $180 per production run and a holding cost of $5 per pizza-year. The pies are frozen immediately after they are produced and stored in a

refrigerated warehouse with a current maximum capacity of 2,000 pies. Annual demand is 37,500 pies per year.

q = 400/day K = €180 h = €5/pizza/year €5/250 = 0.02
D = 37,500/year

37,500/250 = 150/day

$$q = \sqrt{\frac{2KD}{h}} * \sqrt{\frac{r}{r-D}} = \sqrt{\frac{2*180*150}{0.02}} * \sqrt{\frac{400}{400-150}} = 2078{,}46\ldots$$

Capacity is only 2000, so you can produce 2000 per run.

Vervolg?

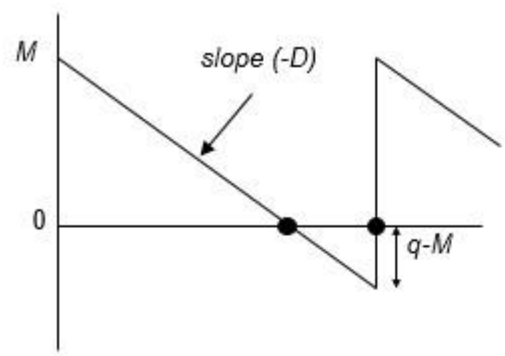
Economic Order Quantity with backorders allowed (15.5)
Backlogging means sales lost (customers accept shortages/later delivery) s = cost of being short one unit for one year q = order quantity

q-M = maximum shortage that occurs under an ordering policy

$$A = \frac{M}{D}$$

$$B - A = \frac{q-M}{D}$$

$$\frac{D}{q}\ cycles/year$$



Maximum inventory = *-q + M + q = M*

Find q and M such that: $\frac{holding\ cost}{year} + \frac{shortage\ cost}{year} + \frac{order\ cost}{year}$ is minimized

$$\text{So}: \; M^* = \sqrt{\frac{2KD}{h}} \cdot \sqrt{\frac{s}{h+s}} = EOQ \cdot \sqrt{\frac{s}{h+s}}$$

$$q^* = M^* \cdot \left(\frac{h+s}{s}\right) = \sqrt{\frac{2KD}{h}} \cdot \sqrt{\frac{h+s}{s}} =$$

$$EOQ \cdot \sqrt{\frac{h+s}{s}}$$

[Example (p. 870)]
K = €50

D = 10,000 frames/year

p = €15

h = 30% per € value of inventory

s = €15 (future business loss)
h = €15 * 0.3 = €4.50/frame/year

$$q* \quad = \sqrt{\frac{2KD}{h}} * \sqrt{\frac{h+s}{s}} = \sqrt{\frac{2 * €50 * 10{,}000}{€4.50}} * \sqrt{\frac{€4.50 + €15}{€15}} = 537.48$$

$$M* \quad = \sqrt{\frac{2KD}{h}} * \sqrt{\frac{s}{h+s}} = \sqrt{\frac{2 * €50 * €10{,}000}{€4.50}} * \sqrt{\frac{€15}{€4.50 + €}} = 413.45$$

Maximum shortage = q^* - M^* = 124.03

When to use EOQ models? Assumptions:

- Constant-rate assumption of demand
- EOQ models may be used if the variability coefficient (VC) is less than 0.2. If VC>0.20 use Wagner-Whitin algorithm or Silver-Meal heuristic for nonconstant demand (periodic)

$$(1) \quad \bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i$$

$$(2) \quad \text{est. var } D = \frac{1}{n}\sum_{i=1}^{n} d_i^2 - \bar{d}^2$$

$$(3) \quad \text{variability coefficient } VC = \frac{\text{est. var } D}{\bar{d}^2}$$

[Exercise 4]

Demand for the next 6 weeks is given in the following table:

Week	1	2	3	4	5	6
Demand	20	50	40	30	40	30

Can we use an EOQ model?

1. d = (20 + 50 + 40 + 30 + 40 + 30) / 6 = 35
2. Est. var D = (1/6) ($20^2 + 50^2 + 40^2 + 30^2 + 40^2 + 30^2$) - 35^2 = 91,667
3. VC = 91,667 / 35^2 = 0,075

VC < 0.20 so an EOQ model can be used.

Criticism on EOQ approach

Assumptions:

- Constant rate demand
- Identifiable ordering, holding and shortage costs
- Linear functions (ordering one item)

Real cost of stock in operations:

- Higher holding costs means EOQ goes down Is stock necessary? To what extent?
- Maybe setup costs can reduced?

STOCHASTIC DEMAND MODELS

Single-period inventory models

○ Newsboy problem (16.1-16.3)

Continuous review models

○ Cost based (r,q) and (s,S) models (16.6)

○ Service level based (16.7)

Periodic review models

○ (R,S) models (16.8)

Single-period inventory models (16.1 – 16.3)

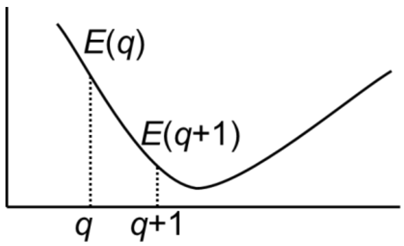
Newsboy problem: cost vs. profit version

Cost version

- Order q
- Probability p(d), demand d occurs
- Cost c(d,q)

$$Expected\ cost = E(q) = \sum_d p(d) * c(d,q)$$

If *E(q)* is convex, then the smallest q^* for which $E(q^* + 1) - E(q^*) \geq 0$ minimizes *E(q)*

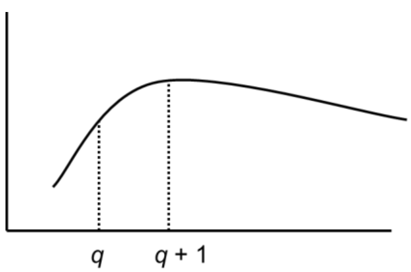


Profit version

- Order q
- Probability p(d), demand d occurs
- Profit c(d,q)

$$Expected\ profit = \tilde{E}(q) = \sum_d p(d) * \tilde{c}(d,q)$$

If $\tilde{E}(q)$ is concave, then the smallest q^* for which $\tilde{E}(q^* + 1) - \tilde{E}(q^*) \leq 0$ maximizes $\tilde{E}(q)$



[Example 1]

Purchase price: €0.20, selling price: €0.25.

Discrete demand: $d \in \{6, 7, 8, 9, 10\}$
so, clearly, q = 6, 7, 8, 9, or 10

Probability: $p(d) = 0.2 \ \forall d$

Profit:

$$\tilde{c}(d,q) = \begin{cases} 5q, & q+1 \leq d \quad (1) \\ 25d - 20q, & q \geq d \quad (2) \end{cases}$$

Situation 1 (understocked):

Extra profit = $c_u q$ = 5

Situation 2 (overstocked):

Extra loss = $-c_o q$ = 20

$$\widetilde{E}(q+1)-\widetilde{E}(q)=$$
$$=c_u P(d \ge q+1)-c_o P(d \le q)$$
$$=c_u-(c_u+c_o)P(d \le q)$$

Find the smallest q satisfying

$$\widetilde{E}(q+1)-\widetilde{E}(q) \le 0$$

$$P(d \le q) \ge \frac{c_u}{c_o+c_u}=\frac{5}{5+20}=\frac{1}{5}$$

So $\quad q^*=6$

[Exercise 5]

A shoe store has to decide how many pairs of size 43 to order for a new model summer shoe. The purchasing and selling prices are €60 and €80, respectively. In case shoes cannot be sold in the spring season, then the shoes are sold during the summer sales for €35 per pair. Assume that all pairs can be sold then. Based on demand for comparable shoes in earlier years, the store owner estimated demand for size 43 for this type of shoe.

Demand	25	30	35	40	45	50	55	60	65
Prob.	0.1	0.1	0.1	0.25	0.2	0.1	0.05	0.05	0.05

Use marginal analysis to determine the optimal order quantity.

P = 80

S = 60

O = 35

c (d, q) is 20q for q+1≤d

c (d, q) is 80d - 60q + 35(q-d) for q≥d, gives: 45d-25q

Find the smallest q satisfying

$$\widetilde{E}(q+1)-\widetilde{E}(q) \le 0$$

$$P(d \le q) \ge \frac{cu}{cu+co}=\frac{20}{20+25}=\frac{4}{9}$$

Looking at the probability table, probability is higher or equal to 4/9 at the first time for d=40. Therefore, q*=40.

Repetitive stochastic demand

- Demand during lead time and possible lead time itself are stochastic
- Safety stock = average level of stock when a replenishment order arrives
- Safety stock is used as protection against possible higher than average demand during lead time
  - Cost or service level approach
  - Probability distribution of lead time usage is important

Continuous review models (16.6)

[Example]

- An importer of sports shoes is not certain about order lead time.
- Analysis of previous orders shows that lead time may be 1-5 weeks.
- Demand rate per week varies between 110-140 pairs of shoes.
- See the table for probabilities of demand rate and lead time.
- When should the importer place a replenishment order if the probability of a stockout during lead time should be less than 10%?

Table 12.5

Matrix of lead time and demand rate probabilities

		Lead time probabilities				
		1	*2*	*3*	*4*	*5*
		0.1	*0.2*	*0.4*	*0.2*	*0.1*
Demand rate	probabilities					
110	*0.2*	110	220	330	440	550
		0.02	0.04	0.08	0.04	0.02
120	*0.3*	120	240	360	480	600
		0.03	0.06	0.12	0.06	0.03
130	*0.3*	130	260	390	520	650
		0.03	0.06	0.12	0.06	0.03
140	*0.2*	140	280	420	560	700
		0.02	0.04	0.08	0.04	0.02

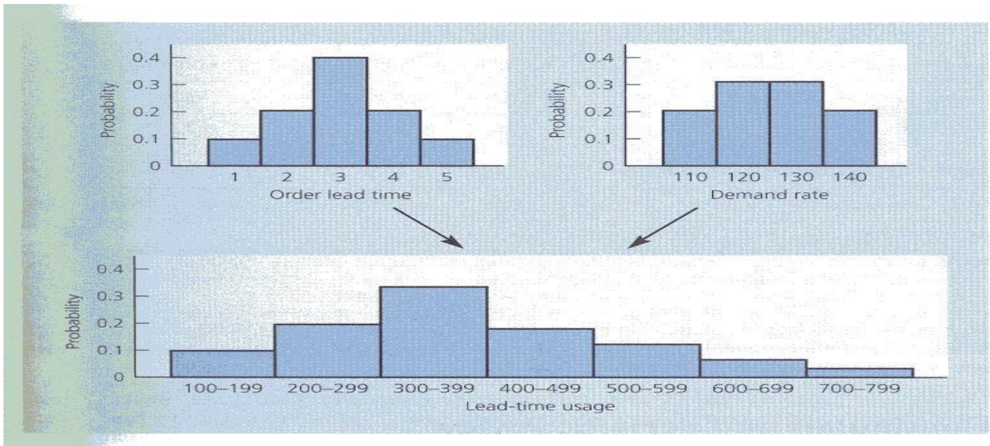


Figure 12.14 The probability distributions for order lead time and demand rate combine to give the lead-time usage distribution

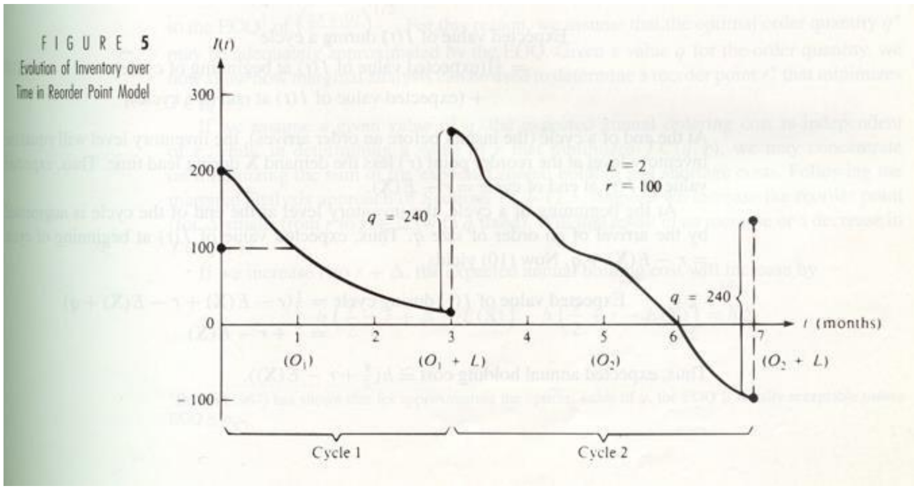
Table 12.6 Combined probabilities

Lead-time usage	100 199	200 299	300 399	400 499	500 599	600 699	700 799
Probability	0.1	0.2	0.32	0.18	0.12	0.06	0.02

Table 12.7 Combined probabilities

Lead-time usage X	100	200	300	400	500	600	700	800
Probability of usage being greater than X	1.0	0.9	0.7	0.38	0.2	0.08	0.02	0

- On-hand stock or physical stock: physically [OH(t)]
- Stockout: customer demand exceeds on-hand stock
- Backorders or backlog: amount of unfilled demand waiting to be filled when replenishment order arrives [B(t)]
- Net stock or inventory level: on hand – backorders
- Safety stock or buffer stock: average (expected) net stock just before a replenishment order arrives
- Inventory position or economic inventory: on hand + on order – backorders committed. Key quantity in deciding when to replenish: when ≤ order point s (r,q) in case of backlogging



Definitions

h = holding cost/unit

K = ordering cost
L = lead time
Q = order quantity

D = stochastic variable representing annual demand, E(D), varD, stdD

CB = cost/unit short

Br = stochastic variable representing the number of stockouts/backorders during a cycle if the reorder point = r

X = stochastic variable representing demand during lead time

- Minimize *TC(r,q)* = the annual expected total cost (excluding purchasing cost)
- *TC(r,q)* = expected annual holding cost +
  expected annual shortage cost +
  expected annual ordering cost

- $$TC(r,q) = h\left(\frac{q}{2} + r - EX\right) + c_B E(B_r)\frac{E(D)}{q} + K\frac{E(D)}{q}$$

Demand during lead time

- L = constant E(X) = L*E(D)
  Var(X) = L*var(D), std(X) = √L * stdD
- L = stochastic variable
  E(X) = E(L)*E(D)

Var(X) = E(L)var(D) +
E(D))^2*var(L) r – EX = safety stock

$$EOQ = q^* = \sqrt{\frac{2K * E(D)}{h}}$$
$$P(X \geq r^*) = \frac{hq^*}{C_B E(D)}$$

[Example]

ED = 1000, stdD = 40.8 boxes of disks per year

L = 2 weeks = 1/26
year K = 50 h = 10

Cb = 20 (all demand backlogged)

Calculate q^*, r^* and $P(X \geq r^*)$
For L = constant

$$EOQ = \boldsymbol{q}^* = \sqrt{\frac{2K * E(D)}{h}} = \sqrt{\frac{2 * 50 * 1000}{10}} = 100$$

$$EX = L * ED = \frac{1}{26} * 1000 = 38{,}46$$

$$\sigma_X = \sqrt{L} * \sigma_D = \sqrt{\frac{1}{26}} * 40.8 = 8$$

$$P(X \geq r^*) = \frac{hq^*}{C_B E(D)} = \frac{10 * 100}{20 * 1000} = 0.05$$

$$P\left(z \geq \frac{r^* - EX}{\sigma_X}\right) = P\left(z \geq \frac{r^* - 38.46}{8}\right) = 0{,}05$$

$$P(z \leq 1.65) = (1 - 0{,}05)$$

$$so\ \boldsymbol{r} * = EX + 1.65 * \ \sigma = 38.46 + 1.65 * 8 = 51.66$$

$$Safety\ stock = r - EX = 13.20$$

For L = variable
E(L) = 2 weeks
stdL = 1 week = 1/52 year

$$\sigma^2{}_X = \frac{1}{26}(40.8)^2 + 1000^2\left(\frac{1}{52}\right)^2 = 433.38$$
$$\sigma_X = 20.83, r^* = 72.83$$

(s, S) in case of backlogging or lost sales (min/max policy)

- Order might be larger than one unit so 'undershoot' could appear
- Good approximation of s and S
  S – s = q* = EOQ
  s* = r* → S* = r* + q*

Service level approach for continuous stochastic (r,q) models with backlogging

- CB is often hard to determine
- Service level measures:
  - SLM1 = expected fraction of all demand that is met on time (fill rate, P2 measure) ○ SLM2 = expected number of cycles per year during which a shortage occurs (P1 measure)
  - SLM3 = *a* if the probability of out-of-stock during a cycle equals 1-*a*, i.e. if *a* is the fraction of cycles per year during which no shortage occurs (often used in practice, not in Winston)

[Example]
ED = 1000, EOQ = 100, r = 30

Lead time demand: 20, 30, 40, 50, 60, each with probability 1/5

Determine SLM1, SLM2 and SLM3

SLM2 and SLM3:

If X is (40, 50, 60) there is a stockout, so the probability of a stockout is 3/5, so SLM3 = 2/5 and since there are 10 cycles per year, SLM2 = 3/5 * 10 = 6 cycles per year.

SLM1:

If X is (20, 30) there is no shortage.

If X = 40, shortage is 10 etc.

So the expected shortage per cycle = 1/5 * 10 + 1/5 * 20 + 1/5 * 30 = 12

And since orders/year = ED/EOQ = 10

The expected shortage per year = 10*12=120

SLM1 = (1000 – 120) / 1000 = 0.88

SLM2 = (1/SLM3) * (ED/Q)

Given SLM1 → what is r*?

EOQ = q*

ED (per year)

Lead time demand X: N(μ_X = EX, $\sigma^2{}_X$)

Reorder point = EX + yσ_X

Expected shortage per cycle E(B_r)

Expected shortage per year is $EB_r * \frac{E(D)}{q}$

1-SLM1 = $\frac{expected\ shortage/year}{expected\ demand/year} = \frac{E(Br)}{q}$

$E(B_r)$ = expected number of shortages per cycle =

$$\int_r^\infty (x-r) f_N(\mu_X, \sigma_X^2)(x)dx$$

E(B_r) can be transformed into a formula containing the (unit) normal loss function NL(y)=

$$\int_y^\infty (z-y)\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}z^2} dz$$

$$EB_r = \int_r^\infty (x-r) f_N(\mu_X, \sigma_X^2)(x)dx$$

$$= \int_r^\infty (x-r)\cdot\frac{1}{\sigma_X\sqrt{2\pi}}\cdot\exp\left[-\frac{1}{2}\left(\frac{x-\mu_X}{\sigma_X}\right)^2\right]dx$$

$$= \left(z = \frac{x-\mu_X}{\sigma_X}, r = \mu_X + y\sigma_X, x = r \leftrightarrow z = y\right)$$

$$= \sigma_X \int_y^\infty (z-y)\cdot\frac{1}{\sqrt{2\pi}}\cdot e^{-\frac{1}{2}z^2} dz$$

So, $1 - SLM_1 = \frac{1}{q^*}\cdot\sigma_X\cdot NL\left(\frac{r-\mu_X}{\sigma_X}\right)$

$r^*:\quad NL\left(\frac{r^*-EX}{\sigma_X}\right) = \frac{q^*(1-SLM_1)}{\sigma_X}$

[Example]

ED = 1000 food processors/year

σ_D = 69.28

K = \$50

h = \$10/year

L = 1 month

$$q^* = 100,$$

$$EX = \tfrac{1}{12} \cdot ED = 83\tfrac{1}{3},$$

$$\sigma_X = \frac{69.28}{\sqrt{12}} = 20$$

99% service level :

$$NL\left(\frac{r - 83\frac{1}{3}}{20}\right) = \frac{q^*(1 - SLM_1)}{\sigma_X} =$$

$$= \frac{100 \cdot (1 - 0.99)}{20} = 0.05$$

$$NL(1.25) = 0.05059$$

$$\longrightarrow \quad r^* = 83\tfrac{1}{3} + 1.25 \cdot 20 = 108.33$$

95% service level :

$$NL\left(\frac{r - 83\frac{1}{3}}{20}\right) = \frac{100 \cdot (1 - 0.95)}{20} = 0.25$$

$$NL(0.34) = 0.2518 \quad \longrightarrow \quad r^* = 90.13$$

80% service level :

$$NL\left(\frac{r - 83\frac{1}{3}}{20}\right) = \frac{100 \cdot (1 - 0.80)}{20} = 1$$

(larger than numbers in table)

$$z = \frac{r - 83\frac{1}{3}}{20} < 0$$

$$NL(z) = -z + NL(-z) = 1 \quad \text{"try"}$$

$$NL(-0.9) = 0.9 + NL(0.9) = 1.004$$

$$\longrightarrow \quad r^* = 65.33$$

x	NL(x)	x	NL(x)
0.00	0.3989	1.25	0.05059
0.01	0.3940	1.26	0.04954
0.34	0.2518	0.89	0.1023
0.35	0.2481	0.90	0.1004
		0.91	0.09860
1.24	0.05165	(4.00)	(0.0000...)

- On average s_o cycles per year stockout
- Given r_o, the stockout probability is: $P(X > r_0) = P(X \geq r_o)$
- Thus $SLM_3 = P(X \leq r_o)$

$$\frac{E(D)}{q^*} \text{ cycles per year}$$

So $SLM_2 = P(X \geq r_0) \cdot \dfrac{E(D)}{q^*}$ cycles per year there is a stockout and r^* follows from $P(X \geq r) \leq \dfrac{s_o q^*}{ED}$

[Example]

$s_o = 2$ cycles a year stockout is allowed

$$P(X \geq r) = \frac{2 \cdot 100}{1000} = 0.2$$

$$P\left(z \geq \frac{r - 83.33}{20}\right) = 0.2 \qquad \begin{aligned} P(z \leq 0.84) &= 0.7995 \\ P(z \geq 0.84) &= 0.2005 \end{aligned}$$

$$r^* = 83\tfrac{1}{3} + 0.84 \cdot 20 = 100.13$$
$$\text{safety stock: } 100.13 - EX = 16.8$$

Periodic review models (16.8)

- (R,S) periodic review model
- On-order inventory = on-hand inventory + inventory on-order
- S = order-up-to level

- R = review period
- (R,S)-policy: every R units on time the on-hand inventory is reviewed and an order is placed to bring the on-order inventory level up to S
- The order-up-to-level S should be enough to fulfil demand during the time interval between inspections plus demand during the lead time

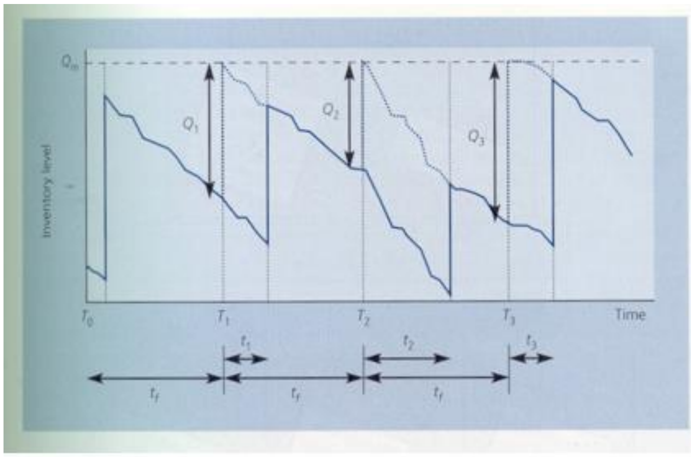


Figure 12.15 A periodic review approach to order timing with probabilistic demand and lead time

$Q_m = S$, $t_f = R$, t_1, t_2, t_3 (variable) lead time

Suppose R is known, what is the value of S that minimizes total expected annual cost?

- Expected annual holding cost $= h\left[S - E(D_{L+R}) + \frac{E(D)R}{2}\right]$
- Expected annual ordering cost $= \frac{K}{R}$
- Expected annual review cost $= \frac{J}{R}$
- Expected annual shortage cost $= c_B(EB_S)/R$

S follows from

- $P(D_{L+R} \geq S) = \frac{Rh}{c_B}$ (backlogging)
- $P(D_{L+R} \geq S) = \frac{Rh}{Rh + c_{LS}}$ (lost sales)

Remarks:

1) Often $R = \frac{EOQ}{E(D)}$ with $EOQ = \sqrt{\frac{2(K+J)\cdot E(D)}{h}}$

2) c_{LS} = shortage cost + lost profit per unit

Exercise 6, 7, 8, 9

Risk pooling

- Reducing variability by aggregating demand. Here: aggregating by location
- Example: replacing two warehouse by one lead to less safety stock and a lower average inventory level needed to reach the same service level

[Example]

- A producer of electronic components has two warehouses, one in Assen and one in Lille from which it serves all customer orders.
- Lead time from manufacturer to each of the warehouses is one week
- Two products: A and B
- Ordering costs €60 per order and inventory costs €0.27 per unit per week for each of the products
- Transporting one unit to a customer costs €1.05
- Service level is 97 percent, i.e. the inventory policy employed by each warehouse is designed in such a way that the probability of a stockout during lead time is 3 percent

- Is centralizing inventory in the neighborhood of Brussels cheaper?

- Transportation cost per unit will rise to €1.10
- Inventory and ordering costs do not change
- Lead time does not change significantly, so it is still one week
- Service level should be 97 percent again

Historical data for product A

Week	1	2	3	4	5	6	7	8
Assen	33	45	37	38	55	30	18	58
Lille	46	35	41	40	26	48	18	55
Total	79	80	78	78	81	78	36	113

Historical data for product B

Week	1	2	3	4	5	6	7	8
Assen	0	2	3	0	0	1	3	0
Lille	2	4	0	0	3	1	0	0
Total	2	6	3	0	3	2	3	0

	Product	Average demand	Standard deviation	Coefficient of variation
Assen	A	39.3	13.2	0.34
Assen	B	1.125	1.36	1.21
Lille	A	38.6	12.0	0.31
Lille	B	1.25	1.58	1.26
Total	A	77.9	20.71	0.27
Total	B	2.375	1.9	0.81

	Product	Average demand	Safety stock	Reorder point	Reorder quantity Q
Assen	A	39.3	25.08	65	132
Assen	B	1.125	2.58	4	22
Lille	A	38.6	22.8	62	131
Lille	B	1.25	3	5	24
Total	A	77.9	39.35	118	186
Total	B	2.375	3.61	6	33

- Some calculations for product A for the warehouse in Assen
- Average demand during lead time (= 1 week) = average demand = 39.3 units (= μ_X)
- Standard deviation of demand during lead time = √L*standard deviation = √1*standard deviation = 13.2 units (= σ_X)
- Q = √(2*K*ED/h) = √(2*60*39.3/0.27) = 132.16
- Service level: P(demand during lead time > reorder point) = P(X > r) = 0.03 = $P(Z > (r - \mu_X)/\sigma_X) = 0.03$
- P(Z < 1.9) = 0.97 , so r = 39.3 + 1.9*13.2 = 39.3 + 25.08 = 64.38
  - Normality assumed
  - More precisely, P(Z < 1.88) = 0.97

- Average inventory level for product A
  - Assen Q/2 + SS = 66+25.08 ≈ 91 units (for 91/39.3 ≈ 2.3 weeks)
  - Lille ≈ 88 units
  - Total 93+39.35 ≈ 132 units (132/77.9 ≈ 1.7 weeks))
  - Centralized stock leads to a 26% decrease in average stock
- Average inventory level for product B
  - Assen Q/2 + SS = 11+2.58 ≈ 14 units (for 14/1.125 ≈ 12.4 weeks)
  - Lille ≈ 15 units
  - Total ≈ 20 units (for 20/2.375 ≈ 8.4 weeks)
  - Centralized stock leads to a 31% decrease in average stock

Conclusions risk pooling

- Centralizing inventory reduces both safety stock and average inventory
- Higher coefficient of variation → greater benefit from centralization (due to a larger impact of safety stock on inventory reduction)
  - Benefits of risk pooling lower in case markets are positively correlated

Vehicle routing

Route planning

- Efficiently and effectively deliver:
  - A group of customers with known demand and location
  - From a fixed number of depots with known locations
  - Using a fixed number of vehicles with known capacities

Assumptions in this course

- **Routing** vs. scheduling
- Fleet characteristics (size, capacities, homogeneity) → **all trucks same size**
- Depot characteristics (number, capacity restrictions) → **one or more depots**
- Customer characteristics (pick-up/deliver, fixed/unknown) → **deterministic demand**
- Road characteristics (directed/undirected network, triangle inequality) → **Euclidian (undirected) with triangle inequality**
- Time restrictions (time windows, number of routes per vehicle) → **no time windows, one route per truck**
- Cost factors
- Optimization criterion (costs, service) → **minimize variable routing costs**
- Other (soft constraints, fixed routes, dynamic planning) → **sometimes constraints on maximum lengths**

Principles for good routing (Ballou)

- Create routes on customer concentrations
- Base division of stops on different days in the week on customer concentrations
- Build routes beginning with the farthest customer
- Pattern of routes: teardrop-shaped, no crossing links

- Plan vehicle with largest capacity first
- Loading and unloading simultaneously in one route, not all loading addresses to the end of routes
- Plan isolated customer separately (separate small vehicles, outsourcing)
- Try to avoid small time windows

Heuristics

Often hard to find optimal solutions for large problems in a reasonable time, therefore applying heuristics ("common sense methods"). Heuristics give good solution in general, but most of the time not optimal ones. Two parts:

1. Construction of a feasible solution
2. Improvement of the solution

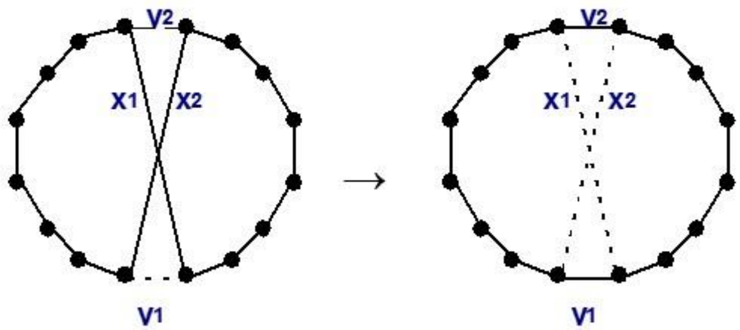
Traveling salesman problem (TSP)

Customers with known demand have to be served by trucks with a limited capacity.

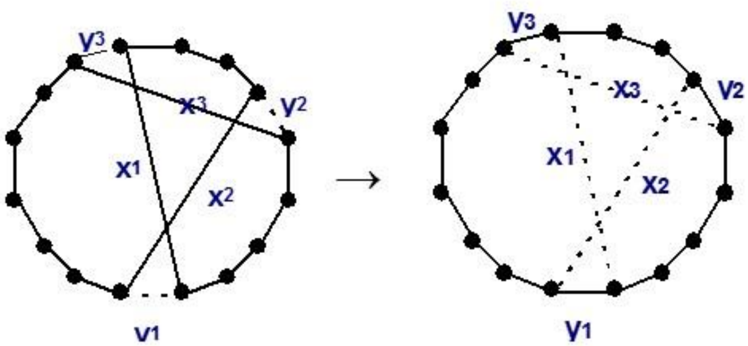
Traveling salesman problem: given n cities that have to be visited once and the travelling distances between these cities → minimize the length of a rout in which the n cities are visited.

- Nearest neighbour
- Farthest insertion

- **k-opt** (improvement heuristic):



Replacing links x1 and x2 by links y1 and y2 yields a shorter route.



Replacing links x1, x2 and x3 by links y1, y2 and y3 yields a shorter route.

Multiple TSP

Standard vehicle routing problem

Multiple TSP with extra restrictions:

- Vehicle capacities
- Possible route length and time

Assumptions

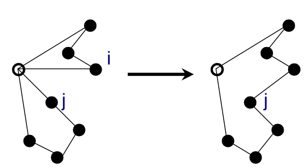
- Homogeneous fleet
- No restrictions w.r.t. route length
- Symmetric distances (from A to B = from B to A)
- Triangle inequality (Euclidian distances)

Solution methods

1. Exact

    a. Dynamic programming
    b. Branch-and-bound

2. A) Construction heuristics $_i$



    a. Savings (Clarke & Wright) Algorithm:
        1. Starting solution: N routes with one customer
        2. Determine possible savings s_{ij}

Savings $s_{ij} = c_{i0} + c_{j0} - c_{ij}$ (savings of combining customer i & j = costs depot-i + costs depot-j – costs i-j)

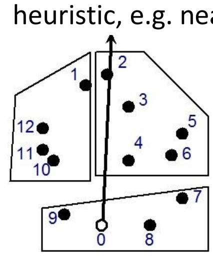
3. Choose largest saving with:
    a) Customer i and j in different routes
    b) Both customers directly connected to depot
    c) Constraints are not violated
4. Remove combinations that are not possible anymore
5. Repeat steps 3 and 4 until no savings are possible anymore

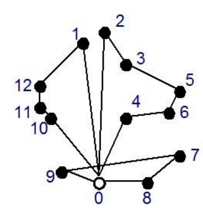
[Example on slides 27 – 32]

b. Cluster first – route second

Gillet and Miller's sweep-algorithm:

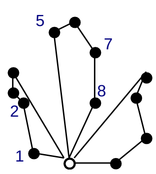
1. Divide customers in clusters by forward or backward sweep after choosing a start location
2. Determine a route for each cluster by solving TSP (possibly by using a heuristic, e.g. nearest neighbour)

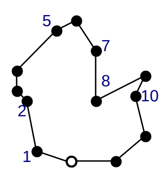




c. Route first – cluster second

1. Construct one big tour visiting all customers
2. Divide this tour into routes





Beasley's method:

1. Solve TSP (possibly by using a heuristic)
2. Find the shortest path in an acyclic network
    - Node i: cut the tour after customer i
    - a_{ij} = costs of cutting after customer j, given previous cut was after customer i: $a_{ij} = c_{0,i+1} + (c_{i+1,\, i+2} + \ldots +) + c_{j,0}$
    - f_k = costs of optimally dividing the first k customers over routes
    - Use (forward) dynamic programming (DP)

$$f_0 = 0, f_i = \min_{0 \le k \le i-1} \{f_k + a_{ki}\}$$

[Example]

	E	B	L	A	U	R	G
E	-	140	120	95	90	110	250
B	140	-	100	50	180	150	360
L	120	100	-	130	210	220	365
A	95	50	130	-	135	100	315
U	90	180	210	135	-	60	190
R	110	150	220	100	60	-	250
G	250	360	365	315	190	250	-

Location customer	Ordersize (m^3)
Brussel	15
Luik	20
Antwerpen	15
Utrecht	25
Rotterdam	10
Groningen	10

Traveling times (m), Eindhoven = depot

Nearest neighbour yields the tour: E-G-U-R-A-B-L-E

Costs by $a_{ij} = c_{0,i+1} + (c_{i+1,\, i+2} + \ldots +) + c_{j,0}$

	G	U	R	A	B	L
E (0)	500	530	610	695	∞	∞
G (1)		180	260	345	∞	∞
U (2)			220	305	400	480
R (3)				190	285	365
A (4)					280	360
B (5)						240

∞ = not feasible route, e.g. due to exceeding truck capacity

fi = min {fk + aki}

$f_0 = 0$

f_1 = min {fk + ak1} = {$f_0 + a_{01}$} = {0 + 500} = 500

f_2 = min {fk + ak2} = {$f_0 + a_{02}$, $f_1 + a_{12}$} = {0 + 530, 500 + 180} = 530*

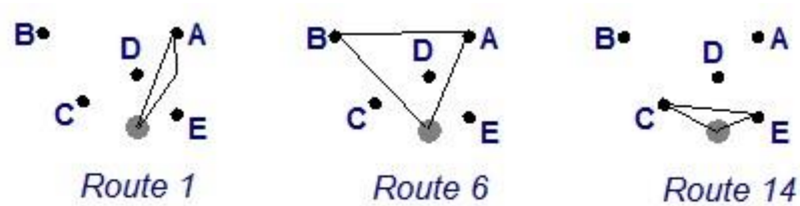
f_3 = min {fk + ak3} = {$f_0 + a_{03}$, $f_1 + a_{13,}$ $f_2 + a_{23}$} = {0 + 610, 500 + 260, 530 + 220} = 610*

etc.

d. <u>Column generation and set partitioning</u>

Column generation: generate a number (not necessarily all) feasible routes with associated costs. Here simple heuristics may be used, but you can also choose e.g. all routes including only one or two customers (slide 39).

Route j is represented by a column a_j, where $a_j(i)$ = 1 if city I is visited on route j and 0 if not.



$$a_1 = \begin{pmatrix} 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}, a_6 = \begin{pmatrix} 1 \\ 1 \\ 0 \\ 0 \\ 0 \end{pmatrix}, a_{14} = \begin{pmatrix} 0 \\ 0 \\ 1 \\ 0 \\ 1 \end{pmatrix}$$

Let c_j denote the costs of route j, and e the column vector $(1, 1, 1, 1, 1)^T$

Set partitioning: choose some of the routes s.t. all customers are visited once and costs are minimized.

Suppose J routes are generated and let x_j = 1 if route is chosen and 0 if not for j=1….J. Then the routes can be found by solving the following ILP problem (called the set partitioning problem):

$$\min \sum_{j=1}^{J} c_j x_j$$
$$s.t.$$
$$\sum_{j=1}^{J} a_j x_j = e$$
$$x_j \in \{0,1\}, j = 1,\ldots,J$$

B) Improvement heuristics/exchange heuristics

Multi-depot vehicle routing

- Sweep (Gillet & Johnson) ○ Assigning customers to depots:
  - Determine distance (costs) of customer i to nearest depot k and to the second nearest depot l: c_{ik}, c_{il}
  - Determine cost ratio: $r_i = c_{ik}/c_{il}$
  - For small values of r_i assign customer to nearest depot
  - r_i larger than given value (e.g. 0.8), assignment criterion is as follows:
- Calculate cost increase s_{ik} if customer i is added to tour of depot k
- Assign customer I to depot with smallest cost increase s_{ik}
  - VRP per depot, for example using the sweep approach, so by using CFRS

Multi depot savings (Tillman & Cain)
Definitions:
ci(k) = distance between customer i and depot k

sij(k) = savings for combination of route with customer i and route with customer j at depot k

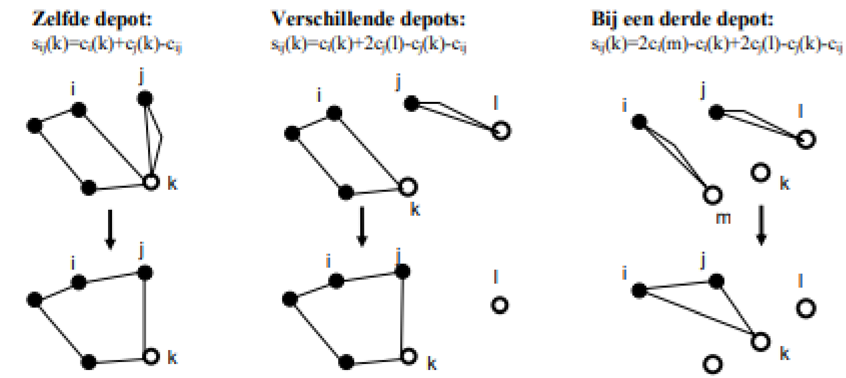
Differences compared to Clarke & Wright:

- Assign customer temporarily to nearest depot → N returns (permanent allocation after first combination)

- Savings $s_{ij}(k) = c^*_i(k) + c^*_j(k) - c_{ij}$ by combining route with customer i and route with customer j at depot k. The modified costs $c^*_i(k)$ are given by:

$$c_i^*(k) = \begin{cases} 2 * \min_t \{c_i(t)\} - c_i(k) & \textit{temporary allocation} \\ c_i(k) & \textit{permanent allocation} \end{cases}$$

  - Adapt the savings when you go from a temporary to a permanent allocation



Figuur 10-1. Drie varianten voor de bepaling van savings

Algorithm:

1. Start solution: N routes with one customer at nearest depot 2.

Determine modified costs $c^*_i(k)$ and the possible savings $s_{ij}(k)$ 3.

Choose the largest savings with:

   a. Customer i and j in different routes
   b. Both directly connected with a depot
   c. Customer i (j) at another depot than k: i (j) should have a temporary allocation
   d. Constraints are not violated

4. For every depot remove combination of routes that are not possible
5. If the allocation of customer i to depot k is now permanent, determine savings $s_{ih}(k)$ again for each customer h. Similar for customer j
6. Repeat step 3, 4 and 5 until no more savings are possible

[Example slides 50 – 59]

Location/allocation (network design)

1. **Network design (long term)**
    - Physical configuration and infra of the supply chain
    - Involves long term sourcing decisions, numbers and locations of plants etc.
2. Inventory positioning
    - Identify stocking points and CODP
3. Resource allocation (short term)
    - Is production and packaging done at the right facility?
    - How much capacity at each facility?

Why network (re)design?

- Changes in the environment, e.g.
    - Customer requirements (e.g. delivery times)
    - Geographical dispersion of customers
    - Ownership (merger, acquisition)
    - Cooperation
- Cost reduction
- Other factors
    - Working environment
    - Infrastructure
    - Availability of transportation
    - Availability of qualified personnel
    - Availability of building ground
        - Technical aspects of tax and customs

Tools for analysis

1. Mathematical optimization techniques (static)
    - Exact algorithms: *find* optimal solutions
    - Heuristics: *find* “good” solutions, not necessarily optimal, see lecture notes, part I
    - Mostly no need for a specified design as input
2. Simulation models: *test* specified design alternatives (possibly outcomes of 1) (dynamic)
    - Also allows taking into account estimated changes over time
    - But you need to rerun after changing an alternative
    - This may take a lot of time
    - Is not an optimization model

Some heuristic rules

- Locate warehouses near demand clusters
- Large deliveries (FTL) separately
- Add warehouse with largest total cost reduction

Location/allocation models

Location: where to locate warehouses and transhipment points?
Allocation: from which location are customers delivered?

Classification of models

Continuous

- One warehouse location (Weber)
- Multiple warehouse locations with customer allocation (Cooper)
- Two-level network

Discrete

- One warehouse location:
  enumeration
- More warehouse locations with customer allocation (exactly K locations,
  UFLP
  [Uncapacitated Facility Location Problem])
- Two-level, multiple-product network

- Number of locations
- Number of levels in network
- Number of product types
- Allowed cost structures
- Capacity constraints
- Demand pattern
- Optimization objective
- Other aspects (‘single sourcing’, clusters of customers assigned to the same warehouse, delivery frequency)

One warehouse location (Weber)
N customers with locations $x_i=(x_{i1}, x_{i2})$

Needed: warehouse location $Y=(y_1, y_2)$

D_i = demand per customer in units

R_i = transportation cost per unit per km

$W_i = R_i * D_i$ = transportation cost per km

$D(X_i, Y)$ = distance from customer I to warehouse Y

Problem: min. $Z(Y) = \sum w_i * d(X_i,Y)$

Euclidian distances

[Example 1]

A catering company delivers lunches to five regular customers. Transportations costs are €5 per km per tour. One tour: 50 lunches. Choose start location (2,4)

Customer number (i)	Location $X=(x_{i1}, x_{i2})$	Demand per week (units)
1	(10, 0)	1000
2	(4, 5)	250
3	(2, 3)	500
4	(0, 14)	650
5	(0, 4)	2000

Since demand should be considered in multiples of 50 units, D = (20, 5, 10, 13, 40). Since R_i = €5/km/tour, W = (100, 25, 50, 65, 200).

Centre of gravity/grid method

$$y_1 = \frac{\sum_{i=1}^{N} w_i * x_{i1}}{\sum_{i=1}^{N} w_i}, y_2 = \frac{\sum_{i=1}^{N} w_i * x_{i2}}{\sum_{i=1}^{N} w_i}$$

$$Y_1 = \frac{(100*10 + 25*4 + 50*2 + 65*0 + 200*0)}{(100+25+50+65+200)} = 2{,}73$$

$$Y_2 = \frac{(100*0 + 25*5 + 50*3 + 65*14 + 200*4)}{(100+25+50+65+200)} = 4{,}51$$

So Y=(2,73; 4,51), with Z=2170

Successive approximation

$$y_1 = \frac{\sum_{i=1}^{N} \frac{w_i * x_{i1}}{d(X_i, Y)}}{\sum_{i=1}^{N} \frac{w_i}{d(X_i, Y)}}, y_2 = \frac{\sum_{i=1}^{N} \frac{w_i * x_{i2}}{d(X_i, Y)}}{\sum_{i=1}^{N} \frac{w_i}{d(X_i, Y)}}$$

Note that, since $Y = (y_1,y_2)$, the optimal solution Y appears in the right-hand side as well as the lefthand side of the formula. So, the formula does not directly give a solution. A solution is obtained by using successive approximation: start with an arbitrary location $Y_0 = (y_{01},y_{02})$, fill in this location in the RHS of the formula. This yields a new solution in the LHS of the formula $Y_1 = (y_{11},y_{12})$. Now fill in Y_1 in the RHS, etc. This yields $Y_2 \rightarrow Y_3 \rightarrow Y_4 \rightarrow$… . Stop if $d(Y_k,Y_{k+1})$ is “small”.

Distances from customers to start location (2,4) are 8.94, 2.24, 1, 10.2 and 2 respectively.

$$Y_1 = \frac{\left(\frac{100*10}{8.94}\right) + \left(\frac{25*4}{2.24}\right) + \left(\frac{50*2}{1}\right) + \left(\frac{65*0}{10.2}\right) + \left(\frac{200*0}{2}\right)}{\left(\frac{100}{8.94}\right) + \left(\frac{25}{2.24}\right) + \left(\frac{50}{1}\right) + \left(\frac{65}{10.2}\right) + \left(\frac{200}{2}\right)} = 1{,}44$$

$$Y_2 = \frac{\left(\frac{100*0}{8.94}\right) + \left(\frac{25*5}{2.24}\right) + \left(\frac{50*3}{1}\right) + \left(\frac{65*14}{10.2}\right) + \left(\frac{200*4}{2}\right)}{\left(\frac{100}{8.94}\right) + \left(\frac{25}{2.24}\right) + \left(\frac{50}{1}\right) + \left(\frac{65}{10.2}\right) + \left(\frac{200}{2}\right)} = 3{,}89$$

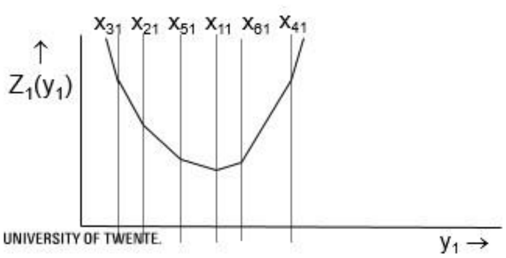
So, Y_0 = (2,4), Y_1= (1.44, 3.89). Now calculate the distances from Y_1 to the customers and continue. Finally, you end up with the optimal location (0,4) with costs z = €1942.

Manhattan distances

Minimize: $\sum w_i * |x_{i1} - y_1| + \sum w_i * |x_{i2} - y_2| = Z_1(y_1) + Z_2(y_2)$

Determine y_1 and y_2 independently

The functions Z_1 and Z_2 are piecewise linear:



Recall:
|a| = a if a ≥ 0
|a| = -a if a < 0

- Sort the customer locations according to x_{i1} … x_{in}
- Determine slope for $y_1 < x_{i1}$: $r_0 = -\sum w_i$

- Determine next slopes $r_i = r_{i-1} + 2w_i$ until the slope becomes ≥ 0

Customer	D_i	w_i	x_{i1}	slope
				-440
5	2000	200	0	-40
4	650	65	0	90
3	500	50	2	190
2	250	25	4	240
1	1000	100	10	440
SUM		440		

Customer	D_i	w_i	x_{i2}	slope
				-440
1	1000	100	0	-240
3	500	50	3	-140
5	2000	200	4	260
2	250	25	5	310
4	650	65	14	440
SUM		440		

The first coordinate of the optimal warehouse location is $y_1 = 0$, the second coordinate is $y_2 = 4$. So Y = (0,4) and the costs are z = €1942. **By accident**, this solution equals the successive approximation solution

- $Z_2(y) = 100|y-0| + 50|y-3| + 200|y-4| + 25|y-5| + 65|y-14|$
- For $y < 0$, $Z_2(y)$ reduces to

$Z_2(y) = 100(0\text{-}y) + 50(3\text{-}y) + 200(4\text{-}y) + 25(5\text{-}y) + 65(14\text{-}y) = \ldots - 440y$ with slope -440

- For $0 \leq y < 3$, $Z_2(y)$ reduces to
  $Z_2(y) = 100(y\text{-}0) + 50(3\text{-}y) + 200(4\text{-}y) + 25(5\text{-}y) + 65(14\text{-}y) = \ldots - 240y$ with slope -240 = -440 + 2*100

Multiple warehouse locations with customer allocation (Cooper) K location's, Cooper's heuristic:

1. Choose K start locations
2. Assign customers to the nearest (w.r.t. costs) warehouse
3. Improve the locations using Weber's algorithm
4. Repeat steps 2 and 3 until the assignment of customers to warehouses does not change anymore

Note that the result highly depends on which start locations are chosen, therefore try different start locations. Also, use an improvement heuristic, e.g. based on reallocation of one or two customers. In step 3 either successive approximation or the centre of gravity method can be used. Apply Cooper's heuristic for small K's, use bisection for large K's.

[Example]

- In the table demand (in units) and location of 5 customers are given. Suppose that the transport costs per unit per km equal €1, so the transport costs per km w_i equal demand D_i.
- Apply Cooper's algorithm for K = 2 locations M_1 and M_2.
  I Start locations M_1 (5,0), M_2 (5,5)
  II Start locations M_1 (3,3), M_2 (7,3)

Customer	Location	Demand
A	(1,0)	9
B	(4,3)	7
C	(5,5)	9
D	(7,4)	6
E	(8,1)	10

I Start locations M_1 (5,0), M_2 (5,5 Iteration 1:

Assign customers to warehouse: A, E → M_1, B, C, D → M_2.

	d(…,M_1)	d(…,M_2)
A (1,0)	4	√41
B (4,3)	√10	√5
C (5,5)	5	0
D (7,4)	√20	√5
E (8,1)	√10	5

New locations: M_1 (4.68, 0.53), M_2 (5.23, 4.09)

Iteration 2:

No changes, so optimal solution found

II Start locations M_1 (3,3), M_2 (7,3)

Iteration 1:

Assign customers to warehouse: A, B → M_1, D, E → M_2, C → ?

	d(…,M_1)	d(…,M_2)
A (1,0)	√13	√45
B (4,3)	1	3
C (5,5)	√8	√8
D (7,4)	√17	1
E (8,1)	√29	√5

Iteration 1:

IIa: Assign C to M_1

New locations: M_1 (3.28, 2.64), M_2 (7.63, 2.13)

Iteration 2:

No changes, optimal solution found

IIb: Assign C to M_2

Iteration 1:

Assign customers to warehouse:

A, B → M_1, C, D, E → M_2

New locations: M_1 (2.31, 1.31), M_2 (6.68, 3.16)

Iteration 2:

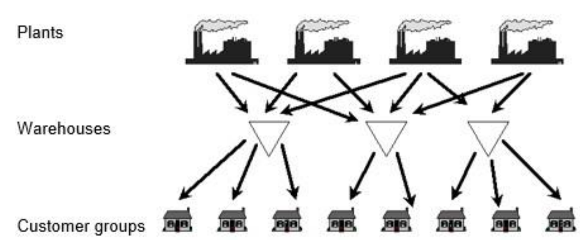
Assign customers to warehouse:

A → M_1, B, C, D, E → M_2

Two-level network

Known locations:

- Plants $U_m = (u_{m1}, u_{m2})$, m = 1,…, M
- Customers $X_i = (x_{i1}, x_{i2})$ i = 1,…, N



D_{mi} : delivery from plant m to customer i (m^3/year)

Transportation costs:

- R_{ki} per m^3 per km (from warehouse k to customer i)
- S_{mk} per m^3 per km (from plant m to warehouse k)

To be determined: warehouse locations $Y_k = (y_{k1}, y_{k2})$, k = 1,…, K

Extra requirement: *we demand that customers are delivered from ONE warehouse (single sourcing).* In case production facilities are not taken into consideration (as before) it is obvious that a customer is allocated to the nearest warehouse. In case production is taken into consideration it might be cheaper to deliver part of the demand from a specific customer via warehouse 1 and another part via warehouse 2. We do not want that, so we explicitly demand that a customer should be allocated to a single warehouse.

Changes in Cooper's algorithm:

- Criterion for assigning customer to warehouse:

  Minimize the sum of all transportation costs for plant → warehouse → customer:

$$\sum_{m=1}^{M} S_{mk} D_{mi} * d(U_m, Y_k) + R_{ki} \left\{ \sum_{m=1}^{M} D_{mi} \right\} * d(Y_k, X_i)$$

- Determining warehouse location, given that customers i_1,…, i_k are allocated to warehouse k. Include terms for the product flows between the plants and the warehouse, so minimize

$$\sum_{m=1}^{M} S_{mk} \left(\sum_{i=i1}^{i=ik} D_{mi} \right) * d(U_m, Y_k) + \sum_{i=i1}^{i=ik} R_{ki} \left\{ \sum_{m=1}^{M} D_{mi} \right\} * d(Y_k, X_i)$$

- Often, also scaling effects in location costs F_k are taken into account, depending on throughput W_k per year , so $F_k = a + bW_k + c\sqrt{W_k}$. In that case the allocation criterion in Cooper's algorithm changes. Marginal costs $D_i\{b + c/(2\sqrt{W_k}\}$ have to be taken into account.
- Background of the formula $F_k = a + bW_k + c\sqrt{W_k}$
  - Fixed location costs per year: a
  - Variable (handling) costs: ~W
  - Inventory costs (space): ~W
  - Seasonal and cycle stock: ~W
    - Safety stock: ~√W
- So the allocation criterion becomes

$$\sum_{m=1}^{M} S_{mk} D_{mi} * d(U_m, Y_k) + R_{ki} D_i * d(Y_k, X_i) + D_i \{ b + c/(2\sqrt{W_k}) \}$$

Use Eilon's algorithm to solve the problem:

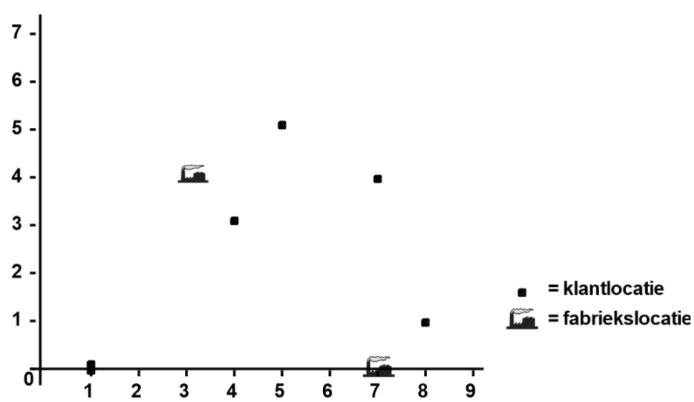
1. Choose a value K for which you are almost sure that it is larger than the number of depots in the optimal solution
2. Choose K initial depot locations, preferably with some kind of intelligence (e.g. in the neighbourhood of customer clusters)
3. Allocate the N customers to the K depots with criterion (3-11) where initially W_k is equally distributed (i.e. every depot gets $1/K^{th}$ of the total volume)
4. After the allocation, determine the new values of W_i per depot
5. Determine per depot with its corresponding customers the optimal depot location using successive approximation or the centre of gravity method
6. Repeat steps 3, 4 and 5 until the allocation does not change anymore. In that case we have found a good solution for the K-depot problem
7. Remove the depot with the smallest throughput Wi and repeat steps 3-6 for the remaining K-1 depots. If the solution with (K-1) depots is better than the one with K depots, continue to decrease K, otherwise choose the solution of the (K-1) depot problem as 'the best one'. We thus decrease K until no further cost reductions can be found.

[Example]

- 5 customers, 2 production facilities, maximum 3 warehouses
- $R_{ki} = 3$, $S_{mk} = 2$
- $F_k = 40 + 3W_k + 15\sqrt{W_k}$
- Start locations for warehouses: $Y_1 = (2,1)$, $Y_2 = (4,4)$, $Y_3 = (8,3)$

Customer i	Coord. (x_{i1}, x_{i2})	Volume factory I: D_{1i}	Volume factory II: D_{2i}
A	(1, 0)	7	2
B	(4, 3)	5	2
C	(5, 5)	6	3
D	(7, 4)	2	4
E	(8, 1)	3	7
TOTAL		23	18

Factory m	Coordinates (u_{m1}, u_{m2})
1	(3, 4)
2	(7, 0)



What are the costs of allocating customer E to warehouse 1 in the first iteration? □ W = 41/3

□ Costs:

- Transport from plants to WH1: $S_{11}*D_{1E}*d(U_1,Y_1) + S_{21}*D_{2E}*d(U_2,Y_1)$

  $= 2*3*3.16 + 2*7*5.10 = 18.96 + 71.40 = 90.36$

- Transport from WH1 to customer E: $R_{1E}*D_E*d(Y_1,X_E) = 3*10*6 = 180$
- Marginal costs for adding customer E to WH1: $D_E*(b + c/(2\sqrt{W_1})) =$

$10*(3 + 15/(2*\sqrt{41}/3)) = 50.33$

- Total costs for allocation of customer E to WH1: 320.66

Discrete location-allocation models

Advantages of discrete models compared to continuous models:

- More possibilities to include a more realistic structure for transportation costs in the analysis
- Easier to include location-dependent location costs
- Unrealistic locations can be excluded in advance Disadvantages:
- Sometimes more effort necessary to get good information
- Risk that you exclude a location that is not as unrealistic as it seems at first

Queuing

I. If ρ<1 then $\sum_{j=0}^{\infty} \rho^j = 1 + \rho + \rho^2 + \rho^3 + \cdots = \frac{1}{(1-\rho)}$

II. If ρ<1 then $\sum_{j=0}^{\infty} j\rho^j = \rho + 2\rho^2 + 3\rho^3 + \cdots = \frac{\rho}{(1-\rho)^2}$

III. $\sum_{j=0}^{n} \rho^j = \frac{1-\rho^{n+1}}{1-\rho}$

$$\sum_{j=0}^{\infty} j\rho^j = \sum_{j=1}^{\infty} j\rho^j = \sum_{j=1}^{\infty} (j+1)\rho^j - \sum_{j=1}^{\infty} \rho^j$$

$$= \frac{d}{d\rho}\left[\sum_{j=1}^{\infty} \rho^{j+1}\right] - \left[\sum_{j=0}^{\infty} \rho^j - 1\right]$$

$$= \frac{d}{d\rho}\left[\sum_{j=1}^{\infty} \rho^{j+1}\right] - \frac{1}{(1-\rho)} + 1$$

$$\sum_{j=0}^{\infty} j\rho^j = \frac{d}{d\rho}\rho\left[\sum_{j=1}^{\infty} \rho^j\right] - \frac{1}{(1-\rho)} + 1$$

$$= \frac{d}{d\rho}\left(\frac{\rho}{1-\rho} - \rho\right) - \frac{1}{(1-\rho)} + 1$$

$$= \frac{(1-\rho)+\rho}{(1-\rho)^2} - 1 - \frac{1}{(1-\rho)} + 1 = \frac{\rho}{(1-\rho)^2}$$

$$Note: \left(\frac{f}{g}\right)' = \frac{f'g - fg'}{g^2}$$

[M/M/1 Example]

Helpdesk:

- 6 requests for help on average per hour ($\lambda = 6$)
- Average service time of the only employee is 7.5 minutes ($\mu = 8$)

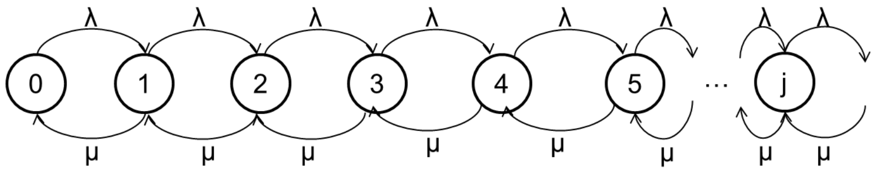
Required:

- L = average number of customers present in the queuing system
- L_q = average number of customers waiting in line
- W = average time a customer spends in the system

- W_q = average time a customer spends in line

Define P_n (n = 0,1,2,…) = probability of having n customers in the system = fraction of the time that n customers are in the system.

λP_n is the average number of transitions per hour from state n to state n+1
μP_n is the average number of transitions per hour from state n to state n-1



Flow balance equations

$$\lambda P_0 = \mu P_1 \rightarrow P_1 = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$$

$$(\lambda + \mu) P_i = \lambda P_{i-1} + \mu P_{i+1} \rightarrow \; P_2 = \rho^2 P_0 \,, P_i = \rho^i P_0 ,\; i = 1,2,\ldots$$

$$1 = \sum_{j=0}^{\infty} P_j = P_0 \sum_{j=0}^{\infty} \rho^j = \frac{P_0}{1-\rho}$$

$$P_0 = (1-\rho) = \frac{1}{4}, P_i = \; (1-\rho)\rho^i, \qquad i = 0,1,\ldots$$

$$utilization = 1 - P_0 = \rho$$

Little's queuing formula: L = λW, where W is the number of customers <u>entering</u> the system per hour

$$L = \sum_{j=0}^{\infty} jP_j = (1-\rho) \sum_{j=0}^{\infty} j\rho^j$$

$$= (1-\rho)\frac{\rho}{(1-\rho)^2} = \frac{\rho}{(1-\rho)} = 3$$

So, the average number of customers in the system equals $L = \rho/(1-\rho) = 3$

$W = L/\lambda = \rho/[\lambda(1-\rho)] = 1/(\mu-\lambda) = 3/6 = 0.5$ hours (average sojourn time)

$W_q = W - 1/\mu = 0.5 - 1/8 = 3/8$ hours (average waiting time)

$L_q = \lambda W_q = 6*3/8 = 9/4$ (average number of customers waiting in queue)

$L = \lambda W$

$L_q = \lambda W_q$

$L_s = \lambda W_s$

$W = W_q + 1/\mu$

$L = L_q + \rho$

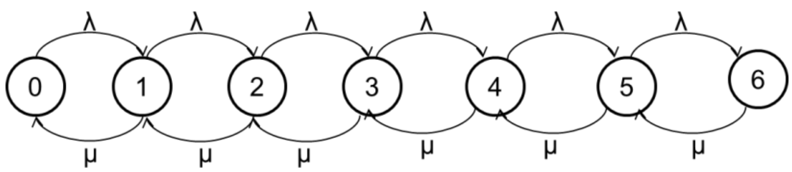
L_s = average number of customers in service

W_s = average time a customer spends in service

[McDonalds example]
Our local MacDonalds' uses an average of 10,000 pounds of potatoes per week. The average number of pounds of potatoes on hand is 5000 pounds. On the average, how long do potatoes stay in the restaurant before being used? We are given that L = 5000 pounds and λ = 10,000 pounds/week. Therefore W = 5000 pounds/(10,000 pounds/week) = 0.5 weeks.

M/M/1/c EXAMPLE (Kendall-Lee notation: M/M/1/GD/c/∞)
Consider the previous helpdesk example. Not assume that capacity is limited: when 6 customers are present, all arrivals are turned away and forever lost to the system:

- $\lambda = 6$
- $\mu = 8$
- $c = 6$



$$\lambda P_0 = \mu P_1 \rightarrow P_1 = \left(\frac{\lambda}{\mu}\right) P_0 = \rho P_0$$

$$(\lambda + \mu) P_i = \lambda P_{i-1} + \mu P_{i+1} \rightarrow \; P_2 = \rho^2 P_0 \,, P_i = \rho^i P_0, \; i = 1,2,\ldots,5$$

$$\lambda P_5 = \mu P_6$$

- This yields: $P_i = \rho^i P_0 \,,\; i = 1, \ldots, 6$

$$1 = \sum_{j=0}^{6} \rho^j P_0 = P_0 \frac{1-\rho^7}{1-\rho}, \qquad P_0 = \frac{1-\rho}{1-\rho^7}$$

$$L = \sum_{j=0}^{6} j P_j = \frac{1-\rho}{1-\rho^7} \sum_{j=0}^{6} j\rho^j = \frac{1-\rho}{1-\rho^7} \left\{ \sum_{j=0}^{\infty} j\rho^j - \sum_{j=7}^{\infty} j\rho^j \right\}$$
$$= \frac{1-\rho}{1-\rho^7} \left\{ \frac{\rho}{(1-\rho)^2} - \sum_{j=7}^{\infty} \rho^7 j \rho^{j-7} \right\}$$

Note that

$$\left\{\sum_{j=7}^{\infty} \rho^7 j \rho^{j-7}\right\} = \rho^7 \sum_{l=0}^{\infty} (l+7)\rho^l = \rho^7 \frac{\rho}{(1-\rho)^2} + 7\rho^7 \frac{1}{1-\rho}$$

So

$$L = \frac{1}{(1-\rho^7)(1-\rho)}[\rho - \rho^8 - 7\rho^7(1-\rho)]$$
$$= \frac{\rho}{(1-\rho^{c+1})(1-\rho)}[1-(c+1)\rho^c + c\rho^{c+1}] = 1.92 \ \ (c=6)$$

For calculating W, Little's formula can be used. How many customers are entering the system per hour? Not λ because sometimes incoming customers are blocked!

On average $\lambda(1-P_6)$ customers are entering the system per hour. So:

$$W = \frac{L}{\lambda(1-P_6)} = \frac{1.92}{5.69} = 0.34\, hour$$

Bulk arrivals: more than one arrival can occur at a given instant

Finite source models: when arrivals are drawn from a small population

Servers in parallel: all servers provide the same type of service and a customer need only pass through one server to complete service

Servers in series: a customer must pass through several servers before completing service

Queue discipline: described the method used to determine the order in which customers are served. Most common queue discipline is FCFS (first come, first served). LCFS: last come, last served. SIRO: service in random order.

We define t_i to be the time at which the *i*th customer arrives and $T_i = t_{i+1} - t_i$ the *i*th interarrival time. We assume that the T's are independent, continuous random variables decribed by the random variable A.

We assume that each interarrival time is governed by the same random variable implies that the distribution of arrivals is independent of the time of day or the day of the week. This is **the assumption of stationary interarrival times.**

Stationary interarrival times are often unrealistic, but we may often approximate reality by breaking the time of day into segments. A negative interarrival time is impossible. This allows us to write

$$P(\boldsymbol{A} \le c) = \int_0^c a(t)dt \quad \text{and} \quad P(\boldsymbol{A} > c) = \int_c^{\infty} a(t)dt$$

We define1/λ to be the mean or average interarrival time.

$$\frac{1}{\lambda} = \int_0^{\infty} ta(t)dt$$

An important question is how to choose **A** to reflect reality and still be computationally tractable. The most common choice for **A** is the **exponential distribution**. An exponential distribution with parameter λ has a density $a(t) = \lambda e^{-\lambda t.}$ We can show that the average or mean interarrival time is given by $E(A) = 1$

Using the fact that $var\ \mathbf{A} = E(\mathbf{A}^2) - E(\mathbf{A})^2$, we can show that $var\ \boldsymbol{A} = \frac{1}{\lambda^2}$

No-memory property: If **A** has an exponential distribution, then for all nonnegative values of *t* and *h:*
$P(\boldsymbol{A} > t + h |\ \boldsymbol{A} \geq \boldsymbol{t}) = \boldsymbol{P}(\boldsymbol{A} > h)$. This implies that if we want to know the probability distribution of the time until the next arrival, it does not matter how long it has been since the last arrival.

Relations between Poisson distribution and exponential distribution

If interarrival times are exponential, the probability distribution of the number of arrivals occurring in any time interval of length *t* follows a Poisson distribution with parameter λ*t* (and the other way around).

If **N***t* (= number of arrivals in interval [0,t)) follows a Poisson distribution with parameter λ*t*, then for n = 0, 1, 2,…

$$P(\boldsymbol{Nt} = n) = \frac{e^{-\lambda t}(\lambda t)^n}{n!}$$

When are interarrival times exponentially distributed?
If the following assumptions hold:

1. Arrivals defined on nonoverlapping time intervals are independent
2. For small Δ*t*, the probability of one arrival occurring between times *t* and *t* +Δ*t* is λΔ*t*+*o(*Δ*t)*

Then **N***t* follows a Poisson distribution with parameter λ*t*, and interarrival times are exponential with parameter λ; that is, $a(t) = \lambda e^{-\lambda t}$.

- This theorem states that if the arrival rate is stationary, if bulk arrives cannot occur, and if past arrivals do not affect future arrivals, then interarrival times will follow an exponential distribution with parameter λ, and the number of arrivals in any interval of length *t* is Poisson with parameter λ*t*.

- Remark: $o(\Delta t)$ is any quantity for which $$\lim_{\Delta t \to 0} \frac{o(\Delta t)}{\Delta t} = 0$$

The Erlang distribution

If interarrival times do not appear to be exponential they are often modeled by an Erlang distribution.

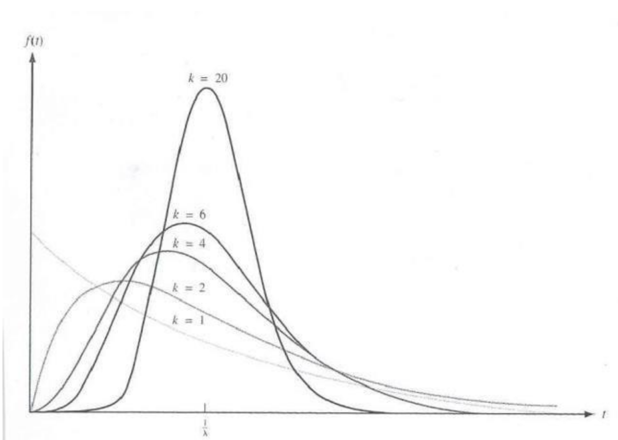
An Erlang distribution is a continuous random variable (call it **T**) whose density function *f(t)* is specified by two parameters: a rate parameter *R* and a shape parameter *k* (*k* must be a positive integer).

Given values of *R* and *k*, the Erlang density has the following probability density function:

$$f(t) = \frac{R(Rt)^{k-1} e^{-Rt}}{(k-1)!} \quad (t \geq 0)$$

Using integration by parts, we can show that if **T** is an Erlang distribution with rate parameter *R* and shape parameter *k*, then $E(\boldsymbol{T}) = \frac{k}{R}$ and $\mathrm{var}(\boldsymbol{T}) = \frac{k}{R^2}$.

An exponential distribution is an Erlang distribution with k=1 and R= *λ*. Remark: An Erlang distribution with shape parameter k and rate parameter R = kλ can be seen as the sum of k independent exponentially distributed random variables with parameter kλ.



- We assume that the service times of different customers are independent random variables and that each customer's service time is governed by a random variable **S** having a density function *s(t)*.
- We let 1/μ be the mean service time for a customer.
- The variable 1/μ will have units of hours per customer, so μ has units of customers per hour. For this reason, we call μ the service rate.

- Unfortunately, actual service times may not always be consistent with the no-memory property. In that case, we often assume that *s(t)* is an Erlang distribution with shape parameter k and rate parameter $k\mu$.

The Kendall-Lee Notation for Queuing Systems

- Standard notation used to describe many queuing systems.
- The notation is used to describe a queuing system in which all arrivals wait in a single line until one of s identical parallel servers is free. Then the first customer in line enters service, and so on.
- To describe such a queuing system, Kendall devised the following notation.
- Each queuing system is described by six characters: 1/2/3/4/5/6

1. Specifies the nature of the arrival process. The following standard abbreviations are used:
   M = interarrival times are independent, identically distributed (iid) and exponentially distributed
   - D = interarrival times are iid and deterministic
   - E_k = interarrival times are idd Erlangs with shape parameter k
   - GI = interarrival times are iid and governed by some general distribution

2. Specifies the nature of the service times:
   - M = service times are iid and exponentially distributed
   - D = service times are iid and deterministic
   - E_k = service times are idd Erlangs with shape parameter k
   - GI = service times are iid and governed by some general distribution

3. The number of parallel servers

4. The queue discipline
   - FCFS = first come, first served
   - LCFS = last come, first served
   - SIRO = service in random order
   - GD = general queue discipline

5. Specifies the maximum allowable number of customers in the system

6. The size of the population from which customers are drawn

Remark/example:

- In many important models 4/5/6 is $GD/\infty/\infty$. If this is the case, then 4/5/6 is often omitted
- M/E_2/8/FCFS/10/∞ might represent a health clinic with 8 doctors, exponential interarrival times, two-phase Erlang service times, an FCFS queue discipline, and a total capacity of 10 patients.

The waiting time paradox

Suppose the time between the arrival of buses at the student center is exponentially distributed with a mean of 60 minutes. If we arrive at the student center at a randomly chosen instant, what is the average amount of time that we will have to wait for a bus?

The no-memory property of the exponential distribution implies that no matter how long it has been since the last bus arrived, we would still expect to wait an average of 60 minutes until the next bus arrived.

PASTA property

Poisson Arrivals See Time Averages

- The distribution of customers in a queuing system when a new customer arrives is equal to the equilibrium distribution of customers in the system
- Poisson arrivals are essential, e.g. for a D|D|1 system, the property does not hold

Birth-death processes

- We define the number of people present in any queuing syste at time t to be the state of the queuing system at time t
- We call π_j the steady state, or equilibrium probability, of state j
- The behavior of $P_{ij}(t)$ before the steady state is reached is called the **transient behavior** of the queuing system, where $P_{ij}(t)$ denotes the probability that j people are present in the queuing system at time t, given that i people were present at time 0.
- A **birth-death process** is a continuous-time stochastic process for which the system's state at any time is a nonnegative integer.

Laws of motion for birth-death processes

- Law 1
  With probability $\lambda_j \Delta t + o(\Delta t)$, a birth occurs between time t and time $t+\Delta t$. A birth increases the system state by 1, to j+1. The variable λ_j is called the **birth rate** in state j. In most queuing systems, a birth is simply an arrival.

- Law 2
  With probability $\mu_j \Delta t + o(\Delta t)$, a death occurs between time t and time $t + \Delta t$. A death decreases the system state by 1, to j-1. The variable μ_j is the death rate in state j. In most queuing systems, a death is a service completion. Note that $\mu_0 = 0$ must hold, or a negative state could occur.

- Law 3
  Births and deaths are independent of each other.

Steady-state probabilities

- Most queuing systems with exponential interarrival times and exponential service times may be modeled as birth-death processes.
- More complicated queuing systems with exponential interarrival times and exponential service times may often be modeled as birth-death processes by adding the service rates for occupied servers and adding the arrival rates for different arrival streams.

- To find π_j relate (for small Δt) $P_{ij}(t+\Delta t)$ to $P_{ij}(t)$, divide by Δt and let $\Delta t \to 0$.
  This yields the flow balance equations, or conservation of flow equations, for a birth-death process.

- We obtain the flow balance equations for a birth-death process:

$$
\begin{aligned}
&(j=0) && \pi_0\lambda_0 = \pi_1\mu_1 \\
&(j=1) && (\lambda_1+\mu_1)\pi_1 = \lambda_0\pi_0 + \mu_2\pi_2 \\
&(j=2) && (\lambda_2+\mu_2)\pi_2 = \lambda_1\pi_1 + \mu_3\pi_3 \\
&\ \vdots && \\
&(j\text{th equation}) && (\lambda_j+\mu_j)\pi_j = \lambda_{j-1}\pi_{j-1} + \mu_{j+1}\pi_{j+1}
\end{aligned}
$$

- Define

$$c_j = \frac{\lambda_0\lambda_1 \ldots \lambda_{j-1}}{\mu_1\mu_2 \ldots \mu_j}$$

- If $\sum_{j=1}^{j=\infty} c_j$ is finite, we can solve for π_0: $$\pi_0 = \frac{1}{1+\sum_{j=1}^{j=\infty} c_j}$$
- It can be shown that if $\sum_{j=1}^{j=\infty} c_j$ is infinite, then no steady-state distribution exists.
- The most common reason for a steady-state failing to exist is that the arrival rate is at least as large as the maximum rate at which customers can be served.

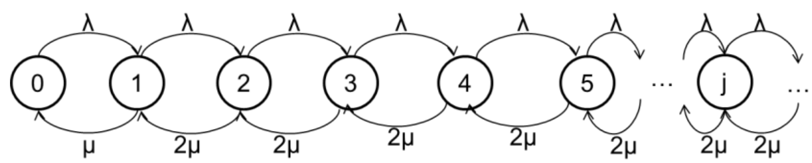
[M/M/s example (Kendall-Lee notation: M/M/s/GD/∞/∞)]

- Consider again the helpdesk example. In case $\lambda = 9$ and $\mu = 10$, then the average number of customers in the system equals $L = \rho/(1-\rho) = 0.9/0.1 = 9$ and the average sojourn time is 1 hour.

- In case this is unacceptable, adding an identical parallel server improves performance of the system.
- We assume that interarrival times are exponential (with rate λ), service times are exponential (with rate μ), and there is a single line of customers waiting to be served at one of the *s* parallel servers.

  If $j \leq s$ customers are present, then all *j* customers are in service; if $j > s$ customers are present, then all *s* servers are occupied, and $j - s$ customers are waiting in line.

- **Now assume that 2 employees are available to answer questions. Then,** ○ 9 requests for help on average per hour (λ = 9) ○ Average service time of an employee is 6 minutes (μ = 10) ○ 2 employees available (s = 2)



Flow balance equations

$$\lambda P_0 = \mu P_1$$
$$(\lambda + \mu)P_1 = \lambda P_0 + 2\mu P_2$$
$$(\lambda + 2\mu)P_i = \lambda P_{i-1} + 2\mu P_{i+1} \text{ , } i = 2, 3, \ldots$$

$$P_1 = \rho P_0 \text{ ; } P_2 = \frac{1}{2}\rho^2 P_0 \text{ ; } P_3 = \frac{1}{2^2}\rho^3 P_0 \text{ ; } P_j = \frac{1}{2^{j-1}}\rho^j P_0, \qquad j = 1, 2, \ldots$$

$$1 = \sum_{j=0}^{\infty} P_j = P_0 \sum_{j=1}^{\infty} \frac{1}{2^{j-1}}\rho^j = P_0 + P_0.2\sum_{j=1}^{\infty}\left(\frac{\rho}{2}\right)^j = P_0 + P_0\left[2\frac{1}{1-\rho/_2} - 2\right]$$

$$= -P_0 + P_0\frac{4}{2-\rho} = \frac{2+\rho}{2-\rho}P_0 \rightarrow P_0 = \frac{2-\rho}{2+\rho}$$

$$P_j = \frac{1}{2^{j-1}}\rho^j P_0 \text{ , } j = 1, 2, \ldots$$

So: a statistical equilibrium exists if $\rho/2 < 1$, i.e. if $\lambda/\mu < s = 2$!

Remark: In Section 20.6 ρ is defined as $\rho = \lambda/(2*\mu)$, thus the formulas on the slides and in the book are not directly comparable!

Average number of customers in the system

$$L = \sum_{j=0}^{\infty} jP_j = P_1 + 2P_2 + 3P_3 + \cdots = \frac{2-\rho}{2+\rho}\left\{\rho + \sum_{j=2}^{\infty} j\left(\frac{1}{2}\right)^{j-1}\rho^j\right\}$$

$$= \frac{2-\rho}{2+\rho}\left\{\rho + 2\sum_{j=2}^{\infty} j\left(\frac{\rho}{2}\right)^j\right\} = \frac{2-\rho}{2+\rho}\left\{\rho + 2\frac{\rho/_2}{\left(1-\rho/_2\right)^2} - 2\,\rho/_2\right\}$$

$$= \frac{4\rho}{(2+\rho)(2-\rho)} = \frac{4*0.9}{2.9*1.1} = 1.129$$

Average time a customer spends in the system

$$W = \frac{L}{\lambda} = \frac{1.129}{9} = 0.125\ hour$$

What is the probability that an arriving customer has to wait?

$$1 - P_0 - P_1 = 1 - (1+\rho)\frac{2-\rho}{2+\rho} = \frac{\rho^2}{2+\rho} = \frac{0.81}{2.9} = 0.28$$

Probability all servers are busy = probability that a customer has to wait (PASTA)!

What is the average number of busy servers = average number of customers in service?

$$L_s = 0P_0 + 1P_1 + 2\sum_{j=2}^{\infty} P_j = 0 + \rho\frac{2-\rho}{2+\rho} + 2\frac{\rho^2}{2+\rho} = \frac{(\rho^2 + 2\rho)}{2+\rho} = \rho = 0.9$$

Average number of busy servers = average number of customers in service!

Finite source models: the machine repair model

- Until now, all the models we have studied have displayed arrival rates that were independent of the state of the system.
- There are two situations where the assumption of the state-independent arrival rate may be invalid:
  1. If customers do not want to buck long lines, the arrival rate may be a decreasing function of the number of people present in the queuing system
  2. If arrivals to a system are drawn from a small population, the arrival rate may greatly depend on the state of the system

- Models in which arrivals are drawn from a small population are called **finite source models**.
- In the machine repair problem, the system consists of *K* machines and *R* repair people.
- At any instant in time, a particular machine is in either good or bad condition.
- The length of time that a machine remains in good condition follows an exponential distribution with rate λ.
- Whenever a machine breaks down the machine is sent to a repair center consisting of *R* repair people.

- Thus, if $j \le R$ machines are in bad condition, a machine that has just broken will immediately be assigned for repair; if $j > R$ machines are broken, $j - R$ machines will be waiting in a single line for a repair worker to become idle.
- The time it takes to complete repairs on a broken machine is assumed exponential with rate μ.
- Once a machine is repaired, it returns to good condition and is again susceptible to breakdown.

- The repair center services the broken machines as if they were arriving at an *M/M/R/GD/K/K* system.
- The machine repair model may be modeled as a queuing system, where the state *j* at any time is the number of machines in bad condition.

- Note that an arrival ("birth") corresponds to a machine breaking down and a customer leaving ("death') corresponds to a machine having just been repaired.
- When the state is j, there are K-j machines in good condition. Thus $\lambda_j = (K-j)\lambda$.
- When the state is j, min (j,R) repair people will be busy.
- Since each occupied repair worker completes repairs at rate μ, the repair rate μ_j is given by

$$\mu_j = j\mu \quad (j = 0,1,\ldots,R)$$
$$\mu_j = R\mu \quad (j = R+1, R+2,\ldots K)$$

- If we define $\rho = \lambda / \mu$, an application of steady-state probability distribution:

$$P_j = \binom{K}{j} \rho^j P_0 \,, (j = 0, 1, \ldots, R)$$

$$= \frac{\binom{K}{j} \rho^j j! \, P_0}{R! \, R^{j-R}} \,, (j = R+1, R+2, \ldots, K)$$

- Using the steady-state probabilities shown on the previous slide, we can determine the following quantities of interest:
  - L = expected number of broken machines
  - L_q = expected number of machines waiting for service
  - W = average time a machine spends broken (down time) ○ W_q = average time a machine spends waiting for service
- Unfortunately, there are no simple formulas for L, L_q, W, W_q. The best we can do is express these quantities in terms of the P_j's:

$$L = \sum_{j=0}^{j=K} jP_j \qquad W = \frac{L}{\lambda_{av}}$$

$$L_q = \sum_{j=R}^{j=K} (j-R)P_j \qquad W_q = \frac{L_q}{\lambda_{av}}$$

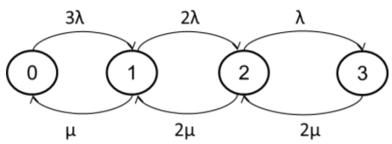
where the average number of arrivals per unit time λ_{av} is given by

$$\lambda_{av} = \sum_{j=0}^{K} P_j \lambda_j = \lambda(K - L)$$

[Example]

- K = 3 machines, break down once in 8 days on average
- R = 2 repair workers, repair takes 2 days on average
- State n = number of broken machines λ = 1/8 per day μ =1/2 per day

ρ = λ/μ = 2/8 = ¼



$$3\lambda P_0 = \mu P_1$$
$$(2\lambda + \mu)P_1 = 3\lambda P_0 + 2\mu P_2$$
$$(\lambda + 2\mu)P_2 = 2\lambda P_1 + 2\mu P_3$$
$$\lambda P_2 = 2\mu P_3$$

$$\longrightarrow$$

$$P_1 = 3\rho P_0 = \binom{3}{1}\rho^1 P_0$$
$$P_2 = 3\rho^2 P_0 = \binom{3}{2}\rho^2 P_0$$
$$P_3 = \frac{3}{2}\rho^3 P_0 = \frac{3!\binom{3}{3}\rho^3 P_0}{2!\,2}$$

$$P_0 + P_1 + P_2 + P_3 = 1 \rightarrow P_0 = \frac{1}{1 + 3\rho + 3\rho^2 + \frac{3}{2}\rho^3}$$

So: $P_0 = \frac{128}{251}\,; P_1 = \frac{96}{251}\,; P_2 = \frac{24}{251}\,; P_3 = \frac{3}{251}$

- Average number of broken machines
$$L = \sum_{j=0}^{3} jP_j = P_1 + 2P_2 + 3P_3 = \frac{153}{251}$$
- Average time a machine is down
$$W = \frac{L}{\lambda_{av}} = \frac{153}{75} > 2\ days!$$

where λ_{av} = average number of "arriving customers"
$$= 3\lambda P_0 + 2\lambda P_1 + \lambda P_2 + 0P_3 = \frac{75}{251}$$
- In general:
$$\sum_{j=0}^{3} P_j\lambda_j = \sum_{j=0}^{3} P_j\lambda(K-j) = \lambda(K-L) = \frac{1}{8}\left(3 - \frac{153}{251}\right) = \frac{75}{251}$$

[M/G/∞/GD/∞/∞ and GI/G/∞/GD/∞/∞ example]

Self service

- There are many examples of systems in which a customer never has to wait for service to begin.
- In such a system, the customer's entire stay in the system may be thought of as his or her service time.
- Since a customer never has to wait for service, there is, in essence, a server available for each arrival, and we may think of such a system as an **infinite-server** (or self-service).
- Using Kendall-Lee notation, an infinite server system in which interarrival and service times may follow arbitrary probability distributions may be written as *GI/G/∞/GD/∞/∞* queuing system.

- Such a system operates as follows:
    - Interarrival times are iid with common distribution **A**. Define $E(\mathbf{A}) = 1/\lambda$. Thus λ is the arrival rate.
    - When a customer arrives, he or she immediately enters service. Each customer's time in the system is governed by a distribution **S** having $E(\mathbf{S}) = 1/\mu$.
- Let L be the expected number of customers in the system in the steady state, and W be the expected time that a customer spends in the system, then $W = 1/\mu$ and $L = \lambda/\mu$.

  In many queuing systems, an arrival who finds all servers occupied is, for all practical purposes, lost to the system.
- If arrivals who find all servers occupied leave the system, we call the system a blocked customers cleared, or BCC, system.
- Assuming that interarrival times are exponential, such a system may be modeled as an *M/G/s/GD/s/∞* system.
- In most BCC systems, primary interest is focused on the fraction of all arrivals who are turned away.
- Since arrivals are turned away only when s customers are present, a fraction π_s of all arrivals will be turned away.
- Hence, an average of $\lambda\pi_s$ arrivals per unit time will be lost to the system.

- Since an average of $\lambda(1-\pi_s)$ arrivals per unit time will actually enter the system, we may conclude that

$$L = L_s = \frac{\lambda(1-\pi_s)}{\mu}$$

- For an *M/G/s/GD/s/∞* system, it can be shown that π_s depends on the service time distribution only through its mean $(1/\mu)$.
- This fact is known as **Erlang's loss formula**.
- In other words, any *M/G/s/GD/s/∞* system with an arrival rate λ and mean service time of $1/\mu$ will have the same value of π_s (cf. Lecture notes OR2):

$$\pi_s = \frac{\rho^s / s!}{1+\rho+\rho^2/2!+\ldots+\rho^s/s!}, \rho = \lambda/\mu$$

Dynamic Programming

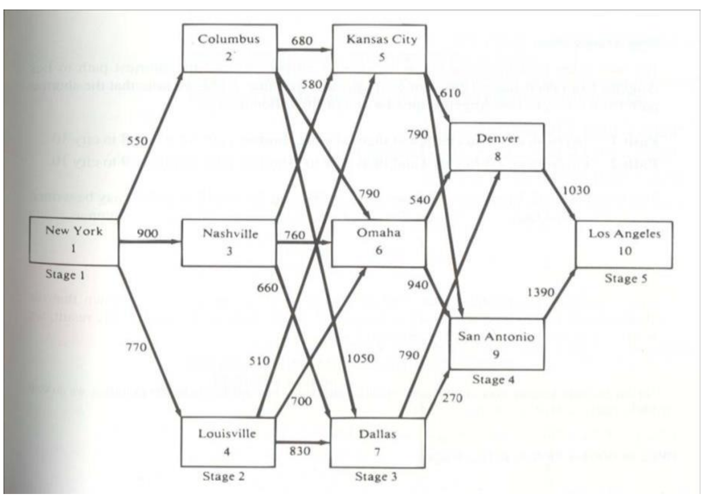
Deterministic dynamic programming

Dynamic programming obtains solutions by working backward from the end of the problem toward the beginning, thus breaking up a large, unwieldy problem into a series of smaller, more tractable problems with the same structure as the original problem, but of a smaller size.

Characteristics:

1. The problem can be divided into stages with a decision required at each stage t=1,…T
2. Each stage had a number of states associated with it (state = information needed to make the optimal decision)
3. The decision chosen at any stage describes how the state at the current stage is transformed into the state at the next stage
4. Gìven the current state, the optimal decision for each of the remaining stages must not depend on previously reached states of previously chosen decisions. This is called the principle of optimality
5. If the states for the problem have been classified into one of T stages, there must be a recursion that relates the cost or reward earned during stages t, t+1, … T to cover the cost or reward during stages t, t+1 … T

[Example network problem]



Costs = distance c_{ij} from city *i* to city *j*

Goal: determine the length of the shortest path from NY(1) to LA (10)

- Phase *n*: beginning of day *n*
- State *i*: current location
- Decision *d*: next city (chosen from possible options $D_n(i)$)
- Direct result r_n: distance between current and next city
- Cost function $f_n(i)$: minimum distance between city *i* and LA in case you start in city *i* at the beginning of day *n*
- Recurrence relation:

$$f_n(i) = \min_{d \in D_n(i)} \{r_n(i,d) + f_{n+1}(d)\}$$

Remark

You can start the backward recursion in two ways:

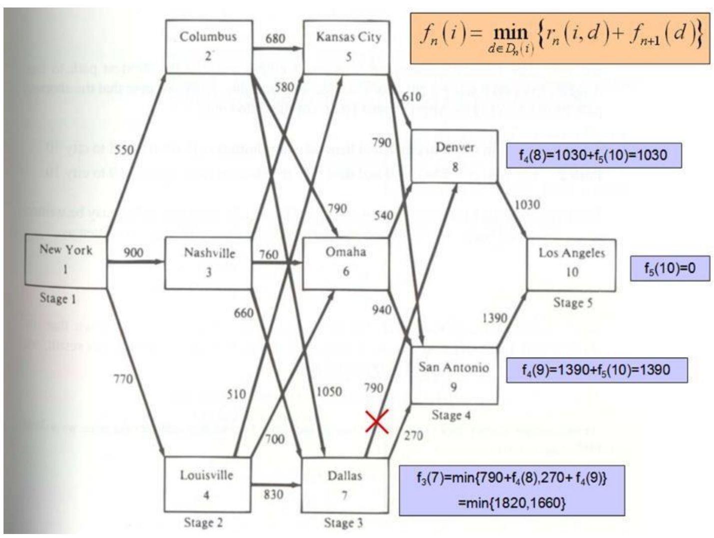
1) Start in phase 4: $f_4(8) = 1030$, $f_4(9) = 1390$ and work backward e.g.

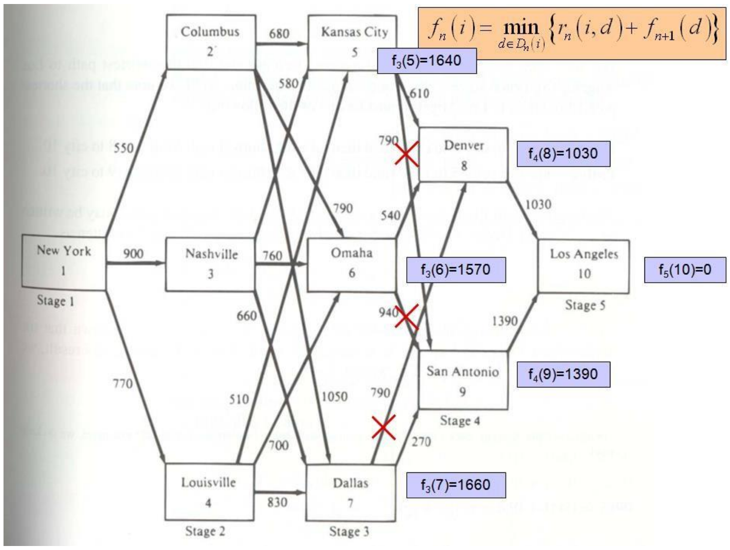
   $f_3(7) = \min\{790 + f_4(8), 270 + f_4(9)\}$

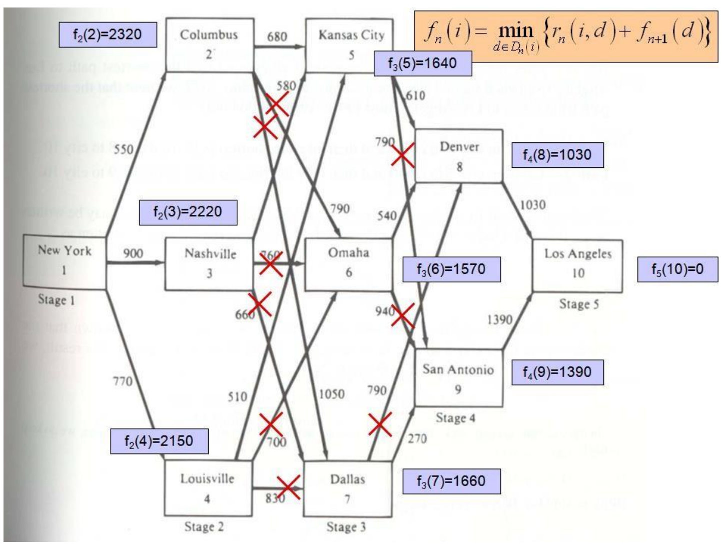
   $= \min\{790 + 1030, 270 + 1390^*\} = 1660$ etc.

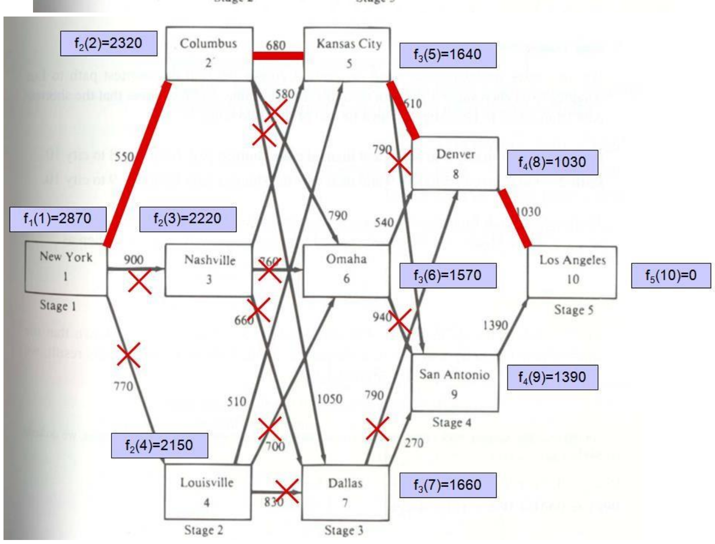
2) Start in phase 5 with $f_5(10) = 0$ and work backward e.g.

   $f_4(8) = 1030 + f_5(10) = 1030$




$f_n(i) = \min_{d \in D_n(i)} \{ r_n(i,d) + f_{n+1}(d) \}$
Columbus 2
Kansas City 5
New York 1
Stage 1
Nashville 3
Omaha 6
Denver 8
Los Angeles 10
Stage 5
San Antonio 9
Stage 4
Louisville 4
Stage 2
Dallas 7
Stage 3
680
580
550
900
760
790
540
610
790
1030
660
940
1390
770
510
1050
790
700
270
830
$f_3(5)=1640$
$f_4(8)=1030$
$f_3(6)=1570$
$f_5(10)=0$
$f_4(9)=1390$
$f_3(7)=1660$


$f_n(i) = \min_{d \in D_n(i)} \{ r_n(i,d) + f_{n+1}(d) \}$
$f_2(2)=2320$
$f_2(3)=2220$
$f_2(4)=2150$
$f_3(5)=1640$
$f_4(8)=1030$
$f_3(6)=1570$
$f_5(10)=0$
$f_4(9)=1390$
$f_3(7)=1660$


$f_2(2)=2320$
$f_1(1)=2870$
$f_2(3)=2220$
$f_2(4)=2150$
$f_3(5)=1640$
$f_4(8)=1030$
$f_3(6)=1570$
$f_5(10)=0$
$f_4(9)=1390$
$f_3(7)=1660$

[Fishery example]
The owner of a lake must decide how many bass to catch and sell each year. If she sells x bass during year t, then a revenue $r(x)$ is earned. The cost of catching x bass during a year is a function $c(x, b)$ of the number of bass caught during the year and of b, the number of bass in the lake at the beginning of the year. Of course, bass do reproduce.

To model this, we assume that the number of bass in the lake at the beginning of a year is 20% more than the number of bass left in the lake at the end of the previous year. Assume that there are 10,000 bass in the lake at the beginning of the first year. Develop a dynamic programming recursion that can be used to maximize the owner's net profit over a T-year horizon.

We define x_t to be the number of bass caught during year t. To determine an optimal value of x_t, the owner of the lake need only know the number of bass (call it b_t) in the lake at the beginning of year t. Therefore, the state at the beginning of year t is b_t. We define $f_t(b_t)$ to be the maximum net profit that can be earned from bass caught during years $t, t+1, \ldots T$ given that b_t bass are in the lake at the beginning of year t.

For the recursion formula, we need to consider 3 aspects:

1. What are the allowable decisions? During any year we can't catch more bass than there are in the lake. Thus, in each state and for all t, $0 \leq x_t \leq b_t$ must hold.
2. What is the net profit earned during year t? If x_t bass are caught during a year that begins with b_t bass in the lake, then the net profit is $r(x_t) - c(x_t, b_t)$.
3. What will be the state during year t+1? The year t+1 state will be $1.2(b_t - x_t)$.

- After year T, there are no future profits to consider, so $f_T(b_T)=\max\{r(x_T) - c(x_T,b_T)\}$ where $0 \leq x_T \leq b_T$.
- For other phases the recursion is $f_t(b_t)=\max\{r(x_t) - c(x_t,b_t)+f_{t+1}[1.2(b_t-x_t)]\}$ where $0 \leq x_t \leq b_t$.

To determine the optimal fishing policy, we begin by choosing x_1 to be any value attaining the maximum in the equation for $f_1(10{,}000)$.

Then year 2 will begin with **$1.2(10{,}000 - x_1)$** bass in the lake.

This means that x_2 should be chosen to be any value attaining the maximum in the equation for ***$f_2(1.2(10{,}000-x_1))$***. Continue in this fashion until optimal values of $x_3, x_4, \ldots x_T$ have been determined.

Remark In this example the number of states equals $10{,}000 * 1.2^{T-1}$. It is an example of the *curse of dimensionality.*

[Production-inventory problem example]

- Demand during periods 1,…,T is known.
- This demand can be met from production in the same period or inventory.
- Constraints on production and storage capacity

How can demand be met at minimal costs, given

- Fixed costs for every production run
- Variable production costs per item
- Inventory costs per item (end of period)

Given:

- Demand d_t in months 1,…, 4 given
- Setup costs per batch: \$3
- Variable production costs: \$1
  → $c(x) = 3 + x$ for $x > 0$, $c(0) = 0$
- Inventory costs per unit: h = \$0.5
- Production capacity $K = 5$
- Storage capacity $S = 4$
- Starting inventory: 0
- Extra variable: production x_t, then $i_t = i_{t-1} - d_t + x_t$

Month	Demand (d_t)
1	1
2	3
3	2
4	4

1. **Stages?**
   Period t = 1,…,T (or T+1)
2. **States?**
   Inventory I (at the beginning of period t)
3. **Decisions?**
   Production in period t
4. **Cost function?** $f_t(i)$ = minimum costs over periods $t,...,T$ starting in state i at stage t
5. **Recursion?**
   $f_t(i) = \min_j\{c_{ij} + f_{t+1}(j)\}$

   Start recursion: $f_{T+1}(0) = 0$: no ending inventory

   End recursion: $f_1(0)$: no starting inventory

Always write down explicitly the possible options j in set $D_n(i)$:

- Phase 4: (demand in month 4 equals 4)

  $f_4(i)=c(4-i)$ en $x_4(i)=4-i$ (calculate for $i=0..4$)

- **Phase 3: (demand in month 3 equals 2)**
  $f_3(i)=\min_x\{0.5*(i+x-2)+c(x)+f_4(i+x-2)\}$ (for $i=0..4$ calculate the best x, where $2 \le x+i \le 6$ en $x \le 5$)
- **Phase 2: (demand in month 2 equals 3)**
  $f_2(i)=\min_x\{0.5*(i+x-3)+c(x)+f_3(i+x-3)\}$ (for $i=0..4$ calculate the best x, where $3 \le x+i \le 7$ en $x \le 5$)
- **Phase 1: (demand in month 1 equals 1)**
  $f_1(0)=\min_x\{0.5*(x-1)+c(x)+f_2(x-1)\}$ where $1 \le x \le 5$ en $x \le 5$)

Resource allocation

Resource-allocation problems, in which limited resources must be allocated among several activities, are often solved by dynamic programming.

Linear programming can be used to solve resource allocation problems if the following 3 assumptions are fulfilled:

1. The amount of a resource assigned to an activity may be any non negative number
2. The benefit obtained from each activity is proportional to the amount of the resource assigned to the activity. (Proportional: r(x) = 3x, not proportional e.g. r(x) = 2 + 3x)
3. The benefit obtained from more than one activity is the sum of the benefits obtained from the individual activities.

In order to use dynamic programming only assumption 3 needs to be fulfilled. In addition, the amount of the resource allocated to each activity must be a member of a finite set.

Problem: execute several activities using a certain resource. What is the allocation that yields maximum profit?

- Invest 6 portions of $1000 in 3 projects
- Investing d_t portions in project t yields NPV $r_t(d_t)$
  MAX $z = r_1(d_1) + r_2(d_2) + r_3(d_3)$

  s.t.

  $d_1 + d_2 + d_3 = 6$

  d_1, d_2, d_3 integer

1. Stages? Project t
2. States? Remaining budget i
3. Decisions? Assign budget j ($j \leq i$) to project t
4. Cost function? $f_t(i)$ = maximum NPV in case budget i can be used for projects $t,...,3$
5. Recursion? $f_t(i) = \max_{0 \leq j \leq i} \{r_t(j) + f_{t+1}(i-j)\}$, $t = 1, 2$
   Start recursion $f_3(i) = \max_{0 \leq j \leq i} r_3(j)$
   End recursion $f_1(6) = ?$

Given:

- Demand d_t during period t (t = 1,…, T)
- Production cost (for producing x units during any period):

$$c(x) = \begin{cases} K + cx, & x > 0 \\ 0, & x = 0 \end{cases}$$

- Inventory level i_t at end of period t incurring holding cost hi_t (i_0 = inventory at the beginning of period 1)

Determine a production level x_t for each period t that minimizes the total cost of meeting (on time) the demands for periods 1, 2, …, T.

Wagner-Whitin

Given:

- Demand d_t during period t (t = 1,…, T)
- Production cost (for producing x units during any period):

$$c(x) = \begin{cases} K + cx, & x > 0 \\ 0, & x = 0 \end{cases}$$

- Inventory level i_t at end of period t incurring holding cost hi_t (i_0 = inventory at the beginning of period 1)

Determine a production level x_t for each period t that minimizes the total cost of meeting (on time) the demands for periods 1, 2, …, T.

Lemma 1:

If production occurs during period t, we must (for some j) produce an amount that exactly suffices to meet the demand for periods t, t+1, …, $t+j$.

Lemma 2:

If it is optimal to produce anything during period t, then $i_{t-1} < d_t$.

So, with the possible exception of the first period, production will occur only during periods in which the beginning inventory is zero

i_0 = 0 (without loss of generality)

f_t = minimum cost incurred during periods $t, t+1, …, T$, given that at the beginning of period t the inventory is zero f_{T+1} = 0

Then

$$f_t = \min_{j=0,1,\ldots T-t} (c_{tj} + f_{t+j+1})$$

with

$$c_{tj} = K + c\left[d_t + d_{t+1} + \ldots + d_{t+j}\right] + \\ + h\left[j \cdot d_{t+j} + (j-1) \cdot d_{t+j-1} + \ldots + d_{t+1}\right]$$

Start with $t = T$, then $t = T - 1$, …, 1.

Remark: Omitting variable production cost yields the same production schedule.

1

D_1 = 1, D_2 = 3, D_3 = 2, D_4 = 4 (total demand 10)

K = 3, c = 1, h = ½, i_0 = 0

$$f_5 = 0$$

$$f_4 = 3 + 1 \cdot 4 + \tfrac{1}{2} \cdot 0 = 7^*$$

[produce for month 4]

$$f_3 = \min(\text{ mth } 3, \text{mth } 3+4)$$

$$f_3 = \min \begin{cases} 3 + 1 \cdot 2 + f_4 = 12 \\ 3 + 1 \cdot (2+4) + \tfrac{1}{2} \cdot 4 + f_5 = 11^* \end{cases}$$

2

$$f_2 = \min(\text{mth } 2, \text{mth } 2+3, \text{mth } 2+3+4)$$

$$f_2 = \min\begin{cases}(3+1\cdot 3+f_3)=17\\(3+1\cdot(3+2)+\tfrac{1}{2}\cdot 2+f_4)=16^*\\(3+1\cdot(3+2+4)+\tfrac{1}{2}\cdot(2+4+4)+\\\quad +f_5)=17\end{cases}$$

3

$$f_1 = \min(\text{mth } 1, \text{mth } 1+2, \text{mth } 1+2+3,\ \text{mth } 1+2+3+4)$$

$$f_1 = \min\begin{cases}(3+1\cdot 1+f_2)=20\\(3+1\cdot(1+3)+\tfrac{1}{2}\cdot 3+f_3)=19\tfrac{1}{2}^*\\(3+1\cdot(1+3+2)+\tfrac{1}{2}\cdot(3+2+2)+\\\quad +f_4)=19\tfrac{1}{2}^*\\(3+1\cdot(1+3+2+4)+\\\quad +\tfrac{1}{2}\cdot(3+2+4+2+4+4)+\\\quad +f_5)=22\tfrac{1}{2}\end{cases}$$

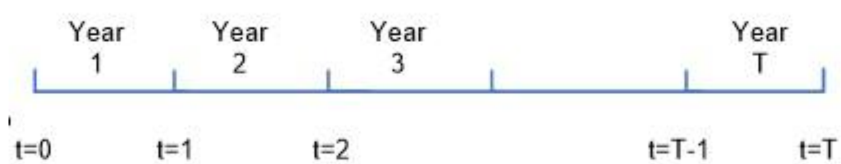
4

Solution: minimum cost 19½ Two
optimal strategies:

1. month 1: produce for months 1,2 (4)
   month 3: produce for months 3,4 (6)

2. month 1: produce for months 1,2,3 (6)
   month 4: produce for month 4 (4)

- Distinguish between time t and year t
- $t = 0$ is the beginning of year 1, $t = 1$ is the end of year 1



[Example equipment replacement problem]

- Purchase of a car with value P
- Maintenance costs m_t during year t (increasing)
- Replacement within 3 years with salvage value s_t after year t (decreasing)
- **Problem:** what is the best replacement strategy over the next T years when a new car is bought at the beginning of year 1?

1. **Stages**? Year t
2. **States**? Age of car at the beginning of year t
3. **Decisions**? Replace car at the beginning of year t? (Yes/No)
4. **Cost function**? $f_t(i)$ = minimum costs over years $t,...,T$ if the present car is i years old at the beginning of year t
5. **Recursion**? $f_t(i)$ = min

replace $\{P- s_i + m_1 + f_{t+1}(1),$

keep (i<3) $m_{i+1} + f_{t+1}(i+1)\}$

Start recursion: $f_{T+1}(i) = -s_i\ (1\leq i\leq 3)$

End recursion: $f_1(0) = ?$

Remark: the car has a salvage value at the end of year T equal to si. This is taken into account by saying that you get that money back at the beginning of the artificial year T+1.

You can also take this salvage value into account by considering the last stage T separately as is done in the fishery and production inventory example. In that case the start of the recursion is

fT(i) = min

replace {P- si + m1 - s1,

keep (i<3) mi+1 - si+1}

Remarks

- In Winston the DP approach presented here is called the alternative recursion. In my opinion the model presented in this lecture is more straightforward and thus easier to grasp.
- There are alternative DP formulations possible for many problems.

Part two

! Disclaimer: always check what you need to study corresponds with the content of the summaries, courses can be changed which could cause changes in study material for your exams

If you made a summary for a course this module you can sent them to education@stress.utwente.nl and depending on how many summaries we have for this course you will receive compensation for your work.

Each specialization has a specific pre-master programme. If you are a student with a technical programme from a Research University and you are admitted to one of the pre-master programmes, you have to take up to 15 EC of courses. Which means part two and therefore these summaries are not relevant for you.

If you are a student with a technical programme from a University of Applied Science or a student with a social science programme from a Research University and you are admitted to one of the pre-master programmes, you have to take up to 30 EC of courses. (two parts)

Courses PLM, HCTM and FEM

- Statistics and Probability Theory for premaster IEM	202001176	5 EC
- Academic Skills for premaster IEM	202000451	4 EC
- Excel/VBA	202000397	3 EC
- Operations Strategy for premaster IEM	202000453	3 EC
- (students with prior knowledge on OS will have to do Project OM for premaster IEM instead)	202000453	3 EC

Summary 1

Course: Operations Strategy for premaster IEM test one
Book: Slack, N., Brandon-Jones, A., & Johnston, R. (2016). Operations Management (8th edition). Pearson Education Australia.
Chapters: 1-9
Year the summary: 2019/2020

Summary 2

Course: Operations Strategy for premaster IEM test two
Book: Slack, N., Brandon-Jones, A., & Johnston, R. (2016). Operations Management (8th edition). Pearson Education Australia.
Chapters: 10, 11, 14 - 18
Year the summary: 2019/2020

Summary 4: Operations Strategy for premaster IEM test one

Chapter 1

Operations management (OM): activity of managing the resources that create and deliver services/products. All organizations have 'operations' that produce some mix of services/products, operations manager responsible. Three core functions in any organization:

- **Marketing** (incl sales) -> communicating organization's services and products to its markets to generate customer requests
- **Product-service development** -> coming up with new and modified services/products to generate future customer requests
- **Operations** -> creation and delivery of services and products based on customer requests

Always some additional support functions, can depend on type of organization. In this book: **operations function**: all activities necessary for day-to-day fulfilment of customer requests within constraints of environmental and social sustainability.

Fundamental that functional boundaries should not hinder efficient internal processes. All organizations have 1 thing in common: make a profit from creating and delivering their products or services. Non-profit -> to serve society in some way. OM uses 'resources to appropriately create outputs that fulfil defined market requirements'. OM is at the forefront of coping with, and exploiting, developments in business and technology.

All operations create and deliver service/products by changing inputs into outputs using an '**input-transformation-output**' process. One set to any operations processes is transformed resources: treated, transformed or converted in the process. Usually mix of: **materials** (transform physical properties, location, possession, store), **information** (transform informational properties, possession, store, location) and **customers** (change physical properties, accommodate, location, physiological state). When customers play role in operations: 'co-production'.

Other input: transforming resources: **facilities** (building, equipment etc) and **staff**.

Facilitating services: e.g. aluminum smelters also delivering technical advice. Most operations produce a mixture of tangible products and intangible services. Whether an operation produces tangible products or intangible services is becoming increasingly irrelevant. In a sense all operations produce service for their customers.

It is critical that operations managers are aware of customers' needs, both current and potential.

All operations consist of a collection of processes interconnecting with each other. **Process**: arrangement of resources and activities that transform inputs into outputs that satisfy customer needs. Each process is an internal supplier and an internal customer for other processes. Any process is made up of a network of resources. Business or operations itself is also part of a greater process: supply network. Process principle can be used at three levels 'levels of analysis': level of operations itself, level of supply network and level of individual processes.

Not just the operations function that manages processes; each function will have its 'technical' knowledge. OM is relevant for all functions, and all managers should have something to learn from the principles, concepts, approaches and techniques of OM -> distinguish between: 1. 'Operations' as a function: part of the organization which creates and delivers services and products for the organizations external customers. 2. ‘Operations’ as an activity: management of the processes within any of the organization’s functions.

Processes are defined by how the organization chooses to draw process boundaries. Operations processes differ in 4 V’s:

- **Volume** of their output

High (e.g. McDonalds): high repeatability, specialization, capital intensive, low unit costs. Low (e.g. local cafeteria): low repetition, each staff member performs more of each task (might be more rewarding), less systemization, high unit costs, less feasible to invest in specialized equipment.

- **Variety** of their output

High (e.g. taxi): flexible, complex, match customer needs, high unit costs. Low (e.g. bus service): well defined, routine, standardized, regular, low unit costs.

- **Variation** in the demand for their output

High (e.g. summer holiday resort hotel): changing capacity, anticipation, flexibility, in touch with demand, high unit costs. Low (e.g. formule 1 hotel): stable, routine, predictable, high utilization, low unit costs.

- Degree of **visibility** which customers have of the creation of their output

High (e.g. ‘bricks and mortar’): short waiting tolerance, satisfaction governed by customer perception, customer contact skills needed, received variety is high, high unit costs. Low (e.g. web-based): time lag between production and consumption, standardization, low contact skills, high staff utilization, centralization, low unit costs.

Some mixed high- and low-visibility: e.g. airports.

Operations and processes can reduce their costs by increasing volume, reducing variety, reducing variation and reducing visibility.

OM activities:

- **Directing** overall strategy of the operation. General understanding of operations and processes and their strategic purpose and performance, together with appreciation of how strategic purpose is translated into reality.
- **Designing** the operation’s resources and processes. Determining the physical form, shape and composition of operations and processes in line with services/products that they create.
- Planning and control process **delivery**. After being designed, the delivery of services/products from suppliers and through the total operation to customers must be planned and controlled.
- **Developing** capabilities of their processes to improve process performance.

Environmental sustainability: extent to which business activity negatively impacts the natural environment. Decisions -> affect utilization of materials both in short and long term recyclability. Design -> proportion energy, labor, materials wasted. Planning and control -> energy, labor, materials wastage. Reducing waste also reduces costs.

Chapter 2

Operations management being able to make or break any business. Understand importance of operations management -> first understand why things can go wrong in operations and their impact. 'Make' the organization in several ways: 1. OM concerned with doing things better, improvement can potentially make operations the driver of improvement for the whole organization. 2. Through continual learning that can come from its improvement activities, OM can build 'difficult to imitate' capabilities -> strategic impact. 3. OM concerned with process, good OM -> best way to produce good services/products.

Operations managers make decisions on economic, social, political and technological environment changes.

How operations can judge its performance at 3 levels:

- Broad, societal level, 'triple bottom line'
- Strategic level of how an operation can contribute to the organization's overall strategy
- Operational level, 5 'performance objectives'

Societal level – Decisions will affect whole variety of 'stakeholders'. Stakeholders: people and groups who have a legitimate interest in the operation's activities, internal or external. Responsibility of the operations function to understand the objectives of its stakeholders. This idea often termed 'corporate social responsibility (CSR)'. Of increasing importance, both ethical and commercial point of view.

Triple bottom line; 'people, plant and profit': organizations should measure not only on traditional economic profit, also on impact their operations have on society and ecological impact on environment -> sustainability. Balance economic, environmental and societal interests. Social bottom line performance: businesses should accept that they bear some responsibility for impact they have on society and balance the external consequences with more direct internal.

Environmental bottom line (=planet): sustainability: extent to which business activity negatively impacts the natural environment.

Economic bottom line (=profit): operations managers must use operation's resources effectively.

How is operations performance judged at a strategic level? 5 aspects are contributing to the 'economic' aspect of the triple bottom line. How operations affect profit.

- OM affects **costs** -> most important aspect how they judge their performance. Efficiency with which operation purchases its transformed, transforming resources , converts those resources will determine costs.
- OM affects **revenue** -> judges on organization's profitability.
- OM affects the **required level of investment** -> How the transforming resources that are necessary to produce the required type and quantity of its products and services.
- OM affects the **risk of operational failure** -> reduce risks and recover quickly with less disruption (resilience)
- OM affects **ability to build the capabilities on which future innovation** is based -> learn from experience, improve over time.

Operations performance objectives:

Quality – consistent conformance to customers' expectations. Major influence on customer satisfaction. Makes internal easier as well -> quality reduces costs (fewer mistakes -> less time needed to correct, less confusion). Quality increases dependability (can give potential for better services and products.

Speed – elapsed time between customers requesting products/services and their receiving them. The faster -> more likely to buy it, more pay, greater benefit they receive. Speed reduces inventories. Speed reduces risks (forecasting). Speed can give the potential for faster delivery of services and products, and save costs.

Dependability – doing things in time for customers to receive products or services exactly when they are needed/promised. Dependability saves time, saves money (ineffective use of time), gives stability ('quality' of time and will be trust in operation). Dependability can give the potential for more reliable delivery of services and products, and save costs.

Flexibility – being able to change the operation in some way.

- **Product/service flexibility**: ability to introduce new/modified products/services
- **Mix flexibility**: ability to produce a wide range or mix of products/services
- **Volume flexibility**: ability to change its level of output/activity to produce different quantities/volumes of products and services over time
- **Delivery flexibility**: ability to change the timing of delivery of its services/products

Some companies have developed their flexibility in such a way that products and services are customized for each individual customer, yet manage to produce in high volume, mass production -> keeps costs down: **mass customization**.

Agility -> responding to market requirements by producing new and existing products and services fast and flexibly.

Flexibility speeds up response, saves time, maintains dependability. Flexibility can give the potential to create new, wider variety, differing volumes and differing delivery dates of services and products, and save costs.

Cost – important objective for OM, even if the organization does not compete directly on price. Productivity: ratio of what is produced (output) to what is required (input). Often partial measures of input/output used so that comparisons can be made: single-factor measure. $P_s = \frac{O}{I_i}$, $P_m = \frac{O}{\sum_i I_i}$. O = output; I = input; i = factor considered.

Improving productivity is to reduce costs. E.g. reducing costs some or all transformed and transforming resource inputs. Or making better use of inputs. All performance objectives affect cost, so an important way to reduce costs is to improve all the other objectives.

Polar representation of performance objectives good way to represent them all.

Performance measurement: business will need to measure how well, or not, it is doing. -> process of quantifying action where measurement means process of quantification and the performance of the operation is assumed to derive from actions taken by its management.

What factors to include as performance measures? Already discussed 3 levels: social, strategic and operational level. 1. Sometimes to these measures are aggregated into composite measures that combine several measures, help to present a picture of the overall performance. 2. All these factors can be broken down to more detailed ones.

Which are the most important performance measures? Often a compromise reached by making sure there is a clear link between the operation's overall strategy, most important KPI's that reflect strategic objectives, and the bundle of detailed measures used to 'flesh out' each KPI.

Multi-dimensioned performance measurement approaches, such as **balanced scorecard**, give a broader indication of overall performance.

Improving the performance of one objective, but cause others to sacrifice. Two views: 1. Emphasizing 'repositioning' performance objectives 2. Emphasizes increasing the 'effectiveness' by overcoming **trade-offs**. Presumably all the operations would ideally like to be able to offer very high variety while still having very high levels of cost efficiency. Operations that lie on the '**efficient frontier**' have performance levels that dominate those which do not: In these circumstances they may choose to reposition themselves at some other point along the efficient frontier. Those with a position on the efficient frontier will generally also want to improve their operations effectiveness by overcoming the trade-off that is implicit in the efficient frontier curve. An operation's improvement path can be described in terms of repositioning and/or overcoming its performance trade-offs. The distinction between positioning on the efficient frontier and increasing operations effectiveness by extending the frontier is an important one. Any business must make clear the extent to which it is expecting the operation to reposition itself in terms of its performance objectives and the extent to which it is expecting the operation to improve its effectiveness in several ways simultaneously.

Chapter 3

Strategic decisions: are widespread in their effect on the organization to which the strategy refers; define the position of the organization relative to its environment; and move the organization closer to its long-term goals. **Strategy**: total pattern of the decisions and actions that influence the long-term direction of the business. 'Operations' is not the same as 'operational'; it does not have a strategic role. **Content** of operations strategy: specific decisions and actions that set the operations role, objectives and activities. **Process**: method used to make specific content decisions.

3 roles of operations:

- **Implementing** business strategy: without effective implementation even the most brilliant strategy will be rendered totally ineffective
- **Supporting** business strategy: beyond implementing, developing the capabilities allow the organization to improve and refine its strategic goals.
- **Driving** business strategy: by giving it a unique and long-term advantage. The operation drives the company's strategy.

Professors Hayes and Wheelright developed 4 stage model, used to evaluate the role and contribution of the operations function.

1. **Stage 1: Internal neutrality**: very poorest level of contribution by operations function. Holding company back from competing, very little positive contribute towards competitive success, its goal 'to be ignored', improve by 'avoiding making mistakes'
2. **Stage 2: External neutrality**: begin comparing itself with similar companies in outside market
3. **Stage 3: Internally supportive**: among the best in their market. They still aspire to be clearly and unambiguously the very best in the market -> achieve by gaining a clear view of the company's competitive or strategic goals and supporting it by developing appropriate operations resources.
4. **Stage 4: Externally supportive**: operations function as foundation for its competitive success. Look at long-term, innovative, creative and proactive '1 step ahead'.

From 1 to 2: implementing strategy. 2 to 3: supporting strategy. 3 to 4: driving strategy.

Four perspectives of operations strategy:

- **Top-down** reflection of what whole group/business wants to do
- **Bottom-up** activity where operations improvements cumulatively build strategy
- **Market requirements perspective**: involves translating market requirements into operations decisions ('outside-in')
- **Operations resource perspective**: involves exploiting the capabilities of operations resources in chosen markets ('inside-out')

Corporate strategy: strategic positioning of a corporation and the businesses with it
Top-down strategies: operations strategies should take its place in hierarchy, main influence whatever the business sees as its strategic direction.
Business strategies: strategic positioning of a business in relation to its customers, markets and competitors, a subset of corporate strategy.
Functional strategy: overall direction and role of a function within the business, subset of business strategy.
Operations strategies should reflect top-down corporate and/or business objectives. Operations strategy should reflect bottom-up experience of operational reality.

Bottom-up strategies - Top-down perspective provides orthodox view of how functional strategies should be put together. However, more complex -> group reviewing its corporate strategy, also take circumstances, experiences, capabilities of businesses into account. Concept of **emergent strategies**: strategy gradually shaped over time and based on experience instead of theoretical positioning. Key virtues required shaping strategy from bottom-up -> ability learn from experience and philosophy of continual and incremental improvement.

Bottom-up and top-down can reinforce each other -> top-down to implement strategy -> bottom-up to exploit capabilities strategically.

Market-requirements-based strategies – Impossible to ensure that operation is achieving right priority between performance objectives without understanding of what market requires. Operations satisfy customers through developing their 5 performance objectives. Key point is whatever competitive factors are important should influence the priority of each other.

Competitive factors: order winners or qualifiers. **Order-winning factors**: directly and significantly contribute to winning business, key reasons to buy product/service, raising performance -> (improve chances) more business. **Qualifying factors**: performance above particular level just to be considered by the customer. Above will not necessarily gain more business. **Less important factors**: neither one of them.

If operation produces for multiple customer groups -> determine competitive factors for each group.

Way of generalizing behavior of customers and competitors: link to **life cycle** of products/services operation is producing. 4 stages:

1. **Introduction stage**

Product/service first introduced, likely to offer new in terms of design/performance, few competitors with same product/service. Need customers unlikely to be well understood, OM need flexibility to cope with changes.

2. **Growth stage**

Competitors may enter growing market. Keeping up with demand could be main operation preoccupation. Quality level ensure company keeps share of the market.

3. **Maturity stage**

Demand starts to level off. Some early competitors may have left, industry probably dominated by few larger companies. Operations expected: costs down (maintain profits, allow price cutting). Cost, productivity issues, dependable supply main concerns.

4. **Decline stage**

Sales will decline with more competitors dropping out, might be residual market, unless shortage of capacity develops, market will continue to be dominated by price competition.

Operations resources perspective – based on resource-based view (RBV). Firms with above average strategic performance likely to have gained their sustainable competitive advantage because of their core competences of their resources -> way organization acquires resources will, over long term, have significant impact on strategic success. Also, understanding and developing capabilities particularly important. This perspective may identify constraints to satisfy some markets but can also identify capabilities which can be exploited in other markets. Very important to be aware of the **intangible resources** such as relationship with suppliers, reputation with customers, knowledge of process technologies, how staff can work together etc. Central issue for OM: ensure its pattern of strategic decisions really does develop appropriate capabilities within its resources and processes. Resources can have particularly influential impact on strategic success if exhibit properties:

- **Scarce**: unequal access to resources, so that not all competing firms have scarce resources can strengthen competitive advantage
- **Not very mobile**: e.g. specialized staff that does not want to move -> process's resources more likely to be retained over time
- **Difficult to imitate** or substitute for: the less tangible and more connected with tacit knowledge embedded within the organization, the more difficult to understand and copy

Distinction operation's structure and infrastructural decisions -> **structural**: classed as primarily influencing design activities. **Infrastructural**: influence workforce organization, planning, control, improvement. Compare with hardware/software. The best and most costly facilities and technology only effective if operation also has appropriate infrastructure which governs day-to-day basis.

How can operations strategy form the basis for operations improvement? 2 models use market requirements and operations capabilities perspectives.

1. '**Line of fit**' between market requirements and operations capabilities

There should be a fit between what an operation is trying to achieve in its markets (market requirements) and what it is good at doing (operations capabilities). In terms of the framework improvement means:

i) **Achieving 'alignment'** – achieving approximate balance between required market performance and actual operations performance.
ii) **Achieving 'sustainable' alignment** – also important whether operations processes could adapt to the new market conditions.
iii) **Improving overall performance** - the more demanding the level of market requirements, the greater the level of operations capabilities will have to be.

Sometimes deviating from the line brings risks, sometimes improvements.

2. **Importance-performance matrix**

More focused -> gain understanding of relative importance customers of various competitive factors. Competitors: points of comparison against which operation can judge its performance. Competitive viewpoint: operations improve their performance, the improvement which matters most is that which takes the operation past the performance levels achieved by competitors. Both importance and performance need to be viewed together to judge the prioritization of objectives:

- **Judging importance to customers** – e.g. order-winning, qualifying and less important factors, within -> strong, medium, weak
- **Judging performance against competitors** – also in point scale

All factors can be put in a matrix. Different zones:

- **'Appropriate' zone** – satisfactory
- **'Improve' zone** – Below lower bound of acceptability
- **'Urgent-action' zone** – Important to customers but performance below that of competitors
- **'Excess?' zone** – High performance but not important for customers

'Process' of strategy -> 'how' strategies are put together. 4 stages:

Formulation – of operations strategy: process of clarifying the various objectives and decisions that make up the strategy, and the links between them. Infrequent activity.

What should the formulation process be trying to achieve?

- Is operations strategy **comprehensive**? – Does it include all important issues?
- Is operations strategy **coherent**? – coherence = choices made in each decision area all direct the operation in the same strategic direction, complementing and reinforcing each other.
- Does operations strategy have **correspondence**? – Decision pursued in each part of the strategy should correspond to the true priority of each performance objective.
- Does operations strategy identify **critical issues**? The more critical, the more attention deserves

Implementation – strategies operationalized or executed. 3 issue important:

- **Clarity of strategic decisions** - crucial for formulation: clarity. The clearer, the easier to implement the intention.
- **Motivational leadership** – to bring sense and meaning to strategic aspirations, maintain sense of purpose, modify plan in light of experience
- **Project management** – breaking up complex plan into set of distinct activities

Monitoring – Track performance make sure changes are proceeding as planned.

Control – involves evaluation of results from monitoring the implementation. Difficult at strategic level because strategic objectives not always clear. Many strategies too complex for that. So, rather than adhering dogmatically to a predetermined plan, it may be better to adapt as circumstances change.

Chapter 4

Innovation: doing something new. **Invention**: novel or unique but does not necessarily imply that the novel device/method has potential to be practical, economic etc. Innovation goes further than invention. Innovation not just the idea, also process of transforming into something that provides a return for organization, customer etc. **Creativity**: ability to move beyond conventional ideas, rules or assumptions, in order to generate new ideas. Innovation creates the novel idea; **design** makes it work in practice.

New ideas usually follow **S-shaped progression**. Begin: often large amounts resources, time, effort needed, relatively small performance improvements. With time, experience and knowledge grow -> performance increases. Idea established, extending performance further becomes increasingly difficult. Again improved -> many s's connected.

Obvious difference between how pattern of new ideas emerges in different operations is rate and scale of innovation. **Radical innovation**: discontinuous, 'breakthrough' change, and **incremental innovation**: smaller, continuous changes. New entrants to market often more radical.

The Henderson-Clark model – why some established companies sometimes fail to exploit seemingly obvious incremental innovations. **Technological knowledge** into 'knowledge of the components of knowledge' and 'knowledge of how the components of knowledge link together' -> '**architectural knowledge**'. **Incremental innovation**: upon existing component and architectural knowledge. **Radical innovation**: changes both component and architectural knowledge. **Modular innovation**: existing architectural knowledge, requires new knowledge for components. **Architectural innovation**: great impact upon linkage of components (architecture), knowledge components same.

Matters -> ability to identify innovations and manage their transformation into effective designs so that they can sustain competitive advantage and/or generate social payback.

Product designers: try achieve pleasing designs meet/exceed customers' expectations, performs well, reliable, manufactured easily and quickly. Service designers: also customers' expectations, but within capabilities and delivered at reasonable cost. Design can add value to any organizations, it can:

- Drive and operationalize innovation
- Differentiate products/services
- Strengthen branding
- Reduce overall costs associated with innovation

Design process as input-transformation-output. The performance of the design process can be assed in much the same performance objectives, now also include sustainability as design objective.

Quality for design process – distinguish high- and low-quality designs in terms of ability to meet market requirements. Distinction between specification quality and conformance quality of designs important. 'Conformance' failures in the design process -> product recalls. 'Specification' quality -> degree of functionality/experience etc.

Speed for design process – depends on industry. Advantages of fast design:

- Early market launch
- Starting design late
- Frequent market stimulation

Dependability for design process – dependable -> minimize design uncertainty. External disturbances to innovation process will remain. Flexibility one of most important ways to ensure dependability.

Flexibility for design process – ability to cope with external/internal change. External when requirements customers change, especially necessary in fast moving markets. Internal -> increasing complexity and interconnectedness of product/service. One way measuring innovation flexibility: compare cost of modifying design with consequences profitability if no changes made. The lower the cost, the higher the flexibility.

Cost for design process – usually in 3 categories: cost buying inputs to process, cost providing labor, other general overhead costs of running process. Delayed completion design -> more expenditure and delayed revenue.

Sustainability for design process – extent to which benefits the triple bottom line (people, planet, profit). Design innovation process important in impacting ethical, environmental and economic well-being of stakeholders. Lots of examples in the book.

Design process involve a number of stages that move an innovation from a concept to a fully specified state. Stages:

1) <u>Concept generation</u>

Innovation can come from sources:

- Ideas from customers
- Listening to customers
- Ideas from competitor activity ('reverse engineering')
- Ideas from staff
- Ideas from R&D

Open-sourcing: products developed by open community, incl people who use it. E.g. Google, Wikipedia. Concept: large communities of people around the world come together and produce software product. Regularly updated and free. **Crowdsourcing**: process of getting work/funding/ideas from a crowd of people. **Parallel-path approach**: utilizing variety of different sources and approaches to generating ideas. **Ideas management**: type of enterprise software that can help operations to collect ideas from employees. Often used to focus ideas of specific organizational targets.

2) <u>Concept screening</u>

Evaluate concepts by assessing the worth or value of design options, assessing against number of design criteria. 3 broad categories:

1. **Feasibility** of design option – can we do it?

   Do we have skills? Do we have organizational capacity? Have financial resources?

2. **Acceptability** of the design option – do we want to do it?

   Satisfy performance criteria? Will our customers want it? Give enough financial return?

3. **Vulnerability** – do we want to take the risk?

   Do we understand full consequences? What could go wrong? 'Downside risk' -> consequences if everything goes wrong.

Uncertainty surrounding the design reduces as the number of alternative designs considered decreases -> **design funnel**. As the design progresses, changes become more expensive.

3) <u>Preliminary design</u>

Objective this stage -> have a first attempt at specifying the individual components of products/services and relationship between them -> constitute final offering. First task: define exactly what will go into product/service -> require information about constituent component parts so package and component structure.

3 approaches to reduce design complexity (and reduce costs): **Standardization** – restrict variety to which has real value for customer (e.g. fast-food restaurants, clothes (sizes)). **Commonality** – using common elements within a product/service simplifies. Likewise, standardizing format of information inputs to a process can be achieved by using appropriately designed forms or screen formats. **Modularization** – designing standardized 'sub-components' which can be put together in different ways. Possible to create wide choice through fully interchangeable assembly of various combinations of smaller number of standard sub-assemblies.

4) Evaluation and improvement

To take preliminary design and subject it to series of evaluations to see if can be improved before tested in market. Several techniques, best known: **quality function deployment** (QFD): to try ensure eventual innovation actually meets needs of customers. QFD matrix: formal articulation of how company sees relationship between requirements of customers (whats) and design characteristics of new product (hows):

- Whats: list of competitive factors which customers find significant, relative importance is scored
- Competitive scores indicate relative performance
- Hows: various dimensions of design which will operationalize customer requirements
- Central matrix (relationship matrix): view of interrelationship betweens whats and hows.
- Technical assessment: contains absolute importance of each design characteristic
- 'Roof' captures information team has about correlations between various design characteristics.

5) Prototyping and final design

Turn improved design into prototype -> tested. Trials can include simulations, but also implementation. **Computer-aided design (CAD)** to create and modify product drawings. Store and retrieve design data quickly, also manipulate design details. Enhance flexibility of design activity, use of standardized libraries of shapes can reduce possibility of errors. **Alpha and beta testing** -> Alpha testing: internal process where developers examine product for errors, private process, in simulated environment. After that -> Beta testing: product released for testing by selected customers, external pilot test in real world, 'field testing' or customer validation.

Benefits of interactive product and service innovation? – Generally considered mistake to separate product and service design from process design. Merging stages of design innovation: 'interactive design'. Main benefit -> reduction in elapsed time for whole design innovation activity -> 'time to market' (TTM). Gives increased competitive advantage, more opportunities to improve performance. 3 factors significantly reduce time:

- Simultaneous development of various stages in overall process

Advantages of stages step-by-step: easy to manage, each stage can focus on its skills and expertise. Main problem: time consuming and costly. Often little need to until the absolute finalization of one stage. Often: 'simultaneous/concurrent engineering'.

- Early resolution of design conflict and uncertainty

A decision, once made, need not totally and utterly commit the organization. Manage to resolve conflict early in design activity, reduce degree of uncertainty within project and reduce extra costs, time. Design process requires strategic attention early, when there is most potential to affect design decisions.

- Organizational structure which reflects the development project

Total process of developing concepts through to market will almost certainly involve personnel from several different areas of the organization. Organizational question which of these 2 ideas – various organizational functions which contribute or design project itself – should dominate the way in which design activity is managed?

- Pure functional organization: all staff associated with design project based unambiguously in their functional groups, no project-based group at all
- Project forms: all individual members of staff who are involved could be moved out of their function to a task force dedicated solely to the project

In between various types with varying emphasis on two aspects of the organization. **Skunkworks**: organizational structure claimed to release design and development creativity of a group. Small team taken out of normal work and granted freedom from their normal management.

Chapter 5

Structure of operation's supply network: shape and form of the network.
Scope of operation's supply network: extent that operation decides to do activities performed by network itself, as opposed to requesting a supplier to do them.
Supply network: setting an operation in the context of all the other operations with which it interacts (suppliers/customers).
On supply side -> operation has suppliers of parts, information, services. These suppliers have their own suppliers etc.:

- First-tier suppliers: group of operations that directly supply the operation
- Second-tier suppliers: supply the first-tier suppliers

On demand side -> operation has customers, might have own set of customers

- First-tier customers: main customers group of operation
- These supply second-tier customers etc.

Its **immediate supply network**: suppliers and customers who have direct contact with an operation. All the operation which form network of suppliers and customers: **total supply network**.

Why look at whole supply network? 3 reasons:

- It helps an **understanding of competitiveness** – immediate customers and immediate suppliers main concern but looking beyond these to understand why they act like they do
- It helps **identify significant links** in the network – some operations contribute more to the performance objectives which valued by customers, so analysis of network needs to understand the downstream and upstream operations which contribute most to end customers service
- It helps **focus on long-term issues** – times when circumstances render parts of a supply network weaker than its adjacent links. Long-term supply network view would involve constantly examining technology and market changes to see how each operation might be affected.

Decisions relating to structure and scope often interrelated. The structure of operations supply network determined by 3 sets of decisions:

1. <u>How should the network be configured?</u>

'Configuring' a network: determining overall pattern. Involves attempting to manage network behavior by reconfiguring the network so as change the nature of the relationships between them, sometimes parts merging. Another trend: 'disintermediation' -> cutting out intermediaries, making direct contact with customers' customers etc.

Co-opetition – business surrounded by suppliers, customers, competitors and complementors. Complementors: your products more valued when also products of complementor. Competitors: your product less valued when also products of competitor. Can be both. Customers and suppliers should have 'symmetric' roles. Harnessing the value of suppliers, just as important as listening to needs of customers. Co-opetition when all parts can be enemies and friends.

Business ecosystem – economic community supported by a foundation of interacting organizations and individuals. Economic community produces goods/services of value to customers, who members of ecosystem. Difference with co-opetition: inclusion in idea of ecosystem of business no or little direct relationship with main supply network, exist because of network.

Dyadic interaction: focus on individual interaction between 2 specific operations to understand them better. Recently, more triads that are basic elements of supply network, complex can be broken down in **triads**. Thinking in triads strategically important: emphasizes dependence organizations are placing on their suppliers' performance when outsource service delivery. 2nd: control that buyer has over service delivery to its customer is diminished. Products/service bypass the buying organization and go directly from provider to customer. 3rd: direct link between service provider and customer can result in power gradually transferring over time from buying organization to supplier that provides service. 4th: becomes increasingly difficult for buying organization to understand what is happening between supplier and customer at day-to-day level. 5th: closeness between supplier and customer could prevent buyer from building important knowledge.

2. <u>What physical capacity should each part of the network have? (long-term capacity decision)</u>

Low occupancy -> high cost per customer. Operating high level of capacity -> longer waiting times, reduced customer service, overtime -> reduce productivity. As nominal capacity increases, lowest cost point at first reduces -> fixed cost and capital costs of constructing do not increase proportionally. These 2 factors: **economies of scale**. Do not go one forever -> **diseconomies of scale** -> complexity costs increase as size increases (communication, coordination) and larger center more likely to be partially underutilized (transport cost).

Advantages small scale in 4 areas:
- Allow businesses to locate near 'hot spots'
- Can respond rapidly to regional customer needs by basing more and smaller units of capacity close to local markets
- Take advantage of potential for HR development by allowing staff greater degree of local autonomy.
- Explore radically new technologies by acting in same way as smaller, more entrepreneurial rival.

In deciding when new capacity to be introduced company can mix 3 strategies:

- Capacity introduced generally to lead demand – always sufficient capacity to meet forecast demand
- Capacity introduced generally to lag demand – demand always ≥ capacity
- Capacity introduced to sometimes lead and sometimes lag demand – **smoothing with inventory**: inventory built up during lead times used to meet demand during lag times

Each advantages and disadvantages

Alternative view -> examining cost implications of adding increments of capacity on break-even basis. Each additional unit capacity -> fixed-cost break. Operation unlikely profitable at low out-puts. Assuming prices > marginal costs -> revenue exceed total costs. However, output equal to capacity may not be sufficient and thus unprofitable.

3. <u>Where should each part of the network be located? (location decision)</u>

Location decisions have effect on operation's costs, ability to serve customers, difficult to undo. Reasons for location decisions:

- **Changes in demand** – customer demand shift, changes volume demand (expand on existing site, larger site, additional location)
- **Changes in supply** – changes in cost, availability (mining, India -> cheap)

Aim location decision to achieve appropriate balance between (different for profit and non-profit organization):

- Spatially variable costs of operation (spatially variable = something changes with geographical location)
- Service the operation is able to provide to its customers
- Revenue potential

Location decision for any operation determined by relative strength of number of factors:

- Labour costs – simple wage costs can be misleading, must take effects productivity difference and different exchange rate in account.
- Labour skills availability – e.g. science parks close to university
- Land costs – trade-off between hotspot and generate enough revenue
- Energy costs – availability of relatively inexpensive energy
- Transportation costs – include transporting inputs form their source to operation and cost of transporting outputs to customers. Proximity to sources of supply dominates decision, also transportation to customers.
- Community factors – derive from social, political and economic environment (tax, corruption, language etc.)
- Suitability of the site itself – intrinsic characteristics can affect ability to serve customers (hotel near beach).
- Image of the location – e.g. fashion design in Milan
- Convenience for customers – important for service to customers, e.g. hospitals close to centers of demand

Scope of operation's activities within network determined by 2 decisions:

1. <u>Extent and nature of the operation's vertical integration</u>

Vertical integration: extent to which an organization owns network of which it is a part. Usually involves organization assessing wisdom of acquiring suppliers/customers. OEMs: original equipment manufacturers. Organization's vertical integration strategy can be defined in following terms:

- Direction of integration – strategy of expanding on supply side of network: **backward** or 'upstream' vertical integration. Expanding on demand side: **forward** or 'upstream' vertical integration. Backward -> to gain cost advantages or prevent competitors gaining control of important suppliers. Forward -> organization closer to its markets, allows more freedom make direct contact customers.
- Extent of process span of integration – some organization deliberately choose not to integrate far, some choose very vertically integrated.
- Balance among vertically integrated stages – amount of capacity at each stage devoted to supplying next stage. Totally balanced -> one stage produces only for next and fully satisfied.

Perceived advantages of vertical integration:

- Secures dependable access to supply or markets – more secure supply or bring business closer to customers. Downstream can give firm greater control over its market positioning.
- May reduce costs – take start-up and learning costs in account. Reduces transportation, loading costs
- May help to improve product/service quality – for specialist or technological advantage, preventing getting in hand of competitors.
- Helps in understanding other activities in the supply network

Perceived disadvantages of vertical integration:

- Creates an internal monopoly
- You cannot exploit economies of scale – specialist suppliers who can serve more than one customer likely to have volumes larger
- Results in loss of flexibility – high proportion of costs -> fixed -> reduction total volume easily below break-even point
- Cuts you off from innovation
- Distracts you from core activities (loss of focus) – vertical integration -> doing more things which can distract from few particularly important things.

2. <u>Nature and degree of outsourcing it engages in</u>

Vertical integration and **outsourcing** are the same thing. No business does everything that is required. Outsourcing -> 'do-or-buy' decision. Many indirect and administrative processes outsourced: business process outsourcing (BPO). Reason: reduce cost. Sometimes gains in quality and flexibility. <u>Difference</u>: vertical integration usually applied to whole operations. Outsourcing: usually smaller sets of activities previously performed in-house.

Assessing the advisability of outsourcing should include how it impacts relevant performance objectives. Also include consideration of the strategic importance of the activity and operation's relative performance.

Outsourcing: deciding to buy in products/services rather than perform the activities in-house.
Offshoring: obtaining products/services form operations that are based outside one's own country. One can do both, very closely related, motives may be similar.

Chapter 6

Process design -> at start important to understand design objectives, when overall shape and nature of process being decided. To do by positioning according to its volume and variety characteristics. Eventually, details analyzed to ensure that fulfils objectives effectively. Design of processes cannot be done independently of the products/services that they are creating. Point of process design -> make sure that the performance of the process is appropriate for whatever it is trying to achieve. Again, include sustainability as operational objective. Design of process judged on quality, speed, dependability, flexibility, costs and sustainability.

'Micro' performance flow objectives used to describe process flow performance:

- **Throughput rate** (flow rate): rate at which items emerge form process, number of items passing through process per unit of time
- **Cycle time** (takt time): reciprocal of throughput rate; time between items emerging from process.
- **Throughput time**: average elapsed time taken for imputs to move through process and become outputs
- **Number of items in process** ('work-in-progress'/in-process inventory) as an average over period of time
- **Utilization of process resources**: proportion of available time that resources within process are performing useful work.

Throughput time = work-in-progress x cycle time (e.g. 10-minute wait = 10 people in system x 1 min per person) → **Little's law**. -> average number of things in system is product of average rate at which things leave system and average time one spends in system.

Throughput efficiency: % of time item being worked on during item being processed. This assumes all 'work content' is needed, be that individual elements may not be considered 'value-added'. -> value-added throughput efficiency. Throughput efficiency = work content/throughput time x 100%.

When transformed resource is information, when information technology used to move, process design: '**workflow'**.

Bottleneck in process: congestion because workload greater than capacity. Will dictate rate at which process can operate. Bottlenecks reduce efficiency, other stages underloaded. **Balancing**: trying to allocate work equally between stages, wasted time as percentage: **balancing loss**. Process design must respect task precedence, precedence diagram: representation of ordering of elements.

Advantages of long thin arrangement of stages:

- Controlled flow of items, easy to manage
- Simple handling, when heavy and difficult to move
- Lower capital requirements, specialist piece of equipment only bought once
- More efficient operation, one person small part of the job

Advantages short fat arrangement:

- Higher mix flexibility, each stage could specialize in different types

- Higher volume flexibility, when volume varies each stage can be closed or opened
- Higher robustness, if one breaks down, others can continue
- Less monotonous work

Important design objectives: to which extent process designs should be standardized. Doing things differently gives degree of autonomy and freedom but can cause confusion, misunderstanding, inefficiency. Practical dilemma: how draw line between processes required to be standardized and those allowed to be different.
Fundamental issues in regard to sustainability:
- Source of inputs
- Quantities and sources of energy consumed in the process
- Amounts and type of waste material, created in manufacturing processes
- Life of product itself
- End of life of the product

Life cycle analysis -> analyses all production inputs, life cycle use of product and final disposal (total energy used and emitted waste).

The design of any process should be governed by the volume and variety it is required to produce. General approaches to designing and managing processes: **process types**.
1. **Project processes** – discrete, usually highly customized products, relatively long timescale to complete. Low volume and high variety. Transforming resources may have to be organized especially for each item (movie production, construction)
2. **Jobbing processes** – high variety and low volumes, each product has to share operation's resources with many others, can be complex, fewer unpredictable circumstances (made-to-measure tailors, furniture restorers)
3. **Batch processes** – wide range volume and variety (special gourmet frozen foods, components which go into mass automobiles)
4. **Mass processes** – high volume and relatively narrow variety, repetitive and largely predictable (frozen food production, television factories)
5. **Continuous processes** – even higher volume and usually lower variety than mass, operate for longer periods of time, relatively inflexible, capital-intensive technologies, smooth flow from part to part (water processing, steel making)

1. **Professional services** – high-contact processes, spend considerable time, high level of customization, people based (management consultants, lawyers, architects)
2. **Service shops** – levels of volume and variety between the extremes (banks, schools, hotels, travel agents)
3. **Mass services** – limited contact, many customer transactions, staff follow set procedures (supermarkets, airport, library, call center). **Script**: using carefully designed enquiry process to cope with high volume.

Common to show volume-variety position in 'product-process' matrix. Many important elements of process design strongly related to volume-variety position of process. Most processes should lie close to diagonal of matrix -> 'natural' diagonal, 'line of fit'. On line -> normally lower operating costs. On the right: lower volumes and higher variety, more flexible than seems -> not taking advantage of ability to standardize activities -> higher costs. Left: position normally for higher volume and lower variety, now 'over-standardized' and too inflexible for their position -> higher costs.

Detailed design of process involves identifying all individual activities and deciding sequence. Can be done using **process mapping** – describing processes in terms of how activities within the process relate

to each other. Many techniques, all identify different types of activity that take place during process and show flow materials/people/information through process. Symbols used to classify different types of activities. See symbols in book!!!
Sometimes too complex -> mapped at more aggregated level: **high-level process mapping**. More detailed level: **outline process map**, in a general way. All activities: **detailed process map**. **Micro-detailed process map**: specify every single motion involved in each activity.

Sometimes useful to map such processes in a way that makes the degree of visibility of each part of the process obvious -> allows those parts of the process with high visibility to be designed so that they enhance the customer's perception of the process. Line of visibility': boundary between activities that the customer could see and those they couldn't. The highest level of visibility: above 'line of interaction' are those activities that involve direct interaction between staff and customer.

Effects of process variability – 2 fundamental types of **variability**:

- Variability in demand for processing at an individual stage within the process, variation in the inter-arrival times of items to be processed
- Variation in time taken to perform activities at each stage

When arrival and process times are variable, sometimes the process will have items waiting to be processed, sometimes process will be idle. Process will have a 'non-zero' average queue. The greater the variability in the process, the more the waiting time utilization deviates from the simple rectangular function. 3 options to process designers wishing to improve waiting time or utilization performance:

- Accept long average waiting times and achieve high utilization
- Accept low utilization and achieve short average waiting times
- Reduce variability in arrival times, activity times, or both, and achieve higher utilization and short waiting times.

Chapter 7

'Layout' of operation or process: how its transforming resources are positioned relative to each other, how its various tasks are allocated to these transforming resources and general appearance of the transforming resources. Layout must start with full appreciation of objectives, will largely depend on strategic objectives, some general:

- Inherent safety – e.g. fire exits clearly marked, pathways clearly defined
- Security – anyone with malicious intent cannot gain access
- Length of flow – flow of materials/information/customers channeled to be appropriate for objectives, e.g. minimizing distance travelled, in supermarkets -> maximize
- Minimize delays – caused by over-long routes, inconvenient placing facilities, bottlenecks
- Reduce work-in-progress – excessive caused by bottlenecks, layout can limit
- Clarity of flow – all flow of materials/customers well signposted, clear to staff/customers
- Staff conditions – layout should provide pleasant working environment
- Communication – layout designed to promote meetings etc.
- Management co-ordination – supervision and communication should be assisted by relative location of staff, use of communication devices and information plants
- Accessibility – all machines, plant or equipment should be accessible so sufficient for inspection, cleaning etc.
- Use of space – all layouts achieve appropriate use of total space (maximize or minimize)
- Use of capital – minimized when finalizing layout
- Long-term flexibility – layouts changed periodically as needs of operation change. Good layout devised with possible future needs in mind.
- Image – layout can help to shape image, appearance of layout can be used as deliberate attempt to establish company's brand.

Most practical layouts are derived from 4 basic types:

1. **Fixed-position layout**

Transformed resources do not move between transforming resources. Instead of materials/information/customers flowing through operation, recipient of the processing is stationary and equipment move. E.g.: motorway construction, open-heart surgery, shipbuilding. Loosely related to project process type.
<u>Advantages</u>: very high product and mix flexibility, product/customer not moved/high variety of tasks for staff
<u>Disadvantages</u>: very high unit costs, scheduling space and activities can be difficult
<u>Objective</u>: maximize transforming resources' contribution to the transformation process

2. **Functional layout**

Resources or processes located together. When products/information/customers flow through operation, will take route from activity to activity according to their needs. Flow pattern complex. E.g.: hospital, machining the parts which go into aircrafts engines, supermarket, library (see book for extensive example). Loosely related to jobbing and batch process.
<u>Advantages</u>: high product and mix flexibility, relatively robust in the case of disruptions, easy to supervise
<u>Disadvantages</u>: low utilization, can have very high WIP, complex flow, with much transportation

Objective: Minimize distance travelled

3. **Cell layout**

Where transformed resources entering the operation are pre-selected to move to one part of the operation in which all the transforming resources, to meet their immediate processing needs, are located. Cell itself may be functional or line layout. Cell layout to bring order. E.g.: computer component manufacture, 'lunch' products area in supermarket, maternity unit in hospital. Service -> department sore with units of particular class. Loosely related to mass process
Advantages: can give good compromise, fast throughput, group work can result in good motivation
Disadvantages: can be costly to rearrange existing layout, can need more plant
Objective: create efficient flow for different component/product families

4. **Line ('product') layout**

Involves locating transforming resources entirely for convenience of transformed resources. Each product/information/customer follows prearranged route in which sequence activities required matches sequence in which processes have been located. 'Flow or product layout'. E.g.: automobile assembly, self-service cafeteria. Still improvements are made to the line layout.
Advantages: low unit costs for high volume, opportunities for specialization of equipment
Disadvantages: can have low mix flexibility, not very robust in case of disruptions, work can be very repetitive
Objective: balancing flow

Many operations combine elements of some or all basic layouts, e.g. hospitals. Restaurant -> kitchen functional layout with various processes grouped together. Traditional service restaurant arranged in fixed-position layout. Buffet restaurant cell-type layout. Cafeteria restaurant, line layout.

Importance of flow will depend on volume and variety. Volume very low and variety high -> fixed-position. Resources in low-volume-high-variety processes should be arranged to cope with irregular flow.

For any particular product/service, the fixed costs of physically constructing a fixed-position layout are relatively small compared with any other way of product or service. The variable costs of producing each individual product/service are relatively high compared to the alternative layout types. The total costs for each layout type will depend on the volume of products or services produced. There is uncertainty about the exact fixed and variable costs of each layout -> decision can rarely be made on cost alone. Different layout types have different fixed and variable which determine the appropriateness of layout for varying volume-variety characteristics.

Look and feel very important, especially in high-visibility operations. Factors deal with physiological aspects of working. Aesthetics reflect culture of organization. It can encourage desired behaviors. Allen curve -> shows powerful negative correlation between physical distance and frequency of communication. Even when sitting closely, more likely to email more often. Look and feel of environment within operation from customer's perspective: '**servicescape'**. Individual factors that influence experience will lead to certain responses: 1. Cognitive (what people think). 2. Emotional (what they feel). 3. Physiological (what their body experiences). Also contain subjective stimuli.

Detailed design in fixed-position layout – location of resources determined on convenience of transforming resources.

Detailed design in functional layout – complex due to different options of flow. For N centers there are N! different ways of arranging. Designer needs essential pieces of information:

- Area required by each work center
- Constraints on shape of the area allocated to each work center
- Degree and direction of flow between each work center -> usually shown on **flow record chart**
- Desirability of work centers being close together or close to some fixed point in the layout

Prime objective usually to minimize costs -> minimizing total distance. **Effectiveness of layout** = $\sum F_{ij}D_{ij}$. F_{ij} = flows in loads per period of time from work center i to work center j. D_{ij} = distance between work center i and work centre j. The lower the score, the better.

Due to complexity now several heuristic procedures to derive a good sub-optimal solution. One of them: **CRAFT** -> when N is large, it is feasible to start with initial layout and then evaluate all different ways of exchanging two work centers. There are $\frac{N!}{2!(N-2)!}$ Possible ways of exchanging 2 out of N work centers. 3 inputs required: matrix of flow between departments; matrix of cost associated with transportation between each department; spatial array initial layout. From these: location of centroids of each department calculated; flow matrix weighted by cost matrix, weighted flow matrix is multiplied by distances between departments -> total transportation costs of initial layout; then calculates cost consequence of exchanging every possible pair of departments.

Detailed design in cell layout – In dividing layout into 4 cells, OM implicitly taken two interrelated decisions regarding: 1. Extent and nature of the cells it has chosen to adopt; 2. Which resources to allocate to which cells. Detailed design of cellular layouts difficult, partly because idea of cell itself is compromise between process and product layout. Simplify -> concentrate either process or product aspects of cell layout.

Concentrate on processes -> could use cluster analysis to find which processes group naturally together.

PFA production flow analysis = one approach to allocate tasks and machines to cells, examines both product and requirements and process grouping simultaneously.

Detailed design in line layout – instead of ‘where to place what’ it is ‘what to place where’. Work tasks allocated to each location. Layout activity very similar to aspects of process design -> see chapter 6! Main decisions:

- What cycle time is needed?
- How many states are needed?
- How should the task-time variation be dealt with?
- How should the layout be balanced? (bottlenecks reduced)
- How should the stages be arranged? (‘long thin’ or ‘short fat’)

Chapter 8

How operations managers deal with process technology is now one of the most important decisions that shape the capabilities of operations. High-volume services have for years understood the value of process technologies. Operations managers do not need to be technologists. Yet they should be able to do three things:

1. Understand the process technology to the extent that they are able to articulate what it should be able to do: What do operations managers need to know about process technology?
2. Evaluate the process technology – should be able to evaluate alternative technologies and share in the decisions of which technology to choose: How does the process technology affect the operation?
3. Must implement the process technology so that it can reach its full potential in contributing to the performance of the operation as a whole: How can operations managers introduce new process technology smoothly?

Process technology: machines/equipment/devices that create/deliver products/services. This is pervasive in all types of operations. Without it many of the products and services we all purchase would be less reliable, take longer to arrive and arrive unexpectedly, only be available in a limited variety, and be more expensive. Process technology has a very significant effect on quality, speed, dependability, flexibility and cost. Facilitating direct inputs: indirect process technology, increasingly important.

Distinguish different types of process technologies by that they process:
Material-processing technologies: include any technology that shapes, transports, stores, or in any way changes physical objects. E.g.: machines and equipment found in manufacturing operations, trucks, warehouse systems, etc.
Information-processing technology: Most common single type of technology within operations, and includes any device which collects, manipulates, stores or distributes information. The use of internet-based technology increases both reach and richness.
Customer-processing technology: Very much in evidence in many services. Human element of service is being reduced with customer-processing technology being used to give an acceptable level of service while significantly reducing costs. Three types of customer-processing technologies:
1. **Active interaction technology** such as automobiles, telephones and ATMs. In all of these, customers themselves are using the technology to create the service.
2. **Passive interactive technologies**: they 'process' and control customers by constraining their actions in some way. E.g. aircraft, cinemas and theme parks.
3. Technologies that are 'aware' of customers but not the other way round. E.g.: security monitoring technologies in shopping malls. The objective of these 'hidden technologies' is to track customers' movement or transactions in an unobtrusive way.
Some technologies process more than one type of resource. -> **Integrating technologies** = the technologies that process combinations of material, people and customers.

Understanding process technology: knowing about the principles behind the technology to be comfortable in evaluating some technical information, capable of dealing with experts in the technology and confident enough to ask relevant questions. In particular the following 4 key questions can help operations to grasp the essentials of the technology:
What does technology do (that is different from other similar technologies)?
How does it do it? = what particular characteristics of the technology are used to perform its function?
What benefits does it give (using the technology give to the operation)?

What constraints or risks does it impose (using the technology place on the operation?
Operations managers should understand enough about process technology to evaluate alternatives.

Implications: natural consequence for operation of adopting the technology, effects. Emerging technologies can have a potentially significant impact on how operations are managed.

Technology-related decision -> choice between alternative technologies. Process technologies can be evaluated in terms of their fit with process tasks, their effect of performance and their financial impact.
<u>Does the processing technology fit the processing task?</u> – High-variety-low-volume require general purpose, high-volume-low-variety more dedicated. In between 3 dimensions:

- **The degree of automation**

The ratio of technological to human effort it employs: the capital intensity of the process technology. Generally processes that have high variety and low volume will employ process technology with lower degrees of automation than those with higher volume and lower variety

- **Scale/scalability**, capacity of technology to process work

The virtues of smaller-scale technology are often the nimbleness and flexibility that is suited to high-variety, lower-volume processing. The advantages of large-scale technologies are similar to those of large-capacity increments (discussed in chapter 6). *Scalability*: ability to shift to a different level of useful capacity quickly and cost-effectively.

- **Coupling/connectivity**, extent to which it is integrated with other technologies

Coupling: linking together of separate activities within a single piece of process technology to form an interconnected processing system. Tight coupling usually gives fast process throughput. Closely coupled technology can be both expensive and vulnerable. Process technology in high-volume, low-variety processes is relatively automated, large-scale and closely coupled when compared to that in low-volume, high-variety processes. Coupling generally more suited to relatively low-variety-high-volume. Higher-variety processing generally requires a more open and unconstrained level of coupling -> different products and services will require a wide range of processing activities.

<u>How does technology improve operation's performance?</u>
A sensible approach to evaluating the impact of any process technology on an operation is to assess how it affects the quality, speed, dependability, flexibility and cost performance of the operation. (Flexibility usually only volume and delivery improve)
<u>Does the technology give an acceptable financial return?</u>
While the benefits of investing in new technology can be spread over many years into the future, the costs associated with investing in the technology usually occur up front. The rate of interest assumed: discount rate. More generally, the present value of €x in n years' time, at a discount rate of r per cent is:
$€\frac{x}{(1+\frac{r}{100})^n}$.

Implementing process technology: organizing all the activities involved in making the technology work as intended. No matter how potentially beneficial technology, remains only prospective until it has been implemented successfully. Four particularly important issues that affect technology implementation:

1. Technology planning in the long-term – technology road mapping

Technology roadmap: approach that provides structure that attempts to assure the alignment of developments in technology. TRM process supports technology development by facilitating collaboration between various activities and contribute technology strategy. Devine technological evolution in advance by planning timing and relationships between elements involved. Benefits: way bring together stakeholders and various perspectives they have.

2. Resource and process ‘distance’

The degree of difficulty in the implementation of process technology will depend on the degree of novelty of the new technology resources and the changes required in the operation’s processes. The less that the new technology resources are understood, the greater their ‘distance’ from the current technology resource base of the operation. The greater the resource and process distance, the more difficult any implementation is likely to be. The more distance -> more difficult to know what has (not) worked and why

3. Customer acceptability

If customers are to have direct contact with technology, they must have some idea of how to operate it. Where customers have an active interaction with technology, the limitations of their understanding of the technology can be the main constraint on its use. The ability of the operation to train its customers in the use of its technology depends on 3 factors: complexity, repetition and variety of tasks performed by customers. If services are complex, higher levels of ‘training’ may be needed. Frequency of use -> payback for the ‘investment’ in training will be greater if customer uses the technology frequently. Training easier if customer is presented with a low variety of tasks.

4. Anticipating implementation problems

The implementation of any process technology will need to account for the ‘adjustment’ issues that almost always occur when making any organizational change. Adjustment issues: losses that could be incurred before the improvement is functioning as intended. But estimating the nature and extent of any implementation issues is notoriously difficult. Murphy’s Law seems to prevail, this law: “if anything can go wrong, it will” -> effect has been identified empirically in range of operations, especially when new types of process technology are involved. New technology rarely behaves as planned and as changes are made their impact ripples throughout the organization. It is recognized that implementation may take some time; therefore allowances are made for the length and cost of a ‘ramp-up’ period. As the operation prepares for the implementation, the distraction causes performance actually to deteriorate. The area of the dip indicates the magnitude of the adjustment costs, and therefore the level of vulnerability faced by the operation.

Chapter 9

Human resources aspects are especially important in the operations function, where most 'human resources' are to be found. People in organization: contribute to human resource strategy; understand organization design; design the working environment; design individuals' and groups' jobs; allocate work times.

'Organizational' culture can apply to a single function. Overcoming cultural differences between different functions -> cultural fragmentation, own subcultures. 3 elements of operations culture:

- Believe – what the people within the operations function accept as self-evident
- Know – understand underlying principles that govern how operations and processes work, only with thorough understand contribute fully to success
- Behave – not fundamentally different from any effective manager

Human resources strategy: overall long-term approach to ensuring that an organization's human resources provide a strategic advantage. It involves 2 interrelated activities:
1. Identifying the number and type of people that are needed to manage, run and develop the organization so that it meets its strategic business objectives.
2. Putting in place the programmes and initiatives that attract, develop and retain appropriate staff.

Dave Ulrich: traditional HR departments often inadequate at fulfilling meaningful strategic role. 4 elements to HR activity:

1. Being a 'strategic partner' to the business
2. Administering HR procedures and processes
3. Being an 'employee champion'
4. Being a 'change agent'

People issues are inter-reliant -> strategic perspective aimed at identifying the relationship between all 4 roles is necessary. 1^{st} step in developing HR strategy to understand overall strategy.

Stress can undermine quality of people's working lives -> effectiveness. Stress: adverse reaction people have to excessive pressures or other types of demand placed on them. Business-related benefits:

- Staff feeling happier, perform better
- Introducing improvements easier when stress is managed effectively
- Employment relations: problems resolved more quickly
- Attendance levels increase and sickness reduces

Organization structure: the way in which tasks and responsibilities are divided into distinct groupings, and how responsibility and co-ordination relationships between the groupings are defined.
There are many valid approaches to describing organizations. The process perspective is a particularly valuable one.
Organizations are machines - resources within organizations can be seen as 'components' in a mechanism whose purpose is clearly understood. Where it is important to impose clarity such a metaphor can be useful, and is the basis of the 'process approach'.
Organizations are organisms: Organizations are living entities. Their behaviour is dictated by the behaviour of the individual humans within them. Useful if parts of the environment change radically. The survival of the organization depends on its ability to exhibit enough flexibility to respond to its environment.

Organizations are brains: like brains, organizations process information and make decisions they balance conflicting criteria, weigh up risks and decide when an outcome is acceptable.
Organizations are cultures: its shared values, ideology, pattern of thinking and day-to-day ritual. A major strength of seeing organizations as cultures is that it draws attention to their shared 'enactment of reality.
Organizations are political systems: Organization, like communities, are governed. Individuals and groups seek to pursue their aims through the detailed politics of the organization. Such a view is useful in helping organizations to legitimize politics as an inevitable aspect of organizational life.

Most organizational designs attempt to divide an organization into discrete parts that are given some degree of authority to make decisions; it allows specialization. 3 approaches to grouping:

1. According to their functional purpose
2. By the characteristics of the resources themselves (similar technologies, similar skills, resources for particular products)
3. By markets which the resources are intended to serve (location, type of customers)

Some pure types of organization:

- **U-form organization** - Clusters resources primarily by functional purpose. Pyramid management structure. Can emphasize process efficiency above customer service and ability to adapt to changing markets. Keeps together expertise an can promote creation and sharing of technological knowledge. Problem: flexibility of deployment
- **M-form organization** – groups either resources needed for each product/service group or those needed to serve a particular geographical market. Separate functions may be distributed throughout different divisions can reduce economies of scale and operating efficiency. Allows each individual division to focus on specific needs of market
- **Matrix forms** – hybrid, combining U and M-form. 2 different structures. Each resource cluster has at least 2 lines of authority. Ensures representation of all interests within the company, can be complex and confusing.
- **N-form organization** – n = network. Resources are clustered into groups as in other organizational forms but with more delegation of responsibility for strategic management of those resources. Little hierarchical reporting. Relative strength of relationships between clusters.

4 types of operations developer role

- **Top down or bottom up** – Top down: programmatic approach, emphasizing implementation of overall strategy. Bottom up: emergent model where individual business operations together contribute to overall building of operations expertise.
- **Market requirements or operations resource focus** – Market requirements: focus on explicit performance achieved by each part of operations function. Operations resource focus: emphasizes way in which each part of operation function develops its competences and successfully deploys them in marketplaces.

How they could work:

- Operations developers as **governors** – sets clear goals for each part, judges performance and wants to know reason why
- Operations developers as **curators** – collecting performance information, disseminating information, adopt best practice elsewhere
- Operations developers as **trainers** – develop clear objectives, devise improvement methodologies that can be customized
- Operations developers as **facilitators** – advise, support and aid development through process of mentoring, relatively long-term approach

Job design involves number of elements:
- What tasks are to be allocated to each person in the operation?
- What is the best method of performing each job?
- How long will it take and how many people will be needed?
- How do we maintain commitment?
- What technology is available and how will it be used?
- What are the environmental conditions of the workplace?

Division of labour = dividing the total task down into smaller parts, each of which is accomplished by a single person or team. Some real advantages in division-of-labour principles:
1. **Promotes faster learning** - easier to learn how to do a relatively short and simple task than a long and complex one.
2. **Automation becomes easier** - dividing a total task into small parts raises the possibility of automating some of those small tasks.
3. **Reduced non-productive work** - In large, complex tasks the proportion of time spent picking up tools and materials, putting them down again and generally finding, positioning and searching can be very high indeed.

There are also serious drawbacks to highly divided jobs:
1. **Monotony**: The shorter the task, the more often operators will need to repeat it.
2. **Physical injury**: The continued repetition of a very narrow range of movements can, in extreme cases, lead to physical injury. The over-use of some parts of the body can result in pain and a reduction in physical capability = repetitive strain injury (RSI).
3. **Low flexibility**: Dividing a task up into many small parts often gives the job design a rigidity which is difficult to alter under changing circumstances.
4. **Poor robustness**: Highly divided jobs imply materials passing between several stages. If one of these stages is not working correctly, the whole operation is affected. On the other hand, if each person is performing the whole of the job, any problems will only affect that one person's output.

<u>Designing job methods – scientific management</u>

Frederick Taylor identified what he saw as the basic tenets of scientific management:
- All aspects of work should be investigated on a scientific basis to establish the laws, rules and formulae governing the best methods of working.
- Such an investigative approach to the study of work is necessary to establish what constitutes a 'fair day's work'.
- Workers should be selected, trained and developed methodically to perform their tasks.
- Managers should act as the planners of the work, while workers should be responsible for carrying out the jobs to the standards laid down.
- Co-operation should be achieved between management and workers based on the 'maximum prosperity' of both.

Scientific management -> 2 fields of study: method study, which determines the methods and activities to be included in jobs; and work measurement, which is concerned with measuring the time that should be taken for performing jobs. Together, these 2 fields are often referred to as work study.

Physiology -> way the body functions and involves two aspects: how a person interfaces with his/her immediate working sea and how people react to environmental conditions. **Ergonomics** sometimes referred to as human factors engineering or just 'human factors'. Both aspects are linked by 2 common ideas:

- There must be a fit between people and the jobs they do. To achieve this fit there are only 2 alternatives. Either the job can be made to fit the people who are doing it, or the people can be made to fit the job. Ergonomics addresses the former alternative.
- It is important to take a 'scientific' approach to job design.

Many ergonomic improvements are primarily concerned with the **anthropometric aspects of jobs**: aspects related to people's size, shape and other physical abilities. Anthropometric data: data which ergonomists use when doing this. Because we all vary in our size and capabilities, ergonomists are particularly interested in our range of capabilities, which is why anthropometric data is usually expressed in percentile terms.

Job design should also take into account the desire of individuals to fulfil their needs for self-esteem and personal development -> motivation theory and contribution to behavioral approach important. Achieves 2 important objectives in job design: 1. Provides jobs which have an intrinsically higher quality of working life and ethically desirable end in itself. 2. More motivation -> better performance, quality and quantity. This involves 2 conceptual steps: 1. Exploring how various characteristics of job affect people's motivation; 2. Exploring how individual motivation towards the job affects performance.

Job rotation: If increasing the number of related tasks in the job is constrained in some way, e.g. by technology, one approach may be to encourage job rotation = moving individuals periodically between different sets of tasks to provide some variety in their activities. When successful, job rotation can increase skill flexibility and make a small contribution to reducing monotony.
Job enlargement = allocation a larger number of tasks to individuals which are as the same type as those in the original job. It may provide a more complete and therefore slightly more meaningful job. People performing an enlarged job will not repeat themselves as often -> marginally less monotonous. Operators repeat themselves less frequently and presumably the variety of tasks is greater, although no further responsibility or autonomy is necessarily given to each operator.
Job enrichment = increasing the number of tasks and allocating extra tasks which involve more decision making, greater autonomy and greater control over the job. The effect is both to reduce repetition in the job and to increase autonomy and personal development. **Horizontal changes** = those which extend the variety of similar tasks assigned to a particular job; **Vertical changes** = those which add responsibilities, decision making or autonomy to the job.
Job enlargement -> movement only in the horizontal scale, job enrichment certainly implies movement on the vertical scale and perhaps on both scales.
Empowerment = extension of the autonomy job characteristic prominent in the behavioural approach to job design. It is usually taken to mean more than **autonomy** = giving staff the ability to change how they do their jobs. The benefits: providing fast responses to customer needs, employees who feel better about their jobs and who will interact with customers with more enthusiasm, promoting 'word-of-mouth' advertising and customer retention. There are costs associated, including higher selection and trainings costs, perceived inequity of service and the possibility of poor decisions begin made by employees.
Teamworking (closely linked to empowerment) = where staff, often with overlapping skills, collectively perform a defined task and have a high degree of discretion over how they actually perform their task. The concept of teamwork, is more prescriptive and assumes a shared set of objectives and responsibilities. The benefits of teamwork can be summarized as:

- Improving productivity through enhanced motivation and flexibility
- Improving quality and encouraging innovation
- Increasing satisfaction by allowing individuals to contribute more effectively

- Making easier to implement technological changes in workplace because teams are willing to share challenges this brings

From an operations management perspective, 3 aspects of flexible working are significant:
Skills flexibility: flexible workforce that can move across several different jobs could be deployed in whatever activity is in demand at the time. Greater emphasis must be placed on training, learning and knowledge management. Defining what knowledge and experience are required to perform particular tasks and translating into training activities are clearly prerequisites for effective multi-skilling.
Time flexibility: Many people, only want to work for part of their time, sometimes during specific parts of the day/week. Bringing both the supply of staff and the demand for their work together is the objective of 'flexible time' or flexi-time working systems.
Location flexibility: The number of jobs which are not location-specific has increased. Location-specific: a job must take place in one fixed location. Many jobs could be performed at any location where there are communication links to the rest of the organization.

<u>Designing the working environment – ergonomics</u>
One aspect of ergonomics is concerned with how a person interfaces with the physical aspects of his or her immediate working area. The immediate environment in which jobs take place will influence the way they are performed. Working conditions:
<u>Working temperature</u>: individuals vary in the way their performance and comfort vary with temperature. Some points regarding working temperatures provide guidance to job designers:
- Comfortable temperature range will depend on type of work
- Effectiveness at performing vigilance tasks reduces at temperatures too warm
- Chances of accidents occurring when temperature too high or too low

<u>Illumination levels</u>: intensity of lighting required to perform any job satisfactorily will depend on the nature of the job.
<u>Noise levels</u>: Noise-induced hearing loss is a well-documented consequence of working environments where noise is not kept below safe limits. In addition to the damaging effects of high levels of noise, intermittent and high-frequency noise can also affect work performance at far lower levels.
<u>Ergonomics in the office</u>: As the number of people working in offices has increased, ergonomic principles have been applied increasingly to this type of work. At the same time, legislation has been moving to cover office technology such as computer screens and key-boards.

At best, any 'measurement' of how long a task will, or should, take will be an estimate. Why this process of estimating work times: '**work time allocation**' -> allocating a time for completing a task because we need to do so for many important operations management decisions. The advantage of structured and systematic work measurement -> gives common currency for the evaluation and comparison of all types of work.
Work measurement: process of establishing the time for a qualified worker (=one who is accepted as having the necessary physical attributes, intelligence, skill, education and knowledge to perform the task to satisfactory standards of safety quality and quantity), at a defined level of performance, to carry out a specified job (=one for which specifications have been established to define most aspects of the job).
<u>Work measurement techniques</u>:
1. **Synthesis from elemental data**: totalling element times obtained previously from studies in other jobs containing the elements concerned or from synthetic data.
2. **Predetermined motion-time systems** (PMTS): technique whereby times established for basic human motions are used to build up the time for a job at a defined level of performance.

3. **Analytical estimating**: technique which is a development of estimating, estimated from knowledge and experience of the elements concerned.
4. **Activity sampling**: a technique in which a large number of instantaneous observations are made over a period of time of a group of machines, processes or workers.

Summary 5: Operations Strategy for premaster IEM test two

Chapter 10

Planning and control: activities that attempt to reconcile the demands of the market and ability of the operation's resources to deliver. It provides the systems, procedures and decisions which bring different aspects of supply and demand together. Customers' perceptions of an operation will partially be shaped by the customer interface of its planning and control system. **Planning** -> formalization of what is intended to happen at some time in the future, no guarantee, more statement of intention. **Control** -> process of coping types of change, e.g. plans redrawn, make adjustments which allow the operation to achieve the objectives that the plans has set, even when assumptions on which the plan was based do not hold true. Planning and control are separate but closely related.

Very long term -> plans concerning what they intend to do, resources needed, objectives hope to achieve, emphasis on planning, demand described in aggregated terms, mainly focus on volume and financial targets. Medium-term planning -> more detailed, overall demand in partially disaggregated manner. Short-term planning -> difficult to make large changes, totally disaggregated.

Low-volume/high-variety: slow customer responsiveness, short planning horizon, major planning decision: timing, detailed control decisions, high robustness.

High-volume/low-variety: fast customer responsiveness, long planning horizon, major planning decision: volume, aggregated control decisions, low robustness.

Plan and control depend on both nature of demand and supply:
Uncertainty in supply and demand – makes planning and control more difficult. Systems should be able to cope with uncertainty in demand.

Dependent and independent demand – **Dependent demand**: predict demand with relative certainty because demand dependent upon some other factor which is known, e.g. car tires in an automobile factory, custom-made dressmaker, high-class restaurant. **Independent demand**: do not have firm 'forward visibility' of customer orders. E.g. tire fitter, governed by type of car arriving, fluctuations.

Responding to demand – 'Design-resource, create and deliver to order': e.g. advertising agency. 'Design, create and deliver to order': e.g. website designer. 'Create and deliver to order': e.g. house builder who has standard designs, or 'make to order'. 'Partially create and deliver to order': e.g. 'assemble to order' computers. 'Create to stock': e.g. preserved food production. 'Collect/choose from stock': e.g. IKEA. See the connecting for these types of operations and volume-variety. Planning and control activity will vary depending on how much work is done before demand is known.

P:D ratios – D: demand time: total length of time customers have to wait between asking for service/product and receiving it. P: total throughput time: how long operation takes to design the service/product, obtain resources, create and deliver it. The larger the P:D ratio, the more speculative the operation's planning and control activities will be.

Planning and control include 4 overlapping activities:

- **Loading**

= amount of work allocated to a work center. Finite loading: approach which only allocates work to a work center up to a set limit, estimate of capacity for the work center. Especially relevant for operations where:

a. It is possible to limit the load (appointment system for hairdresser, general medical practice)
b. It is necessary to limit the load (safety reasons -> amount of people/luggage in aircraft)
c. The cost of limiting the load is not prohibitive (maintain a finite order book at specialist sports car manufacturer does not adversely affect demand)

Infinite loading: approach loading where no limit, tries to cope with it. Relevant for operations where:

a. It is not possible to limit the load (emergency department hospital)
b. It is not necessary to limit the load (fast food outlets designed to flex capacity)
c. The cost of limiting the load is prohibitive (retail bank turned away customers because of a set number of customers -> customers unhappy)

- **Sequencing**

= decisions taken on the order in which the work will be tackled.

The **physical nature** of the inputs being processed may determine the priority of work e.g. operation using paints will use lighter shades before darker ones. Sometimes the mix of work arriving may determine the priority, e.g. jobs that physically fit together may be scheduled together to reduce waste. Sometimes there is **customer priority** sequencing e.g. in emergency room of hospital.
DD: prioritizing by due date: work sequenced according to when it is 'due' for delivery, irrespective of the size/importance, e.g. printing unit when copies are needed. Improve: delivery dependability and average delivery speed, may not provide optimal productivity, can be flexible.
LIFO: last in, first out, e.g. unloading an elevator, more practical.
FIFO: first in, first out, e.g. queues in theme parks.
LOT: longest operation time. Advantage of occupying work centers for long periods, relatively small jobs progressing through the operation will take up time at each work center. This only takes high utilization into account.
SOT: shortest operation time first. When operation become cash constrained. Improving delivery performance, however may adversely affect total productivity and can damage service to larger customers.
Performance objectives often used: meeting 'due date' (dependability), minimizing flow time (speed), minimizing work-in-progress inventory (element of cost), minimizing idle time (element of cost)

- **Scheduling**

= detailed timetable showing at what time or date jobs should start and when they should end e.g. bus schedules. Rapid-response services cannot plan when customers arrive.
Scheduling one of most complex tasks. E.g. Number of possible schedules = $(n!)^m$, n: number of jobs, m: number of machines. So scheduling often not optimal solution but an acceptable one.
Forward scheduling: starting work as soon as it arrives. **Backward scheduling**: starting jobs at the last possible moment. Choice depends largely upon circumstances.
Gantt charts – advantages: provide simple visual representation both of what should be happening and of what is actually happening. Used to 'test out' alternatives. They are not an optimizing tool; merely facilitate by communicating alternative schedules effectively.

Staff rostering: where dominant resource is staff, schedule of work times effectively determines the capacity of the operation itself, main task sufficient number of people working at point in time. E.g. retail shops schedule with demand in mind -> high visibility, have to respond directly to customer demand. Very complex because of vacation, fluctuations in demand, skill etc.

- **Monitoring and control**

= ensure that planned activities really happen. Deviations from plan -> re-planning.

Push system of control -> activities scheduled by means of a central system and completed in line with central instructions. Each center pushes work without considering whether the next can make use of it. Often characterized by idle time, inventory and queues. **Pull system of control** -> the place and specification of what is done are set by the customer workstation, which pulls work from previous. Demand is transmitted back through the stages from the original point of demand by original customer. Pull less likely for inventory build-up -> favored by lean operations.

Drum, buffer, rope concept from theory of constraints and optimized production technology concept -> helps to decide exactly where in a process control should occur. Most do not have same amount of work loaded -> likely to be a part of the process which acting as bottleneck -> should be control point: *drum*. Bottleneck working all the time (not enough capacity) -> sensible to keep *buffer* in front. Communication between bottleneck and input -> input not overproduce: *rope*.

Simple monitoring control model often simplification. Critical questions:

Is there consensus over what the operations objectives should be? – because individual managers have different and conflicting interests. Objectives ambiguous -> strategy has to cope with unpredictable changes in environment.
Are the effects of interventions into the operation predictable? - Relationships between intervention and resulting consequence within process are predictable, which assumes that degree of process knowledge is high.
Are the operation's activities largely repetitive? – non-repetitive -> little opportunity for learning.

These questions can form a decision tree. Any divergence from the conditions necessary for routine control implies a different type of control.

- **Expert control** -> If objectives unambiguous, effects well understood but activity not repetitive. Need to 'network' -> acquiring expertise and integrating expertise into organization.
- **Trial-and-error control** -> if strategic object relatively unambiguous, effects intervention unknown, activity repetitive -> learn through own failures.
- **Intuitive control** -> if objectives relatively unambiguous, effects and control interventions unknown, neither repetitive, learning by trial-error not possible -> intuition.
- **Negotiated control** -> objectives ambiguous -> negotiation to make unambiguous objectives. To some extent dependent on power structures.

Chapter 14

Planning and control systems: information-processing, decision support and execution mechanisms that support operations planning and control activity. Common elements:

- Customer (demand) interface that forms two-way information link between the operation's activities and its customers;

Set of activities that interface with both individual customers and market more broadly. Defines nature of the customer experience. Quality of the service defined by the gap between customers' expectations and their perceptions of they service they receive. Managing of customer expectations is particularly important in the early stages of the experience. It should reflect the operation's objectives. It acts as a trigger function, what is triggered will depend on nature of business.

- Supply interface does same for suppliers;

Provides link between activities of operation and its suppliers. Has long and short term function. Important to manage supplier expectations.

- Set of overlapping 'core' mechanisms that perform basis tasks;

Recognizes difficulties and tries to bring order to complexity by dividing up many interrelated planning and control decisions into sub-problems to reflect the organizational hierarchy, separate different kinds of decisions at different levels in the organization and over different time periods. Gives certain amount of independence to the planners at different levels. Effectiveness -> depend largely on how effective and consistent boundaries between levels of hierarchy are managed.

- Decisions mechanism involving operations staff and information systems that makes or confirms planning and control decisions.

Does systems integrate human with 'automated' decision making? Computer can cope with immense complexity. Humans 'soft' qualitative tasks:

a. Flexibility, adaptability and learning -> deal with ambiguous, incomplete, unstable goals and constraints
b. Communication and negotiation
c. Intuition

This should provide a clue as to what should be appropriate degree of automation built into decision making.

Enterprise resource planning (**ERP**) - series of interrelated decisions about volume (quantity) and timing of materials needed, basis for ERP, called: **materials requirements planning** (**MRP**). ERP systems automates and integrates core business processes. Helps company to 'forward-plan' and understand all implications of any changes to the plan.
MRP popular during 1970's because availability of computer power to drive basic planning and control mathematics. MRP uses bill of materials (**BOM**), with demand information -> form master production schedule (**MPS**). **MRP II** during 1980's -> higher degree of processing power and communication, modelling 'what-if' scenarios. Strength of both: explore consequences of any changes.
ERP -> consequences of decisions in one part reflected in planning and control systems of rest of the organization. ERP fully exploited when a web-integrated ERP widely implemented (allow different ERP systems to communicate).
Benefits of ERP:

- Absolute visibility of what is happening in all parts of the business

- Discipline of forcing business-process-based changes effective mechanism for making all parts of the business more efficient
- Better ‘sense of control’ of operations that will form the basis for continuous improvement
- Enables for more sophisticated communication with customers, suppliers and other business partners -> more accurate and timely information
- Capable of integrating whole supply chains including suppliers’ suppliers and customers’ customers

ERP systems are only fully effective if the way a business organizes its processes is aligned with the underlying assumptions of its ERP system. Usually companies invest in ERP because the prospect of such organizational efficiency is attractive, or because all competitors do so. Effectiveness depends partly on suppliers’ and customers’ ERP systems.

High failure rate for IT projects -> often managerial, implementation or organizational factors -> one of most important: degree of alignment and integration between firm’s overall IT strategy and general strategy of the firm. Impact of some IT is limited: **function IT** -> function IT can be adopted with or without any changes to other organizational structures. **Enterprise IT**: extends across much of or even the entire organization, most common: ERP. **Network IT**: facilitates exchanges between people and groups inside/outside the organization, challenge: learn how to exploit its emergent potential.
Key issues in ERP implementation -> what **critical success factors** (CSFs) should be managed. CSFs: get right in order for ERP system to work effectively. Likely problems in ERP implementation:

- Total cost of underestimated
- Time and effort often underestimated
- Resourcing from business and IT function often higher than expected
- Level of outside expertise required will be more than expected
- Changes required to business processes greater than expected
- Controlling the scope of the project more difficult than expected
- Will never be enough training
- Need for change management often not recognized until it is too late, changes required to corporate culture often underestimated (single biggest failure point)

Supplement to chapter 14

Materials requirements planning (MRP): approach to calculating how many parts/materials are required and what times. This requires data files that can be checked and updated. Calculations based on customer orders and forecast demand.
MPS main input to MRP, contains statement of volume and timing of end products to be made. It is the basis for the planning and utilization of labor and equipment, it determines the provisioning of materials and cash. Should include all sources of demand.
MPS contain statement of demand and currently available stock -> available inventory is projected ahead in time. MPS is ‘**chasing demand**’: MPS increasing as demand increases, aims to keep inventory zero. **Levelled MPS**: averaging the amount required, smooth out peaks, more inventory.
Sales function can load known sales orders against the MPS and keep track of what is available to promise (ATP), ATP line shows maximum that is still available in any one week, against which sales can be loaded.

Bill of materials (BOM): information of what parts are required for each product. Different levels of assembly are shown, finished: level 0. Intended bill of materials: indentation of level of assembly, know required number of each part.
Calculate 'net' requirements (extra requirements so demand is met) -> requires: item master file (contains unique standard identification code), transaction file (record of receipts into stock, from stock and running balance) and location file (identifies where inventory is located).
Most important element of MRP netting process: systematic process of taking planning information and calculating the volume and timing requirements which will satisfy demand. MRP check how many of required parts already available in stock, then generates 'works orders' for net requirements, through BOM next level down etc.
MRP also considers when each of these parts is required, by process: back-scheduling, takes lead time into account.
3 planning routines to check production plans against operations resources:

- Resource requirements plans (RRPs): looking forward in long term to predict the requirements for large structural parts of the operation
- Rough-cut capacity plans (RCCPs): medium to short term, check the MPSs against known capacity bottlenecks, in case capacity constraints are broken, checks MPS and key resources only.
- Capacity requirements plans (CRPs) look at daily effect of works orders issued from MRP on the loading individual process stages.

Chapter 11

Capacity: maximum level of value-added activity over a period of time that the process/operation can achieve under normal circumstances. -> **scale of capacity**. **Processing capabilities**: more important, level of output. **Capacity management**: activity of understanding the nature of demand for products/services, and effectively planning and controlling capacity in the short, medium and long term.
Established **long-term capacity** -> decide on **medium-term**: assessment of demand forecasts (2-18 months), during which time planned output can be varied. **Short-term** -> demand varies and how to respond.
Setting capacity levels over medium/short term in **aggregated terms** -> making overall broad capacity decisions, not concerned with all of the detail of the individual products/services offered. Aggregated: different products/services bundled together in order to get a broad view of demand and capacity, may mean some degree of approximation. Ultimate aggregation measure is money.
Some parts of operation at their capacity 'ceiling' -> capacity constraint for whole operation.
Decision in devising their capacity plans will affect aspects of performance:

- Costs – balance between capacity and demand, under-utilization -> high unit cost
- Revenues – balance capacity and demand, capacity $\geq$ demand -> no revenue lost
- Working capital – affected if build up finished goods prior to demand, demand met but inventory costs
- Quality – affected by capacity plan that involves large fluctuations in capacity levels
- Speed – of response enhanced by build-up of inventories or deliberate provision of surplus capacity to avoid queuing
- Dependability – affected by how close demand levels are to capacity
- Flexibility – especially volume flexibility enhanced by surplus capacity, if demand and capacity balanced then not able to respond to unexpected increase in demand

1st step in managing capacity: measure the aggregate demand and capacity levels and understand changes in these levels for the planning period. 2nd step: determine the operation's base level of capacity from which adjustments up or down will be made -> largely determined by performance objectives, perishability of outputs and degree of variability in demand and supply. 3rd step: identify and select methods of coping with mismatches between demand and capacity. 4th: understand the consequence of different capacity decisions.

1. <u>How are demand and capacity measured?</u>

Understanding demand for products and services – Demand forecasting very important input into the capacity management decisions. 3 requirements from demand forecast, concerned with capacity management:

- Expressed in terms that are useful for capacity management, same units such as machine hours per year or space
- As accurate as possible. To meet demand, output must be decided from a forecast.
- Gives an indication of relative uncertainty. Decisions to operate extra hours e.g. usually based on forecast levels of demand, which could differ from relative.

Understanding changes in demand – Most markets are influenced by some kind of seasonality, e.g. climatic (holidays), festive (gifts). **Seasonality**: describe changes to demand over a period of a year. Can also be weekly/daily. Extent to which an operation will have to cope with very short-term fluctuations is partly determined by how long customers are prepared to wait. E.g. emergency services cannot let their customers wait long.

Better forecasting or better operations responsiveness? – Most organizations find a balance between having better forecasts and being able to cope without perfect forecasts. Capacity management requires combining attempts to increase market knowledge with attempts to increase operations flexibility.

Understanding capacity – Only one type of output, the **output capacity measure** is pretty simple. With a much wider range, output places varying demands on the process so output measures capacity are less useful -> **input capacity measures** frequently used. See table in book for examples.

- *Capacity depends on activity mix* - Capacity is a function of service/product mix, duration and product service specification, e.g. in a hospital it is very hard to predict the output. Some problems caused by variation mix can be partially overcome by using **aggregated capacity measures**.
- *Capacity depends on the duration over which output is required* – Capacity is the output that can be delivered in a defined unit of time. The level of activity and output that may be achievable over a short period is not the same as the capacity that is sustainable on a regular basis. 3 different measures of capacity:
    a. **Design capacity** – theoretical capacity is what technical designers had in mind.
    b. **Effective capacity** – the capacity after planned losses are accounted for.
    c. **Actual output** – capacity after both planned and unplanned losses are accounted for.

    Utilization = $\frac{Actual\ output}{Design\ capacity}$. Efficiency = $\frac{Actual\ output}{Effective\ capacity}$.
- *Capacity depends on the specification of output* – Some operation can increase their output by changing the specification of the product/service. Important to distinguish the most important tasks to increase capacity in short term.
- *Capacity 'leakage'* - = reduction in capacity due to quality problems, delays in delivery, machine breakdown etc. Popular method of assessing this leakage is the overall equipment effectiveness (OEE) measure, calculated: OEE = a x p x q. a: availability of a process, p: performance or speed of a process, q: quality of product/service that process creates. To process effectively all 3 need high levels of performance. OEE represents valuable operating time as a percentage of the capacity something was designed to have.

Understanding changes in capacity – capacity management decisions should reflects both predictable and unpredictable variations in capacity and demand.

2. <u>How should the operation's base capacity be set?</u>

The higher the **base level** of capacity, the less capacity fluctuations is needed to satisfy demand.
Setting demand – Base level of capacity should be related to 3 factors:

1. *Relative importance of the operation's performance objectives*

E.g. high base level -> high cost (lots of underutilization) but also ability to flex output. Trade-off between fixed and working capital -> low base level, low investment but increasing working capital to meet future demand.

2. *The perishability of the operation's outputs*

Demand or supply perishable -> base capacity relatively high level, because cannot be stored.

3. The degree of variability in demand or supply

Variability in either demand or capacity will reduce ability to process its inputs -> reduce effective capacity. -> long throughput times, queue build up so affects inventory levels. So when high level of variability -> relatively high base level of capacity to provide extra capacity.

3. <u>What are the ways of coping with mismatches between demand and capacity?</u>

The nature of capacity management depends on the mixture of predictable and unpredictable demand and capacity variation. 3 options for coping with variation, usually organizations use a combination.

- **Level capacity plan** – capacity is fixed regardless of fluctuations in forecast demand. Can achieve: stable employment patterns, high process utilization, high productivity, low unit costs. But also: create considerable inventory, decisions taken for what immediately sale and what inventory? For e.g. hotel and supermarket a high level of capacity would be necessary at this plan. Low utilization can be at e.g. expensive jewelry and real-estate agents. May let customers wait long.
- **Chase demand plan** – attempts to match capacity closely to varying levels of forecast demand, opposite and more hard than level plan. Usual by operations that not able to store their output such as perishable products or certain services. Where output can be stored, used to minimize inventory. Adjusting capacity by e.g. overtime, part-time staff, outsourcing, hiring, firing, subcontracting.
- **Demand management** – change pattern of demand to bring it closer to available capacity, some methods:

  a. *Constraining customer access* – customers only allowed to access products/services at certain times (reservation systems)
  b. *Price differentials* – adjusting price to reflect demand (vacation in high-season more expensive)
  c. *Scheduling promotion* – varying degree of market stimulation through promotion to encourage demand during low periods (promoting turkeys outside thanksgiving)
  d. *Service differentials* – allowing service levels to reflect demand, if explicitly then customers educated to move to periods of lower demand.

  More radical approach is to create alternative products/services to fill capacity in quiet periods. New products/services should meet criteria: (1) can be produced on same processes, (2) have different demand patterns of existing offerings, (3) they are sold through similar marketing channels. E.g. universities fill lecture rooms with conferences during vacation. Operation must be fully capable of serving both markets, weigh risks against benefits.

Yield management: collection of methods used to maximize operation's potential to generate profit. For operations which have relatively fixed capacities like airlines and hotels. E.g. from data airlines can balance risks of over-booking and under-booking. Hotels can give discounts outside holidays.
Mixed plans – most organizations use mix of approaches. E.g. peak demand been brought forward by offering discounts (manage demand). Capacity adjusted to 2 points in the year to reflect broad changes in demand (chase demand). Adjustment in capacity not sufficient to avoid totally build-up of inventories (level capacity).

4. <u>How can operations understand the consequences of their capacity decisions?</u>

Before operation can choose a capacity plan, it must understand consequences by adopting plans:

- *Consider capacity decisions using cumulative representations*

Assessing capacity plans to first plot demand on a **cumulative basis**. Cumulative representation of demand shows although peak demand in September, due to number of productive days the production starts in August. Also, fluctuation in demand is even greater than seemed. Demand per productive day is more relevant, because productive days represent the time element of capacity. Most useful: plotting demand and capacity on same graph, feasibility and consequences of capacity plan can be assessed. Vertical distance between cumulative demand and production shows surplus or shortage. Meet demand as it occurs -> cumulative production line always above demand.
Chase demand also illustrated on a cumulative representation. But production line have varying gradient, match the demand line. Inventory, carrying costs now zero/low, but now costs with changing capacity levels.

- *Considering capacity decisions using queuing principles*

Cumulative useful if able to story goods as inventory. Not possible -> waiting or **queueing theory**. Accepts that sometimes customers have to wait. Especially when arrival difficult to predict, time to make product/service uncertain, or both. Common set of elements that define queueing behavior:

a. **Source of customers** – 'population', can be finite or infinite. Finite -> probability of customers arriving depends on number of customers already being served.
b. **Arrival rate** – rate at which customers arrive. Some variability so describe in terms of probability distributions.
c. **The queue** – sometimes finite, sometimes infinite
d. **Rejecting** – if customers in system is already maximum -> new customers could be rejected
e. **Baulking** – customer refuses to wait, queue too long
f. **Reneging** – leaves queue before being served
g. **Queue discipline** – determine which order being served, e.g. fifo, lifo.
h. **Servers** – can be parallel, or in series. Human servers will vary in time it takes to serve, so processing time is usually described by a probability distribution.

Dilemma -> how many servers available? Because of variability in processing capability and arrival rate, there will always be queues and idle time. Trade-off between customers waiting and system utilization. Customer reactions to having to queue will be influenced by set of principles: unoccupied time feels longer than occupied,

a. Unoccupied time feels longer than occupied time, Pre-process waits feel longer than in-process waits
b. Anxiety makes wait seem longer, uncertain waits feel longer, unexpected waits feel longer, unfair waits feel longer, solo waiting feels longer, uncomfortable waits feel longer, new or infrequent users feel they wait longer
c. The more valuable the service, the longer customers will 'happily' wait

- *Considering capacity decisions over time*

Capacity management far more dynamic process, involves controlling and reacting to actual demand and actual capacity. At beginning of each period, same type of decisions must be made in the light of the new circumstances. Considerations of capacity strategy usually is the difference between long- and short-term outlook for volume and demand.

Chapter 15

Focus of **lean**: achieve a flow of materials, information or customers that delivers exactly what customers want, in exact quantities, exactly when needed, exactly where required, at the lowest cost.
3 perspectives of lean:

- Lean is a **philosophy** of how to run operations – the involvement of staff in the operation, the drive for continuous improvement, the elimination of waste.
- Lean is a **method of planning and controlling** operations – many ideas are concerned with how many items flow through the operation and how managers can manage this flow. Uses 'pull' in contrast to MRP which uses 'push'.
- Lean is a **set of tools that improve operations** performance – the 'engine' room of the lean philosophy is a collection of improvement tools and techniques that are the means for cutting out waste.

How lean operations consider…
…Flow – Traditional: buffer between stages, makes stages insulated -> paid in terms of inventoy, queues, slow throughput times. Also, when problem in a stage, it is up to that stage to fix it and will not directly be noticed by other stages. In lean: items directed passed to next stage, problem will be seen quickly by entire system so chances to fix it increase. 1. The downstream 'customer' stage signals the need for action, customer 'pulls' item through process. 2. Flow in synchronized matter instead of dwelling in inventory. 3. This affects the motivation to improve because stages are no longer decoupled. 4. This improves exposes waste and encourages its elimination.
…Inventory – Gradually reducing the inventory exposes the worst of the problems which can be resolved, and so on.
…Capacity utilization – in lean process, any stoppage will affect the whole process -> lower capacity utilization, short term. May sound bad but producing just to keep high utilization is pointless.
…The role of people – Lean encourages/requires team-based problem solving, job enrichment, job rotation and multi-skilling. 'Basic working practices' used to implement the 'involvement of everyone' principle:

- **Discipline** – critical for safety of staff, environment and quality, by everyone all the time
- **Flexibility** – should be possible to expand responsibilities to the extent of people's capabilities, barriers (grading systems e.g.) removed
- **Equality** – through uniforms, consistent pay structures, despite position
- **Autonomy** – Delegate responsibility to people involved in direct activities so that management's task becomes one of supporting processes.
- **Development of personnel** – create more members who can support rigors of being competitive
- **Quality of working life (QWL)** – e.g. enjoyment, working area facilities
- **Creativity** – indispensable element of motivation. Not just doing the job but improving how its done, building the improvement in the business
- **Total people involvement** – staff take on more responsibility to use their abilities to the benefit of the company as a whole.

…Improvement – get closer to ideals over time. **Continuous improvement** -> if its aims are set in terms of ideals which individual organizations may never fully achieve, then the emphasis must be on the way in which an organization moves closer to the ideal state.

Most significant: focus on elimination of all forms of waste. **Waste**: any activity that does not add value. Terms conveying 3 causes of waste that should be reduced or eliminated:

- **Muda** - = activities in a process that are wasteful, do not add value to operation/customer. Causes likely to be poorly communicated objectives, inefficient use of resources.
- **Mura** – 'lack of consistency' -> results inn periodic overloading of staff/equipment.
- **Muri** - =absurd/unreasonable. Unreasonable requirements put on a process will result in poor outcomes. Waste can be caused by failing to carry out basic operations planning tasks.

Types of waste:

- *Waste from irregular flow* – barriers that prevent streamlined flow include:
  - Waiting time
  - Transport
  - Process inefficiencies
  - Inventory
  - Wasted motion
- *Waste from inexact supply* – barriers to achieving an exact match between supply and demand include:
  - Over/under-production
  - Early/late delivery
  - Inventory
- *Waste from inflexible response* – symptoms of inadequate flexibility include:
  - Large batches
  - Delays between activities
  - More variation in activity mix than in customer demand
- *Waste from variability* – symptoms of poor variability include:
  - Poor reliability of equipment
  - Defective products/services

Looking for waste - Gemba ('the actual place') term used -> if you really want to understand something, you go to where it takes place. Lean uses 'the Gemba walk' to make problems visible.

<u>Eliminating waste through…</u>

<u>… streamlined flow </u>– Primarily, reconfiguring the layout the layout of a process to aid lean synchronization -> moving from functional- towards cell-based layouts, or cell-based- to line layouts. Towards one that brings more systematization and control to process flow.
Throughput time often taken as measure for waste in a process.
Value stream mapping -> simple but effective approach to understand the flow. It visually maps a product/service production path, records direct and indirect information systems. Distinguishes value-adding and non-value-adding activities. Differs from process mapping:

- Uses broader range of information
- Usually at a higher level (5-10 activities)
- Often has a wider scope, frequently spanning the whole supply chain
- Can often be used to identify where to focus future improvement activities.

Value stream perspective -> working on and improving the 'big picture', rather than optimizing individual processes. (1). Identifying the value stream to map. (2). Physically mapping a process, mapping the information flow that enables the process to occur: 'current state' map. (3). Problems are

diagnosed and changes suggested -> future state map which represents the improved version. Finally, changes are implemented.
Visual management is a lean technique designed to make the current and planned state transparent to everyone. Benefits; it can:

- Act as a common focus for team meetings
- Demonstrate methods for safe and effective working practices
- Communicate to everyone how performance is being judged
- Assess at a glance the current status of the operation
- Understand tasks and work priorities
- Judge your and others' performance
- Identify the flow of work, namely what has been and is being done
- Identify when something is not going to plan
- Show what agreed standards should be
- Provide real-time feedback on performance to everyone involved
- Reduce the reliance on formal meetings.

Important technique used to ensure flow visibility -> simple, highly visual signals to indicate a problem, e.g. an Andon light.
May be possibilities to smooth streamlined flow through use of small machines. Advantages: they can process different products/services simultaneously, system more robust, easily moved -> layout flexibility enhanced, risks of making errors in investment decisions are reduced. Utilization may be lower.

<u>… matching supply and demand exactly</u> – value of supply is always time dependent, something delivered early or late often has less value than something delivered exactly when it is needed. Exact match, often best served by using 'pull control' -> let the downstream stage in a process pull items through the system rather than have them 'pushed'.
Method of operationalizing pull control -> use of kanbans (card or signal). Kanban serves 3 purposes:

- It's an instruction for the preceding process to send more
- It's a visual control tool to show up areas of overproduction and lack of synchronization
- It's a tool for kaizen (continuous improvement), Toyota: 'number of kanbans should reduce over time.'

<u>… flexible processes</u> – flexible processes can significantly enhance smooth and synchronized flow. Responding exactly and directly to customer demand implies they need to be sufficiently flexible to change what and how much they do.
For many processes, increasing flexibility means reducing changeover time, methods:

- Measure and analyze changeover activities
- Separate external and internal activities – internal: cannot be carried out while process in going on, external: can -> identifying and separating, do as much as possible while process is continuing
- Convert internal to external activities – Methods: 1. Pre-prepare activities/equipment, 2. Make changeover process intrinsically flexible, 3. Speed up any required changes of equipment/info/staff.
- Practice changeover routines – practice makes perfect

<u>… minimizing variability</u> – One of biggest causes: variation in quality of items.

Levelling product/service schedules – keeping the mix and volume of flow between stages at an even rate over time. Disadvantages of large batches: large amounts of inventory within units and days very different in what they are expected to produce. Smaller and even batches -> reduce overall level of WIP, regularity of production (easier planning and control).

Levelling delivery schedules – E.g. dispatch smaller quantities of all products in a single truck more frequently -> inventory lower, system can respond quickly to trends

Adopting mixed modelling – principle of levelled scheduling and give a repeated mix of outputs. Ideally, sequence items as smoothly as possible, BACBACBAC -> smooth flow.

5S – Methodology: by eliminating what is unnecessary, and making it clear, work is easier and faster.

- Sort (seiri) – eliminate what is not needed
- Straighten (seiton) – position items such that they can be easily reached whenever needed
- Shine (seiso) – keep things clean and tidy
- Standardize (seiketsu) – maintain cleanliness and order – perpetual neatness
- Sustain (shitsuke) – develop a commitment and pride in keeping to standards

Adopting total productive maintenance (TPM) – TPM aims to eliminate the variability in processes caused by the effect of breakdowns.

Principles of lean are the same for a supply chain as they are for a process. One weakness of lean, especially in context of whole supply network, is when conditions are subject to unexpected disturbance. Less control, outside own operation. For entire network it is more difficult, takes longer to achieve, but just as valuable as for individual operation. How most supply chains traditionally operated -> daily snapshot of their ERP, limited visibility means operations must space out to avoid collisions, thus reducing output or they must fly blind and thus jeopardizing reliability.

<u>Lean and the theory of constraints</u> – **TOC** -> focus attention on capacity constraints/bottleneck of operation. This helps lean in obtaining a smooth flow. Always looking for constraints: optimized production technology (**OPT**). Principles OPT which demonstrate focus on bottlenecks:

- Balance flow, not capacity -> more important to reduce throughput time rather than achieving a notional capacity balance between stages or processes.
- The level of utilization of a non-bottleneck is determined by some other constraint in the system, not by its own capacity.
- Utilization and activation of a resource are not the same. According to TOC, resource is being utilized only if it contributes to the entire process creating more output. A process/stage can be activated in the sense that it is working, but it may only be creating stock or performing other non-value-added activity.
- An hour lost at a bottleneck is an hour lost forever out of the entire system.
- An hour saved at a non-bottleneck is a mirage.
- Bottlenecks govern both throughput and inventory in the system.
- You don’t have to transfer batches in the same quantities as you produce them. Flow will probably be improved by dividing large production batches into smaller ones for moving through a process.
- The size of the process batch should be variable, not fixed.
- Fluctuations in connected and sequence-dependent processes add to each other rather than averaging out.

- Schedules should be established by looking at all constraints simultaneously. Because of bottlenecks and constraints within complex systems, it is difficult to work out schedules according to a simple system of rules.

OPT uses 'drum, buffer, rope' to explain planning and control approach. Steps of TOC:

1. Identify the system constraint
2. Decide how to exploit the constraint – obtain as much capability as possible from the constraint, preferably without expensive changes.
3. Subordinate everything to the constraint – other elements adjusted to the level where the constraint can operate at maximum effectiveness
4. Elevate the constraint – eliminating the constraint, only is 2 and 3 not successful.
5. Start again from step 1

Lean and MRP - JIT scheduling aims to connect the new network of internal and external supply processes by means of invisible conveyors so that parts only move in response to coordinated and synchronized signals derived from end-customer demand. MRP seeks to meet demand by directing that items are only produced as needed. While MRP is excellent at planning, it is weak at control. While lean synchronization may be good at control, it is week on planning. For relatively simple structures and routines, lean and pull are good. Things get more complex -> opportunities for using pull scheduling decrease. Very complex structures require network planning methods (Ch. 19).

Chapter 16

Various reasons to explain the shift towards a focus on improvement:

- There is a perception of increased competitive pressure
- The nature of world trade is changing –> new countries emerging as producers and consumers
- New technology has both introduced opportunities to improve and disrupted existing markets
- Interested resulted in development of many new ideas and approaches to improve, the more ways to improve, the more operations will be improved
- The scope of operations managements has widened to all function of the enterprise

Firms that have improved their competitive position have improved their performance more than their competitors.

Radical/breakthrough change -> the main vehicle of improvement is major and dramatic change in the way the operation works. The impact is abrupt, often high investment of capital. High value on creative solutions and free-thinking.

Continuous/incremental change -> improving by many small incremental steps. Also known as kaizen. The rate of improvement is not important, the momentum is.

Exploitation -> activity of enhancing processes that already exist within a firm, focus on creating efficiencies rather than radically changing resources/processing, benefits tend to be relatively immediate, incremental and predictable.

Exploration -> searching for and recognizing new mindsets and way of doing things, taking risks, innovation, benefits principally long term but can be difficult to predict.

Organizational ambidexterity -> ability of a firm both to exploit and explore as it seeks to improve; improving existing resources/processes while also competing in new technologies/markets where innovation is required.

Four aspects of improvement:

- <u>Elements</u> of improvement approaches – fundamental ideas, 'building blocks' of improvement
- <u>Broad approaches</u> – underlying set of beliefs that form a coherent philosophy and shape how improvement should be accomplished
- Improvement <u>techniques</u> – formal methods that help to produce improvements
- <u>Management</u> – way the process of improvement is organized, resourced and controlled, otherwise not effective

Key elements

<u>Improvement cycles</u> – repeatedly questioning detailed working of process. 2 widely used: **PDCA**: Plan -> Do -> Check -> Act. **DMAIC**: Defining -> Measurement -> Analyzed -> Improving -> Controlled/Check. For both -> cycle starts again.

<u>Process perspective</u> – Advantages: if improvement is not reflected in the process of creating products/services, then it is not really improvement and all parts of the business manage processes.

<u>End-to-end processes</u> – identified customer needs are entirely fulfilled by an 'end-to-end' business process

<u>Evidence-based problem solving</u> – emphasis on scientific method, responding only to hard evidence, using statistical software to facilitate analysis

<u>Customer centricity</u> – see customer as most important part of organization. **Voice of the customer** (VOC) = capturing a customers' requirements, expectations, perceptions and preferences in some depth.

Systems and procedures – some type of system that supports the improvement effort may be needed, improvement system or quality system.
Reduce process variation – some aspect of process performance is measured periodically -> plotted on a timescale. Advantages: check that the performance of the process is acceptable (capable), to check if process performance is changing over time, to check on the extent of variation in process performance.
Synchronized flow – items in a process, operation, supply network flow smoothly and with even velocity
Emphasize education/training – education and training important in motivating all staff towards seeing improvement as a worthwhile activity
Perfection is the goal – current situation calibrated against ‘perfection target’ so visible how much more improvement is possible
Waste identification – central in order to eliminate waste
Include everybody – harnessing the skills and enthusiasm of every person and all parts of the organization
Develop internal customer-supplier relationship – ensure external customers are satisfied -> every part of the organization contributes to external customer satisfaction by satisfying its own internal customers

Broad approaches
Total quality management (TQM) – approach that puts quality at the heart of everything that is done. Stress of following elements:
- Meeting the needs and expectation of customers
- Improvement covers all parts of the organization (group based)
- Improvement includes every person in the organization
- Including all costs of quality
- Getting things ‘right first time’, -> designing in quality rather than inspecting it in
- Developing the systems and procedures which support improvement

Lean – key elements when used as an improvement approach:
- Customer-centricity
- Internal customer-supplier relationships
- Perfection is the goal
- Synchronized flow
- Reduce variation
- Include all people
- Waste elimination

Business process re-engineering (BPR) – all work that does not add value should be eliminated. Advocated radical change rather than incremental. Main principles:
- Rethink business processes in a cross-functional manner which organizes work around the natural flow of information
- Strive for dramatic improvements in performance by radically rethinking and redesigning the process
- Have those who use the output from a process perform the process
- Put decision points where the work is performed, do not separate those how do the work from those who control and manage the work

Controversy of BPR -> only looks at work activities rather tan at people who do the work.

Six Sigma – specification range of any part of a product/service should be 6 the standard deviation of the process. Emphasize drive towards a virtually zero defect objective. General Electric (GE) early

adopter. Now broad improvement concept rather than simple examination of process variation (however, that is still an important part).
Measures to assess performance used:

- A defect – failure to meet customer-required performance
- A defect unit or item
- A defect opportunity – number of different ways a unit/output can fail to meet customer requirements
- Proportion defective – percentage/fraction of units that have one or more defects
- Process yield – percentage/fraction of units defect-free
- Defect per unit (DPU) – average number of defects on a unit of output
- Defects per opportunity – proportion/percentage of defects / total number of defect opportunities
- Defects per million opportunities (DPMO)
- The Sigma measurement – number of standard deviations of the process variability that will fit within the customer specification limits

Following elements often associated with Six Sigma:

- Customer-driven objectives – Six Sigma sometimes 'process of comparing process outputs against customer requirements', in particular it expresses performance in DPMO
- Use of evidence – lots of quantitative evidence
- Structured improvement cycle – in Six Sigma de DMAIC cycle
- Process capability and control
- Process design – latterly Six Sigma proponents also include process design in the collection of elements
- Structured training and organization of improvement – improvement initiatives only successful if significant resources and training are devoted to their management

Master Black Belt -> experts in use of Six Sigma tools and techniques, how such techniques can be used and implemented, seen as teachers to guide improvement projects. Need quantitative analytical skills and organizational and interpersonal skills.
Black Belt -> can take a direct hand in organizing improvement teams. Develop their quantitative analytical skills and act as coaches for green belt.
Green Belt -> work within improvement teams, possibly as team leaders, not in full-time positions, spend about 20% on improvement projects.

Criticisms of Six Sigma:

- Does not offer anything that was not available before, only new is gimmicky martial arts analogy of Black belts etc.
- Too hierarchical in the way it structures its various levels of involvement in improvement activity
- Expensive
- DPMO too onerous

There is a significant overlap between the various approaches to improvement in terms of the improvement elements they contain.
Lean Sigma -> waste reduction, fast throughput time and impact of lean with data-driven, rigor and variation control of Six Sigma.

Techniques
Improvement techniques: ‘step-by-step’ methods and tools that can be used to help find improved ways of doing things;
Scatter diagrams – provide quick and simple method of identifying whether there is evidence of a connection between 2 sets of data. Only evidence of a relationship, not necessarily cause-effect relationship.
Process maps (**flow charts**) – used to give a detailed understanding prior to improvement, quickly shows poorly organized flows. Clarify improvement opportunities and shed light on workings of an operation. Most importantly, highlight problem areas where no procedure exists to cope with a particular set of circumstances
Cause-effect diagrams (**Ishikawa diagrams**) – effective of helping to search for the root causes of problems. Used to identify areas where further data is needed. They provide a way of structuring group brainstorming sessions, often involves identifying possible causes.
Pareto diagrams – distinguish between the ‘vital few’ issues and the ‘trivial many’. Arranging items in order of importance. Based on the fact that relatively few causes explaining majority of effects.
Why-why analysis – stating the problem -> why occurred? -> reasons -> why have reasons occurred?

How to be managed?
Should be clear why improvement is happening and what is consists of -> linking improvement to overall strategy. Generally held that ability to improve its operations performance depends to a large extent on its ‘culture’. Some argue that (legally) ‘copying’ from outsiders can be effective. 3 strategic types of imitators:
- *Pioneer importer*: imitator that is the pioneer in another place
- *Fast second*: rapid mover arriving quickly after an innovator but before other rival imitators take large share of the market
- *The come from behind*: late entrant

Benchmarking is the process of learning from others. Based on ideas that problems are almost certainly shard by processes elsewhere and there is probably another operation somewhere that has developed a better way of doing things. Different types of benchmarking:
- Internal benchmarking – comparison between (parts of) operations within same organization
- External benchmarking – comparison between operations from you and another organization
- Non-competitive benchmarking – against external organizations which don’t compete directly in the same market
- Competitive benchmarking – comparison directly between competitors in the same/similar markets
- Performance benchmarking – comparison between levels of achieved performance in different operations
- Practice benchmarking – comparison between organization’s operations practices and those adopted by another operation

Benchmarking is best practiced as a continuous process of comparison. It does not provide solution but ideas/information. It is not simply copying, more like learning and adapting. It needs some investment but can be advantages in organizing staff at all levels.

Critics of benchmarking -> always limiting themselves to currently accepted methods. Methods appropriate in one operation may not be in another.

There can be no intentional improvement without learning.
Single-loop learning -> when there is a repetitive and predictable link between cause and effect. Error is corrected without questioning/altering the underlying values and objectives of the process.
Double-loop learning -> questions fundamental objectives, ability to challenge existing operating assumptions, remain open to any changes in the competitive environment.

Some tangible causes for implementation failure:
- *Top-management support*

They must understand and believe the benefits, communicate principles and techniques, participate and formulate and maintain a clear 'improvement strategy'. Strategy is necessary to provide goals and guidelines which help to keep effort in line.
- *Senior managers may not fully understand the improvement approach*
- *Avoid excessive 'hype'*

Most new ideas have something to say, but jumping from one fad to another will generate backlash against any new idea, and destroy the ability to accumulate the experience that comes from experimenting with each one.

Chapter 17

Quality -> consistent conformance of customers' expectation. Quality is multi-faceted; its individual elements differ for different operations. Customer's view is what he/she perceives the service/product to be. So quality: degree of fit between customers' expectations and customer perception of the service/product. Customers' expectations and perceptions are influenced by a number of factor, some can be controlled by operation and some can't. Perceived quality is governed by the magnitude and direction of the gap between customers' expectations and their perceptions. If perceived quality gap is such that it does not match expectations, then reasons must lie in other gaps elsewhere:
1. Gap 1: The customer's specification-operation's specification gap

Perceived quality could be poor because of mismatch internal quality and expected quality. Main organizational responsibility: marketing, operations, product/service development.
2. Gap 2: The concept-specification gap

Perceived quality may be poor because of mismatch between service/product concept and way the organization has specified quality internally. Main responsibility: marketing, operations, product/service development
3. Gap 3: The quality specification-actual quality gap

Perceived quality could be poor because of mismatch between actual quality and internal quality specification. Main responsibility: operations
4. Gap 4: The actual quality-communicated image gap

Perceived quality could be poor because there is a gap between the organization's external communication/market image and the actual quality delivered to the customer. Main responsibility: marketing.

The **sandcone theory** -> there is a generic 'best' sequence of improvement. Building up improvement is a cumulative process. *Quality* should be first priority. Next -> internal *dependability* -> next *speed* of internal throughput, only while continuing to improve quality and dependability further etc. Next is *flexibility* -> and last is *cost*.

Achieving conformance to specification requires the following steps:
<u>Step 1</u> – <u>Define the quality characteristics</u>
Generally useful: functionality, appearance, reliability, durability, recovery and contact
<u>Step 2</u> – <u>Decide how to measure each quality characteristic</u>
Lots of characteristics are hard to measure, still try to find a way to break it down and attempt to measure the customer perceptions. Measures used: variables -> can be measured on a continuously variable scale, and attributes -> assessed by judgement and are dichotomous (e.g. right or wrong).
<u>Step 3</u> – <u>Set quality standards for each quality characteristic</u>
Need quality standard against which it can be checked, otherwise they cannot indicate the performance. Quality standard: level which defines boundary between acceptable and unacceptable.
<u>Step 4</u> – <u>Control quality against those standards</u>
Conform to those standards? -> 3 checks.

(1) *Where in the operation should it check?*

(2) *Should it check every service/product or take a sample?*

Advantages of samples; inspecting everything might be dangerous, checking everything might destroy the product/service, checking everything can be time consuming and costly. Also 100% checking may not guarantee that all defects will be identified. **Type I error**: when a decision was made to do something and the situation did not warrant it. **Type II error**: nothing was done, yet a decision to do something should have been taken as the situation did indeed warrant it.

(3) *How should the checks be performed?*

Most common approach as for checking quality of a sample so as to make inferences about all the output from an operation: statistical process control (SPC). SPC looks at variability in performance to check.
<u>Step 5</u> – <u>Find and correct causes of poor quality</u>
<u>Step 6</u> – <u>Continue to make improvements</u>
There is an aspect of quality management that has been particularly important -> **total quality management** (**TQM**).
TQM -> effective system for integrating the quality development, quality maintenance and quality improvement efforts of the various groups in an organization so as to enable production/service at the most economical levels which allow for full customer satisfaction. It is a philosophy of how to approach quality improvement that stresses the 'total' of TQM, and puts quality over everything. Stress the following:

- <u>Meeting the needs and expectations of customers</u>

TQM stresses the importance of starting with an insight into customers needs -> translated into quality objectives and used to drive quality improvement

- Covering all parts of the organization

Powerful concept -> everyone is an internal customer and supplier. Some degree of formality to that concept with service-level agreements (SLAs) -> formal definitions of the dimensions of service and the relationship between two parts of an organization. Type of issues e.g.: response times, range of services, dependability of service supply.
Criticisms of SLAs: 1. 'pseudocontractual' nature can sometimes inhibit rather than encourage improvement. 2. Tend to emphasize the 'hard' and measurable aspects of performance.

- Including every person in the organization

Every person has the potential to contribute to quality.

- Examining all costs which are related to quality (esp. failure costs and getting thing 'right first time')

Usually categorized as:
  - Prevention costs -> costs incurred in trying to prevent problems, failures and errors
  - Appraisal costs -> costs associated with controlling quality to check if problems have occurred during and after creation of the product/service
  - Internal failure costs -> failure costs associated with errors which are dealt with inside the operation
  - External failure costs -> costs associated with an error going out of the operation to a customer

Effective investment in preventing quality errors can significantly reduce appraisal and failure costs. TQM emphasizes prevention -> the more internal and external costs can be reduced -> once confidence has been established, appraisal costs can be reduced. It shifts the emphasis from reactive to proactive -> from inspect-in to design-in (first time right).

- Developing the systems and procedures which support quality and improvement

ISO 9000 provides the set of standardized requirements for a quality management system which should apply to any organization, regardless of size, or private/public sector. Purpose -> provide assurance to purchasers that it has been produced in such a way that meets their requirements. It takes a 'process' approach that focused on outputs from any operation's process. Stresses:
1. Quality management should be customer focused
2. Quality performance should be measured
3. Quality management should be improvement driven
4. Top management must demonstrate their commitment to maintaining and continually improving management systems.

Benefits: to organizations adopting it and customers, provide useful discipline to stick to 'sensible' process-orientated procedures, gaining the certificate shows quality is taken seriously -> marketing benefit.
Criticism: encourages 'management by manual' and over-systemization, expensive and time consuming, cost and time maintaining the certificate are excessive, it is too formulaic.

- Developing a continuous process of improvement

Quality awards ->

The Deming Prize – successfully applied 'company-wide quality control' based upon statistical quality control
The Malcolm Baldrige National Quality Award – like the Deming Prize but in the USA
The EFQM Excellence Model - 14 Western European companies formed the European Foundation for Quality Management (EFQM) -> launched European Quality Award (EQA) for most successful component of TQM in Europe each year. Emphasis now more on customers. The EFQM defines **self-assessment** as 'a comprehensive, systematic, and regular review of an organization's activities and results referenced against a model of business excellence'.
ISO 14000 – largely limited to Europe, it has a three-section environmental management system which covers initial planning, implementation and objective assessment.

Supplement to Chapter 17

Statistical process control (SPC) is concerned with checking a service/product during its creation. Value is to monitor quality of period of time, using control charts -> steps can be taken before there is a problem. Looking for trends.
All processes vary to some extent, if derive from common causes they can never be entirely eliminated but can be reduced. Histogram -> process variation distribution -> normal distribution with 99.7% and within ±3 standard deviations. If that is acceptable depends on the specification range of the operation.
Process capability is a measure of the acceptability of the variation of the process. The simplest measure of capability (C_p) is given by the ratio of the specification range to the natural variation (±3) → $C_p = \frac{UTL-LTL}{6s}$. UTL: upper tolerance limit. LTL: lower tolerance limit. s: standard deviation of the process variability. Usually when $C_p < 1$, the process is not 'capable'. One-sided capabilities:
Upper one-sided index $C_{pu} = \frac{UTL-X}{3s}$
Lower one-sided index $C_{pl} = \frac{X-LTL}{3s}$. Where X is the process average.
When only the lower of the two one-sided indices is used to indicate its capability $(C_{pk}) = \min(C_{pu}, C_{pl})$.

Variation not result of common causes -> then it are assignable causes. To help decide if something is due to a common-cause of assignable-cause, control limits can be added to the control chart. If any points lie outside these limits -> assignable causes, it is 'out of control'. Set limits intuitively or for example, ±3 standard deviations from the mean. If process is stopped when the actual state is in control -> type I error. If process is left alone but actual state is out of control -> Type II error.
High levels of variation reduce the ability to detect changes in process performance.

Attributes have only 2 states. The population mean ($\bar{p}$) may be unknown -> mean can be estimated from the average of the proportion of 'defectives' ($\bar{p}$), from m (≥30) samples each of n (≥100) items:
$\bar{p} = \frac{p_1+p_2+p_3+\cdots+p_n}{m}$. One standard deviation can then be estimated from: $\sqrt{\frac{\bar{p}(1-p^n)}{n}}$. The upper and lower control limits can then be set as: UCL = $\bar{p}$ + 3σ and LCL = $\bar{p}$ - 3σ. LCL cannot be negative so round up to 0.

Commonly used type of control chart is the $\bar{X}$-R chart. The means chart can pick up changes in the average output, the R chart plots the range and gives an indication of whether the variability of the process is changing, even when the process average remains constant.
Calculating the control limits, first is to estimate the population mean and average range using m samples of sample size n.
$\bar{X} = \frac{\bar{X}_1+\bar{X}_2+\cdots+\bar{X}_m}{m}$. $\bar{R} = \frac{R_1+R_2+\cdots+R_m}{m}$.
Control limits of sample means: UCL = $\bar{X} + A_2\bar{R}$, LCL = $\bar{X} - A_2\bar{R}$ (may be negative)
Control limits of range charts: UCL = $D_4\bar{R}$, LCL = $D_3\bar{R}$ (may not be negative)

Other reasons for being out of control: alternating behavior, two points near control limit, apparent trend in one direction, suspiciously average behavior, five points one side of center line, sudden change in level.

Traditionally SPC was seen as very operational, now more operation's strategic capabilities. Process control leads to learning which enhances process knowledge and builds difficult-to-copy process capability.

Chapter 18

Risk management -> identifying that things could go wrong, stopping them going wrong, reducing the consequences when things go wrong and recovering after things have gone wrong. Failure will always occur in operations; recognizing this does not imply accepting or ignoring it. Managing risk and recovery generally involves 4 sets of activities:

1. Understanding what failures could potentially occur and assessing their seriousness

Understand potential sources of risk. A failure to understand failure can be the root cause of a lack of resilience. Failure sources are classified as:

- **Supply failure** -> any failure in the timing/quality of goods/service delivered into an operation. Important source because of increasing dependence on outsourcing in many industries, and emphasis on keeping the supply chain 'lean'
- **Human failures** -> two types:
  - Where key personnel leave, ill, die, cannot fulfil their role
  - People making mistakes
    - Errors = mistakes in judgement, person should do something different
    - Violations = acts that are clearly contrary to defined operating procedure

  Catastrophic failures are often caused by a combination of both.
- **Organizational failure** -> failures of procedures and processes, and failures that derive from a business's organizational structure and culture. Especially failure in the design of processes and failures in the resourcing of processes. This can derive from an organizational culture that minimizes consideration of risk, or lack of clarity in reporting relationships.
- **Technology/facilities failures** -> all IT systems, machines, equipment and buildings. All are liable to failure or breakdown.
- **Product/service design failures** -> in its design stage, a product/service might look fine on paper; only when it has to cope with real circumstances might inadequacies become evident
- **Customer failures** -> customers may 'fail' in that they misuse products/services

- **Environmental disruption** -> all causes outside an operation's direct influence. E.g. political upheaval, hurricanes, fire, fraud etc.
- **E-security** -> trade-off between providing wider access through the internet and the security concerns it generates. 3 developments:
    - Increased connectivity means that everyone has at least the potential to 'see' everyone else -> more available for employees etc
    - There has been a loss of perimeter security as more people work from home or through mobile communications
    - Sometimes unregulated technologies, it takes times to discover all possible sources of risk

Use previous failures to learn about sources of potential risk: post-failure analysis:
- **Accident investigation** – because of their infrequency it's hard to identify new sources of risks in advance
- **Failure traceability** – traced back to the process which produced them
- **Complaint analysis** – analyzing the number and content of complaints over time to understand better the nature of the failure, as the customer perceives it
- **Fault-tree analysis** – made up of branches connected by 2 types of nodes: AND and OR nodes. Branches below an AND node all need to occur for the event above the node to occur. Only one branch below an OR node need to occur for the event above to occur

Estimates of failure based on historical performance can be measured in 3 ways:
- **Failure rates** – how often a failure occurs
    - FR = number of failures/total number of products tested x 100%
    - FR (in time) = number of failures/operating time

  Sometimes failure is a function of time. The curve which describes failure probability of this type: bath-tub curve. It compromises 3 distinct stages:
    - **Infant-mortality**/**early life** – where early failures occur caused by defective parts or improper use
    - **Normal life** stage – when the failure rate is usually low and reasonably constant, caused by random factors
    - **Wear-out** stage – when the failure rate increases as the part approaches the end of its working life and failure is caused by the ageing and deterioration of parts.
- **Reliability** – the chances of a failure occurring, measures the ability to perform as expected over time. With interdependence, a failure in 1 part will cause the whole system to fail. So the reliability of the whole system is $R_s = R_1 \times R_2 \times R_3 \times \ldots \times R_n$. The more interdependent components, the lower its reliability will be.

  Alternative measure is mean time between failures (MTBF) = $\frac{Operating\ hours}{Number\ of\ failures}$ = $\frac{1}{failure\ rate\ (in\ time)}$.
- **Availability** – the amount of available useful operating time = $\frac{MTBF}{MTBF+MTTR}$. MTBF: mean time between failures. MTTR: mean time to repair (average time taken to repair, from time it fails to time it is operational again)

Subjective estimates of failure probability are better than no estimates at all.

Most well-known approach for doing failure mode and effect analysis (**FMEA**), assigning relative priorities to risk -> identify factors that are critical to various types of failure as a means of identifying failures before they happen. 3 key questions;

- What is the likelihood that failure will occur?
- What would the consequence of the failure be?
- How likely is such a failure to be detected before it affects the customer?

Based on a quantitative evaluation -> risk priority number (**RPN**) is calculated for each potential cause.

2. <u>Examine ways of preventing failures occurring</u> – 3 approaches to reducing risk by trying to prevent failure:
    a. **Redundancy** -> having back-up systems or components in case of failure. Calculated:
    $R_{a+b} = R_a + (R_b \times P(failure))$, where R_{a+b} = reliability of component a with its back-up component b. R_a = reliability of a alone. R_b = reliability of back-up component b alone. P(failure) = probability that component a will fail and therefore component b will be needed.
    Industry used 3 main types of redundancy:
        - Hot standby – both primary and secondary systems run simultaneously
        - Warm standby – secondary system runs in the background to the primary
        - Cold standby – secondary system only called upon when primary system fails
    b. **Fail-safeing** -> **poka-yoke** in Japan, human mistakes are to some extent inevitable -> important to prevent them becoming defects. Poka-yokes: simple devices/systems that are incorporated into a process to prevent inadvertent mistakes by those providing a service as well as customers receiving a service.
    c. Maintenance -> how organizations try to avoid failure by taking care of their physical facilities. 3 approaches:
        - **Run to breakdown maintenance** (**RTB**) – allowing facilities to continue operating until they fail.
        - **Preventive maintenance** (**PM**) – eliminate or reduce the chances of failure by servicing the facilities at pre-planned intervals.
        - **Condition-based maintenance** (**CBM**) – perform maintenance only when the facilities require it

Total productive maintenance (TPM) -> productive maintenance carried out by all employees through small-group activities. It is seen as the natural extension in the evolution from run-to-breakdown to preventive maintenance. The five goal of TPM:

1. **Improve equipment effectiveness** by examining all the losses which occur
2. **Achieve autonomous maintenance** by allowing staff to take responsibility for some of the maintenance tasks and for the improvement of maintenance performance
3. **Plan maintenance with a fully worked out approach** to all maintenance activities.
4. **Train all staff** in relevant maintenance skills so that both maintenance and operating staff have all the skills to carry out their roles.
5.
6. **Achieve early equipment management** by 'maintenance prevention' (MP), which involves considering failure causes and the maintainability of equipment during its design, manufacture, installation and commissioning

3. <u>Minimize the negative consequence of failure: failure or risk 'mitigation'</u>

Failure or risk mitigation -> isolating a failure from its negative consequences. The nature of action taken to mitigate failure will depend on the nature of the risk. Some types of mitigation actions that may be generally applicable:

- **Mitigation planning** – ensuring that all possible failure circumstances have been identified and the appropriate mitigation actions identified, may be described in the form of a decision tree, also provides mitigation action in its own right.
- **Economic mitigation** – actions such as insurance against losses from failure, spreading the financial consequences of failure, 'hedging' against failure
- **Containment (spatial)** – stopping the failure physically spreading to affect other parts of an internal or external supply network.
- **Containment (temporal)** – containing the spread of a failure over time, particularly when information about a (potential) failure needs to be transmitted without undue delay
- **Loss reduction** – any action that reduces the catastrophic consequences of failure by removing the resources that are likely to suffer those consequences
- **Substitution** – compensating for failure by providing other resources that can substitute for those rendered less effective by the failure

4. <u>Devise plans and procedures that will help the operation to recover from failures when they do occur</u>

Failure recovery is the set of actions taken to reduce the impact of failure once the customer has experienced its negative effects. Recovery needs to be planned and procedures put in place. Mistakes may be inevitable, dissatisfied customers are not. Failure may turn into positive experiences, good recovery can satisfy customers.

The **complaint value chain** helps to visualize the potential value of good recovery at different stages. Organization need to design appropriate responses to failure that are suitably aligned with the cost and the inconvenience caused by the failure to their customers. Start where failure is recognized:

- **Discover** – discover exact nature of failure. 3 pieces of information needed: What happened? Who will be affected? Why did the failure occur?
- **Act** – 3 actions: (1) tell significant people involved what you are proposing to do. (2) effects of failure need to be contained in order to stop the consequences spreading and causing further failures. (3) some kind of follow-up to make sure that the containment actions really have contained the failure
- **Learn** – revisiting the failure to find out its root cause and then engineering out the causes of the failure so it will not happen again
- **Plan** – identifying all possible failures which might occur and formally defining the procedures which the organization should follow in the case of each type of identified failure.

GO ROCK YOUR STUDIES

GOOD LUCK!