ACADEMIC SKILLS AND LEARNING CENTRE

Mathematics & Statistics Refresher

Introductory Academic Program

Garry Khemka

January 2011

Table of Contents

CHAPTER1: SET THEORY **5**

CHAPTER 2: NUMBER THEORY **11**

CHAPTER 3: PROOFS **13**

CHAPTER 4: PERMUTATIONS AND COMBINATIONS **19**

PERMUTATIONS 19
COMBINATIONS 21

CHAPTER 5: FUNCTIONS **25**

LINEAR FUNCTIONS 25
QUADRATIC FUNCTIONS 28
INDICES/EXPONENTS 30
LOGARITHMS 33

CHAPTER 6: THE DERIVATIVE AND THE DIFFERENTIAL **41**

APPLICATION OF DERIVATIVES 43

CHAPTER 7: SEQUENCES AND SERIES **49**

SEQUENCES 49
SERIES 53
APPLICATIONS OF SERIES AND SEQUENCES 57

CHAPTER 8: LINEAR ALGEBRA **61**

VECTORS 61
MATRICES 63
LINEAR EQUATIONS 68

CHAPTER 9: INTRODUCTION TO STATISTICS **75**

NUMERICAL METHODS 76
GRAPHICAL METHODS 78

CHAPTER 10: PROBABILITY AND RANDOM VARIABLES **85**

DISCRETE PROBABILITY DISTRIBUTIONS 90

CHAPTER 11: REGRESSION MODELLING **101**

NORMAL DISTRIBUTION 101
REGRESSION ANALYSIS 104

LIST OF REFERENCES **116**

Chapter1: Set Theory

Set and Sample Space

Sets are usually defined as a collection of elements. These elements can either be abstract in nature or numerical. Numerical elements are generally defined in the real number space($\mathbb{R}^n$).

Examples: *{1, 2, 3, 4, 5}; S={s: s is a state of Australia}; the names of students in the class; the amount of rainfall (in mm) each day of the last month; etc.*

Sets can include any kind of entity/object (another way of stating elements of a set) as long as they are well defined. A set is generally an "abstract" concept that covers any range of objects that are clustered together and combined according to a definition.

Set Notations:

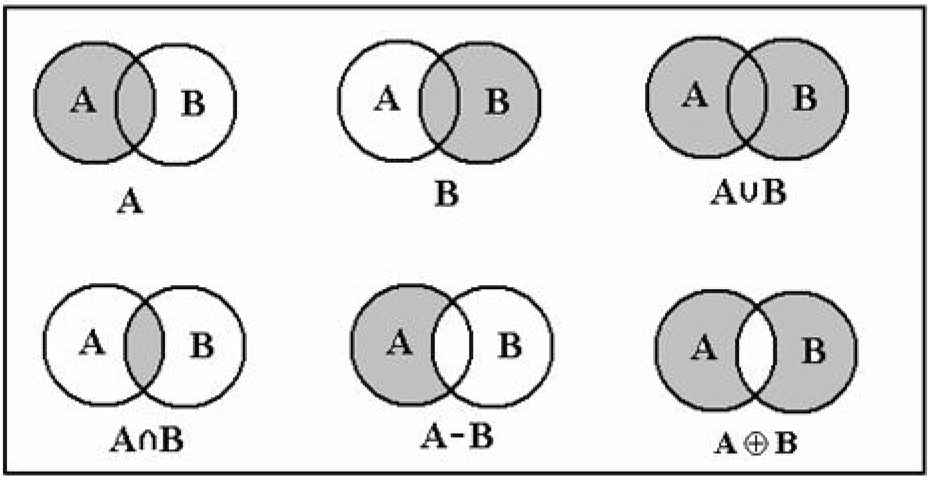
- ϕ denotes an empty set.
- $\omega \in$ A – ω is an element of the set A
- Subset Relation:$\{\omega\} \subset A\, or\, \{\omega\} \subseteq A$: the set consisting of the element $\omega \in A$ is a subset of A. $B \subseteq A - B$ is a subset of A and all members of B are in A $(\{a, b, c\} \subseteq \{a, b, c, d\})$. $B \subset A$ is known as proper subset (All members of A are members of B and there are members in A that are not in B- $\{a, b, c\} \subset \{a, b, c, d\}$ but $\{a, b, c\} \not\subset \{a, b, c\}$).
- $\{\omega: a\ statement\}$: the set of elements ω for which the statement holds. Example: the open interval (a, b) can be defined as $\{\omega: a < \omega < b\}$.
- $A = B$ if A and B contain exactly the same elements (this can be shown by showing $(1)\, A \subset B\, and\, (2)\, B \subset A)$
- ξ: "abstract space" – a non empty set of all elements concerned (also called "universal set"). Example: Draw a card from a deck of 52 cards: $\xi = \{1;\ 2;\ \ldots\ ;\ 52\}$, which is a finite sample space
- $\wp(A)$: The Power Set of A – the set containing all subsets of A ($\wp(\{\mathrm{a,\ b,\ c}\}) =$ $\{\phi\ \ , \{a\}, \{b\}, \{c\}, \{a, b\}, \{a, c\}, \{b, c\}, \{a, b, c\}\}$
- '**/**': Used in various symbols expressing relations to say that the relation does not hold [e.g. ∉, ⊄, etc.]

Operations of Sets

- **Union**: The union of sets A and B is the set of all elements which belong to A or B or both. It is denoted by the symbol "∪". $A \cup B = \{x: x \in A\ or\ x \in B\}$
- **Intersection**: The intersection of sets A and B is the set of all elements which belong to both A and B. It is denoted by the symbol "∩". $A \cap B = \{x: x \in A\ and\ x \in B\}$
- **Difference**: The difference of sets A and B is the set of all elements which belong to A but not B. It is denoted by the symbol "-". $A - B = \{x: x \in A\ and\ x \notin B\}$

- **Complement (w.r.t. ξ)**: If ξ is the universal set then the difference of sets A and ξ is the set of all elements which belong to ξ but not A. It is denoted by the symbol $``A^c" \; or \; "A'". A^c = \{x: x \in \xi \; and \; x \notin A\}$

When we talk about sets, the best way to understand them is using a visual representation. The visual representation of sets can be done via the use of Venn Diagrams. Venn diagrams use a circle to represent a set (like A in the diagram below) and generally a large rectangular box representing the universal set or "ξ". Also, once the universal set is defined all operations have to occur within it. (Note: we have not defined a universal set in the diagram below).



Properties of Sets and the common Laws/Theorems of Set Theory:

1. **Commutative:**
   a. $A \cup B = B \cup A$
   b. $A \cap B = B \cap A$
2. **Associative:**
   a. $(A \cup B) \cup C = A \cup (B \cup C)$
   b. $(A \cap B) \cap C = A \cap (B \cap C)$
3. **Distributive:**
   a. $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$
   b. $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$

4. **Identity: There are elements ∅ and ξ for every A satisfy:**
   a. $A \cap \emptyset = \emptyset$
   b. $A \cap \xi = A$
5. **Complement:**
   a. $A \cup A^c = \xi$
   b. $A \cap A^c = \emptyset$
6. **Transitive:** $If \; A = B \; and \; B = C, then \; A = C$
7. **Equality:** $If \; A = B, then \; A \cup C = B \cup C \; and \; A \cap C = B \cap C$

8. **Also,**
    a. $A \subset B \; means \; A \cap B = A$
    b. $A - B \; means \; A \cap B^c$
9. **Idempotent Laws:**
    a. $A \cup A = A$
    b. $A \cap A = A$
10. **Absorption Laws:**
    a. $A \cup (A \cap B) = A$
    b. $A \cap (A \cup B) = A$
11. **DeMorgan's Laws:**
    a. $(A \cup B)^c = A^c \cap B^c$
    b. $(A \cap B)^c = A^c \cup B^c$
12. **Consistency Principle (These come from the subset relationship):**
    a. $X \subseteq Y \; iff \; X \cup Y = Y \; (iff = if \; and \; only \; if)$
    b. $X \subseteq Y \; iff \; X \cap Y = X$

For further reading:

HRBACEK, K. & JECH, T. J. (1984) ***Introduction to Set Theory,*** **New York, M. Dekker.**

Exercises:

1. **Write down the following sets:**
   a. $\boldsymbol{A} = \{\boldsymbol{x}\text{: }\boldsymbol{x} < 10, x\ \boldsymbol{\epsilon}\ \boldsymbol{N}\}$
   b. $\boldsymbol{B} = \{\boldsymbol{the\ first\ 10\ elements\ of\ the\ fibbonacci\ series}\}$
   c. $\boldsymbol{C} = \{\boldsymbol{x}\text{: }\boldsymbol{x^2} < 10, x\ \boldsymbol{\epsilon}\ \boldsymbol{N}\}$
   d. $\boldsymbol{A} = \{\boldsymbol{the\ names\ of\ the\ seasons}\ \}$
2. **Let** $\boldsymbol{A} = \{\boldsymbol{1}, \boldsymbol{2}, \boldsymbol{3}, \boldsymbol{4}\}\ \boldsymbol{and}\ \boldsymbol{B} = \{\boldsymbol{3}, \boldsymbol{4}, \boldsymbol{5}, \boldsymbol{6}\}$ **find and represent in venn diagram form:**
   a. $A \cup B$
   b. $A \cap B$
   c. $A - B$
   d. A'
3. **Prove DeMorgan's Laws**

Answers:

1. $The\ sets\ are$:
   a. $A = \{1,2,3,4,5,6,7,8,9,\}$
   b. $B = \{1,1,2,3,5,8,13,21,34,55\}$
   c. $C = \{1,2,3\}$
   d. $A = \{Winter, Spring, Summer, Autumn\}$
2. $The\ answers\ are$:
   a. $A \cup B\ =\ \{1,2,3,4,5,6\}$
   b. $A \cap B\ = \{3,4\}$
   c. $A - B = \{1,2\}$
   d. $A'\ -\ cannot\ be\ defined\ unless\ we\ have\ a\ universal\ set$
3. $(A\ \cup\ B)^c =\ A^c \cap\ B^c$
   $Let\ X = (A\ \cup\ B)^c a\ nd\ Y =\ A^c \cap\ B^c. Now\ if\ we\ can\ prove\ that\ X \subset Y\ and\ Y \subset X\ then\ X = Y.$
   $Let\ x \in X, i.e., x\ does\ not\ belong\ to\ either\ A\ or\ B.$
   $Hence, x \in A^c\ and\ x \in B^c\ .$
   $Hence, x \in A^c \cap\ B^c, i.e., x \in Y.$
   $Similarly\ all\ elements\ in\ X\ are\ in\ Y\ and\ vice\ versa.$
   $Hence, proved.$
   $Similarly, you\ can\ prove\ the\ other\ law.$

Chapter 2: Number Theory

A **number** is a tool (mathematical in nature) which is used to represent a measure or to count something. Commonly we use the Hindu Arabic System of numbering.

The following are the different kind of number systems:

- ***Natural numbers***: Natural numbers are also called counting numbers and are represented as $\mathbb{N} = \{1, 2, 3, \ldots.\}$
- ***Whole numbers***: Whole numbers are natural numbers including zero. They are represented as $\mathbb{W}$
- ***Integers***: Integers are whole numbers and negative numbers taken together. They are represented by $\mathbb{Z} = \{\ldots., -3, -2, -1, 0, 1, 2, 3, \ldots\}$
- ***Rational numbers***: Any number that can be represented as an m/n fraction is called a rational number. The numerator is an integer and the denominator is again another integer (excluding zero). They are represented by $\mathbb{Q}$.
- ***Real numbers***: The real numbers include all of the measuring numbers. They are represented by $\mathbb{R} = (-\infty, \infty)$
- ***Irrational numbers***: All the numbers in the set of real numbers that are not rational are irrational numbers. (Note integers are rational numbers).
- ***Complex numbers***: Complex numbers are numbers that can be represented in the form $a + bi, where\ i = \sqrt{-1}$, and a and b are generally real numbers. i is known as an imaginary number/unit. They are denoted by the letter $\mathbb{C}$.

Chapter 3: Proofs

The property of logical implication or implication is of primary importance in mathematical reasoning. This leads us to the concept of proofs.

A proof is a method of establishing a statement by logical operating and reasoning on a collection of assumed or previously established statements.

There are several ways to proof a statement:

1. ***Direct Method***: A proof by direct method involves setting up certain logic/reasoning or statements from something already defined (generally from the question itself) and then combining them together.

   Example: *If we have:*

   $$\boldsymbol{n}! = \mathbf{1} \times \mathbf{2} \times \mathbf{3} \times \ldots \times \boldsymbol{n}$$

   Then prove that $\mathbf{0}! = \mathbf{1}$*. (n! Is pronounced n factorial)*

   If we look at factorials then we see that:

   $$\boldsymbol{n}! = \boldsymbol{n}(\boldsymbol{n} - \mathbf{1})!$$

   By making $\boldsymbol{n} = \mathbf{1}$,

   $$\mathbf{1}! = \mathbf{1}(\mathbf{0}!)$$

   For this statement to be true we must have;

   $$\mathbf{0}! = \mathbf{1}$$

   Hence, proved.

2. ***Indirect method***
   a. **Transposition**: Proof by transposition establishes the conclusion "$if\ p\ then\ q$" by proving the equivalent *contra-positive (or counter-positive)* statement "$if\ not\ q\ then\ not\ p$".
   Example:
   For Natural Numbers (also called counting numbers $\{1, 2, 3, \ldots\ldots\}$*) prove that if* x^2 *is odd then* x *is odd.*
   A natural number is either even or odd. In order to prove the above statement by transposition, we assume that x is even and that the conclusion is false. If x is even, then x times x or x^2 is even as the product of two even numbers is even (can itself be a proof!). This contradicts our premise that x^2 is odd. Thus we were wrong in assuming that x is even, hence x must have been odd!
   b. **Contradiction or *reductio ad absurdum* (RAA):** In order to do a proof by contradiction, we take the opposite of what we are asked to prove and show that the opposite cannot be true (logically or mathematically or intuitively).

Since the opposite cannot be true hence the actual statement must be true. This method is perhaps the most prevalent of mathematical proofs.

Example: *Prove that $\sqrt{2}$ is not a rational number.*

In order to prove the above statement we assume that $\sqrt{2}$ is a rational number and hence can be expressed as a fraction $\frac{\boldsymbol{m}}{\boldsymbol{n}}$. Then we have:

$$\sqrt{2} = m/n$$
$$2 = m^2/n^2$$
$$m^2 = 2n^2$$

Thus we get that m must be a multiple of 2. So we can write $m = 2q$

$$4q^2 = 2n^2$$
$$2q^2 = n^2$$

Which must mean that n is a multiple of 2. Thus, we have both m and n as multiples of 2, which cannot be true as the fraction m/n to be a rational number must not have any common divisors, i.e. must be reduced to its lowest factors. Hence, we conclude that $\sqrt{2}$ cannot be a rational number and must be an irrational number.

3. ***Mathematical Induction***: Mathematical induction in its simplicity involves proving using the "rule of induction". The rule of induction involves the following steps:
    i. Given a rule/theorem, we first show that the statement is true when $n = 1$
    ii. Then , we assume that the law is true for a certain $n = p$
    iii. Then we try to show that, if the statement is true for n=p then the statement holds for $n = p + 1$
    iv. Given the first three steps, if it is true for $n = 1$ then it should be true for $n = 2$. Again, if it is true for $n = 2$ then it should be true for $n = 3$ and so on. Hence the statement has to be true $for\ all\ n$ $(also\ represented\ as\ \forall\ n)$.

The mathematical induction proof generally is employed to proof statements that can run to infinity.

Example:

Prove that for all natural numbers n, $1 + 2 + 3 + 4 + \ldots + n = (n)(n+1)/2$

$$Let\ n = 1. Then\ 1 + 2 + 3 + 4 + \ldots + n\ is\ actually\ just\ 1;\ that\ is:$$
$$\frac{(n)(n+1)}{2} = \frac{(1)(1+1)}{2} = \frac{(1)(2)}{2} = 1.$$
$$Then\ we\ get\ 1 = 1, so\ it\ is\ true\ at\ n = 1.$$

$$Let\ n = p.$$
$$Assume\ that, for\ n = p, the\ formula\ works;\ that\ is, assume\ that:$$

$$1 + 2 + 3 + 4 + \dots + p = \frac{(p)(p+1)}{2}$$

$$\begin{aligned} &Let\, n = p + 1. \\ &Then\, 1 + 2 + 3 + 4 + \dots + p + (p + 1) \\ &= [1 + 2 + 3 + 4 + \dots + p] + p + 1 \\ &= \left[\frac{(p)(p+1)}{2}\right] + p + 1 \\ &= \left[\frac{(p)(p+1)}{2}\right] + \frac{2(p+1)}{2} \\ &= \frac{(p)(p+1)}{2} + \frac{2(p+1)}{2} \\ &= \frac{(p+2)(p+1)}{2} \\ &= \frac{\big((p+1)+1\big)(p+1)}{2} \end{aligned}$$

So the equation works for 1, some p and p+1. If p=1 then it works for 2. If k=3 then it works for 3 and so on. Hence it works for all natural numbers.

For further reading:

LEWIS, J. P. (1969) ***An Introduction to Mathematics for Students of Economics*** **London, Macmillan.**

MILLER, C. D. & HEEREN, V. E. (1969) ***Mathematical Ideas: An Introduction,*** **Glenview, Ill., Scott Foresman.**

Exercises:

1. **Using principle of mathematical induction prove the following statement**

$$(\,5^n - \,1\,)\ \boldsymbol{is\ divisible\ by}\ 4$$

2. **Prove that if x is irrational and y is rational then $x - y$ is irrational.**

Answers:

1. When n=1;

$$5^1 - 1 = 4$$

Which is divisible by 4.
Let the expression be true for n=k,
Then

$$5^k - 1 = say\ 4a, which\ is\ divisible\ by\ 4.$$

Then for n=k+1, we have

$$\begin{aligned} 5^{k+1} - 1 &= 5^k \times 5 - 1 \\ &= (4a - 1)5 - 1 \\ &= 20a + 5 - 1 \\ &= 20a + 4 \\ &= 4(5a + 1) \end{aligned}$$

which is divisible by 4.
Since it holds for n=1, it holds for n=2 and so on.
Hence proved.

2. Assume that it is rational. Then $x - y$ can be expressed in terms of a fraction. However, x cannot be expressed in terms of a fraction (as it is irrational). If we take a rational number out of it then still the remaining cannot be expressed in terms of a fraction. Hence, $x - y$ has to be irrational.

Chapter 4: Permutations and Combinations

Permutations and Combinations come under the field of discrete mathematics and is an important tool used in statistics. They are sometimes also called counting systems as they give us ways to calculate the actual number of outcomes in a particular situation (also, called sample space).

Permutations

Consider the following example:

Suppose we have three balls coloured red (R), green (G) and blue (B). If they are all inside a bag then what are the different ways you can take them out one at a time.

Suppose you pull out B in the first go. Then on the second go you can either pull out G or R. Similarly, if you pull out G on the first go, then on the second go you can either pull out B or R, and if you pull out R on the first go, then on the second go you can either pull out B or G. Hence, you have three ways of getting the first ball. Given the first ball you can either pull out either of the two in the second chance. So, you a further two ways of getting the second ball and, similarly, given the first two balls on the third try, you have one way of getting the final ball.

Writing them out, we get the different combinations to be:

$$BGR\ BRG\ RBG\ RGB\ GBR\ GRB$$

In numerical terms, there are $3 \times 2 \times I = 6$ ways in which we can extract the three balls from the bag.

Similarly, if we had 4 balls, the number of ways we would be able to extract them would be $4 \times 3 \times 2 \times 1 = 24$.

Now, if you notice that we have a pattern emerging here. From Chapter 3, we know that the term $4 \times 3 \times 2 \times 1$ is called $4!$ (or $4\ factorial$).

If we were to repeat the above exercise by increasing the number of balls by one at each turn we will find that in every case, if we start with n number of balls, then the number of possible ways that the balls can be extracted would $n!$.

This rule is called ***permutation****, and applies whenever we have n different objects arranged in sequence (or line).*

Note: the word <u>sequence</u> is of major importance, and the above definition only applies when the sequence of a particular set of events is of importance.

Now, consider that you have a bag with 10 balls each numbered from 1 through to 10. You are given three chances to extract three balls from the bag. How many ways can you extract three balls from the bag?

The above formulation does not give us a straightforward way to answer this question, but at a closer look we find that we do have the tools supplied to us to answer this question.

$Given\ 10\ balls, we\ can\ extract\ the\ first\ ball\ in\ 10\ different\ ways, leaving\ 9\ in\ the\ bag.$

$Given\ 9\ balls, we\ can\ extract\ the\ first\ ball\ in\ 9\ different\ ways, leaving\ 8\ in\ the\ bag.$

$Given\ 8\ balls, we\ can\ extract\ the\ first\ ball\ in\ 8\ different\ ways, leaving\ 7\ in\ the\ bag.$

$$And\ then\ we\ stop!$$

We do not repeat the process as we have finished the number of tries given to us.

Hence, we see that the number of ways we can choose three balls from a bag of 10 is $10 \times 9 \times 8$.

Expressing this in terms of factorials we have:

$$10 \times 9 \times 8 \times \frac{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1}{7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1} = \frac{10!}{(10-3)!} = {}^{\mathbf{10}}\boldsymbol{P}_{\mathbf{3}}$$

The final term in the above expression is the permutation term and gives us the different permutations of 3 objects being selected from an original of 10.

Formally,

$the\ number\ of\ different\ permutations\ of\ r\ objects\ which\ can\ be\ made\ from\ n$

$distinct\ objects\ is$

$$\frac{(\boldsymbol{n})!}{(\boldsymbol{n}-\boldsymbol{r})!} = {}^{\boldsymbol{n}}\boldsymbol{P}_{\boldsymbol{r}}$$

Example:

How many different five letter words can you form from all the alphabet letters (Assuming that it is irrelevant that the words actually make sense)?

We have 26 letters. In order to choose five from them we apply,

$$^{26}\boldsymbol{P}_{5} = \frac{\mathbf{26}!}{\mathbf{21}!} = \mathbf{7,893,600}$$

(It is altogether different matter that most of those words do not make sense!).

Example:

How many ways can you pick 5 balls from a bag of 10 if you replace the ball before you choose the next one?

As we replace the ball after every go, we still have 10 balls to choose from in the next round.

Hence, the number of ways we can choose 5 balls are: $10 \times 10 \times 10 \times 10 \times 10 = 10^5$

Combinations

Permutations and combinations differ from each other in the aspect that in case of permutations the order or sequence of events is important while in case of combinations we do not care about the order or the sequence of events.

For example, if you are playing cards, then given the fact that you have a said number of cards in your hand (depends on the game you are playing, generally 13 in most cases), you would not worry about the order or the sequence in which you received those cards but rather the fact that you were dealt that particular set of cards and none other.

Let's revert to the example of the three balls we discussed above. We found that the different ways we could have the three balls were given by (also known as the sample space):

$$BGR\ BRG\ RBG\ RGB\ GBR\ GRB$$

Now, in case of combinations the sequence in which the three balls were extracted from the bag is irrelevant. Hence, all the six permutations above combine to form one single combination – that is, there is only one way we can extract three balls from three in a bag (given we do not care about the sequence) – because the elements of each of the above permutation are different only in the sense of an order and nothing else.

Now, consider the case of 10 balls (numbered 1 to 10) and we are to draw three balls from them. We found that we had $^{10}\boldsymbol{P}_3$ ways of extracting them (note that $\boldsymbol{P}$ denotes sequence is important). However, each of the three balls can be arranged among themselves in 3! ways (denoting the order – comes from the first example). Since, combinations do not care about the ordering in order, to arrive at the number of combinations we divide our number of permutations by 3!. Thus, the number of combinations we have is $\frac{^{10}\boldsymbol{P}_3}{\mathbf{3!}}$ and is denoted by $^{10}\boldsymbol{C}_3$.

Generalising this we can say that, that the number of combinations of r objects selected from an original of n distinct objects is given by

$$\frac{n!}{(n-r)! \times r!} = {}^{n}\boldsymbol{C}_{r} = \binom{n}{r}$$

Example:

How many three card hands can we have from a standard 52 card deck?

We need 3 cards from a 52 card deck. The sequence in which we receive then is unimportant. Hence, the number of combinations that we can have is

$$\frac{52!}{(52-3)! \times 3!}$$

Example:

Bridge is a game of cards. In this game it is very important that one actually counts the number of cards in each suit (there are four suits in a deck – spades, hearts, diamonds and clubs) that the opponent has and how are they distributed. Say you count that two opponents have five cards between them. How many different ways can the cards are distributed among the two of them?

Now, there are five cards that the opponents have. If you consider a player sitting on the left of you (well it is important who has how many cards), he can have either 0 or 1 or 2 or 3 or 4 or 5 with him.

Then the different combinations that can arise are:

$$\boldsymbol{{}^5C_0 + {}^5C_1 + {}^5C_2 + {}^5C_3 + {}^5C_4 + {}^5C_5 = 32}$$

Hence, there are 32 different combinations in which he can have the cards. (Once we know his distribution then his partner will just have the opposite and hence, we do not need to calculate the partner's combinations – the above formula in effect captures both of them).

For further readings:

EPP, S. S. (2004) *Discrete Mathematics with Applications,* Belmont, CA, Thomson-Brooks/Cole.

MAY, W. G. (1967) *Foundations in Modern Mathematics,* Mass., Blaisdell.

Exercises:

1. **You have three bags which you can use to carry your books to university. Given that there are five working days in a week, in how many different ways can you carry your books to university in a week?**
2. **Take the word "MATHEMATICS". How many words can you form if you take all the letters at the same time?**
3. **You are hosting a dinner party. In how many ways can you sit your five friends and yourself around a six-seater table?**
4. **How many ways can you dial a six digit number from your mobile phone?**

Answers:

1. You can choose 1 bag from 3 in $\boldsymbol{^3C_1}\ ways = 3\ ways$.
$$You\ can\ do\ this\ on\ each\ of\ the\ 5\ days\ of\ a\ working\ week.$$
$$Hence\ there\ are\ 3 \times 5\ ways\ of\ carrying\ your\ books\ to\ university.$$
2. There are 11 letters in the word "MATHEMATICS". So number of different words using all the letters is $\boldsymbol{^{11}P_{11}}$**.**
3. 5! Ways (Circular permutations)
4. There are 10 digits. The sequence matters in a phone number. Hence the number of ways is $\boldsymbol{^{10}P_6}$**.**

Chapter 5: Functions

Linear Functions

Most of you have come across formulae in your life.

Consider the formula:

$$y = 32 + \frac{9}{5}x$$

This formula is actually the formula to convert degrees Celsius (x) to degrees Fahrenheit (y), i.e. if we put a value for x (°C), it will give us a corresponding value in y (°F).

The general form of expressing a function is $y = f(x)$ where y and x are variables of a function and each value of x gives us a unique value of y (one-to-one correspondence).

Definition: *A function from a set X to a set Y is a rule that assigns to each element of the set X exactly one element of the set Y.*

At this point let us introduce certain kinds of subsets of $\mathbb{R}$ called intervals.

Let a and b be two real numbers with $a < b$.

1. $\{x: a < x < b\}$ *is called an open interval and is denoted by* (a, b)
2. $\{x: a \leq x \leq b\}$ *is called a closed interval and is denoted by* $[a, b]$
3. $\{x: a < x \leq b\}$ *and* $\{x: a \leq x < b\}$ *are called half open interval and denoted by* $(a, b]$ *and* $[a, b)$ *respectively. They may be called open closed and closed open intervals respectively.*

The numbers a and b are called endpoints of these intervals. Either a or b or both can be infinite. (Note, infinite itself is not a real number, but a concept).

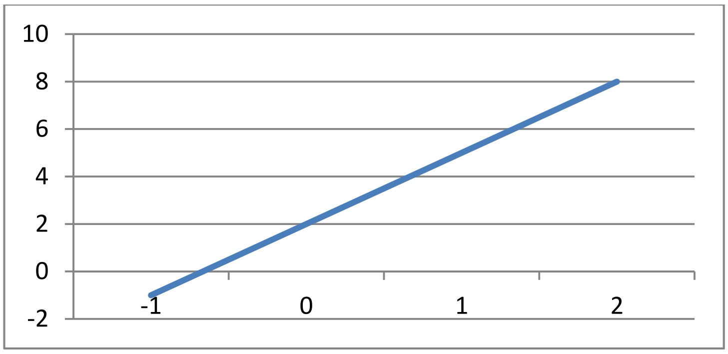
Now in a function; $y = f(x)$ where there is a correspondence from set X to set Y, the set X is called ***domain*** of the function and the set Y is called the ***range*** of the function.

In our above example, $f(30)$ means the value of y when $x = 30$. This gives us the value of y as 86. Thus we come to another way of representing a function: through a table of values.

We can draw up a table of values by evaluating $f(-10), f(-5), f(0), f(5)$ *and* $f(10)$.

x	-10	-5	0	5	10	*Domain*
$y = f(x)$	14	23	32	41	50	*Range*

If we plot these values in a graph we get a straight line through the point $(0, 32)$.



The example is an example of a linear function and the graph of linear functions is a straight line (a linear function is a function with x^1. If there is an x^2 then it is called a quadratic function – all of them belong to the family of polynomial functions – more of it later).

Any function of the form,

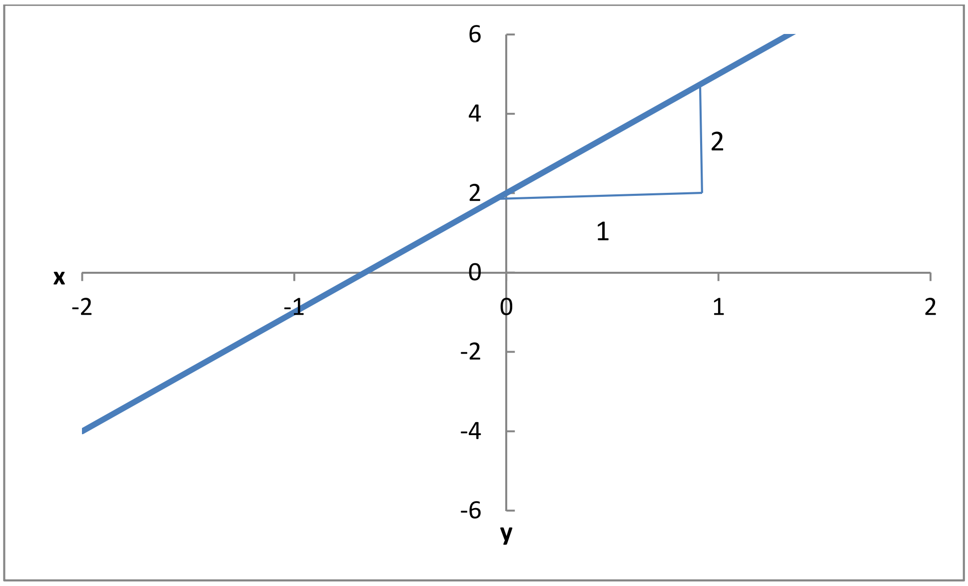
$$y = mx + b \; where \; m \; and \; b \; are \; constants$$

will have a straight line as its graph.

If we have a function of the form:

$$y = 2x + 2$$

the graph of the function looks like:



From the graph we see that when $x = 0$, we have $y = 2$, and when $x = 1$ the value of $y = 4$. That is, an increment of $1 - 0 = 1$ unit in the value of x will result in an increment of $4 - 2 = 2$ units in the value of y. This will always be the case for this function. (Pick a few values of x and try it.)

In any linear function of the form $y = mx + b$, we have two components which are necessary to define the function and to plot it.

The term b is called the intercept of the function. It gives us a value y when $x = 0$. In other words, it gives us the point in the graph where the line crosses the $y - axis$ (the vertical axis).

The term m is called the slope or gradient of the function and it gives the rate of change of y for every unit change of x. That is, if x increased by 1, then how much would y change by? The slope is represented by the ratio:

$$slope = \frac{change\ in\ y}{change\ in\ x}$$

If we are given two points on the line (x_1, y_1) and (x_2, y_2) the slope m can be worked out as follows:

$$m\ = \frac{change\ in\ y}{change\ in\ x}$$

$$= \frac{y_2 - y_1}{x_2 - x_1}$$

For example: if we have two points $(1,4)$ and $(5,8)$, then the slope m is given by:

$$m\ = \frac{y_2 - y_1}{x_2 - x_1}\ = \frac{\{8\ - 4\}}{\{5 - 1\}}$$

$$= 4/4$$

$$=\ 1.$$

For a straight line (or a linear function), we can completely define the equation of the line from two points. Consider the two points mentioned above: $(1,4)$ and $(5,8)$.

We know the slope is 1, i.e. $m = 1$.

The equation of the line is $y\ =\ 1x + b$.

We can determine b by substituting either point into the equation.

$$4\ =\ 1 \times (1) +\ b\ =\ 1 + b$$

which gives $b\ =\ 3.$

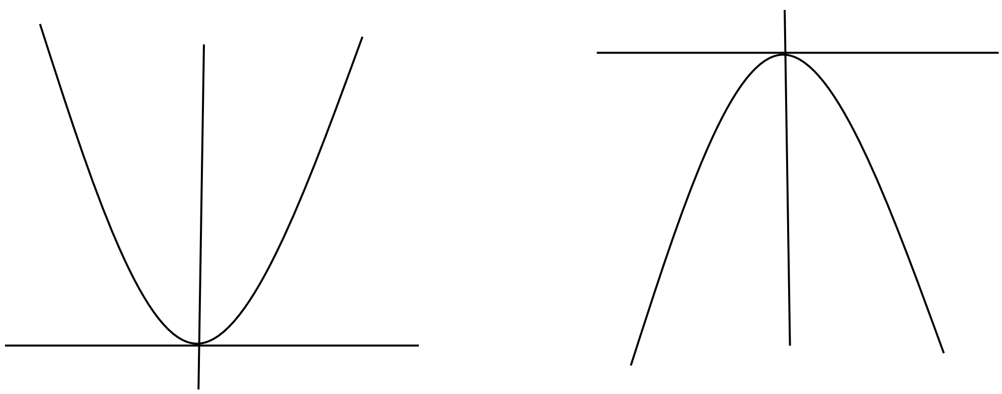
Therefore, the equation of the straight line going through the points $(1,4)$ and $(5,8)$ is

$y\ =\ 1x\ +\ 3\ or\ y = x + 3.$

Quadratic Functions

A quadratic function has an equation of the form $y = ax^2 + bx + c$ where $a \neq 0$, b and c are constants.

The graph of a quadratic is always a parabola and looks like the following:



For the general quadratic function of the form $y = ax^2 + bx + c, if\ a > 0$ we get a parabola represented in the first figure, while if $a < 0$, we get a parabola shown in the second one.
From the above figures it is evident that a quadratic function will have a vertex which can either be maximum or minimum - $if\ a > 0$ we get a parabola minimum, while if $a < 0$, we get a maximum. (We will discuss how to find the maximums and the minimums in our discussion of derivatives).
(Note that both linear and quadratic functions are members of the family of polynomial functions.

A polynomial function looks like: $y = f(x) = a + bx + cx^2 + dx^3 + \ldots\ldots + px^n$)

An important point of note with respect to quadratic functions is that when $y = 0$, the quadratic function becomes a quadratic equation:

$$ax^2 + bx + c = 0$$

A quadratic equation can be solved in many ways. One of the ways to solve a quadratic equation is to use the following relationship:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a}$$

The relationship will give us the two values of x that solve the equation – and generally you have to decide with proper arguments which of these two values would you choose/accept as an acceptable one. (Generally in the world of business we tend to reject the value of x that is negative).

Indices/Exponents

We have come across a term 10^3 in the solution to one of our earlier examples. Any expression of this kind is employing the use of indices.

The term 10^3 is the same as writing $10 \times 10 \times 10$, i.e. multiplying 10 by itself three times and in words is said as "10 raised to the power of 3". In the expression 10^3 we call 10 the base and 3 the index/exponent/power. We have special names for something that is raised to the power of either two or three and we call it "squared" or "cubed" respectively. In this example we have "10 cubed".

In general form indices are written as x^n, where x is the base and n is the index, and it means multiply x by itself n times.

Whenever we are dealing with indices we follow the following rules:

1. Multiplying two indices with the same base gives us the same base with the indices being added together.
$$x^n \times x^m = x^{n+m}$$
2. Division of two indices with the same base gives us the same base with the denominator index being subtracted from the numerator index.
$$\frac{x^n}{x^m} = x^{n-m}$$
3. When a number in exponential form is itself raised to another exponent then the indices get multiplied together (and the base remains the same)
$$(x^m)^n = x^{mn}$$
4. We define anything (except 0) raised to the power of 0 as being equal to 1. (in order to keep the second rule consistent when n=m)
$$x^0 = 1$$
Note: 0^0 is undefined.
Also, given that $x^0 = 1$, then we arrive at the fact that
$$x^{-1} = \frac{1}{x}$$
Hence, negative powers just become a denominator with a positive power (and the numerator is 1).
5. A base that is raised to a fraction m/n is equivalent to taking n^{th} root of that base and raising the entire thing to the power of m.
$$x^{\frac{m}{n}} = \left(\sqrt[n]{x}\right)^m$$
Note: if x >0 then $\sqrt[n]{x}$ is a positive number.
If $x < 0$ we need to look at separately at the cases where n is even and where n is odd.

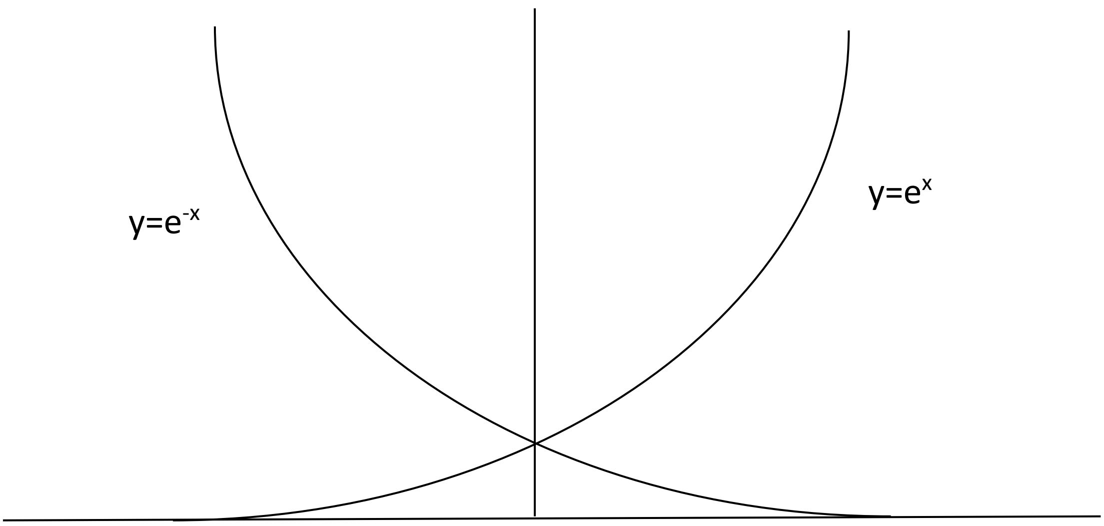
   If n is *even* and $x < 0$, $\sqrt[n]{x}$ cannot be defined, because raising any number to an even power results in a positive number.

If n is *odd* and $x < 0$, $\sqrt[n]{x}$ can be defined. It is a negative number, the n^{th} root of b.

6. If two numbers are multiplied by each other and raised to an exponent, it is the same as raising each number to the exponent first and then multiplying them with each other.
$$(xy)^n = x^n y^n$$
7. If a rational number is raised to an exponent, it is the same as raising the numerator and denominator individually to the exponent.
$$\left(\frac{x}{y}\right)^n = \frac{x^n}{y^n}$$

Any discussion of indices is not complete without mention of the number e. The number e is sometimes called **Euler's number** (after the Swiss mathematician Leonhard Euler) and is of extreme importance in any field where mathematics is used. The numerical value of e is 2.71828 (to five decimal places) and any function of the form $y = e^x$ is called an exponential function (not to be confused with exponents). It can also be represented as $y = exp(x)$.

The graph of the function looks like this (when e is raised to either the negative or the positive power of x)



Examples:

$$(6 \times 4)^2 = 6^2 \times 4^2 = 36 \times 16 = 576$$

$$(4x)^{\frac{1}{2}} = 4^{\frac{1}{2}} x^{\frac{1}{2}} = 2x^{\frac{1}{2}} = 2\sqrt{x}$$

$$(-40)^{\frac{1}{3}} = (-8 \times 5)^{\frac{1}{3}} = (-8)^{\frac{1}{3}} \times (5)^{\frac{1}{3}} = -2 \times 3\sqrt{5}$$

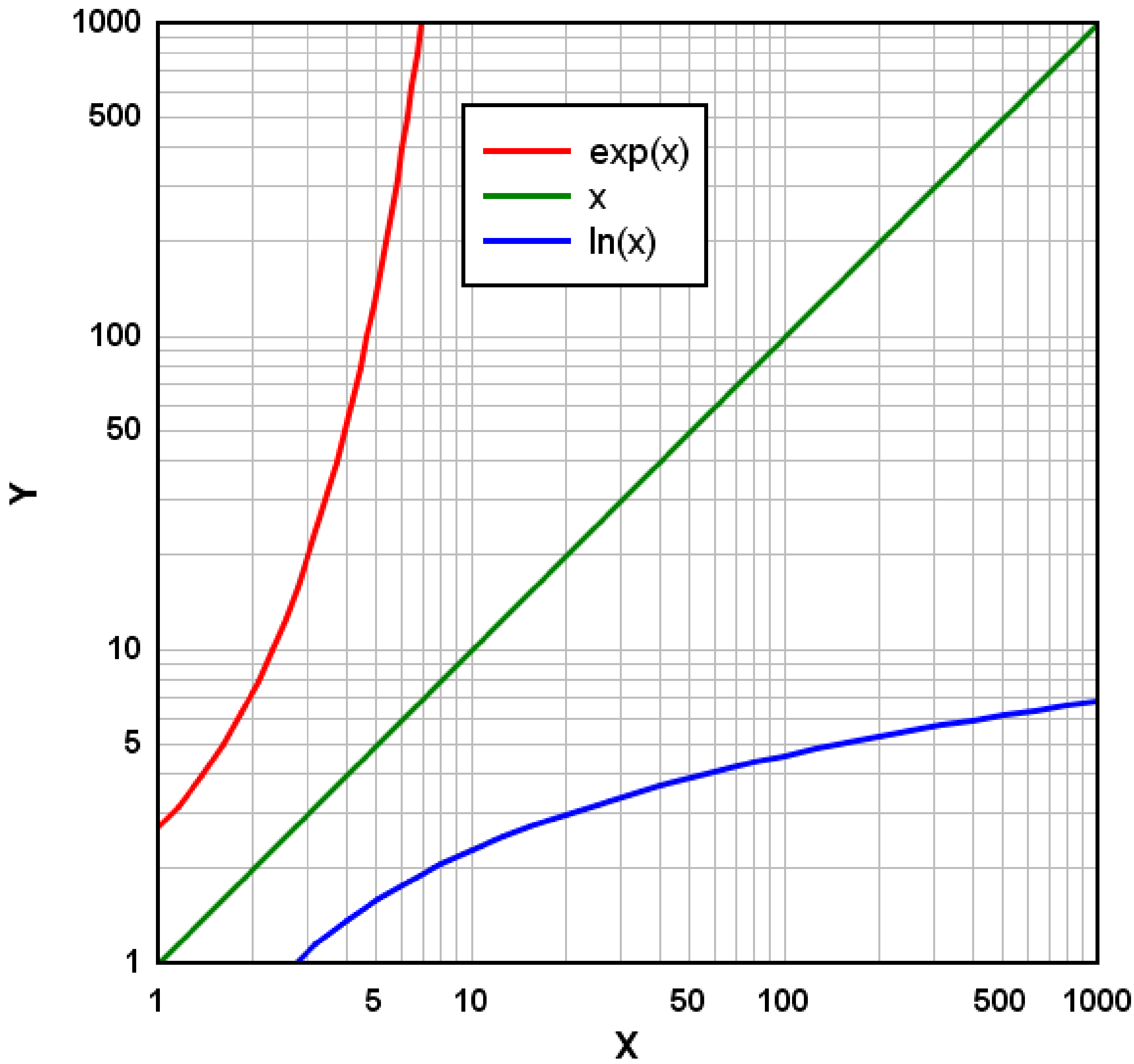
$$\left(\frac{2}{3}\right)^{3} = \frac{2^{3}}{3^{3}} = \frac{8}{27}$$

$$\left(\frac{4}{7}\right)^{-2} = \frac{1}{\left(\frac{4}{7}\right)^{2}} = 1 \times \frac{7^{2}}{4^{2}} = \frac{49}{16}$$

$$\left(-\frac{27}{8}\right)^{-\frac{1}{3}} = \left(-\frac{8}{27}\right)^{\frac{1}{3}} = \frac{(-8)^{\frac{1}{3}}}{27^{\frac{1}{3}}} = -\frac{2}{3}$$

Logarithms

Logarithms are, in effect, the reverse of indices/exponents. In order to understand this concept let us have a look at the following figure which gives a representation of $x, exp(x)\ and\ ln(x)$



Consider the exponent 10^3. Here 10 is the base and 3 is the exponent. The numerical value of the expression is 1000. That is, we have:

$$10^3 = 1000$$

$$or, \sqrt[3]{1000} = 10$$

$$or, \log_{10} 1000 = 3$$

All the above three equations mean the same thing and are expressing it in different ways. The final equation is read as: "log to the base 10 of 1000 is equal to 3".

Thus, we come to the definition of logarithms:

Given a positive number x $(base)$, logarithms give us the $power$ (y) to which x has to be raised to give z.

That is:

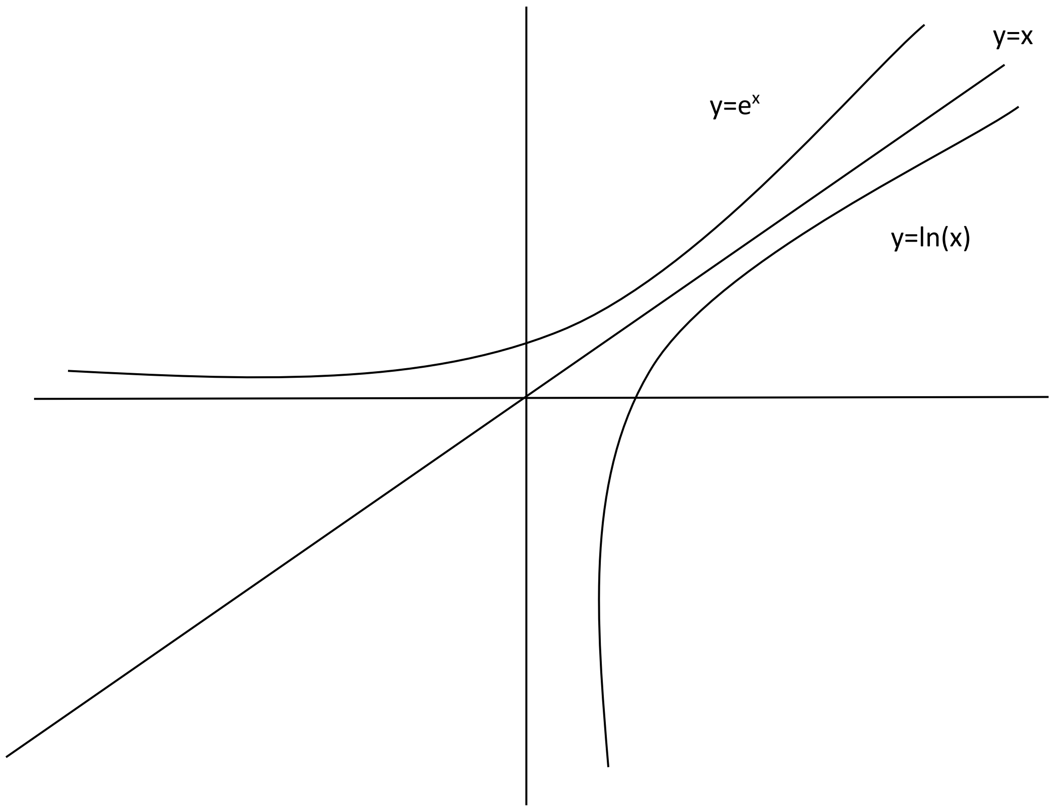
$$log_x z = y$$

$$where,\ x^y = z$$

Before, we move on let us discuss some of the restrictions of the logarithm function:

a. The number z in the above equation can never be negative – logarithm is incapable of dealing with negative numbers.
b. Also, if $z = 0$ then the logarithm function is undefined.
c. If $z = 1$ then irrespective of x the value of y is always 0.

The following figure explains the above conditions (and those of the exponentials as well). (Note: all the three functions have a one to one correspondence, i.e. one value of x gives us only one value of y). The graph shows the relationship between exponential, linear and logarithmic functions.



The two most commonly used bases of logarithms are $10\ and\ e$ (or exp). The base 10 is generally used because of the ease of expressing (and understanding) the logarithm functions in that particular form (and probably because the number 10 to mathematicians is what the number 5 is to statisticians!) Logarithms to the base e are called natural logarithms (represented by "ln") and are the ones you are going to be using most of the time during your university degree (especially in the College of Business and Economics).

Natural logarithms and e^x are of special importance in the finance literature with respect to continuous compounding of the interest rates. $ln(1+i)$ for any period gives us the continuous compounded interest rate for that period and in you are given the continuous compounded return then you use $e^{\delta} - 1 = i$, the interest rate.

Rules of logarithm functions:

1. The following rules are called cancelling exponents and follow from the definition of logarithms ($x \epsilon \mathbb{R}$ and $x > 0$):
    a. $\log_{10} 10^x = x$.
    b. $10^{log_{10} x} = x$
2. Adding two logs with the same base is the same as multiplying the inner numbers together.
$$\log_{10} xy = \log_{10} x + \log_{10} y, \qquad x, y > 0$$
3. Subtracting two logs with the same base is the same as dividing the second inner number from the first.
$$\log_{10} \frac{x}{y} = \log_{10} x - \log_{10} y, \qquad x, y > 0$$
4. If the inner number can be expressed as an exponential then the following holds:
$$\log_{10} x^n = n \log_{10} x, \qquad x > 0$$
5. In order to change the base of a logarithm the following hold:
    a. $\log_b x = \log_b a \times \log_a x, \; x > 0; a, b > 1$
    b. $\log_b x = \frac{\log_a x}{\log_c b}, \; x > 0; a, b > 1$
6. Also,
$$x^{\log_b y} = y^{\log_b x}$$

Examples:

$\log_{10} 2^4 = 4 \log_{10} 2$

$\ln 12 + \ln 9 - \ln 4 = \ln \frac{12 \times 9}{4} = \ln 27$

$\log_{10} 100 = 2$

$\log_7 9 = \frac{\log_{10} 9}{\log_{10} 7} = \log_{10} 9 \times \log_9 7$

For further readings:

MAY, W. G. (1967) *Foundations in Modern Mathematics,* Mass., Blaisdell.

EPP, S. S. (2004) *Discrete Mathematics with Applications,* Belmont, CA, Thomson-Brooks/Cole.

MILLER, C. D. & HEEREN, V. E. (1969) *Mathematical Ideas: An Introduction,* Glenview, Ill., Scott Foresman.

Exercises

1. ***Find the slope and intercept of the line*** $2x + 3y - 6 = 0$.
2. ***Find the equation of the line through the point*** $(0, 4)$ ***with*** $m = -3$
3. ***Find the slope of the line through*** $(-3, 1)$ ***and*** $(2, 4)$.
4. ***Find the equation of the above line.***
5. ***Find the equation of the line which is parallel to the line*** $y = 2x - 5$ ***and passes through the point*** $(1/2\ , -1)$.
6. $3^{n+2}/3^{n-2}$
7. $\sqrt{}(16/x6)$
8. $\left(a^{\frac{1}{2}} + b^{\frac{1}{2}}\right)^2$
9. **Solve:** $2e^x + e^{-x} = 3$
10. **Calculate** log_{10} **of the following numbers.**
    a. ***10000***
    b. ***1/100***
11. **Simplify:**
    a. $\log_{10} 10^{10^{2x}}$
    b. $log_{10}\, x^3 - 2.5\, log_{10}\, y$
    c. $2\, log_3(x + y) - 3\, log_3(xy) + log_3\, x^2$
    d. $log_{10}\, 0.1 \times log_6\, x - 2\, log_6\, y + log_6\, 4 \times log_4\, e$
    e. $ln(e^{-2.4}\, x^6)$
    f. $log_3 \frac{x^3 y^2}{27z^{1/2}}$
12. **Solve:** $ln(x) + 2 = -3ln(x) + 10$

Answers:

1. Rewriting the equation, we get

$$2x + 3y - 6 = 0$$
$$3y = -2x + 6$$
$$y = -\left(\frac{2}{3}\right)x + 2.$$

Hence, $m = -2/3$ and intercept is 2.

2. The slope of the line is −3 so $y = -3x + b$. and the line passes through $(0,4)$ so when $x = 0,\ y = 4$, so 4 is the intercept:

$$y = -3x + 4.$$

3. Let $(x_1, y_1) = (-3, 1)\ and\ (x_2, y_2) = (2, 4)$. Then

$$m = \frac{y_2 - y_1}{x_2 - x_1}$$
$$= \frac{4 - 1}{2 - (-3)} = \frac{3}{5}$$

4. The point $(-3, 1)$ lies on the line $y = 3/5x + b$. *Hence:*

$$1 = \left(\frac{3}{5}\right) * (-3) + b$$
$$1 = -\frac{9}{5} + b$$
$$b = 1 + \frac{9}{5}$$
$$= \frac{14}{5}$$

So the equation of the line is $y = \frac{3}{5x} + \frac{14}{5}$.

5. The slope of the line $y = 2x - 5$ is 2. A line parallel to this line will have the same slope; $m = 2$. The point $(\ 1/2\ , -1)$ lies on the line so starting with $y = 2x + b$ we get,

$$-1 = 2\left(\frac{1}{2}\right) + b$$
$$b = -2.$$

So the equation of the line is $y = 2x - 2$.

6. $3^{n+2}/3^{n-2} = 3^{n+2-(n-2)} = 3^4 = 81$

7. $\sqrt{\frac{16}{x^6}} = \left(\frac{16}{x^6}\right)^{\frac{1}{2}} = \frac{16^{\frac{1}{2}}}{x^{6\times\frac{1}{2}}} = \frac{4}{x^3}$

8. $\left(a^{\frac{1}{2}} + b^{\frac{1}{2}}\right)^2 = \left(a^{\frac{1}{2}}\right)^2 + 2a^{\frac{1}{2}}b^{\frac{1}{2}} + \left(b^{\frac{1}{2}}\right)^2 = a + 2a^{\frac{1}{2}}b^{\frac{1}{2}} + b$

9. $2e^x + e^{-x} = 3$

$\Rightarrow (2e^x + e^{-x})e^x = 3e^x$

$\Rightarrow 2e^{2x} + 1 = 3e^x$

$\Rightarrow 2u^2 + 1 = 3u\ (where, u = e^x)$

$\Rightarrow 2u^2 - 3u + 1 = 0$

$\Rightarrow u = 1, u = ½$

$\Rightarrow e^x = 1, or\ e^x = 1/2$

$If, e^x = 1\ then, \quad ln(e^x) = ln\,1 \Rightarrow x = 0$

$If\ e^x = 1/2, then\ ln(e^x) = ln(1/2) \Rightarrow x = -ln2$

10. *The answers are*

a) $\log_{10} 10^3 = 3$
b) $\log_{10} 10^{-2} = -2$

11. Answers are:

a) $\log_{10}10^{10^{2x}} = 10^{2x}$

b) $\log_{10} x^3 - 2.5\log_{10} y$

$= \log_{10} x^3 - \log_{10} y^{2.5}$

$= \log_{10}\frac{x^3}{y^{2.5}}$

c) $2\log_3 x + y - 3\log_3 xy + \log_3 x^2$

$= \log_3(x+y)^2 - \log_3(xy)^3 + \log_3 x^2$

$= \log_3\frac{((x+y)^2\,x^2)}{(xy)^3}$

d) $\log_{10} 0.1 \times \log_6 x - 2\log_6 y + \log_6 4 \times \log_4 e$

$= -1 \times \log_6 x - \log_6 y^2 + \log_6 e$

$= log_6\frac{e}{xy^2}$

e) $ln(e^{-2.4}x^6)$
$= -2.4 + 6lnx$

e) $log_3 \frac{x^3y^2}{27z^{1/2}}$
$= 3\log_3 x + 2\log_3 y - \log_3 27 - \left(\frac{1}{2}\right)\log_3 z$
$= 3\log_3 x + 2\log_3 y - 3 - \left(\frac{1}{2}\right)\log_3 z$

12. $ln(x) + 2 = -3ln(x) + 10$
$\Rightarrow 4ln(x) = 8$
$\Rightarrow ln(x) = 2$
$\Rightarrow e^{ln(x)} = e^2$
$\Rightarrow x = e^2$

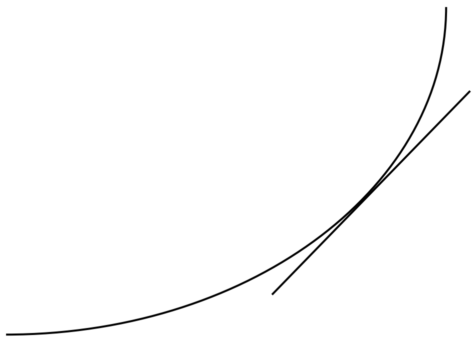
Chapter 6: The Derivative and the Differential

Derivatives in their most general form are used to express the rate of change of a variable with respect to another variable. In the previous chapter we discussed straight line equations where the slope of the line measured the change in y with respect to a unit change in x. Hence, we can say that the slope of a line is the derivative of the line.

If we have two variables $x\ and\ y$, in the functional form $y = f(x)$ (i.e. x is the independent variable and y is the dependent variable), then we define derivatives of y with respect to x as

$$\frac{dy}{dx} = \frac{df}{dx} = \frac{change\ in\ y}{change\ in\ x}$$

If we consider a polynomial curve, then derivative of the curve at any particular point measures the slope of the tangent at that point. This is represented in the figure below:



The rules of derivatives/differentiation are:

1. The derivative of a constant is 0:

$$If\ y = a, \qquad then \frac{dy}{dx} = \frac{d(a)}{dx} = 0$$

2. Derivative of an exponent:

$$if\ y = ax^b, \qquad then \frac{dy}{dx} = a(bx^{b-1})$$

3. Derivative of the exponential: The derivative of the exponential is unique in its nature because the derivative of an exponential is the exponential itself.

$$if\ y = e^x, \qquad then \frac{dy}{dx} = e^x$$

$$if\ y = e^{ax}, \qquad then \frac{dy}{dx} = ae^{ax}$$

4. Derivative of a logarithm;

$$if\ y = \ln x, \qquad then \frac{dy}{dx} = \frac{1}{x}$$

5. If y is the sum of two functions then derivative of y is the derivative of each of the individual functions

$$if\ y = f(x) + g(x), \qquad then \frac{dy}{dx} = \frac{df}{dx} + \frac{dg}{dx}$$

Example:

$$y = x^2 + e^x, \qquad then \frac{dy}{dx} = \frac{d}{dx}(x^2) + \frac{d}{dx}(e^x) = 2x + e^x$$

6. The product rule:

$$if\ y = u(x).v(x), \qquad then \frac{dy}{dx} = u(x).v'(x) + v(x).u'(x)$$

Example:

$$if\ y = xln(x), \qquad then \frac{dy}{dx} = x.\frac{1}{x} + \ln(x).1 = 1 + \ln(x)$$

7. The quotient rule:

$$if\ y = \frac{u(x)}{v(x)}, \qquad then \frac{dy}{dx} = \frac{v(x).u'(x) + u(x).v'(x)}{v(x)^2}$$

Example:

$$if\ y = e^x.\ln x, \qquad then \frac{dy}{dx} = \frac{lnx.e^x + e^x.\frac{1}{x}}{(lnx)^2}$$

8. The chain rule (applied to composite functions):

$$if\ y = f\big(g(x)\big), \qquad then \frac{dy}{dx} = f'\big(g(x)\big) \times g'(x) = \frac{dy}{dg} \times \frac{dg}{dx}$$

Example:

$$if\ y = \ln(3x + 4x^2), \qquad then \frac{dy}{dx} = \frac{1}{3x + 4x^2} \times (3 + 8x)$$

Note: with the application of the chain rule the first and the most important step to solving the questions is the recognition of the composite function. Once that hurdle is crossed the rest should fall into place.

Example:

Find the slope of the tangent of the curve $y = 3x^2 + 4x + 5$ at the point where $x = 2$.

The first step is to differentiate the above function:

$$\frac{dy}{dx} = 6x + 4$$

Now, we plug in the value of x to the above $f'(x)$ to get the value of the slope of the tangent.

Putting $x = 2$, we get

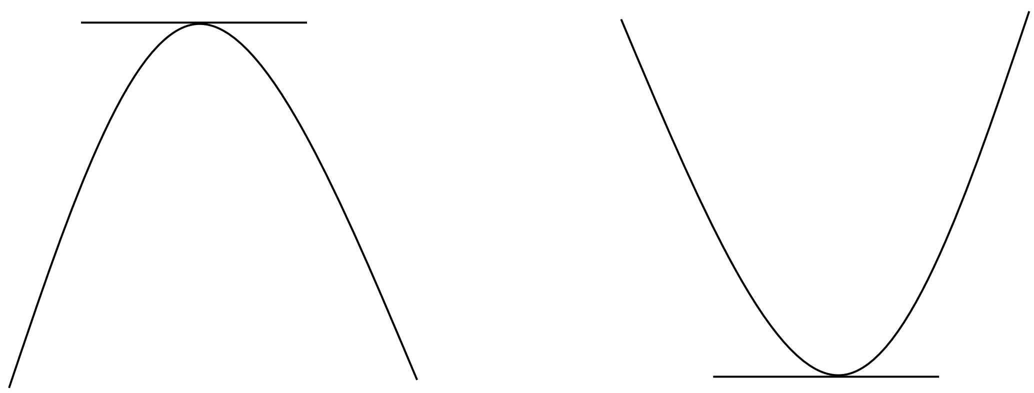
$$\frac{dy}{dx} = 12 + 4 = 16$$

Application of Derivatives

In the previous chapter we came across the terms "maximum" and "minimum" in our discussion of quadratic functions. In various economic and scientific studies, we are at different points interested in finding the maximum and/or minimum points of a given model. (There are generally two kinds of maximums and minimum points – global and local. Global signifies the maximum/minimum over the entire range and local signifies the maximum/minimum between two endpoints. The phenomenon occurs when you have a curve that has many crests and troughs with some bigger than the other. In all further discussion we will only be concerned with the global maximum and minimum points).

We know that derivatives give us the slope of the tangent to a curve at a particular point. We can use this to find the slope of the tangent at a particular point on the curve or to find a particular point on the curve where we know the slope to be equal to a particular value. We will use this analogy to help find the maximum and minimum points of a curve.

Consider the following diagrams.



In each of the diagrams we see vortex (maximum in the first and minimum in the second). The line that is tangent to the vertex in either case is horizontal. We know that the slope of a horizontal line is 0. Hence, the derivative of the curve at the vertex should be equal to 0. Using this principle, if we differentiate the function representing the curve and set it equal to 0, the resulting value(s) should give us the maximum (and/or minimum) points.

So the process is:

- $Let\ y = f(x).$
- $Find\ f'(x).$
- $Solve\ for\ the\ value(s)\ of\ x\ that\ satisfies\ f'(x) = 0.$

Now, we may have a curve that has both maximum and minimum points, so in order to find out whether the value(s) of x that we came up in step 3 above corresponds to maximum or minimum we differentiate $f'(x)$ again with respect to x. In the resulting expression we plug in the values of x and follow these rules:

- If the sign of the expression is negative then the resulting x value corresponds to a maximum; and
- If the sign of the expression is positive then the resulting x value corresponds to a minimum.

Examples:

Consider the following production model:

$$y = 1.2k^2 - 0.8k$$

Find the minimum production possible for a firm with this production function.

In order to find the maximum value, we differentiate the above expression and set it equal to zero:

$$f'(x) = 2.4k - 0.8 = 0$$

$$\Rightarrow k = \frac{0.8}{2.4} = 0.3$$

To check whether this value corresponds to a minimum, we differentiate it again and check the sign for $k = 0.3$:

$$f''(k) = 2.4 > 0, \qquad hence, minimum.$$

Find the maximum and minimum points of the following function:

$$y = 2x^3 - 4x^2 + x - 1$$

Differentiating it once we get:

$$\frac{dy}{dx} = 6x^2 - 8x + 1 = 0$$

The solution to the above quadratic equation is:

$$x = \frac{-b \pm \sqrt{b^2 - 4ac}}{2a} = \frac{8 \pm \sqrt{-8^2 - 4 \times 6 \times 1}}{2 \times 6} = \frac{8 \pm \sqrt{40}}{12} = (0.139, 1.193)$$

To find the maximum and the minimum, we differentiate the expression again:

$$f''(x) = 12x - 8$$

Setting $x = 0.139$, we get

$$f''(x) = -6.324 < 0, hence\ maximum$$

Setting $x = 1.193$, we get

$$f''(x) = 6.324 > 0, hence\ minimum$$

The above discussion is part of the branch of mathematics that is employed in Optimisation Theory. Optimisation theory is of great use especially in the field of economics, where resources are scarce and the economists want to find the optimum use of the resources that reduces waste. From a firm's perspective, optimisation theory is employed to minimise costs and maximise production and revenue at the same time. Similarly, from a consumer's perspective it can be seen as a problem of minimising expenditure and maximising utility from a given bundle of goods. Optimisation theory is of use in the financial world as well, when as an investor you want to maximise your return while minimising your risk. Hence, you decide upon the optimum level of risk – return payoff. In the field of engineering, say with respect to building a bridge, you would like to find the optimum number of strengthening pillars to be constructed such that it does not compromise the strength of the bridge and at the same time does not waste resources.

All these examples are an insight into the most common real word application of derivatives.

For further readings:

STEWART, J. (2007) *Essential Calculus,* Belmont, CA, Thomson Brooks/Cole.

Exercises:

1. **Differentiate the following functions:**
   a. $\boldsymbol{f(x) = x^4}$
   b. $\boldsymbol{y = 5x^2 - 2\sqrt{x}}$
   c. $\boldsymbol{y = 2x^{-7} + \frac{3}{x^2}}$
   d. $\boldsymbol{f(x) = (4x^3 + 2)(1 - 3x)}$
   e. $\boldsymbol{h(x) = \frac{x^2-1}{x^3+4}}$
   f. $\boldsymbol{f(t) = \sqrt{t2 - 5t + 7}}$
   g. $\boldsymbol{f(x) = e^{x^2+x^3}}$
   h. $\boldsymbol{f(x) = \log_e(7x^{-2})}$
2. **Find the equation of the line tangent to the curve $\mathbf{y = \sqrt[3]{x}}$ when $\mathbf{x = 8}$.**
3. **If $\boldsymbol{r = \left(t + \frac{1}{t}\right)(t^2 - 2t + 1)}$, find $\boldsymbol{\frac{dr}{dt}}$ when $\boldsymbol{t}$ = 2.**
4. **Find the maximum and the minimum of the function $\boldsymbol{f(x) = x^4 - 2x^2}$ $\boldsymbol{for\ -1 \leq x \leq 2}$**

Answers:

1. The answers are:
   a. $\frac{dy}{dx} = 4x^3$
   b. $\frac{dy}{dx} = 10x - x^{\frac{1}{2}}$
   c. $\frac{dy}{dx} = -14x^{-8} - 6x^{-3}$
   d. $\frac{dy}{dx} = 12x^2(1-3x) - 3(4x^3+2)$
   e. $\frac{dy}{dx} = \frac{(x^3+4)2x-(x^2-1)3x^2}{(x^3+4)^2} = \frac{-x^4+3x^2+8x}{(x^3+4)^2}$
   f. $\frac{dy}{dt} = \frac{d}{dt}(t^2-5t+7)^{\frac{1}{2}} = \frac{1}{2}(t^2-5t+7)^{-\frac{1}{2}}(2t-5)$
   g. $\frac{dy}{dx} = (2x+3x^2)e^{x^2+x^3}$
   h. $Write\ f(x) = \log_e 7 - 2\log_e x \Rightarrow f'(x) = -\frac{2}{x}$

2. When $x = 8$ we have $y = \sqrt[3]{8} = 2$, so the point (8,2) is on the line. Now

$$\frac{dy}{dx} = \frac{1}{3}x^{-\frac{2}{3}}$$

$$and, \frac{dy}{dx} = \frac{1}{12}, \qquad when\ x = 8$$

Hence, the equation of the tangent is:

$$y = \frac{1}{12}x + b$$

Substituting, $x = 2\ and\ y = 8$ we get,

$$2 = \frac{1}{12}8 + b \Rightarrow b = \frac{4}{3}$$

The equation is therefore,

$$y = \frac{1}{12}x + \frac{4}{3}$$

3. The $f'(t)$ is

$$\frac{dr}{dt} = (1-t^{-2})(t^2-2t+1) + \left(t+\frac{1}{t}\right)(2t-2)$$

Substituting t=2, we get

$$\left(1-\frac{1}{4}\right)(4-4+1) + \left(2+\frac{1}{2}\right)(4-2) = \frac{23}{4}$$

4. Differentiating we get,

$$f'(x) = 4x^3 - 4x = 0$$

$$\Rightarrow 4x(x^2 - 1) = 0$$
$$\Rightarrow x = (-1,0,1)$$

Differentiating it again we get,

$$f''(x) = 12x^2 - 4$$
$$f''(-1) = 8 > 0, hence\ minimum$$
$$f''(0) = -4 > 0, hence\ maximum$$
$$f''(1) = 8 > 0, hence\ minimum$$

Chapter 7: Sequences and Series

Sequences

Consider the following numbers:

$$1,1,2,3,5,8,13, \ldots$$

The above numbers are at first glance just random numbers clubbed together. However, there is a method to the madness. Every number is a sum of the preceding two numbers (except the first number which is set as one, the first natural number). For example, the second number is the sum of the first number and the number preceding it (as there is no number preceding it we assume that number is 0), so the second number is

$$1 + 0 = 1$$

Similarly, the third number is

$$1 + 1 = 2$$

And so on.

This kind of a group of numbers which have a pattern associated with it is known as a sequence. Thus, *a sequence is an ordered list of numbers and every member of the sequence is called a term*. For example, in the above sequence, the first term is 1, the second term is 1, the third term is 2 and so on. The above sequence has a special term known as the Fibonacci Series (it is the wrong terminology because it's a sequence which is different from a series, but well we cannot really control the past!) Also, most of the time sequences do not have names (except special ones like the one above). Instead, we represent sequences by the mathematical relationship between them.

Consider the following sequence:

$$1,5,9,13,17,21, \ldots$$

If we represent b_1 as the first term, b_2 as the second term and so on, then we have

$$b_1 = 1; b_2 = 5; b_3 = 9; \ldots$$

Also, note that

$$b_2 - b_1 = 4; b_3 - b_2 = 4; b_4 - b_3 = 4$$

That is, the difference between the consecutive terms is 4.

This can be represented as:

$$b_n = 1 + (n - 1) \times 4$$

$$or, b_n = b_1 + (n - 1) \times 4;\ \ where\ b_1 = 1$$

The above formulation completely defines the series that we are considering.

Suppose you were asked to find the 40th term of the above sequence, then it would be:

$$b_{40} = 1 + (40 - 1) \times 4 = 157$$

Similarly, for the Fibonacci Series we would have (if we represent the terms by a):

$$a_n = a_{n-1} + a_{n-2}; a_1 = 1 \; and \; a_2 = 1$$

(Note that we cannot find the nth term in a Fibonacci Series as easily as we can for the previous one).

Also, note that the inputs of sequences are generally always Natural Numbers and sometimes Whole Numbers and Integers (will be specified in the context of the example).

Example:

Find the first, second and the tenth term of the following sequence:

$$q_n = 4 + (n^2 - 2n)$$

Given the above formulation, we have:

$$q_1 = 4 + (1^2 - 2 \times 1) = 4 + 1 - 2 = 3$$

$$q_2 = 4 + (2^2 - 2 \times 2) = 4 + 4 - 4 = 4$$

$$q_{10} = 4 + (10^2 - 2 \times 10) = 4 + 100 - 20 = 84$$

The above sequence is an example of one of those sequences that are very hard to identify if you did not have the formulation. Such sequences are very hard to tackle and are beyond the scope of this course. However, we will discuss two of the most common form of sequences – arithmetic and geometric sequences.

Arithmetic sequences

Arithmetic sequences are sequences where the next term can be generated by adding or subtracting a certain numeral from the previous one. For example, consider the following sequence of numbers:

$$2,4,6,8,10,12 \ldots$$

If we denote the terms by the letter 'd' then the formulation of the sequence would be:

$$d_n = d_{n-1} + 2$$

That is, if we add 2 to the last term then we get the next term. Another formulation would be:

$$d_n = 2 + (n - 1) \times 2$$

The above formulation is the way in which arithmetic sequences are represented.

The general form of an arithmetic sequence is:

$$a_n = a_1 + (n - 1) \times d$$

Where a_n is the nth term of the sequence; a_1 is the first term of the sequence and d is the common factor. For example, in the previous example we have $a_1 = 2\ and\ d = 2$.

Similarly, the second example we considered is an example of an arithmetic sequence.

Example:

Express the following in terms of formulation of arithmetic sequences (also, called the expression for the nth term of the sequence) and find the 26th term of the sequence:

$$1,7,13,19, \dots$$

Looking at the sequence we can see that the common difference between the terms is 6, i.e. $d = 6$.

The first term is 1, i.e. $a_1 = 1$

Hence the sequence can be represented as:

$$a_n = 1 + (n - 1)6$$

Now the 26th term of the sequence is:

$$a_{26} = 1 + (25)6 = 151$$

Geometric Sequences

Geometric sequences differ from arithmetic sequences in the sense that instead of adding or subtracting a factor to arrive at the next term, we multiply or divide a factor with the previous one to arrive at the next term.

Consider the following sequence:

$$2,4,8,16,32, \dots$$

We see that each term is a two times the previous one, hence, the next term would be 64.

If we were to represent the above in term of a formula for the nth term, then we would have something like:

$$b_n = 2 \times 2^{n-1}$$

Similarly, the formulation of the following series

$$2, 1, \frac{1}{2}, \frac{1}{4}, \dots$$

would be:

$$g_n = 2 \times \frac{1}{2}^{n-1}$$

That is, we multiply the first term by the common factor raised to the power of $(n - 1)$.

The general form of a geometric sequence is:

$$a_n = a_1 r^{n-1}$$

Where a_n is the nth term of the sequence; a_1 is the first term of the sequence and r is the common factor. For example, in the previous example we have $a_1 = 2 \; and \; r = \frac{1}{2}$.

Example:

Find the expression for the following sequence:

$$25, -5, 1, -0.2, \ldots.$$

Also, find the 5th and the 10th terms in the above sequence.

We see that the above sequence is a decreasing sequence and there is no common factor which we can subtract the previous term to get the next one. However, if we divide the first term by -5 then we get the second term, and if we divide the second term by -5 then we get the third term. Voila! We have found our factor. So we have

$$r = -\frac{1}{5} \;\; and \; a_1 = 25$$

Hence,

$$a_n = 25 \times \left(-\frac{1}{5}\right)^{n-1}$$

So the 5th term is

$$a_5 = 25 \times \left(-\frac{1}{5}\right)^4 = 0.04$$

And the 10th term is

$$a_{10} = 25 \times \left(-\frac{1}{5}\right)^9 = -0.0000128$$

Note: As mentioned earlier, arithmetic and geometric sequences are only two forms of sequences and cannot be used to express all the different forms of sequences (For example: the Fibonacci Series cannot be expressed in terms of either geometric or arithmetic sequences). We will now move on to the concept of a series.

Series

A series differs from a sequence in the sense that while sequences give you the terms of a particular pattern, series actually give you the sum. So, if you replace the commas in all the above examples and replace them with an addition sign, we get a series from the sequences.

For example:

$$1,5,9,13,17,21,\ldots$$

is the expression of a sequence, while

$$1+5+9+13+17+21+\cdots$$

is the expression of a series.

A sequence generally goes on to infinity, however a series may have a defined number of terms at times, and sometimes will be to infinity.

Note that we had the following expression for the n^th^ term of the above sequence:

$$b_n = 1 + (n-1) \times 4$$

Now the sum of the first 10 terms would be a series and would be expressed as:

$$\sum_{i=1}^{10} 1 + (i-1) \times 4$$

The sigma/summation symbol Σ means that we add up all the term for the given number of "i's" by inserting that particular value of i for each subsequent term. The top number of the summation tells us when to stop, and if that number is infinity, then we keep on going forever!

If we are to expand the above term then we get:

$$\sum_{i=1}^{10} 1 + (i-1) \times 4 = [1+0\times 4] + [1+1\times 4] + [1+2\times 4] + \cdots + [1+9\times 4]$$

$$= 1+5+9+13+17+21+25+29+33+37 = 190$$

Arithmetic Series

The above is an example of an arithmetic series. An arithmetic series is basically the sum of the terms of an arithmetic sequence. Given the expression of the nth term of an arithmetic sequence:

$$a_n = a + (n-1)d$$

The sum of an arithmetic series can be represented by the following formula:

$$S_n = n\frac{a_1 + a_n}{2}$$

Where S_n is the sum of the first n terms; a_1 is the first term; a_n is the n^{th} term; and n is the total number of terms.

For example, in the above series, we know that our first term is 1 and our 10^{th} term is 37. The sum of the first 10 terms is:

$$S_{10} = 10 \times \frac{1 + 37}{2} = 10 \times \frac{38}{2} = 10 \times 19 = 190$$

Example:

Using the formula for arithmetic series find the sum of the first 20 terms of the following series:

$$12, 9, 6, 3, 0, \ldots$$

We see that the common factor in the above series is -3 (because the series is decreasing it has to be a negative factor). So our 20^{th} term is:

$$a_{20} = 12 + (-3)(19) = -37$$

Hence, the sum of the first 20 terms is:

$$S_{20} = 20 \times \frac{12 - 37}{2} = 10 \times -25 = -250$$

Geometric Series

A geometric series again is similar to arithmetic series. Here as well we add up the terms of the geometric sequences to arrive at the geometric series.

The following represents a geometric series:

$$10 + 5 + 2.5 + 1.25 + 0.625 + \cdots$$

The common factor in the above sequence is ½. So, the expression of the n^{th} term of the above sequence is

$$a_n = 10 \times \left(\frac{1}{2}\right)^{n-1}$$

Now, if we wanted the sum of the first 10 terms of the above sequence, then the sum would be:

$$\sum_{i=1}^{10} 10 \times \left(\frac{1}{2}\right)^{i-1} = 10 + \left[10 \times \frac{1}{2}\right] + \left[10 + \left(\frac{1}{2}\right)^2\right] + \cdots$$

$$= 10 + 5 + 2.5 + 1.25 + 0.625 + 0.3125 + 0.15625 + 0.078125 + 0.0390625 + 0.01953125 = 19.98046875$$

This is tedious and time consuming. What happens if you were asked to calculate the sum of all the terms of the above sequence?

So, there is actually a very simple formula to calculate the sum of the n terms of a geometrics series:

Given the expression for the nth term of a geometric sequence as

$$a_n = ar^{n-1}$$

The sum of n terms is given by

$$S_n = \frac{a(1-r^n)}{1-r};\ \ when\ |r| < 1$$

$$S_n = \frac{a(r^n-1)}{r-1};\ \ when\ |r| > 1$$

Also, the formula to calculate the sum of an infinite series is:

$$S_\infty = \frac{a}{1-r}, \qquad |r| < 1$$

At this point we will introduce the concept of a Convergence and Divergence:

Consider the following two series:

1. $10 + 5 + 2.5 + 1.25 + 0.625 + \cdots$
2. $1 + 5 + 9 + 13 + 17 + 21 + \cdots$

The first series keeps on decreasing and goes towards zero as the $n \to \infty$. That is, as n increases the n^{th} term is smaller than the before and closer to zero. Such series where the terms approach a particular number as $n \to \infty$ are called converging series.

The second series is an example of a diverging series. Here the terms get closer and closer to $\infty\ as\ n \to \infty$. Any such series where the nth term gets closer to $\pm\infty\ as\ n \to \infty$ is called a diverging series.

Now, the sum of a diverging series for an infinite number of terms is always infinite.

However, the sum of a converging series for an infinite number of terms always converges to a particular number. The condition for a geometric series to be converging is

$$|r| < 1$$

That is, the absolute value of a common factor is always less than 1. This is the reason why we only have one formulation for infinite terms.

Also, an arithmetic series is never converging, it is always diverging! Can you explain why?

Example:

Find the sum of the first 20 terms of the following series. Also, find S_∞ if possible.

$$1 + \left(-\frac{1}{3}\right) + \frac{1}{9} + \cdots$$

The above series is clearly a geometric series. The common factor of the sequence is $-\frac{1}{3}$ (we can find this simply from the first two terms). Hence the sum of the first 20 terms is:

$$S_{20} = \frac{1 \times \left[1 - \left(-\frac{1}{3}\right)^{20}\right]}{1 - \frac{1}{3}} = 1.5$$

We have, $|r| = \frac{1}{3} < 1$, hence the condition of convergence is satisfied.

Hence,

$$S_{\infty} = \frac{1}{1 - \left(-\frac{1}{3}\right)} = \frac{1}{\frac{4}{3}} = \frac{3}{4} = 0.75$$

Applications of Series and Sequences

Series and sequences are used very commonly in finance and economics. One of the most common areas where they are employed is in relation to interest rates.

Interest is defined as the charge for the use of money, and the rate at which it is charged is called the interest rate.

Generally there are two common types of interest rates:

1. *Simple Interest*
2. *Compound Interest*

Simple interest is calculated only on the principal (or initial) amount borrowed. For example, if you borrow $1000 at 5% interest rate, then the interest for the first period is

$$I = 0.05 \times 1000 = \$50$$

Similarly, for the second and all consecutive periods, the interest would be the same.

Simple interest is calculated by:

$$SI = P \times nr$$

Where, SI is the total interest, P is the principal, n is the number of time periods, and r is the interest rate.

Example:

Calculate the simple interest accrued over 10 years for a loan of $5000 at 10% interest rate.

Applying the above formula we have:

$$SI = 5000 \times 10 \times 0.10 = \$5000!$$

Note that simple interest is an application of an arithmetic series.

Compound interest is the concept of interest on interest. Here, you receive/pay interest on not only the principal amount, but also on all the accumulated interest as well.

Compound interest is calculated as:

$$CI = P(1 + r)^t - P$$

where, P=principal or present value, r=nominal interest rate, and t=time of investment.

The formula given above is extremely important in financial literature as it can be expressed slightly differently to give us the present value of one or many future payments.

$$FV = PV(1 + r)^t$$

$$or, \qquad PV = \frac{FV}{(1 + r)^t}$$

where FV=future value and PV=Present value.

For example, given r=8%, how much money do you need now to make a payment of $5000 in 5 years' time.

We have,

$$PV = \frac{FV}{(1+r)^t} = \frac{5000}{(1+0.08)^5} = \$3402.92$$

Similarly, if you were to receive payments of $1000, $3000, $2000, $1000 at the end of each of the next four years, find the present value of those payments.

$$PV = \frac{1000}{(1+0.08)^1} + \frac{3000}{(1+0.08)^2} + \frac{2000}{(1+0.08)^3} + \frac{1000}{(1+0.08)^4} = \$5820.64$$

For further readings:

EPP, S. S. (2004) *Discrete Mathematics with Applications,* Belmont, CA, Thomson-Brooks/Cole.

BRAILSFORD, T., HEANEY, R. & BILSON, C. (2006) *Investments: Concepts and Applications,* South Melbourne, Vic., Thomson.

Exercises:

1. **Find the first five terms and the 15^{th} term of the arithmetic sequence**

$$a_n = 3 + (n - 1)(\frac{1}{2})$$

2. **Find an explicit formula for the following sequence: $1, 3, 5, 7, 9, \ldots$**
3. **Calculate $\sum_{k=4}^{7} 5(k + 2)$.**
4. **Define the arithmetic sequence $a_1 = 6$; $a_{n+1} = a_n - 2$ explicitly.**
5. **What is the sum of the first 45 terms in the arithmetic sequence $a_1 = 3$; $a_n + 1 = a_n + 5$?**
6. **What is the common ratio *r* in the geometric sequence $.5, 1.5, 4.5, 13.5, 40.5, \ldots$?**
7. **What is the limit of the infinite series $2 + 1 + \frac{1}{2} + \frac{1}{4} + \ldots$?**
8. **Find the present values of \$1000 arising at the end of each of the next three years and \$2000 at the end of each year from year 4 to 7. Interest rate is 10% effective.**

Answers:

1. $a_n = 3 + (n-1)(\frac{1}{2})$

$$a_1 = 3 + (1-1)\left(\frac{1}{2}\right) = 3$$

$$a_2 = 3 + (2-1)\left(\frac{1}{2}\right) = 3.5$$

$$a_3 = 3 + (3-1)\left(\frac{1}{2}\right) = 4$$

$Similarly, a_4 = 4.5 \; and \; a_5 = 5$

$$a_{15} = 3 + (15-1)\left(\frac{1}{2}\right) = 10$$

2. $a_n = 2n - 1$

3. $\sum_{k=4}^{7} 5(k+2) = 5 \sum_{k=4}^{7}(k+2) = 5(6+7+8+9) = 150$

4. $a_n = -2n + 8$

5. We have, $a_n = 5n - 2$

$$\therefore S_{45} = \sum_{n=1}^{45} 5n - 2 = 5 \times \frac{45 \times 46}{2} - 45 \times 2 = 5085$$

6. r = 3

7. $\sum a_1 r^{n-1} = \frac{a_1}{1-r} = \frac{2}{\frac{1}{2}} = 4.$

8. $PV = \frac{1000}{1.10} + \frac{1000}{1.1^2} + \frac{1000}{1.1^3} + 2000\left(\frac{1}{1.1^4} + \frac{1}{1.1^5} + \frac{1}{1.1^6} + \frac{1}{1.1^7}\right) = \$7249.99$

Chapter 8: Linear Algebra

Vectors

A vector is an ordered list of numbers generally in the $\mathbb{R}^n$Euclidian space. It is similar in concept to sets but different in the sense that a vector is an ordered array representing some kind of mathematical relationship (and cannot be an abstract relationship). A vector can be represented as

$$\boldsymbol{v} = (v_1\, v_2\, \ldots\, v_n) = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix}$$

A vector is always represented by letters in **bold**. The first set is known as a row vector and the second is known as a column vector. A column vector can become a row vector if we transpose it and the transpose of a vector $\boldsymbol{v}$ is written as $\boldsymbol{v^T}$**.**

The rules of vector operations are:

1. Addition: Two vectors can be added together if they are of the same length (i.e., the number of elements in both the vectors is the same).
$$\boldsymbol{u} + \boldsymbol{v} = (u_1 + v_1\, u_2 + v_2 \ldots u_n + v_n)$$
2. Scalar Multiplication: Multiplying a vector by a scalar quantity (basically a numeral) is the same as multiplying every element of the vector by the same scalar.
$$c\boldsymbol{v} = (cv_1\, cv_2\, \ldots\, cv_n)$$

3. Vector Product or Dot Product: Two vectors can be multiplied together if they are of the same length
$$\boldsymbol{u} \cdot \boldsymbol{v} = u_1v_1 + u_2v_2 + \cdots u_nv_n = \sum_{1=1}^{n} u_iv_i$$

Example:

Let

$$\boldsymbol{a} = (1\ 4\ 7\ 10\ )$$

$$\boldsymbol{b} = (2\ 5\ 9\ 3)$$

Find:

a) $3\boldsymbol{a} + 4\boldsymbol{b}$
b) $3\boldsymbol{a}.4\boldsymbol{b}$

$$\boldsymbol{3a} = (3\ 12\ 21\ 30\ )$$

$$\boldsymbol{4b} = (8\ 20\ 36\ 12)$$

Hence,

a) $3\boldsymbol{a} + 4\boldsymbol{b} = (11\ 32\ 57\ 42)$
b) $3\boldsymbol{a}.4\boldsymbol{b} = 3.8 + 12.20 + 21.36 + 30.12 = 1380$

Matrices

A matrix is a set of real numbers arranged in an ordered set of columns and rows. If the number of rows is m and the number of columns is n, then the matrix has $m \times n$ elements. For example:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & \cdots & a_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} & \cdots & a_{mn} \end{pmatrix}$$

$\boldsymbol{A}$ is a matrix with m rows and n columns, i.e. an $m \times n$ matrix.

Another way of looking at a matrix is thinking of it as a combination of vectors. The above matrix $\boldsymbol{A}$ may be thought of as a combination of either n column vectors (with each vector having m elements) or m row vectors (with each vector consisting of n elements).

Note: this shows that a vector can be considered as a matrix with only one row or one column:

$$\boldsymbol{v} = (v_1\, v_2\, \dots\, v_n) = \begin{pmatrix} v_1 \\ v_2 \\ \cdot \\ \cdot \\ v_n \end{pmatrix}$$

Matrices follow the following rules:

a) **Matrix Addition**: Two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ can be added together as long as they have the same number of rows and columns. In order to add the matrices we simply add each element of matrix $\boldsymbol{A}$ to the corresponding element in matrix $\boldsymbol{B}$ to get the resulting matrix.

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} a_{11} + b_{11} & \cdots & a_{1n} + b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} + b_{m1} & \cdots & a_{mn} + b_{mn} \end{pmatrix}$$

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} 1 & 2 \\ 2 & 1 \end{pmatrix}$$

$$\boldsymbol{A} + \boldsymbol{B} = \begin{pmatrix} 2 & 4 \\ 5 & 5 \end{pmatrix}$$

b) **Matrix Subtraction**: Matrix subtraction is the same as matrix addition and in order to perform this operation they must be conformable for subtraction, i.e. they must have the same numbers of rows and columns.

$$\boldsymbol{A} - \boldsymbol{B} = \begin{pmatrix} a_{11} - b_{11} & \cdots & a_{1n} - b_{1n} \\ \vdots & \ddots & \vdots \\ a_{m1} - b_{m1} & \cdots & a_{mn} - b_{mn} \end{pmatrix}$$

c) **Scalar Multiplication**: Multiplication of a matrix by a scalar quantity s is simply multiplying every element of the matrix by that scalar quantity. (Same as vectors).

$$\boldsymbol{sA} = \begin{pmatrix} sa_{11} & \cdots & sa_{1n} \\ \vdots & \ddots & \vdots \\ sa_{m1} & \cdots & sa_{mn} \end{pmatrix}$$

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ 3 & 4 \end{pmatrix} \quad \boldsymbol{s} = 2$$
$$\boldsymbol{sA} = \begin{pmatrix} 2 & 4 \\ 6 & 8 \end{pmatrix}$$

d) **Matrix Multiplication**: Multiplication of two matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ with each other must conform to the following rule: the number of columns in $\boldsymbol{A}$ must be equal to the number of rows in $\boldsymbol{B}$. If A is a $p \times q$ matrix and $\boldsymbol{B}$ is a $q \times r$ matrix then the resulting matrix $\boldsymbol{C}$ has $p \times r$ number of elements. Every element of $\boldsymbol{C}$ is calculated as:

$$c_{ij} = a_{i1} \times b_{1j} + a_{i2} \times b_{2j} + \ldots + a_{ik} \times b_{kj}$$

Example:

$$\begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \times \begin{pmatrix} A & B \\ C & D \end{pmatrix} = \begin{pmatrix} aA + bC & aB + bD \\ cA + dC & cB + dD \\ eA + fC & eB + fD \end{pmatrix}$$

Note: Every element of the resulting matrix is similar to the dot product of two vectors.

e) **Transpose of a Matrix**: Similar to vectors, transpose of a matrix refers to replacing the columns by rows, and replacing rows by columns. For example, if $\boldsymbol{A}$ is a $m \times n$ matrix, then $\boldsymbol{A^T}$ is a $n \times m$ matrix.

Example:

$$\boldsymbol{A} = \begin{pmatrix} a & b \\ c & d \\ e & f \end{pmatrix} \; then,\; \boldsymbol{A^T} = \begin{pmatrix} a & c & d \\ b & e & f \end{pmatrix}$$

Note that Transpose Matrices have the following properties:

i) $(\boldsymbol{A} + \boldsymbol{B})^{\boldsymbol{T}} = \boldsymbol{A^T} + \boldsymbol{B^T}$
ii) $(\boldsymbol{A^T})^{\boldsymbol{T}} = \boldsymbol{A}$
iii) $(\boldsymbol{sA})^{\boldsymbol{T}} = \boldsymbol{sA^T}$
iv) $(\boldsymbol{AB})^{\boldsymbol{T}} = \boldsymbol{B^T A^T}$

Similar to Set Theory, matrices have the following laws/properties:

a) *Associative Property of Matrices:*

$$(\boldsymbol{A} + \boldsymbol{B}) + \boldsymbol{C} = \boldsymbol{A} + (\boldsymbol{B} + \boldsymbol{C})$$

$$(\boldsymbol{A}\,\boldsymbol{B})\,\boldsymbol{C} = \boldsymbol{A}\,(\boldsymbol{B}\,\boldsymbol{C})$$

b) *Commutative Property of Matrices:*

$$\boldsymbol{A} + \boldsymbol{B} = \boldsymbol{B} + \boldsymbol{A}$$

But

$$\boldsymbol{AB} \neq \boldsymbol{BA}$$

This is because $\boldsymbol{AB}$ may be conformable for multiplication but $\boldsymbol{BA}$ may not be. Even if both of them were conformable, the resulting matrix would not necessarily be the same.

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

$$\boldsymbol{AB} = \begin{pmatrix} 2 & 3 \\ -2 & 2 \end{pmatrix} \quad \boldsymbol{BA} = \begin{pmatrix} 1 & 7 \\ -1 & 3 \end{pmatrix}$$

c) *Distributive Property of Matrices:*

$$\boldsymbol{A}(\boldsymbol{B} + \boldsymbol{C}) = \boldsymbol{AB} + \boldsymbol{AC}$$

$$(\boldsymbol{A} + \boldsymbol{B})\boldsymbol{C} = \boldsymbol{AC} + \boldsymbol{BC}$$

There are certain special kinds of matrices known as square matrices. These matrices are special in the sense that they have the same number of rows and columns. For example if a square matrix has n number of rows, then the dimensions of the matrix are $n \times n$ and is called a *matrix of the order* n.

Now, if a square matrix has all the elements as zero, except on the right diagonal, then such a matrix is called a diagonal matrix.

Example:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 0 \\ 0 & 3 \end{pmatrix}$$

A diagonal matrix that has all the elements as 1 on the right diagonal is called an identity matrix. We denote an identity matrix as $\boldsymbol{I_n}$ where n denotes the order of the matrix.

Examples:

$$\boldsymbol{I}_2 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}$$

$$I_3 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

Inverse of a matrix: The inverse of a square matrix $\boldsymbol{A}$ is such a matrix that satisfies the following rule:

$$\boldsymbol{A}\,\boldsymbol{A}^{-1} = \boldsymbol{A}^{-1}\boldsymbol{A} = \boldsymbol{I}_n$$

where, $\boldsymbol{A}^{-1}$is the inverse of $\boldsymbol{A}$**.** Note**,** the above formulation shows that $\boldsymbol{A}$ is a $n \times n$ matrix.

Example: Let

$$\boldsymbol{A} = \begin{pmatrix} 1 & 2 \\ -1 & 3 \end{pmatrix} \quad \boldsymbol{B} = \begin{pmatrix} 2 & 1 \\ 0 & 1 \end{pmatrix}$$

Since,

$$\mathbf{AB = BA = I_n}$$

We conclude that **B** is the inverse of **A**.

Procedure to Find $\boldsymbol{A}^{-1}$:

Let us take the following matrix:

$$\boldsymbol{A} = \begin{pmatrix} 1 & 3 & 3 \\ 1 & 4 & 3 \\ 1 & 3 & 4 \end{pmatrix}$$

In order to find the inverse, write down the entries of $\boldsymbol{A}$, but in a double wide matrix with an identity matrix in the other space:

$$\begin{pmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 1 & 4 & 3 & 0 & 1 & 0 \\ 1 & 3 & 4 & 0 & 0 & 1 \end{pmatrix}$$

Now we need to do matrix row operations to convert the left hand matrix into an identity matrix.

Subtracting R_1 from R_2 and R_1 from R_3 we get,

$$\begin{pmatrix} 1 & 3 & 3 & 1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$$

Now subtracting $3R_2$ from R_1 we get,

$$\begin{pmatrix} 1 & 0 & 3 & 1 & -3 & 0 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$$

Now subtracting $3R_3$ from R_1 we get,

$$\begin{pmatrix} 1 & 0 & 0 & 7 & -3 & -3 \\ 0 & 1 & 0 & -1 & 1 & 0 \\ 0 & 0 & 1 & -1 & 0 & 1 \end{pmatrix}$$

Thus, we have converted the left hand matrix into an identity matrix and the resulting matrix on the right side is our inverse matrix. Thus,

$$\boldsymbol{A^{-1}} = \begin{pmatrix} 7 & -3 & -3 \\ -1 & 1 & 0 \\ -1 & 0 & 1 \end{pmatrix}$$

Note: If a matrix $\boldsymbol{A}$ has an inverse (i.e. the inverse exists) then matrix $\boldsymbol{A}$ is called non-singular and invertible.

Determinants

Determinants are used to define whether a matrix is invertible and non-singular. In order to be non-singular and invertible, the determinant of that matrix must be non-zero.

For a 2×2 matrix:

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}$$

We define the determinant as:

$$\begin{vmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{vmatrix} = a_{11}a_{22} - a_{12}a_{21}$$

For a 3×3 matrix A as

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{12} & a_{13} \\ a_{21} & a_{22} & a_{23} \\ a_{31} & a_{32} & a_{33} \end{pmatrix}$$

The determinant of **A** is

$$\det(\boldsymbol{A}) = a_{11}(a_{22}a_{33} - a_{23}a_{32}) - a_{12}(a_{21}a_{33} - a_{23}a_{31}) + a_{11}(a_{21}a_{32} - a_{22}a_{31})$$

Example: Find the determinant of the following matrix:

$$\begin{pmatrix} 1 & 1 & 1 \\ 0 & 2 & 3 \\ 5 & 5 & 1 \end{pmatrix}$$

Using the above formula we have:

$$det = 1(2 \times 1 - 3 \times 5) - 1(0 \times 1 - 3 \times 5) + 1(0 \times 5 - 2 \times 5) = -8$$

Since the value ≠0, the inverse of that matrix also exists.

Linear Equations

A linear equation is generally of the following form:

$$a_1x_1 + a_2x_2 + \cdots + a_nx_n = b$$

a_i are the parameters or coefficients and x_i are the unknowns or variables.

We are interested in solving a system of linear equations (also known as simultaneous equations) like:

$$2x + 3y = 9$$

$$3x - y = 2$$

Generally we have a system of n equations with m variables like:

$$a_{11}x_1 + a_{21}x_2 + \cdots + a_{n1}x_n = b_1$$

$$a_{12}x_1 + a_{22}x_2 + \cdots + a_{n2}x_n = b_2$$

$$a_{13}x_1 + a_{23}x_2 + \cdots + a_{n3}x_n = b_3$$

$$. \quad . \quad . \quad .$$

$$. \quad . \quad . \quad .$$

$$a_{1m}x_1 + a_{2m}x_2 + \cdots + a_{nm}x_n = b_m$$

Generally, a system of n equations with m unknowns is inconsistent when $m \neq n$ and has either no solutions or an infinite number of solutions. In most of the courses in the College of Business and Economics, you are not expected to solve such equations. From here on we will only discuss situations where $m = n$.

There are couple of methods to solve a system of linear equations:

- Using elementary equation operations
- Matrix methods

Elementary Equation Operations

Elementary equation operations refer to using certain rules which we can apply to equations such that the system of linear equations undergoes a transformation, while maintaining the equality of the equations. These rules/operations are the following:

1. Interchanging two equations,
2. Multiplying two sides of an equation by a constant, and
3. Adding equations to each other

Interchanging two equations: If we have the following system of equations

$$a_{11}x_1 + a_{21}x_2 = b_1$$
$$a_{12}x_1 + a_{22}x_2 = b_2$$

Then rewriting the equations by changing their positions does not change the system, i.e. we still have the same system of linear equations.

$$a_{12}x_1 + a_{22}x_2 = b_2$$
$$a_{11}x_1 + a_{21}x_2 = b_1$$

Multiplying two sides of an equation by a constant: If we multiply both sides of an equation by the same constant (i.e. multiply the entire equation by a constant), then the equation itself does not change and hence, the system of linear equation does not change either.

$$a_{11}x_1 + a_{21}x_2 = b_1$$
$$ca_{12}x_1 + ca_{22}x_2 = cb_2$$

The above two equations are the same (think of it as a similar analogy to fractions).

Adding equations to each other: If we have a system of equations like:

$$a_{11}x_1 + a_{21}x_2 = b_1$$
$$a_{12}x_1 + a_{22}x_2 = b_2$$

Then adding the two equations together is consistent with the system (and generates another equation in the system). The same law applies for subtraction.

$$(a_{11} + a_{12})x_1 + (a_{21} + a_{22})x_2 = b_1 + b_2$$

Example:

In order to solve a system of linear equation like:

$$x + 3y = 3 \ldots eqn. (1)$$
$$2x + y = 8 \ldots eqn. (2)$$

We first multiply equation 1 by 2, we get:

$$2x + 6y = 6 \ldots eqn. (3)$$

Now, if we subtract equation 2 from equation 3, we get,

$$5y = -2$$

$$or, \qquad y = -\frac{2}{5}$$

Thus we have a value of y that solves the two equations. In order to get a value of x, we substitute the value of y in any of the equations. Substituting the value of y in equation 1 we get,

$$x - \frac{6}{5} = 3$$

$$or, x = \frac{21}{5}$$

Thus, we have the values of x and y that solve the two equations.

Matrix Methods (Cramer's Rule):

A system of linear equations can be represented as

$$\boldsymbol{Ax} = \boldsymbol{b}$$

where, $\boldsymbol{A}$ is the matrix of coefficients, $\boldsymbol{x}$ is the vector of the variables, and $\boldsymbol{b}$ is a vector of the constants.

Example:

For a system of equations

$$a_{11}x_1 + a_{21}x_2 = b_1$$

$$a_{12}x_1 + a_{22}x_2 = b_2$$

We have

$$\boldsymbol{A} = \begin{pmatrix} a_{11} & a_{21} \\ a_{12} & a_{22} \end{pmatrix}; \boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix}; \boldsymbol{b} = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}$$

For us to be able to solve the above system using Cramer's Rule we first have to find whether the inverse exists or not, i.e. we have to show

$$\det(\boldsymbol{A}) \neq 0$$

To solve this system, we find $\boldsymbol{A_j}$ from $\boldsymbol{A}$ by replacing the j^{th} column of $\boldsymbol{A}$ by $\boldsymbol{b}$**.**

Example:

$$\boldsymbol{A_1} = \begin{pmatrix} b_1 & a_{12} & a_{13} \\ b_2 & a_{22} & a_{23} \\ b_3 & a_{32} & a_{33} \end{pmatrix}$$

Then the unique solution $\boldsymbol{x} = (x_1, x_2, \dots, x_n)$ to the $n \times n$ system $\boldsymbol{Ax} = \boldsymbol{b}$ is

$$x_j = \frac{\det(\boldsymbol{A_j})}{\det(\boldsymbol{A})}$$

Example:

Find the solution of the following system:

$$-2x + 5y = 1$$
$$x + 2y = 4$$

Solution:

We have

$$\boldsymbol{A} = \begin{pmatrix} -2 & 5 \\ 1 & 2 \end{pmatrix} \; and \; \boldsymbol{b} = \begin{pmatrix} 1 \\ 4 \end{pmatrix}$$

Thus,

$$\boldsymbol{A_1} = \begin{pmatrix} 1 & 5 \\ 4 & 2 \end{pmatrix} \; and \;\; \boldsymbol{A_2} = \begin{pmatrix} -2 & 1 \\ 1 & 4 \end{pmatrix}$$

And,

$$\det(\boldsymbol{A}) = -9, \qquad \det(\boldsymbol{A_1}) = -18, \qquad \det(\boldsymbol{A_2}) = -9,$$

Hence,

$$x = \frac{-18}{-9} = 2 \;\; and \;\; y = \frac{-9}{-9} = 1$$

For further readings:

LAY, D. C. (2006) *Linear Algebra and its Applications,* Boston, Pearson/Addison-Wesley.

Exercises:

1. **Let**

$$A = \begin{pmatrix} 3 & 0 \\ -1 & 2 \\ 1 & 1 \end{pmatrix};\ B = \begin{pmatrix} 1 & 5 & 2 \\ -1 & 1 & 0 \\ -4 & 1 & 3 \end{pmatrix};\ C = \begin{pmatrix} -3 & -1 \\ 2 & 1 \\ 4 & 3 \end{pmatrix};\ D = \begin{pmatrix} 4 & 2 \\ -1 & 0 \end{pmatrix}$$

Compute wherever possible:

$$A + B;\ A + C;\ AB;\ BA;\ CD;\ DC;\ D^2$$

2. **Solve the system:**

$$x + y + z = 2$$
$$2x + 3y - z = 8$$
$$x - y - z = -8$$

3. **If** $A = \begin{matrix} 4 & -3 \\ 1 & 0 \end{matrix}$**; use the fact** $A^2 = 4A - 3I_2$ **to prove by mathematical induction that**

$$A^n = \frac{3^n - 1}{2}A + \frac{3 - 3^n}{2}I_2 \ if\ n \geq 1$$

Answers:

1. The matrices which are defined are:

$$A + C; BA; CD; D^2.$$

$$A + C = \begin{pmatrix} 0 & -1 \\ 1 & 3 \\ 5 & 4 \end{pmatrix}$$

$$BA = \begin{pmatrix} 0 & 12 \\ -4 & 2 \\ -10 & 5 \end{pmatrix}$$

$$CD = \begin{pmatrix} -14 & 3 \\ 10 & -2 \\ 22 & -4 \end{pmatrix}$$

$$D^2 = \begin{pmatrix} -14 & 4 \\ 8 & -2 \end{pmatrix}$$

2. We use elementary equation operations to solve this (note elementary equation operations is a much easier method to solve linear equations with a few unknowns):

$Adding\ equation\ 1\ and\ 3\ we\ get;$

$$2x = -6$$

$$\Rightarrow x = -3\ (eqn4)$$

$So\ we\ get\ eqn1\ as$

$$y + z = 5\ (eqn5)$$

$And\ eqn\ 2\ as$

$$3y - z = 14\ (eqn\ 6)$$

$Adding\ eqn\ 5\ and\ 6, we\ get$

$$4y = 19\ \ or, y = \frac{19}{4}$$

$Substituting\ in\ eqn5\ we\ get,$

$$z = \frac{1}{4}$$

3. Let p_n be

$$A^n = \frac{3^n - 1}{2}A + \frac{3 - 3^n}{2}I_2$$

Then we have p_1 as:

$$A = \frac{3-1}{2}A + \frac{3-3}{2}I_2 = A$$

Which is true. Assuming that p_n holds then we have p_{n+1} as

$$A^{n+1} = A_n.A = \left\{\frac{3^n-1}{2}A + \frac{3-3^n}{2}I_2\right\} = \frac{3^n-1}{2}A^2 + \frac{3-3^n}{2}A$$

$$= \frac{3^n-1}{2}(4A-3I_2) + \frac{3-3^n}{2}A$$

$$= \frac{(3^n-1)4 + (3-3^n)}{2}A - \frac{(3^n-1)(-3)}{2}I_2$$

$$= \frac{(4.3^n-3^n)-1}{2}A - \frac{(3-3^{n+1})}{2}I_2$$

$$= \frac{3^{n+1}-1}{2}A - \frac{(3-3^{n+1})}{2}I_2$$

Which is true. Hence proved.

Chapter 9: Introduction to Statistics

When we use the word '*Statistic*' we mean data and how to represent them in a form that is easily understood by a layperson. In other words, we are talking about summarizing a data set, as well as interpreting the relationships between the different components of the data set (may be hidden or apparent).

Consider a situation where you might want to find a measure of the IQ of a group of students. You might record their IQ, height, weight, hair colour, sex, etc. You would have one observation for each of the fields (variables) for every student in the group. Such a collection of entries (observations) together combine to form a *data set*.

Statistics is used to make inferences called *statistical inferences* – the problem of determining the behaviour of a large population by studying the behaviour of a small sample from that population.

For example – if one were to study the pattern of stock market movements with economic cycles, the population would be all the economic cycles ever recorded (and it can be argued that the unobserved ones should be part of the population as well because they were there whether they were recorded or not) and all the stock market price movements ever recorded. However, if you were to test this proposition (known as hypothesis testing) you would take the economic cycle dates for a certain period (say 10 years) and corresponding stock market prices. The data that are being used to test the hypothesis is called the sample and the results that you get are used to make inferences about the entire population of economic cycles and stock market cycles.

Thus we can say that:

- <u>Population</u> is the whole of something, e.g., all female students of ANU, all people who play the piano, etc.
- <u>Sample</u> is a subset of the population and the set of individuals are drawn from the population, e.g., students in this class is a sample of all students in ANU.

If we have population we can get parameters – true values like the centre and spread of the population.

If we have a sample we can get statistics – these are values that estimate the parameters, e.g. sample centre and sample spread are used to estimate population centre and population spread (and we use inference to make that estimation).

Variables are generally of two kinds:

a) *Quantitative Variables*: Such variables have a numerical value assigned to them and these numbers have a numerical meaning (i.e. we can perform meaningful mathematical operations on them). Examples include daily rainfall in a particular area, heights of a set of individuals, etc. Quantitative variables can be either *discrete* (belongs to $\mathbb{Z}$) or *continuous*(belongs to $\mathbb{R}$) in nature.

b) *Qualitative (or categorical) Variables*: Such variables define the individual entries in a data set to different groups. These groups are made depending on certain inherent characteristics of the observations. For example, we might want to group or observations with respect to the sex of the observations, or with respect to geographical locations.

Whenever we have a data set the first step we perform with the data set is to find the *Descriptive Statistics* of the data set. Descriptive Statistics refers to summarizing the data set to present certain salient features of each variable (generally quantitative) in the data set. These salient features help us understand the distribution of the variable, i.e. the shape or the pattern of the variable. This can be performed using both numerical and graphical methods.

Numerical Methods

The numerical methods employed in order to understand the distribution of a variable are those that define the centre and the spread of the variable. In situations where we have two variables we also include a measure of the general relationship between the two variables.

The centre of a distribution is generally measured in the following three ways:

1. *Mean*: This is the arithmetic average of the different observations in a sample. If the different observations are labelled as $x_1, x_2, \dots, x_n$ then the sample average is represented as $\bar{x}$ (the corresponding symbol for the population is μ). The mean is calculated as

$$\bar{x} = \frac{1}{n}(x_1 + x_2 + \dots + x_n)$$

$$or, \qquad \bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

2. *Median*: This refers to the middle point of the data set such that the data set is distributed equally on both sides of this value (i.e. it divides the data set in the middle). This differs from the mean in the sense that it does not depend on the actual values of the observations, but on the order of the observations.
   In order to find the median of a set of observations we first have to order the observations from lowest to highest. Then we count the total number of observations (n).
   - If n is odd, then the median is the $\frac{n+1}{2}$ th observation in this ordered list.
   - If n is even, then the median is the average of the $\frac{n}{2}$ th and $\frac{n+2}{2}$ th observations in the ordered list.
3. *Mode*: This is the observation which occurs with the maximum frequency in the data set.

The spread of a distribution is captured in three ways as well:

1. *Standard Deviation*: This measures the spread of the data with respect to the mean and captures the average distance of every observation from the mean. It is calculated as
$$s = \sqrt{\frac{\sum(x_i - \mu)^2}{n-1}} = \sqrt{\frac{\sum {x_i}^2 - n\mu^2}{n-1}}$$
The square of the standard deviation is called the variance (and is the one that is commonly calculated).
2. *Inter-Quartile Range*: This measures the spread of the data in a sense that is analogous to median. Median captures the middle of the data set and divides the data set at the point where the there is 50% of the distribution on either side. The median is also called the 50th percentile.

   In order to find any percentile we use the following formulae:
$$L_p = \frac{(n+1)P}{100}$$
   Where L_p is the location of the p^{th} percentile (i.e., which observation) and n is the number of data points.

   The 25th and the 75th percentiles are called the lower and upper quartiles of a distribution. The Inter-Quartile range is basically the difference between the two quartiles $(upper - lower)$.
3. *Range*: Range refers to the total spread of the data and is simply the difference between the maximum and minimum values of the observations.

Example:

Consider the following data: 5,7,1,2,4,4

$$mean = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{6}(5+7+1+2+4+4) = \frac{23}{6} = 3.83\dot{3}$$

In order to find the median we have to arrange the observation in an increasing order:

$$1,2,4,4,5,7$$

There are six observations, so the median is the average of the 3rd and the 4th observations.

$$median = \frac{4+4}{2} = 4$$

Mode is the observation that occurs the maximum number of times. In this case:

$$mode = 4$$

And,

$$sd = \sqrt{\frac{\sum x_i{}^2 - n\mu^2}{n-1}} = \sqrt{\frac{1}{5}[(25+49+1+4+16+16) - 6 \times 3.833^2]} = \sqrt{3.8055}$$
$$= 1.95$$

(Note: Variance (or standard deviation) of an investment, can be used as a measure of risk e.g. on profits/return. Larger variance ➔ larger risk. Usually, higher rate of return, higher risk

Example – 2 funds over 10 years

A	8.3	-6.2	20.9	-2.7	33.6	42.9	24.4	5.2	3.1	30.5
B	12.1	-2.8	6.4	12.2	27.8	25.3	18.2	10.7	-1.3	11.4

$$\bar{x}_A = 16\%, \qquad s_A^2 = 280.34(\%)^2$$
$$\bar{x}_B = 12\%, \qquad s_B^2 = 99.37(\%)^2$$

Fund A: higher risk, but also higher average rate of return)

Graphical methods

There are various graphical techniques that we employ in order to explain the distribution of a variable. Most of these techniques require the use of a statistical package and hence, we will not discuss the ways to actually perform them in this class (but feel free to come and see me if you need to understand the underlying techniques of constructing them). Remember the saying “a picture speaks a thousand words”. In accordance, graphical tools are very useful and important in statistics as they show us different patterns that we might not be able to capture through numerical methods alone.

(Note all the following graphs are developed using software called MINITAB, but if any other software were to be employed we would get similar results for the same data set).

When we consider categorical data the most common graphical tolls are *bar-charts* and *pie-charts.*

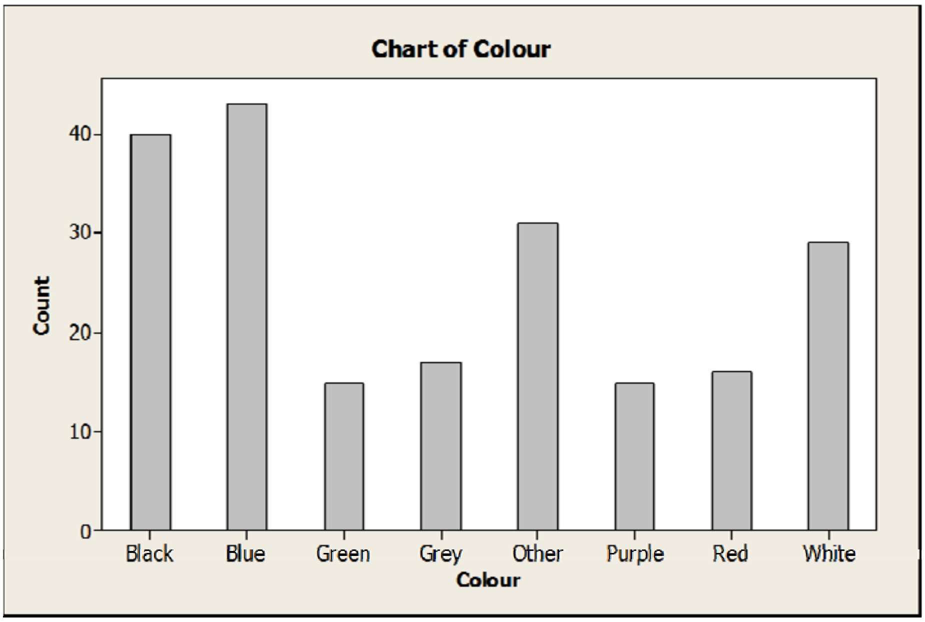
Example:

We are interested in the colour of the socks that students wear to ANU. We ask a group of students to tell us the colour of their socks and we record their responds.

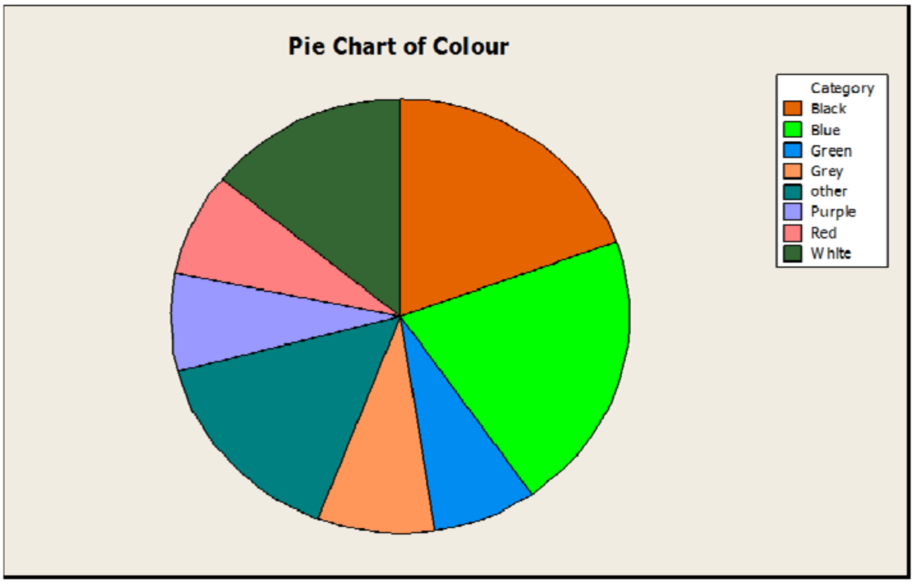
Finally we arrive at:

$$\begin{array}{cc} Colour & Count \\ Black & 40 \\ Blue & 43 \\ Green & 15 \\ Grey & 17 \\ Other & 31 \\ Purple & 15 \\ Red & 16 \\ White & 29 \\ N = & 206 \end{array}$$

The bar chart would look like



And the pie chart of the same data set would look like

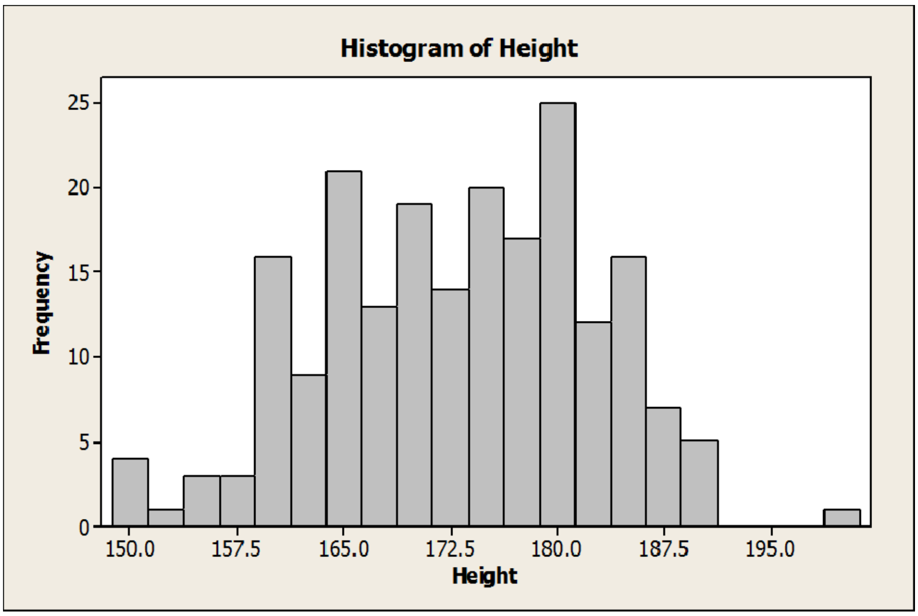


In case of continuous variables the common graphical tools employed are the *histogram*, *cumulative frequency distributions* (also called *Ogives* and are related to the concept of percentiles) and the *box-plot*.

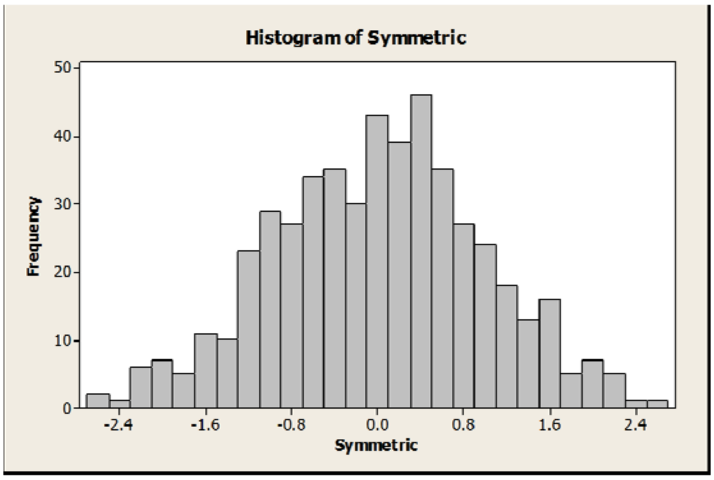
In the above example, we also asked the students their heights. The descriptive statistics of their heights are:

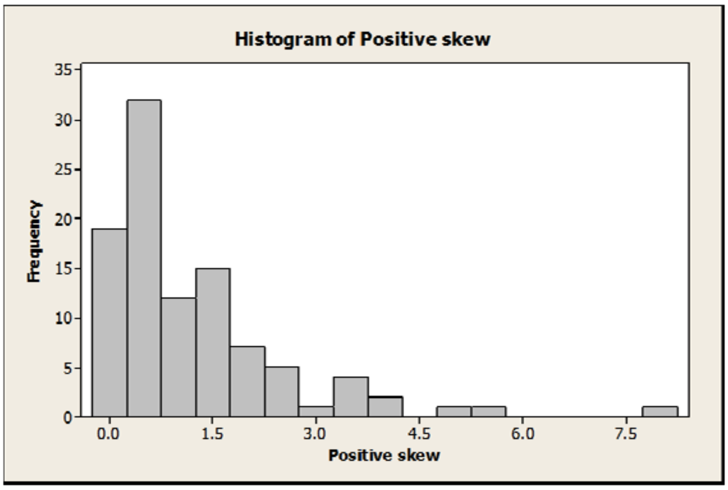
Variable	N	Mean	Median	StDev	Minimum	Maximum	Q1	Q3
Height	206	172.97	173.50	9.45	150.00	200.00	165.00	180.00

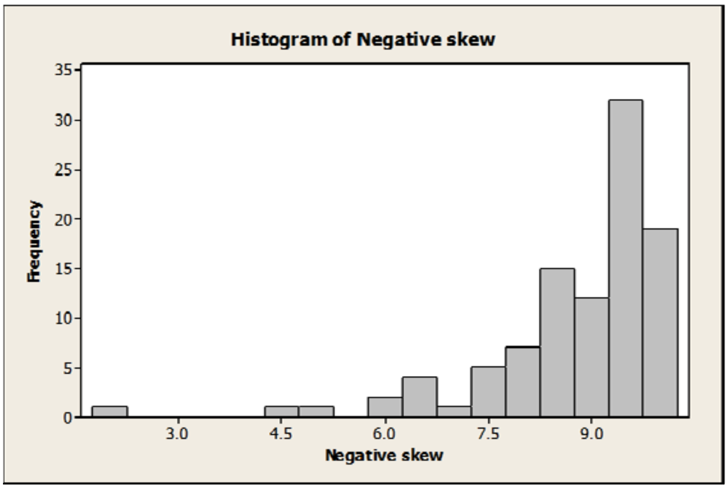
The histogram of the heights look like:

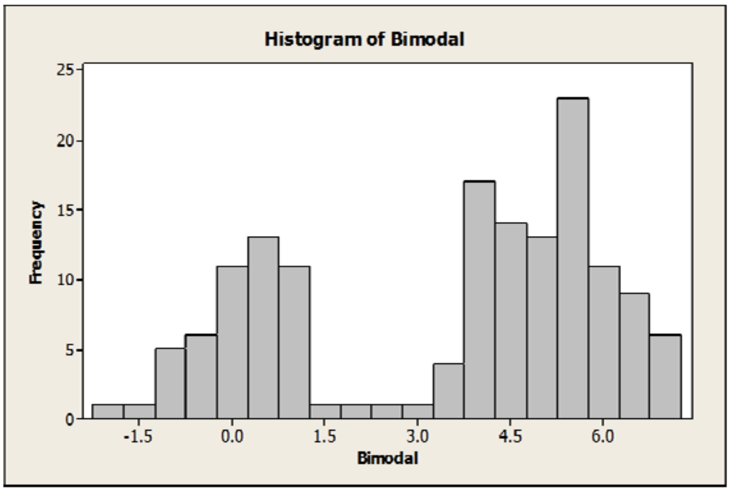


Histograms are the most commonly used graphical tool to visualise the shape of a distribution. A distribution can have the following shapes:

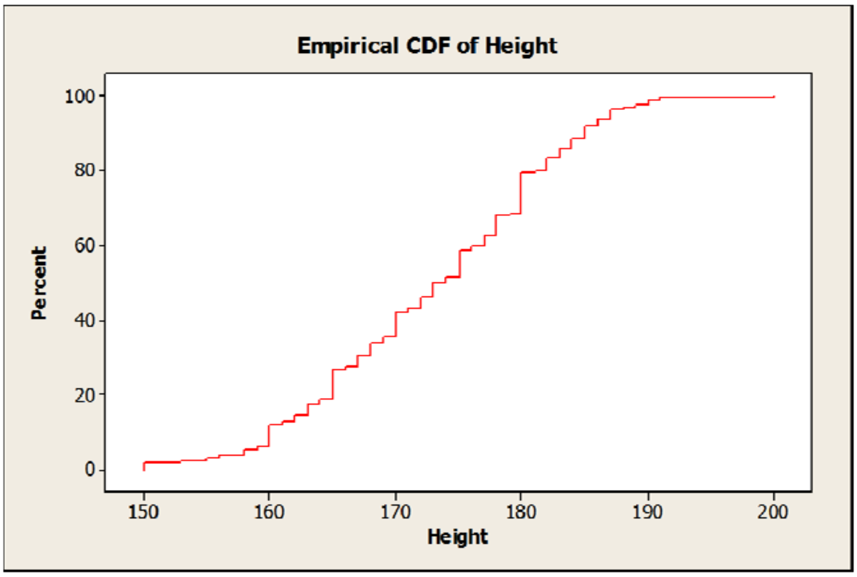




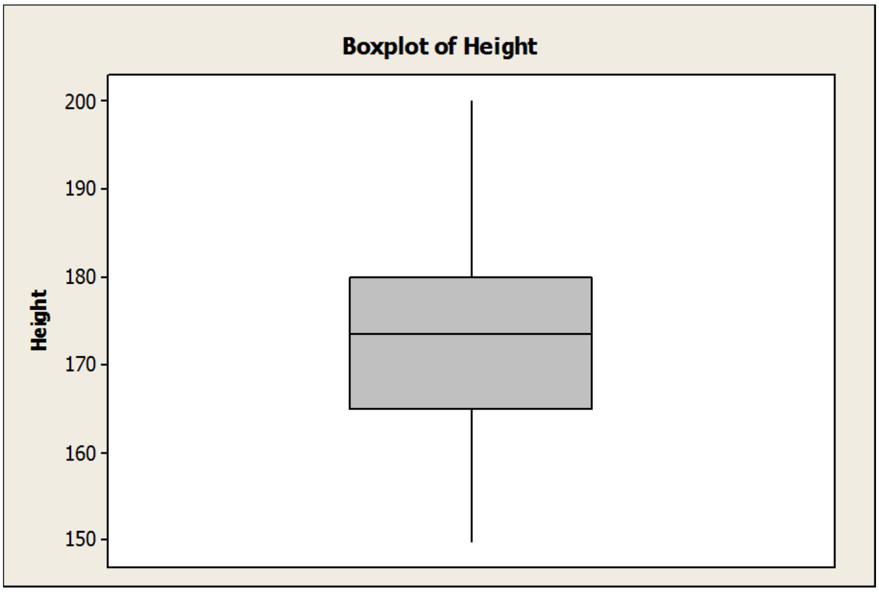




The Empirical CDF of the distribution looks like:



And finally the boxplot looks like:



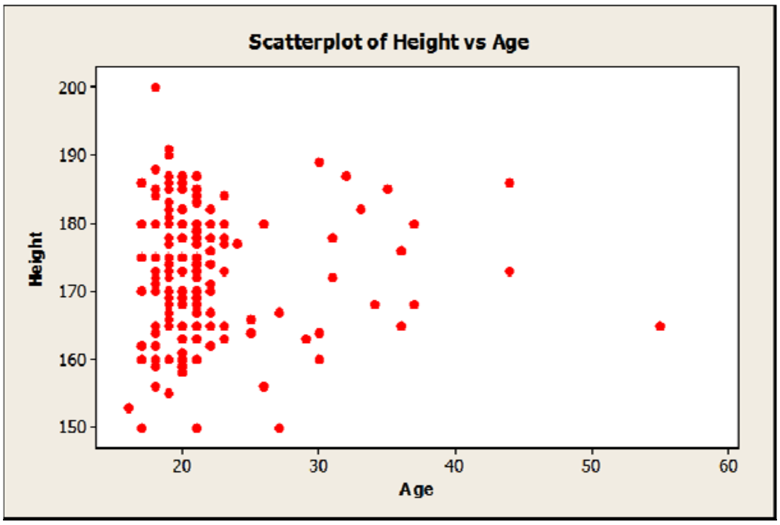
Until now we have only considered cases where we have one variable. In situations where we have more than one variable, like height and weight of individuals, we would typically perform any and all of the above operations for both the variables. In addition we would also find:

1. *Covariance* – Measures the linear relationship between *X* and *Y* – sign indicates direction of slope, but magnitude is dependent on units of measurement (so cannot indicate strength of relationship)
$$cov(X,Y) = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{n-1} = \frac{1}{n-1}\left[\sum X_iY_i - n\bar{X}\bar{Y}\right]$$
2. *Correlation (or Coefficient of Correlation)* – Measures strength of linear relationship between *X* and *Y* and is bounded between -1 and 1.

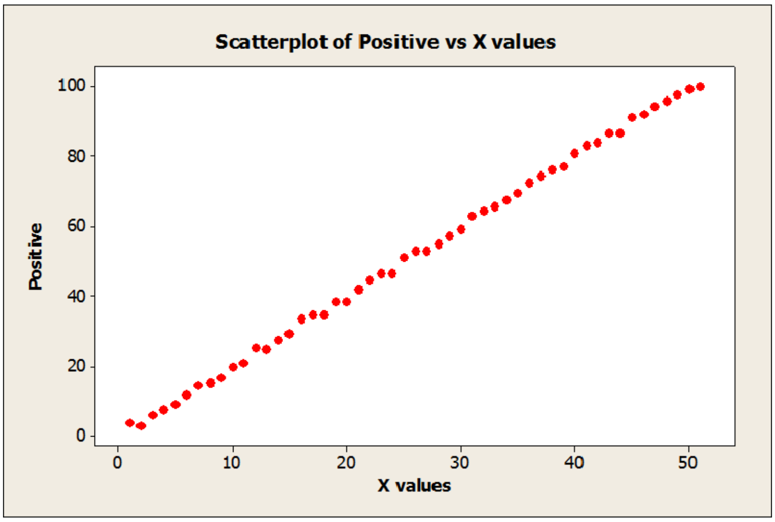
$$\rho = \frac{cov(X,Y)}{\sigma_X\sigma_Y} = \frac{cov(X,Y)}{s_Xs_Y}$$

- If $r = -1$, perfect negative linear relationship
- If $r = +1$, perfect positive linear relationship
- If $r = 0$, no linear relationship

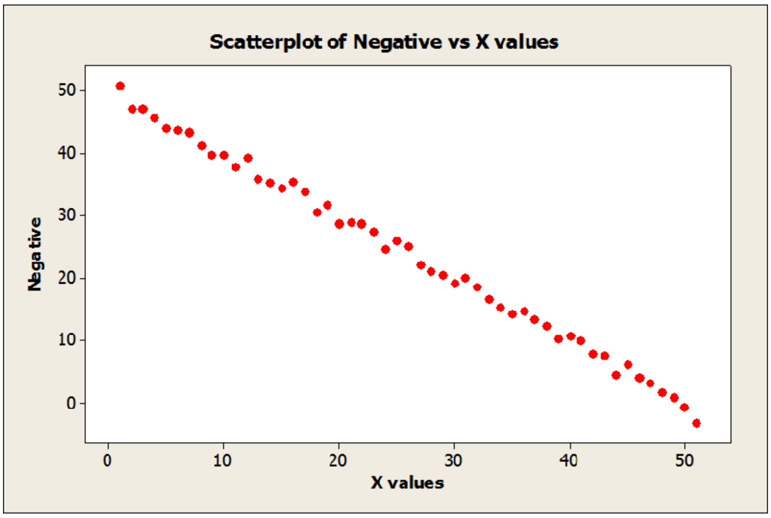
The best way to visualize the relation between two variables is via a scatter plot.



If $cov > 0$, then as X increases, Y increases; as X decreases, Y decreases (positive slope)



If $cov < 0$, then as X increases, Y decreases; as X decreases, Y increases (negative slope)



Chapter 10: Probability and Random Variables

A *random experiment* is a process that results in a number of possible outcomes, none of which can be predicted with certainty; e.g.

- Roll a die: outcomes: 1, 2, 3, 4, 5, 6.
- Toss a coin: outcomes: Heads, Tails
- Drive a car around a racing track: outcomes: {have an accident}, {don't have an accident}

The sample space of a random experiment is a list of all possible outcomes. Outcomes must be *mutually exclusive* and *collectively exhaustive*.

- No two outcomes can both occur on any one trial
- Each trial must result in one of the outcomes in the sample space

E.g. roll a die: sample space: $S = \{1, 2, 3, 4, 5, 6\}$.

An event is a collection of one or more simple (individual) outcomes or events.

E.g. roll a die; event A = odd number comes up. Then $A = \{1, 3, 5\}$.

In general, use sample space $S = \{E_1, E_2, \dots, E_n\}$ where there are *n* possible outcomes.

Probability of an event E_i occurring on a single trial is written as $P(E_i)$. The probability assigned to each simple event E_i must satisfy

1. $0 \le P(E_i) \le 1, for\ all\ i$
2. $\sum P(E_i) = 1$

$$Probability\ of\ an\ event\ = \frac{number\ of\ favourable\ outcomes}{total\ number\ of\ outcomes}$$

Note: for the sample space S, $P(S) = 1$

Complement of an event is the set of all outcomes not belonging to that event, e.g. if A = {1, 3, 5}, then the complement of A' = A^c = {2, 4, 6}

Consider two events, A and B.

- $P(A\ or\ B) = P(A\ U\ B) = P(A\ union\ with\ B) = P(A\ occurs, or\ B\ occurs, or\ both\ occur)$
- $P(A\ and\ B) = P(A \cap B) = P(A\ intersection\ with\ B) = P(A\ and\ B\ both\ occur)$
- $P(\bar{A}) = P(Ac) = P(A\ complement) = P(A\ does\ not\ occur)$
- $P(B|A) = P(B\ occurs\ given\ that\ A\ has\ occurred)$
- $P(A) = P(A \cap B) + P(A \cap Bc)$
- $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$

- $If\ (A\ and\ B) = \Phi\ (the\ null\ or\ empty\ set);\ i.e. A\ and\ B\ are\ mutually\ exclusive, then\ P(A\ and\ B) = 0.$

  E.g. if A is the event that we roll a 1, and B is the event that we roll a 2, then A and B together cannot happen – they are mutually exclusive.

$$P(A\ and\ B)\ =\ 0$$

- Conditional probability that A occurs, given that B has occurred:

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)}$$

- Two events are independent if

$$P(A|B) = P(A)\ or\ P(B|A) = P(B)$$

  Note: If A and B are independent, then

  $P(A\ and\ B)\quad = \ P(A) * P(B)$ $\boldsymbol{Note: only\ if\ independent}!$

  $Then\ P(A|B) = \ [P(A\ and\ B)]/P(B)$

  $= [P(A) * P(B)]\ /P(B)$

  $= P(A)$

Joint Probabilities:

Probabilities	Mutual Fund outperforms market	Mutual fund does not outperform market
Top-20 MBA program	0.11	0.29
Not top-20 MBA program	0.06	0.54

P(Mutual fund outperforms AND top-20 MBA)=0.11

P(Mutual fund outperforms AND not top-20)=0.06

P(Mutual fund not outperform AND top-20)=0.29

P(Mutual fund not outperform AND not top-20)=0.54

- Let events be known as follows:

  A_1=Fund manager graduated from a top-20 MBA program

  A_2=Fund manager graduated from a not top-20 MBA program

  B_1=Fund outperforms the market

  B_2=Fund does not outperform the market

Probabilities	B_1	B_2
A_1	0.11	0.29
A_2	0.06	0.54

Joint probabilities:

$P(B_1 \text{AND } A_1)=0.11$
$P(B_1 \text{ AND } A_2)=0.06$
$P(B_2 \text{ AND } A_1)=0.29$
$P(B_2 \text{ AND } A_2)=0.54$

Marginal probabilities:

Computed by adding across rows or down columns and are named so because they are calculated in the *margins* of the table

Probabilities	B_1	B_2	Totals
A_1	0.11	0.29	0.40
A_2	0.06	0.54	0.60
Totals	0.17	0.83	1.00

$P(A_1)=P(A_1 \text{ and } B_1)+P(A_1 \text{ and } B_2)=0.11+0.29=0.40$
$P(A_2)=P(A_2 \text{ and } B_1)+P(A_2 \text{ and } B_2)=0.06+0.54=0.60$
$P(B_1)=P(B_1 \text{ and } A_1)+P(B_1 \text{ and } A_2)=0.11+0.06=0.17$
$P(B_2)=P(B_2 \text{ and } A_1)+P(B_2 \text{ and } A_2)=0.29+0.54=0.83$

Complement Rule:

Given an event A and its complement, $\bar{A}$, so that $A + \bar{A} = S$;

$$Know\ that\ P(S) = 1;$$
$$so\ P(A) + P(\bar{A}) = 1;$$
$$\therefore\ P(\bar{A}) = 1 - P(A)$$

Interpretation: either A happens, or it doesn't

Multiplication Rule:

Since we know that:

$$P(A|B) = \frac{P(A\ and\ B)}{P(B)}\ and\ P(B|A) = \frac{P(A\ and\ B)}{P(A)}$$
$$P(A\ and\ B) = P(A|B) \times P(B)$$
$$= P(B|B) \times P(A)$$

Example: Coin Toss

An unbiased coin is tossed twice.

S={HH, HT, TH, TT}; all equally likely (prob=$^1/_4$)

Find the probability of at least 1 tail:

$Let\ event\ A\ =\ at\ least\ 1\ tail$

$A = \{HT, TH, TT\}$

$P(A) = 3/4$

$Alternatively: P(A) = 1 - P(\bar{A}) = 1 - P(no\ tails) = 1 - 1/4 = 3/4$

Example:

A store sells 2 brands of a particular product – one expensive and the other inexpensive. Let expensive brand be called "A", inexpensive brand be called "B". A survey of 1000 sales gives the following:

	Brand		
Gender	**A**	**B**	**Total**
Male	**132**	**147**	**279**
Female	**516**	**205**	**721**
Total	**648**	**352**	**1000**

- $P(customer\ is\ male) = 279/1000 = 0.279$

- $P(purchase\ brand\ A) = 0.648$
- $P(purchase\ brand\ A\ AND\ female) = 0.516$
- $P(purchase\ brand\ A\ GIVEN\ female)$

  $= P(purchase\ brand\ A\ AND\ female)/P(female)$

  $= [516/1000]/[721/1000]$

  $= \frac{516}{721}$
- $If\ independent, P(brand\ A\ GIVEN\ female) = P(brand\ A)$

  $LHS = 516/721$

  $RHS = 648/1000$

  $LHS \neq RHS$

  $Therefore\ gender\ and\ brand\ purchased\ are\ NOT\ independent!$

Imagine tossing three unbiased coins.

$$S = \{HHH, HHT, HTH, THH, HTT, THT, TTH, TTT\}$$

We have, 8 equally likely outcomes.

Let X = number of heads that occur. Then, X can take values 0, 1, 2, 3. Actual value of X depends on chance – call it a random variable (r.v.)

Definition: a random variable is a function that assigns a numeric value to each simple event in a sample space

- Denote random variables (X, Y, etc) in upper case
- Denote actual realised values (x, y etc) in lower case

Example: X is the random variable that can take values 0, 1, 2, 3. Actually perform experiment, find the pattern HTT. Then x=1.

There are two types of random variables:

- A discrete random variable has a countable number of possible values, e.g. number of heads, number of sales etc. Does not have to be finite; but values can be strictly ordered.
- A continuous random variable has an infinite number of possible values – number of elements in sample space is infinite as a result of continuous variation e.g. height, weight etc. Values cannot be strictly ordered, as a different instrument may result in new values between those already observed.

Discrete probability distributions

Definition: A table or formula listing all possible values that a discrete r.v. can take, together with the associated probabilities.

E.g. for our toss of three coins example:

x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8

If x is the value taken by a r.v. X, then p(x)=P(X=x)= sum of all the probabilities associated with the simple events for which X=x.

If a r.v. X can take values x_i, then

1. $0 \leq P(x_i) \leq 1, for\ all\ i$
2. $\sum P(x_i) = 1$

Example: Coin tossing again,

We have 3 coins, let X= # heads, then

$p(0) = P(X = 0) = P(no\ heads)\ = \ P(TTT) = 1/8$

$p(1) = P(X = 1) = P(1\ head)\ = \ P(HTT, THT, TTH) = 3/8$

$p(2) = P(X = 2) = P(2\ heads)\ = \ P(HHT, HTH, THH) = 3/8$

$p(3) = P(X = 3) = P(3\ heads)\ = \ P(HHH) = 1/8$

x	0	1	2	3
P(X=x)	1/8	3/8	3/8	1/8

Note that we have $\sum P(x_i) = 1$.

- What is the probability of, at most, one head?

$$P(X \leq 1) = \ P(X = 0\ or\ X = 1) = \ p(0) + p(1) = \ 1/8 + 3/8\ = \frac{1}{2}$$

- What is the probability of, at least, one head?

$$P(X \geq 1)\ = \ p(1) + p(2) + p(3)\ = \ 3/8\ + \ 3/8\ + \ 1/8\ = \ 7/8, OR$$

$$P(X \geq 1) = 1 - P(X = 0) = 1 - \tfrac{1}{8} = \tfrac{7}{8}$$

Expected value or mean of a discrete random variable, *X*, which takes on values *x* with probability $p(x_i)$ is:

$$\mu = E(X) = \sum x_i \times p(x_i)$$

Note: We are using Greek letters; knowing the probability distribution is equivalent to knowing about the population and having all the information

So for the coin toss example:

$$\mu = E(X) = \sum x_i \times p(x_i)$$

$$= 0 \times \frac{1}{8} + 1 \times \frac{3}{8} + 2 \times \frac{3}{8} + 3 \times \frac{1}{8} = \frac{12}{8} = 1.5$$

If X and Y are random variables, and c is any constant, then the following holds:

- $E(c) = c$
- $E(cX) = cE(X)$
- $E(X - Y) = E(X) - E(Y)$
- $E(X + Y) = E(X) + E(Y)$
- $E(XY) = E(X) * E(Y)\ only\ if\ X\ and\ Y\ are\ independent$

Examples:

1. Let $Z = 2X - 7$

   $If\ E(X) = 3, then\ E(Z) = E(2 * X - 7)$

   $= E(2 * X) - E(7)$

   $= 2 * E(X) - 7 \quad = -1$

2. $Let\ Z = 3X + 2Y - 2XY + 3, with\ E(X) = 3, E(Y) = 5, X\ and\ Y\ independent$

   $E(Z) = E(3X + 2Y - 2XY + 3)$

   $= 3E(X) + 2E(Y) - 2E(X) * E(Y) + 3$

   $= 3 * 3 + 2 * 5 - 2 * 3 * 5 + 3$

   $= 9 + 10 - 30 + 3$

   $= -8$

Variance measures spread/dispersion of distribution. Let X be a discrete random variable with values x_i that occur with probability $p(x_i)$, and $E(X) = \mu$. The variance of X is defined as

$$\sigma^2 = E[(X-\mu^2)] = \sum[(x_i-\mu)^2 \times p(x_i)]$$

$$\begin{aligned}\sigma^2 &= E\left[(X-\mu)^2\right] \\ &= E\left[X^2\right]-\mu^2 \\ &= \sum_{\text{all } x_i}\left[x_i^2 \times p(x_i)\right]-\mu^2\end{aligned}$$

1. $\mathrm{Var}(\mathrm{X}) = \sigma^2(\mathrm{X}) = \sigma\mathrm{X}^2$
2. Variance is always non-negative (positive or zero): notice that it is a sum of squared things times probabilities – always non-negative
3. The positive square root of the variance is the standard deviation.

So for the coin toss example:

$$\begin{aligned}V(X) &= \sum_{\text{all } x_i}\left[x_i^2 \times p(x_i)\right]-\mu^2 \\ &= \left[0^2\times\tfrac{1}{8}+1^2\times\tfrac{3}{8}+2^2\times\tfrac{3}{8}+3^2\times\tfrac{1}{8}\right]-1.5^2 \\ &= 0.75 \\ \text{Std Dev}(X) &= \sqrt{0.75} = 0.866 \text{ (to 3dp)}\end{aligned}$$

If X and Y are r.v.s and *c* is a constant,

1. $V(c) = 0$
2. $V(cX) = c^2V(X)$
3. $V(X+c) = V(X)$
4. $V(X+Y) = V(X) + V(Y)\ if\ X\ and\ Y\ are\ independent$
5. $V(X-Y) = V(X) + V(Y)\ if\ X\ and\ Y\ are\ independent$

Examples:

1. $If\ V(Z) = 6\ and\ Y = 7 * Z - 3, then$

   $V(Y) = V(7 * Z - 3)$

   $= V(7 * Z)$

   $= 49 * V(Z)$

$= 49 * 6$

$= 294$

2. $If\ Z = 3X - 2Y, and\ X\ and\ Y\ are\ independent\ with\ V(X) = 2\ and\ V(Y) = 1, then$

   $V(Z) = V(3X - 2Y)$

   $= V(3X) + V(2Y)$

   $= 9V(X) + 4V(Y)$

   $= 9 * 2 + 4 * 1$

   $= 22$

3. *Let X be a random variable with the pdf described below.*

x	*-2*	*-1*	*0*	*1*	*2*
p(x)	*0.1*	*0.2*	*0.2*	*0.2*	*0.3*

a) *Find the mean and variance of X*

b) *Find the mean and variance of* $Y = 2 * X + 5$

x	*-2*	*-1*	*0*	*1*	*2*
p(x)	*0.1*	*0.2*	*0.2*	*0.2*	*0.3*
x.p(x)	*-0.2*	*-0.2*	*0*	*0.2*	*0.6*
x^2.p(x)	*0.4*	*0.2*	*0*	*0.2*	*1.2*

a. $\mu_x = E(X) = \sum x_i \times p(x_i) = 0.4$

   $\sigma_x^2 = \sum x_i^2 \times p(x_i) - \mu_x^2 = 2 - 0.4^2 = 1.84$

b. $Y = 2X + 5$

$E(Y) = E(2X + 5) = 2E(X) + 5 = 5.8$
$V(Y) = V(2X + 5) = 4V(X) = 7.36$

Exercises:

1. **For a period of 2 working weeks, I record the time it takes me to get to work, from turning my car ignition on at home to turning it off in the car park. The times (in minutes) are:**

$$31.0, 29.2, 28.5, 27.6, 21.1, 23.4, 31.4, 36.8, 23.4, 24.3.$$

   (a) Determine the mean and median of these observations.

   (b) Discuss the appropriateness of the "mode" as a measure of location for this data.

   (c) Find the sample variance and sample standard deviation of the data.

   (d) Find the upper and lower quartiles of the data.

2. **A vendor at a local sports field must determine whether to sell ice cream or soft drink at today's game. The vendor estimates that if the weather is warm, he will make a profit of \$60 selling soft drink or \$90 selling ice cream. If the weather is cool, he estimates the profits will be \$50 selling soft drinks or \$30 selling ice cream. His investigations on the Bureau of Meteorology website lead him to believe that the probability of a cool day is 40%, and that there is 60% chance of a warm day.**

   i. Compute the expected profit and standard deviation of the profit for selling soft drink.

   ii. Compute the expected profit and standard deviation of the profit for selling ice cream.

   iii. Compute the covariance of selling soft drink and ice cream.

   iv. Do you think the vendor should sell soft drink or ice cream? Explain your answer.

   v. How would you describe the relationship between selling soft drink and selling ice cream?

3. **An investor is given the following information on two stocks.**

Stock	A	B
Mean return	0.15	0.20
SD of Return	0.11	0.23

   i. If the investor can only choose one stock, which should she choose if

      a. Her goal is to maximise expected return?

      b. Her goal is to minimise risk?

The investor decides to invest in a portfolio comprising stocks A and B and has the choice between 25%A and 75% B; 50% A and 50% B, and 75%A and 25% B.

ii. Calculate the expected return and standard deviation of returns for each portfolio if it is known that the correlation between the stocks is 0.4.

iii. Based on your answers in part (ii), which portfolio should she choose to maximise expected return? Which should she choose to minimise risk?

iv. If the correlation between the stocks is -0.4, what would be your new answers to part (ii) and (iii)?

Answers:

1. Let time to get to work be X (in minutes).

a) $mean\ = \frac{1}{n}\sum x_i = \frac{1}{10}(31 + 29.2 + \cdots + 24.3) = \ 27.67$

To obtain median need to order data smallest → largest.

ORDERED DATA: $21 \cdot 1, 23 \cdot 4, 23 \cdot 4, 24 \cdot 3, 27 \cdot 6, 28 \cdot 5, 29 \cdot 2, 31 \cdot 0, 31 \cdot 4, 36 \cdot 8.$

$Median\ of\ 10\ observations\ =\ (10 + 1)\ /\ 2 = 5 \cdot 5^{th}\ obs$

$5th\ obs\ =\ 27 \cdot 6$

$6th\ obs\ =\ 28 \cdot 5$

$\therefore Median = (27 \cdot 6 + 28 \cdot 5)\ /\ 2 = 28.05$

b) Mode is the most frequently occurring observation. The mode is most appropriate with discrete data, so is not very useful in describing the central tendency or location of this (continuous) data.

c) $variance = s^2 = \frac{\sum {x_i}^2 - n\mu^2}{n-1} = \frac{1}{9}(31^2 + 29.2^2 + \cdots + 24.3^2 - 27.67^2) = 22.487$

$s = \sqrt{var} = \sqrt{22.487} = 4.742$

d) We know, $L_p = \frac{(n+1)P}{100}$ will give the pth percentile.

$Q_1 = 25^{th} percentile$

$$\therefore l_{25} = \frac{11 \times 25}{100} = 2.75^{th} observation.$$

$2nd\ obs\ = 23 \cdot 4$

$3rd\ obs\ = 23 \cdot 4$

$2.75^{th}\ obs\ 23.4$

$Q_3 = 75^{th} percentile$

$$\therefore l_{25} = \frac{11 \times 75}{100} = 8.25^{th} observation.$$

$8th\ obs\ = 31 \cdot 0$

$9th\ obs\ =\ 31 \cdot 4$

$8.25th\ obs\ =\ 31 \cdot 1\ (\frac{1}{4}\ of\ the\ way\ from\ 31\ \cdot\ 0\ to\ 31 \cdot\ 4)$

2. Let profits of selling soft drink be SD, and profits for selling ice cream be IC (in dollars).

Weather	Probability	SD Profits	IC Profits
Cold	0.40	50	30
Warm	0.60	60	90

i) $E(SD) = 0.4 \times 50 + 0.6 \times 60 = \$56$

$var(SD) = \ 0.4 \times 50^2 + 0.6 \times 60^2 - 56^2 = \$^2 24$

$std\ dev\ (SD) = \sqrt{var(IC)} = \sqrt{24} = \$4.90$

ii) $E(IC) = 0.4 \times 30 + 0.6 \times 90 = \$66$

$var(IC) = \ 0.4 \times 30^2 + 0.6 \times 90^2 - 66^2 = \$^2 864$

$std\ dev\ (IC) = \sqrt{var(IC)} = \sqrt{864} = \$29.39$

iii) $cov(SD, IC) = E\{(X - \mu_X)(Y - \mu_Y)\}$

$$= 0.4(50 - 56)(30 - 66) + 0.6(60 - 56)(90 - 66) = \$^2 144$$

iv) The answer will depend on the vendor's attitude to risk. While selling ice cream gives a higher expected profit, it also has a much higher risk as shown though the much larger (approx 6 times larger) standard deviation. Thus, if risk averse, the vendor may choose to sell soft drink because while the expected profit is lower, the risk is also much lower.

v) The covariance between soft drink and ice cream sales is positive, indicating a positive relationship. That is, on a warm day, profits from soft drinks are higher and profits from ice cream are also higher (than on a cool day).

3. i)(a) To maximise expected return, go with stock *B* as mean return is higher.

(b) To minimise risk, go with stock *A* as standard deviation of returns is much smaller.

(ii) $E(aX + bY) = aE(X) + bE(Y)$

$$var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab\rho\sigma_X\sigma_Y$$

Portfolio1: 25%A and 75%B

$E(portfolio\ 1)\ = E(0 \cdot\ 25\ A +\ 0 \cdot 75B)\ =\ 0 \cdot 25 \times 0 \cdot 15 +\ 0 \cdot 75 \times 0 \cdot 2\ =\ 0 \cdot 1875.$

$var(portfolio1) =\ var(0.25A +\ 0.75B)$

$$= 0.25^2 \times 0.11^2 + 0.75^2 \times 0.23^2 + 0.25 \times 0.75 \times 0.4 \times 0.11 \times 0.23$$

$$= 0.0343$$

$\therefore sd(portfolio1) = \sqrt{0.0343} = 0.1852$

Similarly,

$E(portfolio\ 2) = 0.175$

$sd(portfolio2) = 0.1460$

$E(portfolio\ 3) = 0.1625$

$sd(portfolio2) = 0.1179$

iii) For maximum expected return she should choose portfolio 1.

To minimise expected risk she should choose portfolio 3.

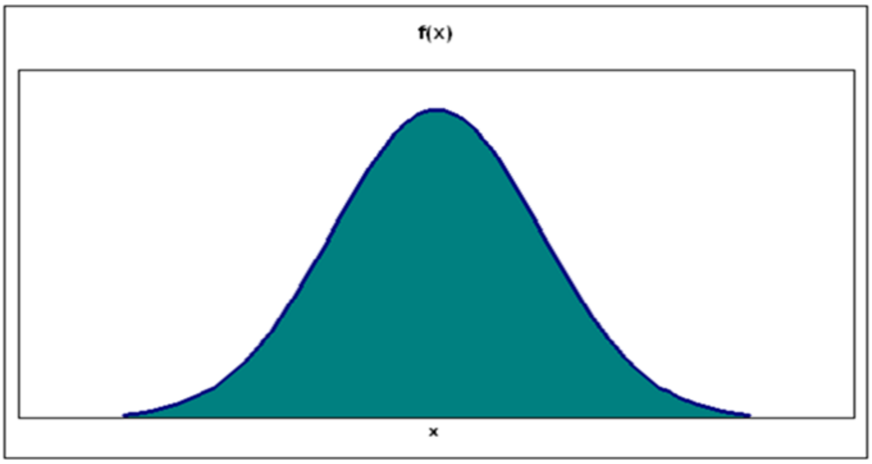
iv) If ρ = –0·4, expected returns stay the same but variances and standard deviations will change

Chapter 11: Regression Modelling

Before we go on to Regression modelling we will have a brief discussion about Normal Distribution which is one of the most commonly used distributions in statistics.

Normal Distribution

A distribution of a variable is the pattern it exhibits. The histogram helps us understand the distribution of the variable. Now, in a histogram, if we replace the bars with a line joining the points, then we get a density curve. A density curve generally looks like:



Where x values are on the horizontal axis and the probability values are on the vertical axis. The probability values for a density curve is generally represented by $f(x)$.

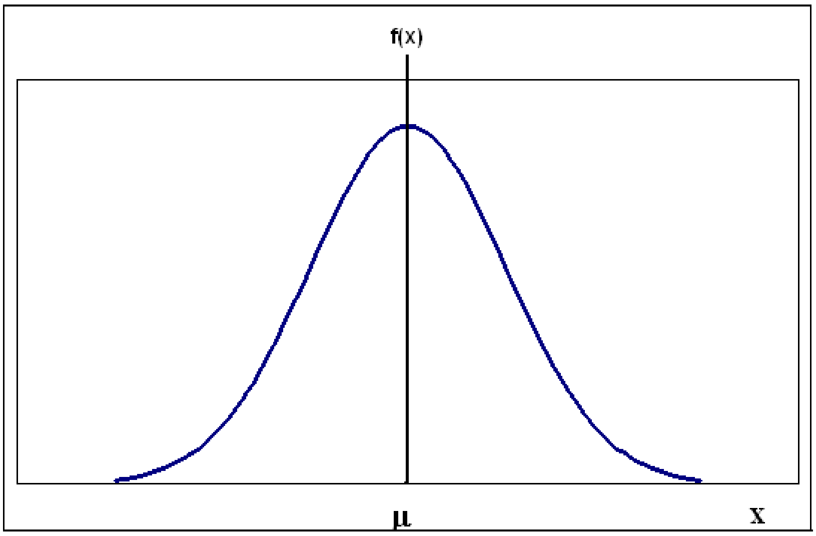
Note: we denote the density function by $p(x)$ for discrete variables and $f(x)$ for continuous variables. Again, in the case of discrete variables, we call them probability mass functions, and in the case of continuous variables, we call them probability density functions.

In statistics there are certain common/general distributions that help us categorize the variables. These pre-defined distributions have characteristics that make them unique in nature and have behaviour and interpretations which are explained in the statistical literature. The reason for the existence of these distributions is that we tend to approximate the distributions of our actual variables (from the data we have) into one of the pre-defined distributions. If we are able to do that then our life becomes much easier as we then know how our variable will behave in different situations (statistically). Also, it helps us a great deal in interpretation of the results.

The normal distribution is one such pre-defined/special distribution. It is applicable to continuous data, and the general form of the probability density function (pdf) of a normal distribution is given by:

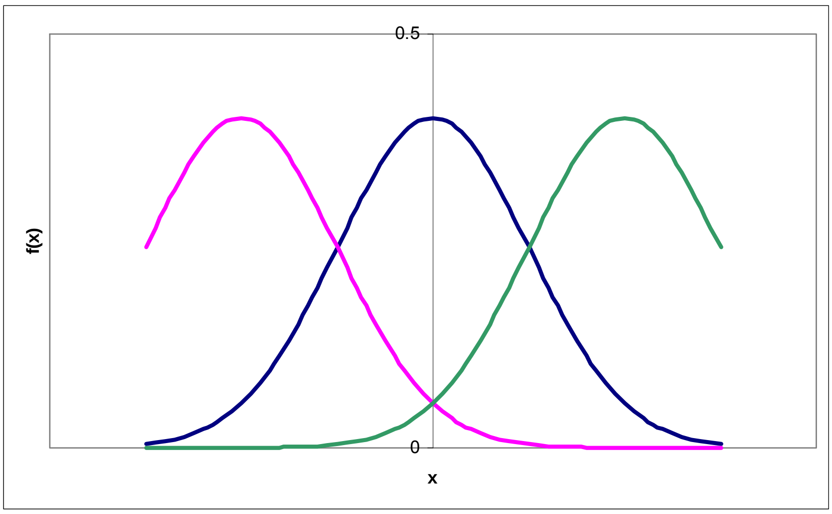
$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\frac{1}{2}\frac{(x-\mu)^2}{\sigma}\right\}, \qquad -\infty < x < \infty$$

The normal distribution is of such importance in the statistical literature because, given the mean (μ) and the standard deviation (σ), we can completely define the distribution. (Most distributions require more than just two numbers to define their density functions completely). The reason we can do this is because the normal distribution is symmetric about the mean (μ), i.e. the density curve is bell shaped, reaches the maximum at $x = \mu$, and tends to 0 as $x \to \pm\infty$. The density curve of a normal distribution looks like:

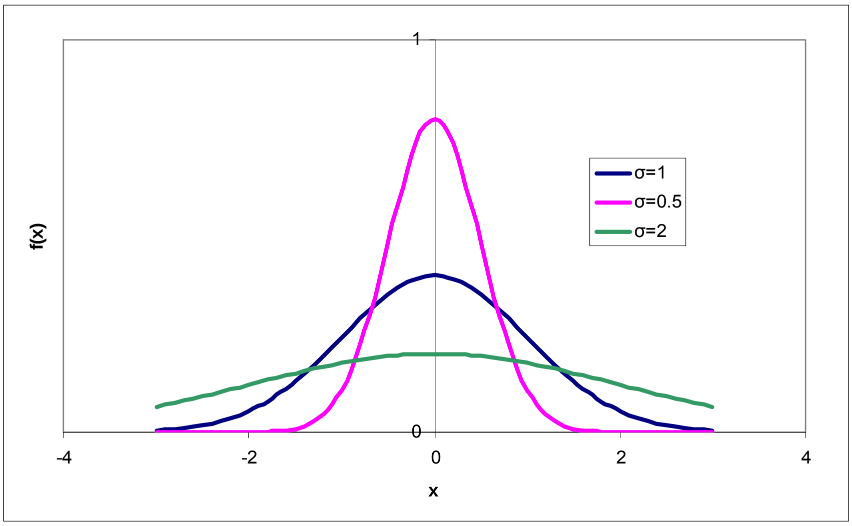


Normal Distribution Notations/Specialities:

1. E(X) = μ; V(X) = σ².
2. Area under curve = 1
3. Different means – shift curve up and down x-axis



4. Different variances – curve becomes more peaked or more squashed



5. Shorthand notation: X~N(μ, σ²).
6. A special property of the normal distribution is that 68% of the entire distribution (all data points) will fall between one standard deviation of the mean ($\mu \pm \sigma$) and 95% of the distribution will fall between two standard deviations of the mean ($\mu \pm 2\sigma$).

Regression Analysis

When do we use regression analysis?

A regression analysis helps investigate whether and how variables are related:

- "Whether" – does the value of one variable have any effects on the values of another?
- "How" – as one variable changes, does another tend to increase or decrease?

The data would generally be:

- One *continuous* response variable (called $\boldsymbol{y}$ - dependent variable, response variable)
- One or more *continuous* explanatory variables (called $\boldsymbol{x}$ - independent variable, explanatory variable, predictor variable, regressor variable)
- That is, in the simplest case we have *n* individuals, and have measured two things about each, such as height and weight on a person – measurements occur in pairs.

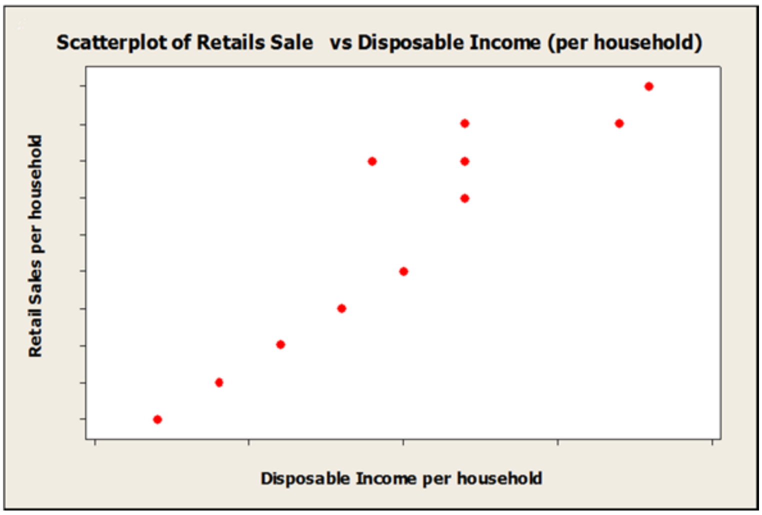
Regression analysis develops an equation that represents the relation between the variables.

Regression analysis can be one or many of the following:

- Simple linear regression – straight line relationship between y and x (i.e. one explanatory variable) (we will only discuss this)
- Multiple linear regression – "straight line" relationship between y and $x_1, x_2, .., x_k$ where we have k explanatory variables
- Non-linear regression – relationship not a "straight line" (i.e. y is related to some function of x, e.g. $log(x)$)
- *And many more*!

Suppose we have n pairs of observations - y = retail spending per household; x = disposable income per household. Each observation is labelled as (x_i, y_i) – pair of observations from i^{th} household in study.

The scatter plot looks like:



Objective of regression analysis:

- To model this relationship – find a model to fit to the data (In this case, a linear model)
- Express random variable *Y* in terms of random variable *X*
- Interested in predicting values of *Y* when *X* takes on a specific value, e.g. what will retail sales be when disposable income is $35,000.

Note: In reality values won't be exact – households which have the same disposable income will still have different retail sales amounts. So, more reasonable is to ask what would be the *expected value* of retail sales when disposable income is $35,000; i.e. $E(Y|X = 35{,}000)$

In general we use, $E(Y|X = x)$

If we add in an assumption of linearity we get

$$E(Y|X = x) = \beta_0 + \beta_1 x$$

where, β_0 and β_1 are coefficients that determine the straight line.

Example:

Imagine we know the coefficients (somehow), and have

$$E(Y|X = x) = 5000 \pm 0.4x$$

So β_0 =5000 and β_1 =0.4.Hence, if retail spending = 35,000 we have

$$E(Y|X = 35000) = 5000 + 0.4 \times 35000 = \$19{,}000$$

That is, expected retail sales will be $19,000.

In general, for any pair of observations:

$$E(Y_i|X = x_i) = \beta_0 + \beta_1 x_i$$

In practice, however, the observed value will (almost always) differ from the expected value. This difference is denoted by the letter epsilon, ε_i , and is called the error term. We assume that the mean of the errors will always be zero.

$$\varepsilon_i = Y_i - E(Y_i|X = x_i) = Y_i - \beta_0 + \beta_1 x_i$$

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

The last line of the previous equation is called the population or true regression line, where β_0 and β_1 are constants to be estimated, and ε_i is a random variable with mean = 0.

Interpretation: example; response of particular retail sales to a particular value of disposable income will be in two parts – an expectation $(\beta_0 + \beta_1 x)$ which reflects the systematic relationship, and a discrepancy (ε_i) which represents all the other many factors (apart from disposable income) which may affect sales.

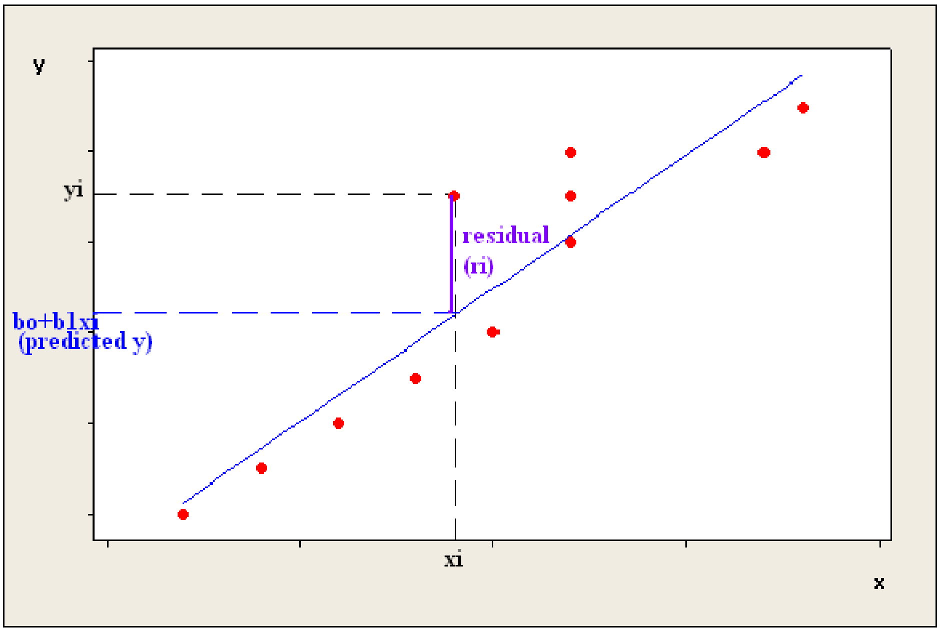
In practice, the population regression line has to be estimated. The process we are going to discuss is known as the "Least Squares Estimation".

$$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$$

$$is\ estimated\ by\ \hat{y} = b_0 + b_1 x$$

How good is the estimate?

Look at the distance between the points (x_i, y_i) and the line:



The vertical distance between the observed point and the fitted line is called the residual. That is,

$$r_i = y_i - (b_0 + b_1 x_i)$$

where, r_i estimates ε_i, the error variable.

<u>*The process*</u>: Want to determine values of b_0+b_1 that best fit the data – choose the values which minimise the sum of the squared differences between observed and estimated, i.e. choose values of slope and intercept which minimise

$$\sum_{i=1}^{n}(y_i - \hat{y})^2 = \sum_{i=1}^{n}(r_i)^2$$

This is the least squares method (involves calculus to do them, but the actual method is beyond the scope of this course).

We can show that the residual sum of squares is minimised by the solution:

$$\widehat{\beta_1} = b_1 = \frac{cov(x,y)}{s_x^2}$$

$$\widehat{\beta_0} = b_0 = \bar{y} - b_1\bar{x}$$

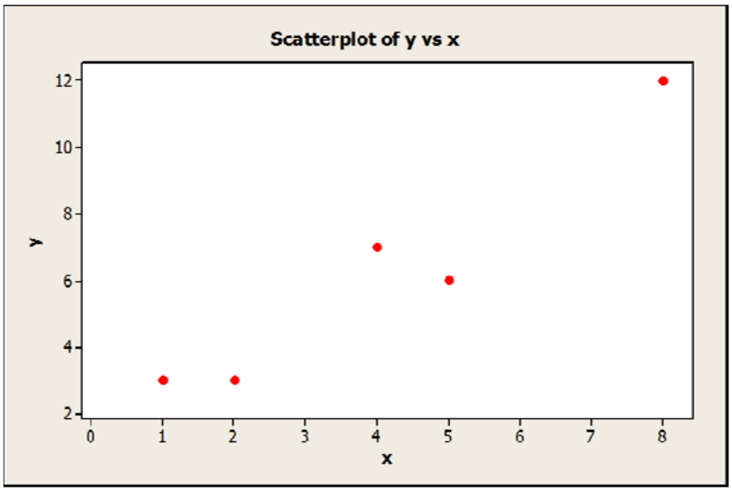
Example:

The investment in certain share portfolios (x) and the value after a year (y) in $\$000$ are given in the table below:

x	1	2	4	5	8
y	3	3	7	6	12

Fit a regression line onto these data.

Firstly, we look at the scatter plot.



We want to find b_0 and b_1 for

$$\hat{y} = b_0 + b_1 x$$

Given that $cov(x, y) = 9.75, var(x) = 7.5, x\ average = 4, y\ average\ = 6.2$

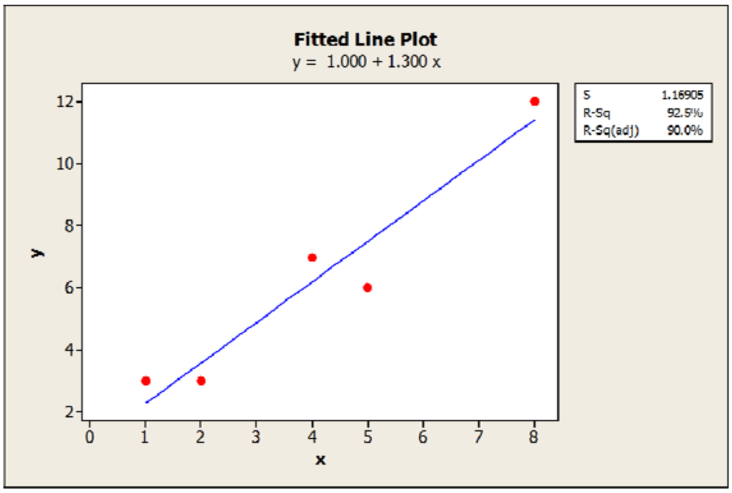
$$\widehat{\beta_1} = b_1 = \frac{cov(x, y)}{s_x^2} = \frac{9.75}{7.5} = 1.3$$

$$\widehat{\beta_0} = b_0 = \bar{y} - b_1 \bar{x} = 6.2 - 1.3 \times 4 = 1$$

So, least squares regression line is

$$\hat{y} = 1 + 1.3x$$

Plotting the line on the scatter plot we get:



Interpretation:

- x is the investment($000), y is the value after a year ($000)
- $b_0\ = 1$ is the intercept, $b_1\ = 1.3$ is the slope coefficient
- Slope: for each extra $1000 invested, value after a year is expected to increase by $1300.

- Intercept: as model only fitted in range $x = 1\ to\ x = 8$, no interpretation can be given (refers to what happens when x=0); model implies that if $0 is invested, value after a year is $1000 – obviously not sensible ➔ demonstrates the danger of extrapolating.
- General rule – can't determine the value of y for a value of X outside our sample range of x.

Assessing the model:

Several methods can be used to assess the model as to how well it fits the data. If fit is poor, discard the model and fit another

- Different shape, e.g. not a straight line – could mean fitting a quadratic curve, cubic curve, etc; or fitting something completely different
- Different predictors

In our fitting we assume the errors have a particular distribution – that is, $\varepsilon \sim N(0, \sigma_\varepsilon^2)$

- Normal distribution
- Mean = 0
- Constant variance = σ_ε^2
  - If σ_ε^2 is small, then small spread of observations around fitted line
  - If σ_ε^2 is large, then observations have wide spread around fitted line
- Errors associated with any two y values are independent

If all these assumptions are satisfied then we accept the model.

Determine the strength and significance of association:

This is measured by R^2 – coefficient of determination. This measures the proportion of total variation explained, i.e.

$$R^2 = \frac{explained\ variation}{total\ variation} = (correlation\ coefficient)^2$$

Will be between 0 and 1; a value close to 1 indicates most of the variation in y is explained by the regression equation

Note: $r = \pm\sqrt{R^2}$

For example if R^2 value is 25% then it says that only 25% of the variation in y is explained by x (using that particular model).

Regression Diagnostics:

We cannot say that model fits data well unless assumptions about errors are met:

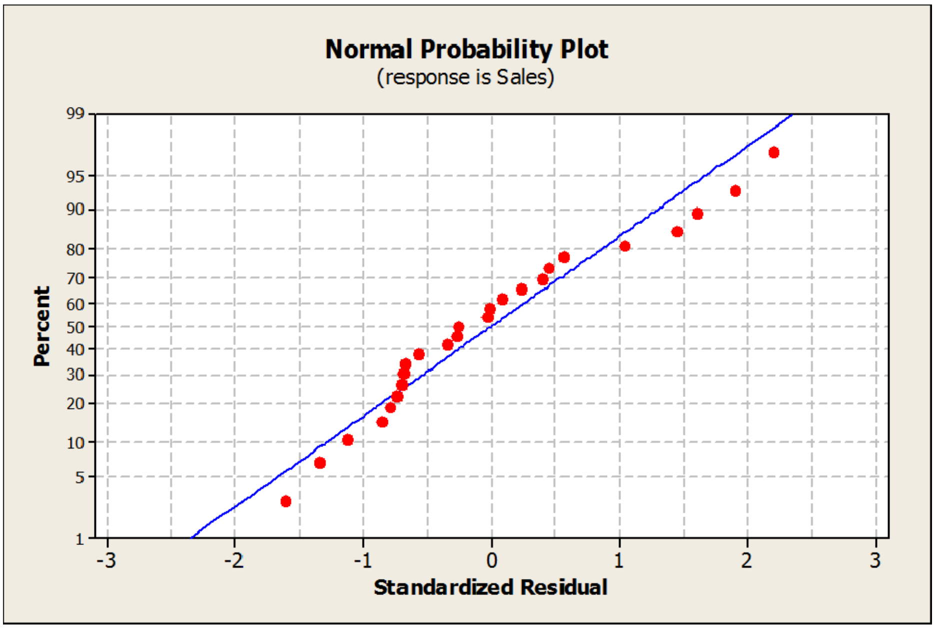
- Independence
- Normally distributed
- Zero mean, constant variance
- (Note that a zero mean of residuals is ensured by the estimation process)

In order to do so we examine residuals (estimates of errors) to see if assumptions are met. This is done using graphical techniques

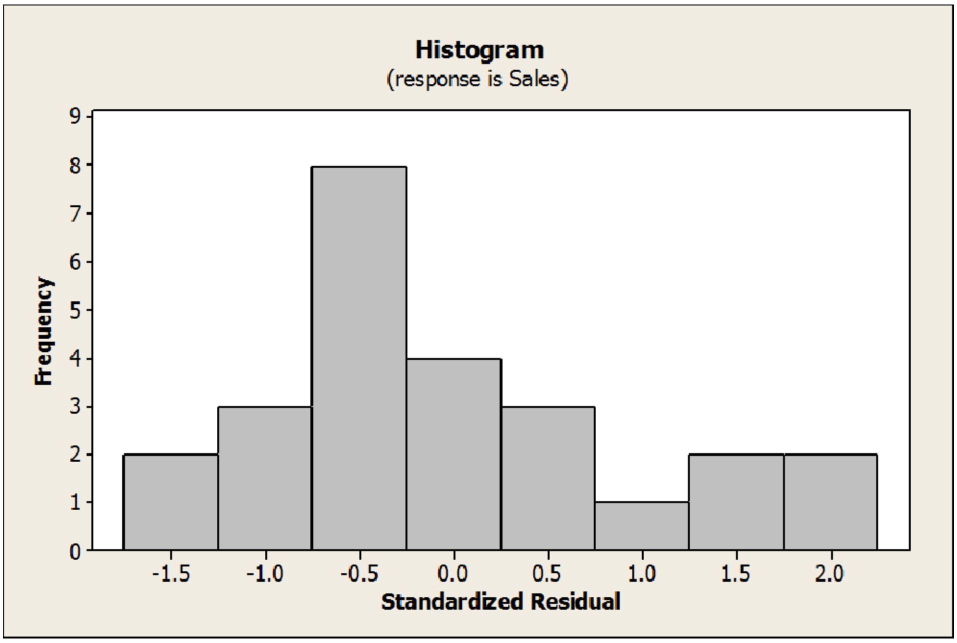
- Assess normality from histogram, normality plot
- Assess independence and variance from scatter plots of residuals vs. fitted values, predictor values, order

Example data:

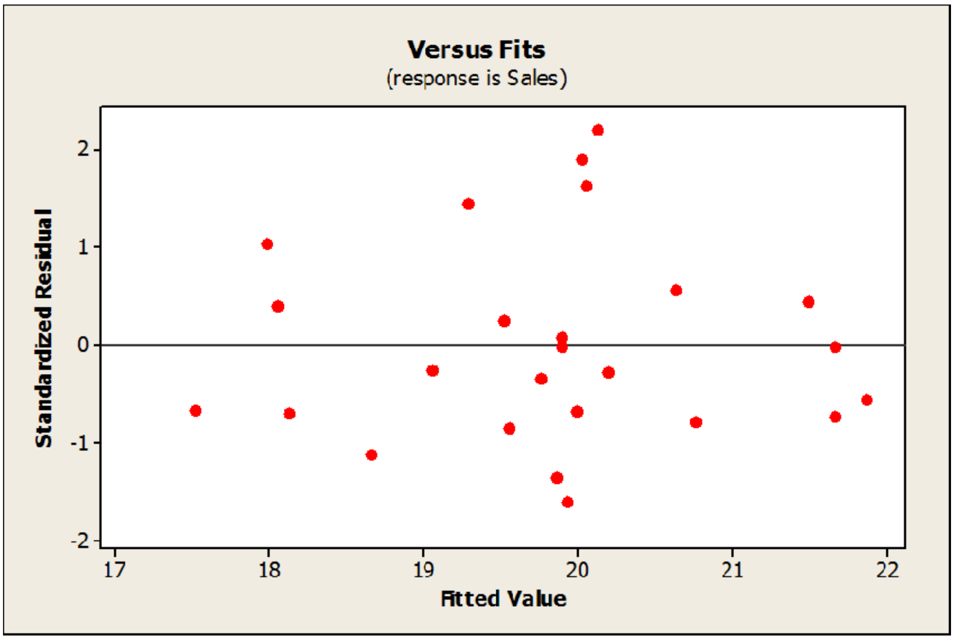
Normal probability plot: Normality is shown by points being close to line



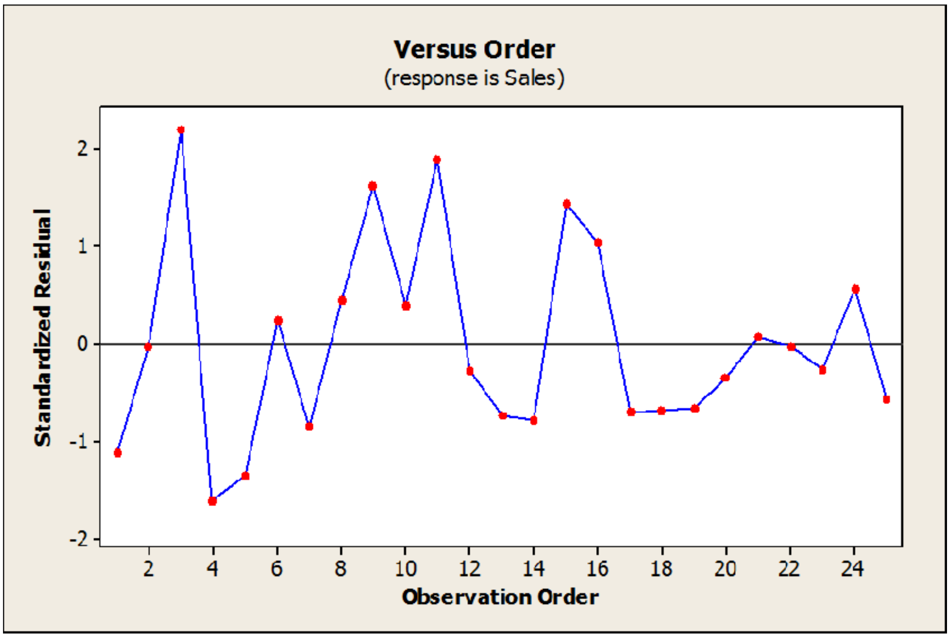
Histogram of standardised residuals: Normality is shown bell shape in histogram



Residuals vs fitted values: Randomness indicates independence; equal spread indicates constant variation

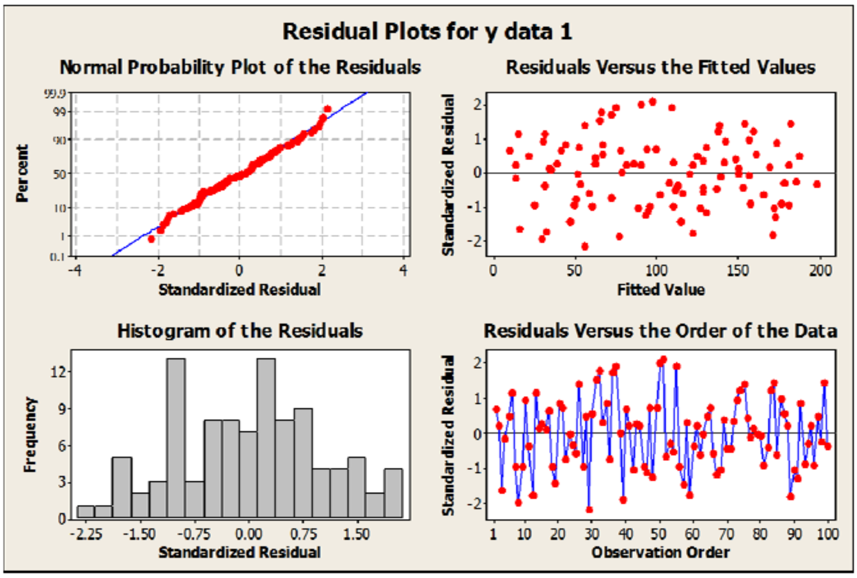


Residuals vs order: Randomness indicates independence; equal spread indicates constant variation (if error terms are not correlated then the problem of auto-correlation arises)

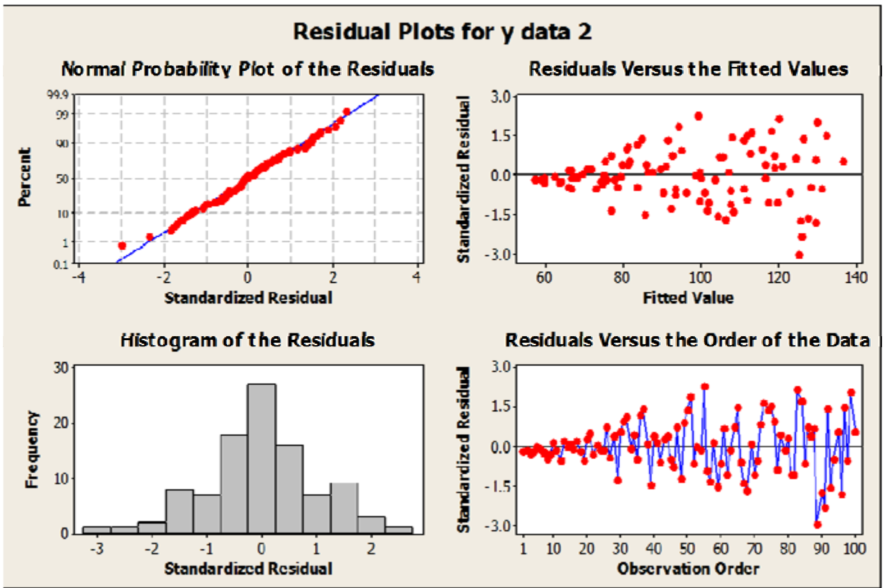


Note:

- If variation is constant (residuals show constant spread around zero), it is called *<u>homoscedastic</u>*

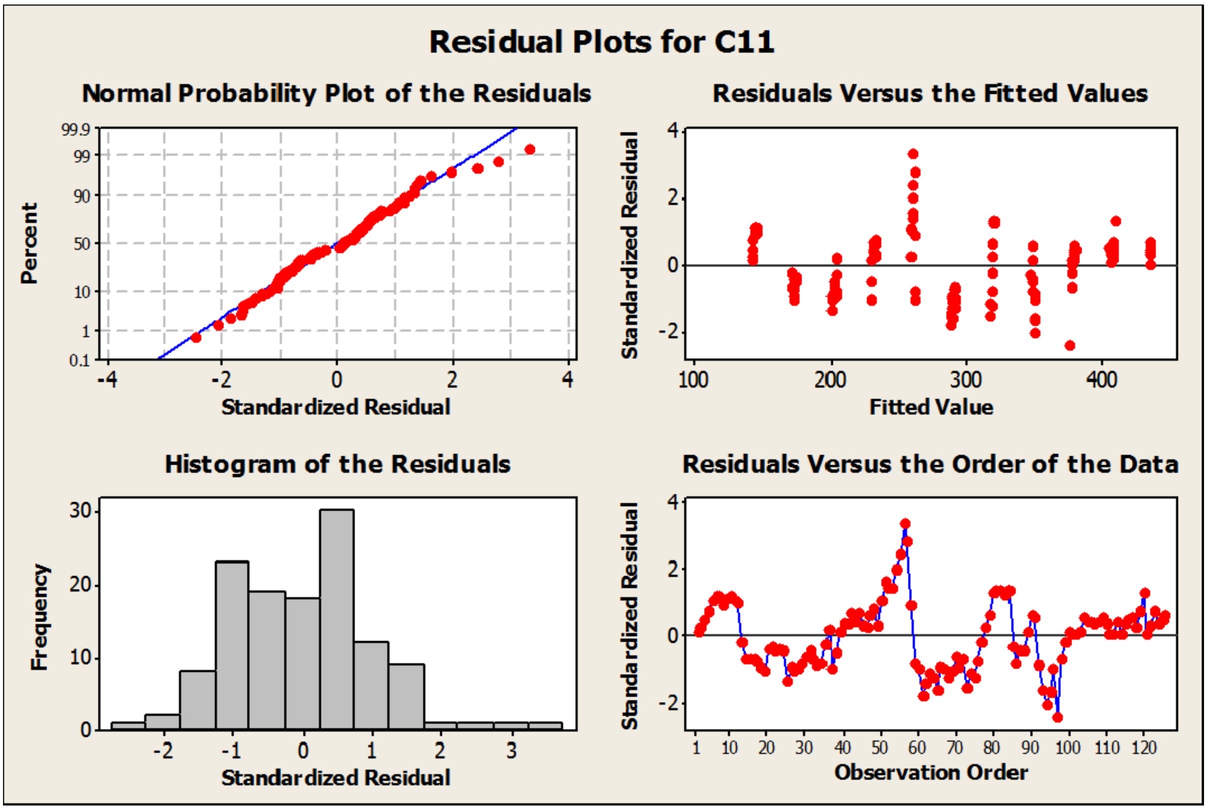


- If variation is non-constant (residuals show varying spread around zero), it is called *<u>heteroscedastic</u>*



Example:

Look at the following graph to state whether the assumptions of the errors are met.



The errors seem to satisfy the condition of normality. However, the error terms seem to be correlated over time. This is known as the problem of *autocorrelation in the error terms*. Autocorrelation often happens with economic and financial data.

Note: In order to carry out any kind of regression analysis we use computer programs. There are various kinds of computer programs that you may use and they vary from department to department and course to course (The outputs produced above have been generated using MINITAB). The discussion on regression modelling here has been very basic and we have not even touched the tip of the iceberg. If you need any help with regression modelling for your assignments/courses, feel free to come and see me and we can discuss how you would go about finding a solution to your specific query.

For further readings in Statistics:

KELLER, G. (2008) *Statistics for Management and Economics,* Mason, OH, South-Western Cengage Learning.

Please note that all the graphs in Chapters 9-11 have been generated using various data sets used in the College of Business and Economics' course *"STAT 1008: Quantitative Research Methods".*

List of References

BRAILSFORD, T., HEANEY, R. & BILSON, C. (2006) *Investments: Concepts and Applications,* South Melbourne, Vic., Thomson.

EPP, S. S. (2004) *Discrete Mathematics with Applications,* Belmont, CA, Thomson-Brooks/Cole.

HRBACEK, K. & JECH, T. J. (1984) *Introduction to Set Theory,* New York, M. Dekker.

KELLER, G. (2008) *Statistics for Management and Economics,* Mason, OH, South-Western Cengage Learning.

LAY, D. C. (2006) *Linear Algebra and Its Applications,* Boston, Pearson/Addison-Wesley.

LEWIS, J. P. (1969) *An Introduction to Mathematics for Students of Economics* London, Macmillan.

MAY, W. G. (1967) *Foundations in Modern Mathematics,* Mass., Blaisdell.

MILLER, C. D. & HEEREN, V. E. (1969) *Mathematical Ideas: An Introduction,* Glenview, Ill., Scott Foresman.

STEWART, J. (2007) *Essential Calculus,* Belmont, CA, Thomson Brooks/Cole.