Department of Mathematics
Faculty of Engineering,
University of Moratuwa

Methods of Mathematics - Probability and Statistics

MA1020 (Level 1/Semester 2)

by

Dr. T S G Peiris
Department of Mathematics

Note: In these handouts only the important points are given. It is necessary that the students to attend all the classes to acquire more details and to expose in tackling different statistical problems related to engineering applications.

Course Content:

Introduction to probability using set theory, Conditional probability and independence, Applications of Bayes theorem, Discrete and continuous random variables, Properties of the probability distributions (Binomial, Normal, Standard Normal, Student's t, Poisson and Exponential) and their applications, Descriptive statistics and Introduction to Minitab for data analysis

Duration: 7-8 weeks (2 hours/week)

Learning Outcomes:

Upon successful completion of this course, students should be able to

- Apply the knowledge on the fundamental probability concepts to various applications
- Use of probability distributions for various engineering application
- Compute various statistical indicators in practical problems
- Apply descriptive statistics for decision making
- Use of inferential statistics for decision making

- Interpret the results of data analysis

Methodology: Lecturers and tutorials

Scheme of Evaluation: Assignments + Mid semester examination – 25%

End of semester examination - 75%

Recommended Readings:

- Mathematics for Engineers – J M J A Cooray
- Business Statistics Concept and Application (Many books on, "Business Statistics" are available in the library. All of those have very good practical applications)

1. BASIC DEFINITIONS IN SET THEORY

1.1 Set: Any well defined collection of objects is called a set.

1.2 Element: The objects comprising the set are called its elements.

If A is set and p is an element of A then it is denoted by $p \in A$.

Eg. 9 is an element of a set $A = \{1, 3, 5, 7, 9\}$ consisting of odd numbers less than 10.

1.3 Universal Set: An universal set is a set which contains all objects, including itself and is denoted by U.

1.4 Null Set: The set contains no elements is called null set and is denoted by ϕ

1.5 Set Operations:

Let A and B be arbitrary sets.

- The **union** of A and B is denoted by $A \cup B$, is the set of elements belong to A **or to** B.

  $$A \cup B = \{x / x \in A \ \ or \ \ x \in B\}$$

- The **intersection** of A and B is denoted by $A \cap B$, is the set of elements belong to A **and to** B.

  $$A \cap B = \{x / x \in A \ \ and \ \ x \in B\}$$

- The difference between A **and** B (or the relative complement of B with respect to A) is said to A **complementary** B.

  $$A - B = \{x / x \in A \ \ and \ \ x \notin B\}$$

- If $A \cap B$ is ϕ (A and B do not have common element) then A and B is said to be **disjoint** set.

1.5 Laws of Set Theory

Associative law: $(A \cup B) \cup C = A \cup (B \cup C)$ and

$(A \cap B) \cap C = A \cap (B \cap C)$

Commutative law: $(A \cup B) = (B \cup A)$ and

$(A \cap B) = (B \cap A)$

Distributive law: $A \cup (B \cap C = (A \cup B) \cap (A \cup C)$

Identity law: $A \cap \phi = \phi$ and $A \cup \phi = \phi$

Idempotent law: $A \cap A = A$ and $A \cup A = A$

De Morgan's law: $(A \cup B)^c = A^c \cap B^c$ and

$(A \cap B)^c = A^c \cup B^c$

2. FUNDAMENTAL PRINCIPAL OF COUNTING

2.1 Factorial Notation

Factorial n is denoted as n! and is defined as $n! = 1.2.3........(n-1).n$

Note: If some procedure can be performed in n_1 different ways and a second procedure can be performed in n_2 ways , third procedure can be performed in n_3 ways, and so forth then the number of ways the procedure can be performed in the order indicated is $n_1 \times n_2 \times n_3............$

2.2 Permutation

An arrangement of set of n objects in a given order is called a permutation. An arrangements of any $(r \le n)$ objects from n objects taken at a time is denoted by

$$^nP_r = n(p,r) = \frac{n!}{(n-r)!}$$

2.3 Permutation with repetitions

The number of permutations of n objects of which n_1 are alike, n_2 are alike and n_3 are alike is given by $\dfrac{n!}{(n_1!n_2!n_3!)}$

Eg The number of different signals each consists of 8 flags in a vertical line formed from a set of 4 indistinguishable red flags and 3 indistinguishable white flags and one blue flag is $8!/4!3!$.

2.4 Combinations

A combination of n objects taken r at a time is denoted by $c(n,r)$ where

$$^nC_r = c(n,r) = \frac{p(n,r)}{r!} = \frac{n!}{(n-r)!r!}$$

Eg. The number of committees of 3 can be formed from 8 persons = $^8C_3 = \frac{8!}{3!(8-3)!} = 56$.

Comparison between combinations and permutations of the four letters a, b, c and d taken 3 at a time

Table 1

Combinations	Permutations
abc	abc, acb, bca, bac, cab, cba
abd	abd, adb, bad, bda, dab, dba
acd	acd, adc, cad, cda, dac, dca
bcd	bcd, bdc, cbd, cdb, dbc, dcb

3. PROBABILITY THEORY

3.1 Probability

Probability is the likelihood or chance that a particular event will occur.

Eg. Chance of picking a black card from a deck of cards;
Chance of winning 50 over game;
Chance of selecting Electronics stream

In a scientific way, probability is defined as a study of random (nondeterministic) experiments.

The theory of probability makes some sense to find the mathematical foundation (numerical measure) for uncertainty. It enables us to make decisions under condition of uncertainty. The theory of probability is useful in day to day life and has many applications in all the fields of engineering. There are three approaches to the subject of probability. They are ,

(i) prior classical probability,
(ii) empirical classical probability
(iii) subjective probability.

Probability theory is based on the paradigm of a random experiment.

3.2 Random Experiment

It is a process which is conducted repeatedly under homogenous environment. That is, an experiment whose outcome cannot be predicted with certainty, before the experiment is run. We usually assume that the experiment can be repeated infinitely under essentially the same conditions.

3.3 Random Variable

In reality all variables are non deterministic. Thus the variables whose exact value can not be predicted (determined) are known as random variables. Thus for a given experiment many number of random variables can be defined.

3.4 Trial

The performance of a random experiment is called a trial. Many random variables can be associated in a trial

Eg. Throwing a die and throwing a coin three times are trials.

3.5 Event

Outcome of a trail is an event.

Let A be an event that two or more heads appear consecutively from an experiment of throwing a coin three times. Then $A = \{HHH, HHT, THH\}$.

3.6 Sample Space - S

The all possible outcome of a given experiment (or random variable) is known as sample space and is generally denoted by S. The values which belong to S are known as elements in the sample space.

Let the random variable X = the number appears by tossing a die, then

$$S = \{1,2,3,4,5,6\}$$ S

3.7 Sample point

A particular outcome of the experiment is known as sample point. An event may consist of one or more sample points. In the above example the number of sample points in A is three and is denoted by $n(A) = 3$.

3.8 Simple Event

An event is known as simple event if it corresponds to a single possible outcome.

Eg. In tossing a die the chance of getting 3 is a simple event.

3.9 Compound (Joint) Event

An event is known as compound event if it corresponds to more than a single possible outcome. In tossing a die the chance of getting an odd number is a compound event.

3.10 Favourable Event

The number of outcome due to a desired event.

3.11 Mutually Exclusive Event

Events are mutually exclusive if they cannot happen at the same time.

- If we toss a coin, either heads or tails might turn up, but not heads and tails at the same time.
- In a single throw of a die, we can only have one number shown at the top face. The numbers on the face are mutually exclusive events

If A and B are mutually exclusive events then the probability of A happening **OR** the probability of B happening is $P(A)+P(B)$. That is $P(A \cup B)=P(A)+P(B)$.

3.12 Equally likely Events

Two or more events are said to be equally likely if the chance of their happening is the same.

- Obtaining 1or 2 or 3 by throwing an unbiased die

3.13 Independent Event

Two or more events are said to be independent if its happening does not influence by the happening of other events. That is, A occurs does not affect the probability of B occurring.

- Choosing a marble from a jar **AND** landing on heads after tossing a coin
- Attending to Maths class and playing a tennis game

3.14 Probability of an event

Probability of an event A in S is defined as $P(A)$ and equals to

$$P(A)=\frac{\text{number of possible points of the event}}{\text{numberof all possible points in S}}=\frac{n(A)}{n(S)}$$

- A spinner has 4 equal sectors colored yellow, blue, green and red. After spinning the spinner, what is the probability of landing on each color?

The possible outcomes of this experiment are yellow, blue, green, and red.

$$P(yellow)=\frac{\text{number of ways to land on yellow}}{\text{total number of colors}}=\frac{1}{4}$$

Find $P(blue)$, $P(green)$ and $P(red)$?

3.15 Axioms of Probability

Let S be a random sample space and A be an event within S. Then

(1) $0 \le P(A) \le 1$

(2) $P(S) = 1$

(3) The sum of the probabilities of all simple events must be 1.

(4) If A and B are mutually exclusive events then $P(A \cup B) = P(A) + P(B)$.

(5) If $A_i (i = 1, 2, ...n)$ are mutually exclusive events then $P(\cup A_i) = \sum_i P(A_i)$

Useful Theorems in Probability

Theorem 1: If ϕ is the empty set then $P(\phi) = 0$

Theorem 2: If A^c is the complementary event of A then $P(A^c) = 1 - P(A)$.

Theorem 3: If $A \subset B$ then $P(A) \le P(B)$

Theorem 4: If A and B are any two events then $P(A - B) = P(A) - P(A \cap B)$

Addition Theorem

If A and B are any two events then probability that at least one of them occurs (that is A or B occurs) is denoted by $P(A \cup B)$ and given by,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

3.16 Simple (Marginal) Probability

Example 1. Suppose that you as the President of a company is interested in studying the intension of 1000 households to purchase a big screen televisions in the next 12 months. As a follow-up study the following results were observed in a survey. What is the probability that a household is planning to purchase a big screen television in the next 12 months? What is the probability that a household is planning to purchase a big screen television and actually purchases the television in the next 12 months?

Table 2

Planed to purchase	Actually purchased	
	Yes	No
Yes	200	50
No	100	650

$p(\textit{planned to purchase}) = \dfrac{\text{number who planned to purcahse}}{\text{total numer of households}}$ = **250/1000=0.25**

Note: Simple probability is also called marginal probability as the total number of success (those who planned to purchase) can be obtained from the appropriate margin of contingency table.

3.17 Joint Probability

Joint probability refers to the situation involving two or more events.

Eg. P(planned to purchase and actually purchased a big screen TV).

$$P(\textit{planned to purchase} \,\&\, \text{actualluy purcahsed})$$

$$= \frac{\textit{number} \text{ who planned to purcahse and acttually purcahsed}}{\text{total numer of households}}$$

$$= \frac{200}{1000} = 0.20$$

Example 2. Suppose in the follow up study the following additional information was obtained from the 300 households who actually purchased a big screen TV.

Table 3

Purchased HDTV	Purchased DVD	
	Yes	No
Yes	38	42
No	70	150

3.18 Conditional Probability

Let B be an any event with $P(B) > 0$. The prob. that an event A occurs once B has occurred is known as the conditional prob. of A given that B occurred and is denoted by

$P(A/B)$ and is given by $P(A/B) = \dfrac{p(A \cap B)}{p(B)}$ ………………………. (1)

Eg. If a pair of pair of fair dice is tossed then the prob. that the sum is 6 given that one dice has 2 is 11 /36. (Discuss in the class).

3.19 Multiplication Theorem for Conditional Probability

For any two events A_1 and A_2, from equation (1) it is clear that

$P(A_1 \cap A_2) = P(A_1) \cdot p(A_2 / A_1)$ and $P(A_2 \cap A_1) = p(A_2) \cdot p(A_1 / A_2)$

Thus the same concept can be extended to for any events A_1, A_2, ….., A_n so that

$$P(A_1 \cap A_2 \cap .... A_n) = [P(A_1)][P(A_2 / A_1)][P(A_3 /(A_1 \cap A_2)].... ....[\mathrm{P}\,(\mathrm{A}_n / A_1 \cap A_2 \cap ..... A_{n-1})]$$

3.20 Partitions and Baye's Thoerm

Suppose the events $A_1, A_2, ...A_n$ be partitions of the sample space S s.t. $S = \bigcup_{i=1}^{n} A_i$.

B be an event s.t. $B = S \cap B = (A_1 \cup A_2 \cup ....\cup A_n) \cap B$

$$= (A_1 \cap B) \cup (A_2 \cap B) \cup ....\cup (A_n \cap B)$$

$$P(B) = \sum_{i=1}^{n} p(A_i \cap B) = \sum_{i=1}^{n} p(A_i).p(B / A_i)$$

Bayes' Theorm

$$(P(A_i / B) = \frac{p(A_i \cap B)}{p(B)} = \frac{p(A_i)\, p(B / A_i)}{\sum_{i=1}^{n} p(A_i).p(B / A_i)}$$

3.21 Independence

An event B is said to be an independent event of an event A if the probability that B occurs is not influenced by whether event A has or has not occurred.

That is $P(B) = P(B / A)$

We know from eq. (1) that $P(B / A) = \frac{p(B \cap A)}{p(A)}$

If A and B are independent $P(B / A) = P(B)$.

Thus it is clear if A and B are two independent events then $P(A \cap B) = P(A) \times P(B)$.

4. Properties of Random Variables

In mathematics, random variables are used in the study of probability. They were developed to assist in the analysis of games of chance, stochastic events, and the results of scientific experiments by capturing only the mathematical properties necessary to answer probabilistic questions. There are two types of random variables; discrete and continuous depending on the type of measurement of the random variable.

4.1 Discrete Random Variables

A discrete random variable is one which may take on only a countable number of distinct values such as 0,1,2,3,4,......etc. Discrete random variables are usually counts.

Examples (a) the number of children in a family

(b) the number of calls received in an interval $(0, t)$

(c) number of cards passed at a traffic light in a specified period

4.2 Probability Distribution of a discrete random variable

The probability distribution of a discrete random variable is a list of probabilities associated with each of its possible values. It is also sometimes called the probability function or the probability mass function.

Eg. Consider the tossing of pair of fair dice. Then the possible outcome is

$$S = \{(1,1), (1,2), ...(1,6), ...(6,1), (6,2), ...(6,6)\}$$

Thus $n(S) = 36$. Let X be a RV such that $X = \max(a, b)$ here (a, b) is the outcome of the pair of dice. Then the possible values that X can have are , {1,2,3,4,5,6}=*S(x)* (say) and $n(X(s))= 6$. Then,

$P(X = 1) = P\{(1,1)\} = 1/36 = f(1)$ (say)

$P(X = 2) = P\{(1,2), (2,2), (2,1)\}$ Thus $f(2) = 3/36$ (say)

Similarly $f(3) = 5/36$, $f(4) = 7/37$, $f(5) = 9/36$, $f(6) = 11/36$.

Thus we can form a table given below and it is called as probability distribution of *X*.

Table 4

x_i	1	2	3	4	5	6
$f(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

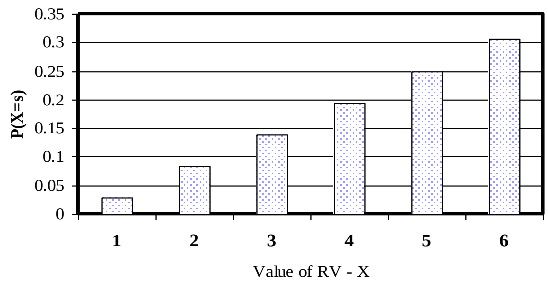


Figure 1. Probability distribution of X (probability histogram)

4.3 Properties of distribution of a RV

Let X be a rv on a sample space S with finite image set, $X(S) = \{x_1, x_2, x_3, \ldots x_n\}$. Let the function f on $X(S)$, denoted by $f(x_i) = P(X = x_i / i = 1,2\ldots n)$ is given by the table below.

Table 5

x_1	x_2	….			x_n
$f(x_1)$	$f(x_2)$				$f(x_n)$

The distribution f satisfies the conditions (a) $f(x_i) \geq 0$ and

$$\text{(b)} \sum_{i=1}^{n} f(x_i) = 1$$

4.4 Cumulative Probability:

It is defined as the probability of observing less than or equal a given number of success (see Table 6).

Eg. Let Y be a RV of the above experiment such that Y = sum of (a, b).

Then $Y = \{2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

Table 6. Probability distribution and cumulative probability distribution of Y

Y_i	2	3	4	5	6	7	8	9	10	11	12
$P(Y_i)$	1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36
Cum Pr	1/36	3/36	6/36	10/36	15/36	21/36	26/36	30/36	33/36	35/36	36/36

4.5 Distribution of *Y*

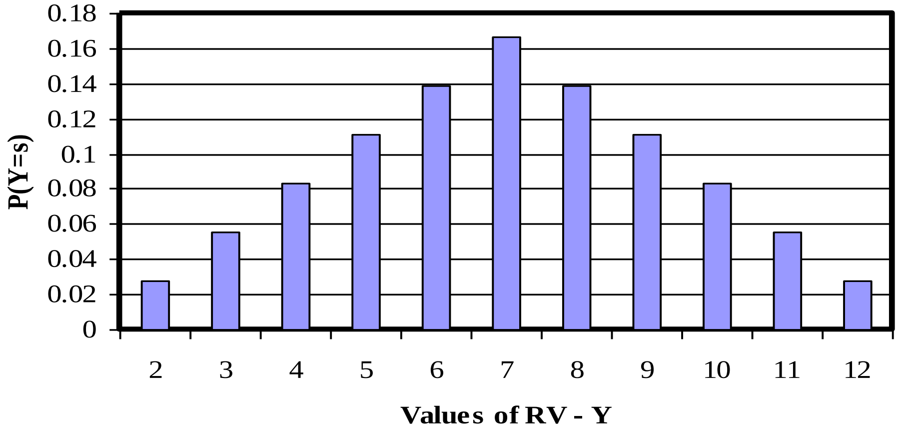


Figure 2. Probability distribution of *Y*

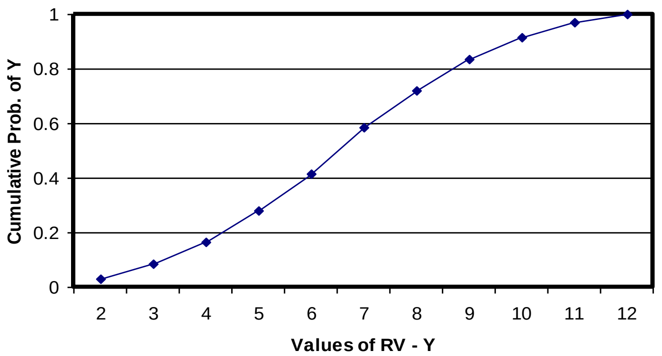


Figure 3. Cumulative distribution of *Y*

4.6 Continuous Random Variable

A ***continuous random variable*** is one which takes an infinite number of possible values. Continuous random variables are usually measurements.

Examples (a) Z score of the first year students in UOM
(b) Leaf areas of the 4th leaf of a given plant

Note: Continuous random variable is not defined at specific values. Instead, it is defined over an interval of values, and is represented by the area under a curve.

Thus if X is a continuous RV with pdf of $f(x)$ then $P(a \le x \le b) = \int_a^b f(x)dx$.

As for discrete case $f(x) > 0$ and $\int_{-\infty}^{\infty} f(x)dx = 1$.

5. PARAMETERS OF A DISTRIBUTION

In order compare different distribution various parameters (statistical indicators) have been defined. The physical meaning of each of the indicator is explained in the class.

5.1. Expected value (mean) - μ

In probability theory the expected value (or expectation, or mean) of a discrete random variable is the sum of the probability of each possible outcome of the experiment multiplied by the outcome value. Thus,

If *X* is RV with a distribution of *f*(*x*) then the mean or expected value of *X* is denoted by *E*(*X*) (μ_X) and is given by *E*(*X*) = $\sum_{i=1}^{n} x_i f(x_i)$ if *X* is a discrete RV and

$$= \int_{-\infty}^{+\infty} x f(x) dx \quad \text{if } X \text{ is a continuous RV}$$

Eg. (a) Consider rv *X* = max (*a*,*b*) where (*a*,*b*) is the outcome of tossing two fair dices. Then the pdf of *X* is given by (as shown above) Table 7.

Table 7 – Pdf of *X*

x_i	1	2	3	4	5	6
$f(x_i)$	1/36	3/36	5/36	7/36	9/36	11/36

Hence *E*(*X*) = μ_X = 1(1/36) + 2(3/36) + 3(5/36) + 4(7/36) + 5(9/36) + 6(11/36) = 4.47

Eg. If *X* is continuous rv with pdf *f*(*x*) where $f(x) = kx^2(1-x)$, 0< x < 1
= o otherwise

Then it can be shown that k =12 and *E*(*X*)=3/5 using the property of $\int_{-\infty}^{\infty} f(x).dx = 1$

Properties of *E*(*X*)

- If *c* is a constant then *E*(*c*) = *c*
- If *X* and *Y* are random variables such that $X \le Y$ the $E(X) \le E(Y)$
- $E(X + Y) = E(X) + E(Y)$
- $E(X + c) = E(X) + c$
- $E(aX) = aE(X)$

Note: Proofs are discussed in the class.

5.2. Variance of a distribution (σ^2) – *Var*(*X*)

$Var(X) = \sum_i (x_i - \mu)^2 f(x_i) = E[(X-\mu)^2]$, where $E(X) = \mu$ when X is discrete

$= \sum_i x_i^2 f(x_i) - \mu^2$ = $E(X^2) - [E(x)]^2$

$V(X) = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2$ when X is continuous

Thus using Table 7 above

$E(X^2) = 1^2 \times (1/36) + 2^2 \times (3/36) + 3^2 \times (5/36) + 4^2 \times (7/36) + 5^2 \times (9/36) + 6^2 \times (11/36)$

= 701/36 = 21.9

$E(X)$ = 4.47 (showed)   Thus $E(X^2)$ = 19.98

$V(X)$ = 21.9 – 19.88 = 1.99

Properties of *V*(*X*)

- $V(X+k) = V(X)$,
- $Var(aX) = a^2V(X)$

5.3 Standard deviation – σ

It is defined as the square root of variance. This indicator has more benefits than the variance in interpreting results.

Remark: Let Y be a random variable with mean μ ad standard deviation σ. Then the standardized random variable Z is defined as $Z = \frac{Y-\mu}{\sigma}$ so that that $V(Z) = 1$ and $E(Z) = 0$. This is very common transformation in statistics and would be very useful in all applications. More details are discussed later in the class.

5.4. Covariance between *X* and *Y*- γ_{xy}

If *X* and *Y* are two RVs then the extent to which two random variables vary together (co-vary) is measured by an indicator known as Covariance and it is denoted by Cov(*X*,*Y*) and given by

$$\text{Cov}(X,Y) = E\{[X - E(X)][Y - E(Y)]\}$$
$$= E(XY) - E(X)\times E(Y)$$

$$E(XY) = \sum x_i y_i h(x_i, y_i) - \mu_X \mu_Y \quad \text{if } X \text{ and } Y \text{ are discrete}$$

$$= \iint h(x, y)dxdy \ - \ \mu_X \mu_Y \quad \text{if } X \text{ and } Y \text{ are continuous}$$

Note:

Positive covariance: It indicates that higher than mean values of one variable tend to be paired with higher than mean values of other variable.

Negative covariance: It indicates that higher than mean values of one variable tend to be paired with lower than mean values of other variable.

Zero covariance: If the two random variables are independent then the covariance will be zero.

However, covariance is zero does not imply that two variables are independent

Note :

$V(X + Y) = V(X) + V(Y) + 2\text{Cov}(X,Y)$.

Example:

A pair of fair dice is tossed. Let $X = \max(a, b)$ and $Y = a + b$ where (a, b) is any ordered pair belongs to *S*.

Table 8

		Y											Sum
		2	3	4	5	6	7	8	9	10	11	12	
X	1	1/36	0	0	0	0	0	0	0	0	0	0	**1/36**
	2	0	2/36	1/36	0	0	0	0	0	0	0	0	**3/36**
	3	0	0	2/36	2/36	1/36	0	0	0	0	0	0	**5/36**
	4	0	0	0	2/36	2/36	2/36	1/36	0	0	0	0	**7/36**
	5	0	0	0	0	2/36	2/36	2/36	2/36	1/36	0	0	**9/36**
	6	0	0	0	0	0	2/36	2/36	2/36	2/36	2/36	1/36	**11/36**
Sum		1/36	2/36	3/36	4/36	5/36	6/36	5/36	4/36	3/36	2/36	1/36	

$h(3,5) = p(X=3 \text{ and } Y=5) = 2/36$

$E(XY) = 1\times2\times1/36 + 2\times3\times2/36 + 2\times4\times1/36 + \ldots\ldots\ldots\ldots +6\times12\times1/36$
$=1232/36 = 34.2$

$E(X) = \mu_X = 4.47, \quad \sigma_X = 1.4$

$E(Y) = \mu_Y = 7.0$ and $\sigma_Y = 2.4$

Thus Cov $(X,Y) = 2.9/(1.4) \times (2.4) = 2.9$

Properties of the Covariance

If *X*, *Y*, *W*, and *V* are real-valued random variables and *a*, *b*, *c*, *d* are constant ("constant" in this context means non-random), then the following can be proved easily.

<u>Notes:</u>

(a) Cov $(X, a) = 0$,
(b) Cov (X, Y) = Cov (Y, X)
(c) Cov $(aX, bY) = ab$Cov (X, Y)
(d) Cov $(X + a, Y+b)$ = Cov (X, Y)
(e) Cov $(aX+bY, cW+ dV) = c$Cov$(X,W) + ad$Cov$(X,) + bc$Cov$(Y, W) + bd$Cov(Y,V)

<u>Note:</u> Proofs of the axioms and their applications are discussed in the class

Limitation

Because of the number represent by Cov(X,Y) depends on the units of the data it is difficult to compare COV among different data sets, having different scales. As a solution to that another very useful indictor, known as correlation coefficient has been introduced.

5.5 Correlation Coefficient - ρ_{xy}

The correlation coefficient ρ_{xy} between two <u>random variables</u> *X* and *Y* with <u>expected values</u> μ_X and μ_Y and <u>standard deviations</u> σ_X and σ_Y is defined as:

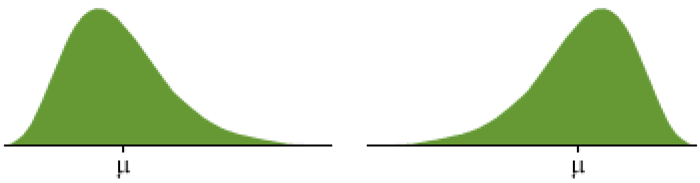
$$\rho_{xy} = \frac{\gamma_{xy}}{\sigma_x \sigma_Y} = \frac{E(x-\mu_X)(Y-\mu_Y)}{\sigma_x \sigma_Y} = \text{Corr}(X,Y)$$

$$\rho_{xy} = \frac{E(XY) - E(X)E(Y)}{\sqrt{E(X^2) - [E(X)]^2}\sqrt{E(Y^2) - [E(Y)]^2}}$$

5.6 Skewness -Sk: It is a measure of symmetry of a pdf. $Sk = [E(X-\mu_X)^3] \ / \sigma_X^3$

Skewness can come in the form of "negative skewness" or "positive skewness", depending on whether data points are skewed to the left (negative skew) or to the right (positive skew) of the data average.

1. **negative skew**: The left tail is longer; the mass of the distribution is concentrated on the right of the figure. It has relatively few low values. The distribution is said to be *left-skewed*.
2. **positive skew**: The right tail is longer; the *mass* of the distribution is concentrated on the left of the figure. It has relatively few high values. The distribution is said to be *right-skewed*.



(+) ve (right) skewed (-) ve (left) skewed

mode < median < mean mean < median < mode

Figure 4 Negative skewness and positive skewness

5.7 Kurtosis – K.

It is a measure of the tallness or flatness ("peakedness") of the pdf. It is a measure of whether the data are peaked or flat relative to a normal distribution. $K = [E(X - \mu_X)^4] \,/ \sigma_X^4$.

Estimation of the above population parameters based on a sample

Table 9

Parameter	**Population**	**Estimator from a sample**
Mean	μ	$\bar{x} = \sum_{i=1}^{n} x_i / n$
Variance	σ^2	$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$
Standard Deviation	σ	***s***
Covariance	γ	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)}$
Correlation coefficient	ρ	$r_{XY} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_X s_Y}$
Skewness	***Sk***	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^3}{(n-1)s^3}$
Kurtosis	***K***	$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^4}{(n-1)s^4}$

Note: More details with applications are discussed in the class.

6. DESCRIPTIVE STATISTICS (Statistical Indicators)

One important use of descriptive statistics is to summarize a collection of data in a clear and understandable way. Collected data may be in either ungrouped form or grouped form. In statistics there are various types of descriptive statistics. Such statistics are known as "statistical indicators". The statistical indicators play an important role in statistical data analysis.

6.1 Indicators to measure central tendency

Table 10

Parameter	Estimator	
	Ungrouped	Grouped
Arithmetic mean	$\bar{x} = \sum_{i=1}^{n} x_i / n$	$\bar{x} = \sum_{i=1}^{n} f_i x_i / \sum_{i=1}^{n} f_i$
Weighted mean	$\bar{x} = \frac{\sum_{i=1}^{n} w_i x_i}{\sum_{i=1}^{n} w_i}$ $= \sum_{i=1}^{n} \lambda_i x_i$ such that $\sum_{i=}^{n} \lambda_i = 1$	$\bar{x} = \frac{\sum_{i=1}^{n} f_i w_i x_i}{\sum_{i=1}^{n} w_i f_i}$
Median	$\frac{n+1}{2}$ *ranked observation*	$L_1 + \frac{N/2 - (\sum f)_1}{f_{median}}$ L_1=lower class boundary of the median class N=Total freq, c=size of med. class f_{med} = freq. of the median class $(\sum f)_1$ =sum of freq. of all classes lower than the median class
Mode	value of the data series that appears most frequently	$L_1 + (\frac{\Delta_1}{\Delta_1 + \Delta_2}) * c$

6.2 Indicators to Measure Dispersion

The indicators of dispersion are important for describing the spread of the data. Some of such indicators are as follows:

(a) **Range** Max(x_i) - Min(x_i)

(b) **Mean Deviation**: $MD = \frac{\sum (x_i - \bar{x})}{n}$

(c) **Mean Absolute Deviation**: MAD = $\frac{\sum_{i=1}^{n} |x_i - \bar{x}|}{n}$

(d) **Sample variance**: $s^2 = \frac{\sum (x_i - \bar{x})^2}{(n-1)} = \frac{\sum x_i^2 - n\bar{x}^2}{(n-1)}$

(e) **Standard deviation**: s = $\sqrt{\frac{\sum (x_i - \bar{x})^2}{(n-1)}}$

(f) **Standard error** $SE = \frac{s}{\sqrt{n}}$

6.3 Indicators to Measure Percentiles

The *p*th percentile is a value so that roughly *p* percentage of the data are smaller and (100-*p*) percent of the data are larger. A percentile is a measure of relative standing against all other observations in the sample.

First quartile (Q_1): The sample 25th percentile (P_{25})

Second quartile (Q_2): The sample 50th percentile (P_{50}) “median”

Third quartile (Q_3): The sample 75th percentile (P_{75})

Inter quartile range: $Q_3 - Q_1$

Desired Percentile: $L_p = (n+1)\frac{p}{100}$

6.4 Indicator to measure relative dispersion

Coefficient of Variation (*CV*) is a relative measure that indicates the magnitude of variation relative to the magnitude of the mean.

Coefficient of variation (*CV*) = $\frac{sd}{\bar{x}} * 100\%$

Eg. A manufacturer of television tubes has two types of tubes namely *A* and *B*. Mean life time tubes *A* and *B* are 1495 hrs and 1875 and *SD* of tubes are 280 hrs and 380 hrs respectively.

The CV of *A* = $\frac{280}{1495} * 100$ =18.7% and CV of *B* = $\frac{310}{1875} * 100$ = 16.9%

6.5 Indicator to measure linear association between variables

Correlation Coefficient is one of the most common and most useful statistical indicators to describe the association (degree of linear relationship) between two variables.

Pearson correlation coefficient

If *X* and *Y* are two random variables (<u>continuous</u>) following the same bivariate distribution and the paired values of *X* and *Y* of a sample of size *n* is (x_1, y_1), (x_2, y_2), ….. (x_n, y_n) then the correlation coefficient between *Y* and *X* is defined by

$$r_{XY} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{(x_i - \bar{x})^2 (y_i - \bar{y})^2}} = \frac{s_{XY}}{\sqrt{s_{XX} s_{YY}}} = \mathbf{r\ (say)} \quad -1 \leq r \leq +1$$

$$= \frac{\sum x_i y_i - n(\bar{x})(\bar{y})}{\sqrt{(x_i^2 - n\bar{x}^2)}\sqrt{(y_i^2 - n\bar{y}^2}}.$$

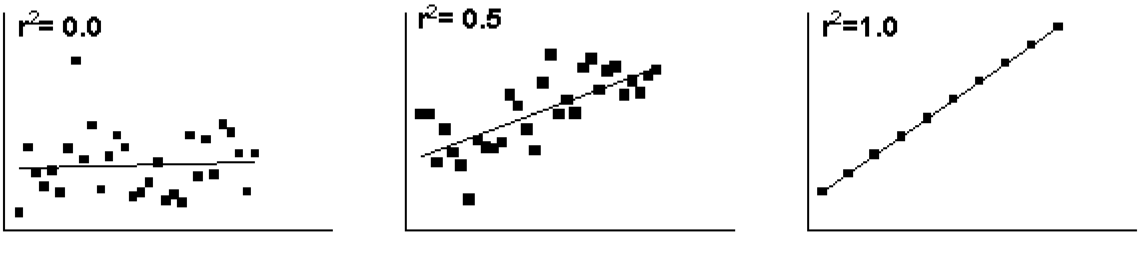


Figure 5

Table 11

number	*x*	*y*		
1	20	35		
2	25	45		
3	30	50		
4	35	65		
5	40	60		
6	45	65		
7	50	70		
8	60	65		

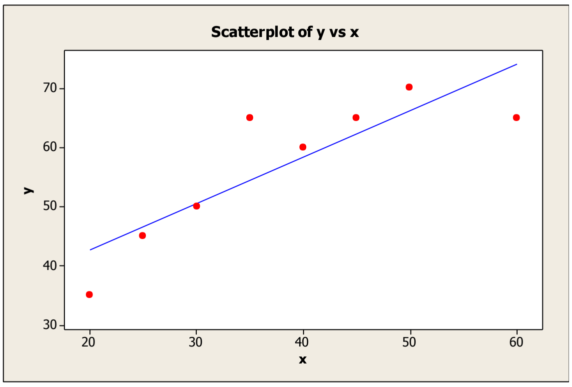


Figure 6

Note1:
Positive correlation: If *x* and *y* have a strong positive linear correlation, r is close to +1. An *r* value of exactly +1 indicates a perfect positive fit That is, the relationship between *x* and *y* variables is such that *y* increases as *x* increases.

Negative correlation: If *x* and *y* have a strong negative linear correlation, r is close to -1. An *r* value of exactly -1 indicates a perfect negative fit. Negative values indicate a relationship between *x* and *y* such that as values for *x* increase, values for *y* decrease.

No correlation: If there is no linear correlation or a weak linear correlation, *r* is close to 0. A value near zero means that there is a random, nonlinear relationship between the two variables.

Note 2: A correlation greater than 0.8 is generally described as *strong*, whereas a correlation less than 0.5 is generally described as *weak*. These values can vary based upon the "type" of data being examined and size of the sample. **THIS IS VERY SUBJECTIVE CRITERIA.**

7. Chebyshev’s Theorm and Empirical Rule

7.1 Chebyshev’s Theorm:

For any number $k > 1$, at least $1-\frac{1}{k^2}$ of the values for any distribution lie within k standard deviations of the mean.

$$P(-k\sigma \le X \le k\sigma) \le \frac{1}{\mathrm{k}^2}$$

7.2 Empirical Rule:

For a normal distribution (mode = median = mean):

68% of data lies within 1 standard deviation of the mean.

95% of data lies within 1 standard deviation of the mean.

99.7% of data lies within 1 standard deviation of the mean.

Ex.

a) According to Chebyshev’s theorem, at least what % of any set of observations will be within 1.8 standard deviations of the mean?

b) The mean income of a group of a sample observations is Rs. 500 and the sd=Rs. 40. According to Chebyshev’s theorem at least what % of the income will be lie between Rs. 400 and 600.

c) HebysIf a group of data has a mean of 54 and a standard deviation of 78.5, what is the interval that should contain at least 93.8% of the data?

d) Given a data set comprised of 4117 measurements that is bell-shaped with a mean of 862. If 99.7% of the data lies between 580 and 1144 then what is the standard deviation?

e) Given a group of data with mean 40 and standard deviation 15, at least what percent of the data will fall between 10 and 70?

8. Discrete (Binomial -*B*, Poisson- *P*) and Continuous (Normal- *N*. Exponential - *E*) Distributions

8.1. Binomial Distribution

The binomial distribution describes the behavior of a count variable if the number of observations *n* (say) is fixed and each observation has only two outcomes ("success" or "failure").

Further, each observation is assumed to be independent. The probability "success" = *p* is the same for each outcome. Then we say *Y* is distributed binomially.

ie. *Y* has a binomial distribution.

Examples.

- The engineer is interested in the number of break down buses in a sample of 100 lot.
- The doctor studies the number of survivors' vs deaths after treatment for a sample of 200 patients
- A teacher may interest how many heads occurs when by throwing 60 coins

The **conditions for** binomial distributions are:

- The experiment consists of n identical trials,
- each trial has only one of the two possible mutually exclusive outcomes, success or a failure
- The probability of each outcome does not change from trial to trial

If $Y \sim B(n, p)$ then $P(Y = r) = C_r^n (p)^r (1\text{-}p)^{n-r}$

p = Probability of success, *n* = number of trials

Ex. 1. A family has 6 children. Find the probability that there are (a) 3 boys and 3 girls and (b) fewer boys than girls. The probability that any child being a boy = ½.

Let Y = Number of boys in a family then $Y \sim B(6,1/2)$

(a) $P(\text{3 boys}) = C_r^6 \; (1/2)^3 \times (1\text{-}1/2)^3 = 5/16$

(b) $P(\text{fewer boys than girls}) = P(\text{no boys}) + P(\text{1 boy}) + P(\text{2 boys})$

$$= C_0^6 \; (1/2)^0 \times (1\text{-}1/2)^6 + C_1^6 \; (1/2)^1 \times (1\text{-}1/2)^5 + C_2^6 \; (1/2)^3 \times (1\text{-}1/2)^3$$

$$= 1/64 + 6/64 + 15/64 = 11/32$$

Ex. 2. A multiple choice test has four possible answers to each of 16 questions. A student guesses the answer to each question, i.e., the probability of getting a correct answer on any given question is 0.25. What is probability that at least 14 questions be correct?

Let Y = Number of correct answers. Then $Y \sim B(16,1/4)$

$P(Y>14) = P(Y=14) + P(Y=15) + P(Y=16)$

$$= C_{14}^{16}\,(1/40^{14}(3/4)^2 + C_{15}^{16}\,(1/40^{15}(3/4)^1 + C_{16}^{16}(1/40^{16}(3/4)^0$$

Properties of Binomial Distribution

Mean = np, Variance = $np(1\text{-}p) = npq$ where $p + q = 1$ and $SD = \sqrt{np(1-p)}$

Proof:: Let Y be a rv distributed $B(n,p)$

Expectation – $E(Y)$

$$\begin{aligned}
E(Y) &= \sum_{k=0}^{n} k \times P(Y=k) \\
&= \sum_{k=0}^{n} k \times \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=0}^{n} k \times \frac{n(n-1)!}{k(k-1)!(n-k)!} p^k (1-p)^{n-k} \\
&= \sum_{k=1}^{n} k \times \frac{n(n-1)!}{k(k-1)!(n-k)!} (p * p^{k-1})(1-p)^{n-k} \\
&= np \sum_{k=1}^{n} \frac{(n-1)!}{(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k} \\
&= np \sum_{s=0}^{m} \frac{(m)!}{(s)!(m-s)!} p^{s-1} (1-p)^{m-s} \quad (\, m = n - 1 \text{ and } s = k-1) \\
&= np \times 1 = np
\end{aligned}$$

Variance – $V(Y)$

$V(Y) \quad = \quad E(Y^2) - [E(Y)]^2$

$$E(Y^2) = \sum_{k=0}^{n} k^2 \times \frac{n!}{k!(n-k)!} p^k (1-p)^{n-k}$$

$$= np \sum_{k=0}^{n} k^2 \times \frac{(n-1)!}{k(k-1)!(n-k)!} p^{k-1} (1-p)^{n-k}$$

$$= np \sum_{s=0}^{n-1} (s+1) \times \frac{(n-1)!}{s!(n-s-1)!} p^s (1-p)^{n-s-1} \quad \text{(let s = k-1)}$$

$$= np \sum_{s=0}^{n-1} (s) \times \frac{(n-1)!}{s!(n-s-1)!} p^s (1-p)^{n-s-1} + np \sum_{s=0}^{n-1} \frac{(n-1)!}{s!(n-s-1)!} p^s (1-p)^{n-s-1}$$

$= np(n\text{-}1)p + 1 = np(np + q)$

[First term is the mean of $B(n\text{-}1, p)$ and second term is sum of probabilities of $B(n\text{-}1,p)$

$V(Y) = np(np+q) - n^2 p^2 = npq$]

<u>Note:</u>

If $X \sim B(n, p)$ and $Y \sim B(m, p)$ and X and Y are independent then $X + Y$ is also a binominal distribution with $(n+m, p)$ parameters.(Proof is not required)

8.2. Normal Distribution

The normal distribution is the most important family of continuous probability distributions in statistics which is widely applicable in all fields. The distribution is defined by two parameters, namely mean ("average", μ) and variance ("variability", σ^2). The normal distribution, is known as the Gaussian distribution and is denoted by $Y \sim N(\mu, \sigma^2)$.
PDF of the Normal distribution is given by

$$f_x(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(x-\mu)^2}{2\sigma^2}] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{[1]}$$

Standard normal distribution

In general all normal random variables are converted to the standard normal. If $X \sim N(\mu,\sigma^2)$, then $Z = \frac{X-\mu}{\sigma}$ is a standard normal distribution.

It can be shown that $E(Z) = 0$ and $V(Z) = 1$. Thus it is written as $Z \sim N(0,1)$.

An important consequence is that the distribution function of a general normal distribution is given by,

$$\frac{1}{\sqrt{2\pi}} \exp[-\frac{x^2}{2^2}] \quad \ldots\ldots\ldots\ldots\ldots\ldots\ldots \text{[2]}$$

One of the useful properties of the std. normal distribution is shown below.

Note: If $Y \sim N(\mu, \sigma^2)$ then $E(Y) = \mu$ and $V(Y) = \sigma^2$

Proof:

$$E(Y) = \int_{-\infty}^{+\infty} y^* \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(y-\mu)^2}{2\sigma^2}].dy$$

Let $t = \frac{y-\mu}{\sigma}$ then $dy = \sigma\, dt$

$$E(Y) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t + \mu) e^{-\frac{t^2}{2}} dt$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt + \mu \int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$= 0 + \mu$$

[it can be easily shown that $\int_{-\infty}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = 1$ and $\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt = 0$]

$$= \mu$$

$$E(Y^2) = \int_{-\infty}^{+\infty} y^2 * \frac{1}{\sqrt{2\pi\sigma^2}} \exp[-\frac{(y-\mu)^2}{2\sigma^2}].dy$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} (\sigma t+\mu)^2 e^{-\frac{t^2}{2}} dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} dt + \frac{2\mu\sigma}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t e^{-\frac{t^2}{2}} dt + \frac{\mu^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} e^{-\frac{t^2}{2}} dt$$

$$= \frac{\sigma^2}{\sqrt{2\pi}} \int_{-\infty}^{+\infty} t^2 e^{-\frac{t^2}{2}} dt + \mu^2$$

Integrating the above integral by parts let $u = t$ and $dv = t e^{\frac{-t^2}{2}} .dt$ it can be shown that above integral equals to σ^2.

Thus $E(Y^2) = \sigma^2 + \mu^2$ and $V(Y) = E(Y^2) - [E(Y)]^2 = \sigma^2$

Applications of Normal Distribution

(1) (a) $P(Z \le 1.2) = \int_{-\infty}^{1.2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$

$$= \int_{-\infty}^{0} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx + \int_{0}^{1.2} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$= 0. 5000 + 0.3849$

$= 8849$

(b) $P(Z \ge 1.13) = 1 - P(Z < 1.13)$

$$= 1 - \int_{-\infty}^{1.13} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1 - [0.5000 + \int_{0}^{1.13} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

$= 1 - [0.5000 + 0.3708] = 0.1293$

(c) $P(-1.37 \le Z \le 2.01) = P(-1.37 \le Z \le 0) + P(0 \le Z \le 2.01)$

$= P(0 \le Z \le 1.37) + P(0 \le Z \le 2.01) = 0.4147 + 0.4778 = 0.8925$

(2) Let T be the temperature (^oF) in May in a given year and distributed normally with mean 68 and SD 6. Find the probability that the temperature is between 70 & 80.

$T \sim N(68, 6^2)$

$$P(70 \le T \le 80) = P\left(\frac{70-68}{6} \le \frac{T-68}{6} \le \frac{80-68}{6}\right) = P(0.33 \le Z \le 2.0)$$

$$= P(0 \le Z \le 2.0) - P(0 \le Z \le 0.33) = \phi(2.0) - \phi(0.33)$$

$$= 0.4772 - 0.1293 = = 0.3479$$

(3) The radius of the nails of a sample of 800 is normally distributed with men 66 mm and variance 25. Find the number of nails with radius between 65 and 70 mm.

$$P(65 \le R \le 70) = P\left(\frac{65-66}{5} \le Z \le \frac{70-66}{5}\right) = P(-0.20 \le Z \le 0.80)$$

$$= P(-0.20 \le Z \le 0) + P(0.0 \le Z \le 0.8) = 0.0793 + 0.2881 == 0.3674$$

Required nails = 800 × 0.3674 = 294

Normal Approximation for Binomial

If n is large enough ($n > 30$) the skewness of the distribution is not too great and in such situation if $Y \sim B(n, p)$ then for large n, $Y \sim N(np, npq)$

Note: But to use the normal approximation to calculate this probability, we apply the ***continuity correction*** in converting discrete to continuous variable.

Ex.

1. A fair die is tossed 180 times. Find the probability that the face 6 will appear between 29 and 32 inclusive.

Let Y = No. of times six appear

Thus $Y \sim B(180, 1/6)$

Using normal approximation to Binomial distribution,

$Y \sim N(np, npq)$ where n = 180, $p = 1/6$ and $q = 1\text{-}p = 5/6$

$E(Y) = np = 180 \times 1/6 = 30$, $V(Y) = npq = 180 \times 1/6 \times 5/6 = 25$

$SD(Y) = 5$

$$P\,(29 \le Y \le 32) = P\left(\frac{29-30}{5} \le \frac{Y-30}{5} = Z \le \frac{32-30}{5}\right)$$

$$= P\,(-0.2 \le Z \le 0.4) = = 0.1554 + 0.0793 = 0.2347$$

With continuity correction we are interested

$$P\,(28.5 \le Y \le 32.5) = P\left(\frac{28.5-30}{5} \le \frac{Y-30}{5} = Z \le \frac{32.5-30}{5}\right)$$

$$= P\,(-0.3 \le Z \le 0.5)$$

$$= 0.1179 + 0.1915 = 0.3094$$

2. Consider a group of voters in a first year undergraduates in the University of Moratuwa. The true proportion of voters who favor candidate A is equal to 0.40. Given a sample of 200 voters, what is the probability that more than half of the voters support candidate A?

3. The grades on a short quiz in statistics were 0, 1, 2, .. ,10 points depending on the number of 10 questions. The mean grade = 6.7 and sd =1.2. Assuming the grades to be normally distributed determine (a) % of students scoring 6 points and (b) maximum grade of the lower 10% and (c) the minimum grade of highest 10% in the class.

8.3. Poisson Distribution

The Poisson distribution is also a discrete distribution which is used to model the number of events occurring within a given time interval.

Examples: (1) Y = number of accidents between time a and time b .

Since Poisson is a discrete distribution, the probability distribution of Poisson variable is given by,

$$P(Y = y) = \frac{e^{-\lambda}\lambda^{y}}{y!} \text{ for } y = 0,1,2, \text{ and is denoted by } Y \sim P(\lambda)$$

λ is the average number of events in the given time interval and it is known as the "shape parameter of the distribution".

<u>Properties of Poisson distribution</u>

If $Y \sim P(\lambda)$ then $E(Y) = \lambda = V(Y)$

<u>Proof:</u>

$$\boldsymbol{E(Y)} = \sum_{K=0}^{\infty} k \times \frac{\lambda^{K} e^{-\lambda}}{k!} = \lambda \sum_{K=1}^{\infty} \frac{\lambda^{k-1} e^{-\lambda}}{(k-1)!} = \lambda \sum_{s=0}^{\infty} \frac{\lambda^{s} e^{-\lambda}}{s!} \quad \text{(let k-1 = s)}$$

$$= \lambda \times 1 \ \left(\sum_{s=0}^{\infty} \frac{\lambda^{s} e^{-\lambda}}{s!} = 1\right) = \lambda$$

$$\boldsymbol{E(Y^2)} = \sum_{K=0}^{\infty} k^2 \times \frac{\lambda^{K} e^{-\lambda}}{k!} = \lambda \sum_{K=1}^{\infty} k \times \frac{\lambda^{K-1} e^{-\lambda}}{(k-1)!} = \lambda \sum_{s=0}^{\infty} (s+1) \times \frac{\lambda^{s} e^{-\lambda}}{s!} \quad (k\text{-}1 = s)$$

$$= \lambda \sum_{s=0}^{\infty} s \times \frac{\lambda^{s} e^{-\lambda}}{s!} + \lambda \sum_{s=0}^{\infty} \frac{\lambda^{s} e^{-\lambda}}{s!} = \lambda E(Y) + \lambda * 1 = \lambda^2 + \lambda$$

$$\boldsymbol{V(Y)} = E(Y^2) - E(Y)]^2 = \lambda^2 + \lambda - \lambda^2 = \lambda$$

<u>Relation between $B(n, p)$ and $P(\lambda)$</u>

In Binomial if n is large and p is close to 0 then the event is called a rare event. In practice we shall consider an event to be rare if $n \geq 50$ while np close to 5. In such situations Binomial is closely approximated by $\lambda = np$.

<u>Note:</u>

Since there is a relationship between Binomial & Normal there is a relation between Poisson and Normal. It is given by ,

$$P(\lambda) \approx N(\lambda, \lambda) = N(\mu, \sigma^2)$$

Eg. 1. If a probability that an individual suffer a bad reaction from injection of a given type is 0.001 determine the probability that out of 2000 individuals (a) exactly 3 and (b) more than 2 will suffer bad reaction.

Let Y = number of individuals suffer bad reaction

As N is large and p is small it can be assumed that $Y \sim P(\lambda)$ where $\lambda = np$ =2000*0.001 = 2.

$P(Y=3) = \frac{2^3 e^{-2}}{3!}$ = 0.1804

Assuming $Y \sim B$ (2000, 0.001), $P(Y=3) = {}^{2000}C_3 \times (0.001)^3(1-0.001)^{1997}$ = 0.1805

8.4. Exponential Distribution

This is continuous distribution. This is useful in many occasions in particularly in business. It is widely used in waiting line (or queuing) theory to model the length of time between arrivals in process such as customers at Bank ATMs, customers at a fast food restaurant, patients entering to a accident ward. It is defined using a single parameter λ.

Thus if $X \sim$ Exp(λ) then $f(x) = \lambda e^{-\lambda}$.

H/E: Prove that $E(X) = \frac{1}{\lambda}$ and $V(X) = \frac{1}{\lambda^2}$

Ex. Suppose that customers arrive at bank's ATM at the rate 20 per hour. If a customer has just arrived, what is the probability that the next customer arrive within 6 minutes ?

$X \sim E(\lambda) = E(20)$

$P(X < 0.1) = 1 - e^{-20(0.1)}$ =0.8647

Note: More exercises on applications are done in the class

9. STATISTICAL INFERENCES

It is not possible to find parameters (mean, variance etc) of a population due to obvious reasons. Thus we have to compute a value (or range) that represents a ``good" guess for the true values of the parameter to make conclusions (inferences) on the population based on sample. There are two types of estimators namely

(a) Point estimator and

(b) interval estimator

9.1 Point Estimator

The point estimate of a parameter is a single number based on the sample data that we can consider to be the most plausible value of parameter.

9.2 Interval estimator

An estimate of a population parameter given by two numbers at a given confidence is called the interval estimator.

For example:

We want to know the average salary of chemical engineering graduate. So we selected 25 people at random. The mean annual income is 60,000/=. This is a point estimate. Using an interval estimate we say that the mean annual income is between 40,000 and 85,000/= with 95% confidence.

How do we judge the confidence?

Let θ *and* σ^2 be the population mean and variance (irrespective of the distribution). Then these two parameters are estimated by

$$\hat{\theta} = \text{sample mean} = \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \text{ and } \hat{\sigma}^2 = sample\,\text{var}\,aince = s^2 = \frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n}$$

How accurate these estimators?

Basically if we know the population parameters the sample estimates should be very close to the population parameters. Let θ be the population mean. Then we want $\theta - \bar{x} \rightarrow 0$ if $\bar{x}$ the mean of the sample. But difference samples can be obtained. So $\bar{x}$ is random variable. Thus we want $\theta - E(\bar{x}) \rightarrow 0$

Note: The difference is known as the 'bias' of the parameter estimated based on sample.

Thus to obtain more precise estimator for the population parameter "bias" should be zero. That is estimator should be unbiased.

9.3 Unbiased estimator

If a statistic τ is used as an estimator for population parameter θ and $E(\tau) = \theta$ then τ is said to be as an unbiased estimator.

Note 1. $\bar{x}$ is an unbiased estimator for the population mean μ.

Proof: $E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right) = \frac{1}{n}\sum_{i=1}^{n} E(x_i) = \frac{1}{n}\sum_{i=1}^{n}\mu = \mu$

Note 2: Sample variance s^2 is not an unbiased estimator for the population variance σ^2

Proof: $s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n} = \frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n}$

$$E(s^2) = E\left(\frac{\sum_{i=1}^{n} x_i^2 - n\bar{x}^2}{n}\right) = \frac{E(\sum_{i=1}^{n} x_i^2)}{n} - \frac{E(n\bar{x}^2)}{n} = \frac{\sum_{i=1}^{n} E(x_i^2)}{n} - \frac{nE(\bar{x}^2)}{n}$$

$$= \frac{\sum V(x_i) + \sum [E(x_i)]^2}{n} - \frac{n[V(\bar{x}) + \{E(\bar{x})\}^2}{n} = \frac{n\sigma^2 + n\mu^2}{n} - \frac{n\frac{\sigma^2}{n} + n\mu^2}{n}$$

($\because$ Central Limit Theorem)

[V($\bar{x}$) =V$\left(\frac{\sum x_i}{n}\right) = \frac{1}{n^2}\sum V(x_i) = \frac{(n-1)\sigma^2}{n}$. Hence $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n}$ is not an unbiased estimator for σ^2.

Thus E[$\sum_{i=1}^{n}(x_i - \bar{x})^2$] = $(n-1)\sigma^2$

Thus $\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{(n-1)}$ is an unbiased estimator for the population variance σ^2.

9.4 Central Limit Theorem (without proof)

For large n (n> 30) the distribution of mean is approximately normal with mean μ and variance $\frac{\sigma^2}{n}$ (irrespective of population and mean and variance are finite). Thus, the Central Limit theorem is the foundation for many statistical procedures. The distribution of an average will tend to be Normal as the sample size increases, regardless of the distribution from which the average is taken.

9.5 Confidence Interval

Instead of point estimator we compute range to represent parameter of the population. We construct this interval with some confidence. Thus it is known as, "confidence interval". That is, a range within which the population value is likely to fall. "Likely" is usually taken to be "95% of the time," and the range is called the **95% confidence interval**. The values at each end of the interval are called the **confidence limits.**

Confidence Interval for mean under Normal

<u>Case 1:</u> σ is known

To explain how confidence intervals are constructed, we are going to work backwards and begin by assuming characteristics of the population.

Let $X \sim N(\mu, \sigma^2)$ then $\overline{X} \sim N(\mu, \frac{\sigma^2}{n})$

$$Z = \frac{\overline{X} - \mu}{\sigma / \sqrt{n}} \sim N(0,1)$$

Based on standard normal assumption we know that an actual sample statistic lying in the interval $\mu \pm \sigma$, $\mu \pm 2\sigma$ and $\mu \pm 3\sigma$ ($\mu = 0$ and $\sigma^2 = 1$) about 68%, 95% and 99% confidence. Thus we use this concept to compute CI for mean instead of point estimator. As $P(-1.96 \leq Z \leq +1.96) = 95.0\%$ (as shown below)

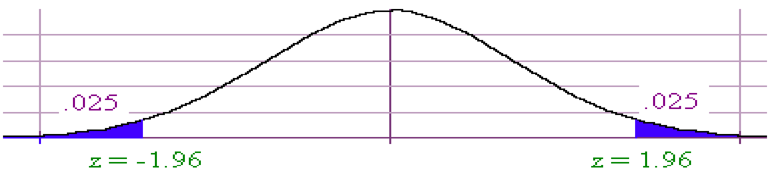


Figure 7

$$P(-1.96 \le \frac{\overline{X}-\mu}{\sigma/\sqrt{n}} \le +1.96) = 95\% \Longrightarrow P(\mu - 1.96\frac{\sigma}{\sqrt{n}} \le \overline{X} \le \mu + 1.96\frac{\sigma}{\sqrt{n}}) = 95\%$$

Thus 95% CI for mean is given by $\overline{X} \pm 1.96\frac{\sigma}{\sqrt{n}}$

So 99% CI for mean is given by $\overline{X} \pm 2.58\frac{\sigma}{\sqrt{n}}$ and 90% CI for mean is given by $\overline{X} \pm 1.65\frac{\sigma}{\sqrt{n}}$

Thus α **%** CI for mean (when σ is known) is given by $\overline{X} \pm \lambda_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

Example 1: Suppose that we found that the mean mark (out of 20) of 50 students in Mid-term test is 12 with a standard deviation of 6. What can we conclude about the average marks of students with a 95% confidence level?

In this example n (> 30) is large and we can approximate to normal.

Let X = marks of students

Assume X~ N(μ,σ^2) and thus 95% CI for the population mean is given by

$$\bar{x} \pm 1.96\frac{\sigma}{\sqrt{n}} = 12 \mp 1.96 \times \frac{6}{\sqrt{50}} = 12 \mp 1.66 = [10.35 - 13.66] = [10,14]$$

Thus we are 95% confidence that interval for the mean marks is between 10.35 and 13.66. We can also say that $\pm$1.66 is the **margin of error**.

Example 1. The blood cholesterol levels of a population of teachers have mean 202 and SD 14. If a sample of 30 teachers is selected approximate the probability that the sample mean of their blood cholesterol level will lie between 198 and 206. Repeat it for sample size of 64 (Class Exercise)

Case 2: Confidence Interval when population variance is not known

When σ is unknown (i.e. σ estimated by a sample variance, s) and thus 100(1- α)% CI for mean is given by $\overline{X} \pm \lambda_{\alpha/2} \frac{s}{\sqrt{n}}$

Case 3: Confidence Interval when population variance is not known and *n* < 30

When σ is unknown (i.e. σ estimated by a sample variance) and sample size is small (< 30) α % CI for mean is given by $\overline{X} \pm t_{\alpha/2,n-1} \frac{s}{\sqrt{n}}$

Where $t_{\alpha/2,n-1}$ = α % value of the t distribution at n-1 degree of freedom

Note : In this case it is assumed that $\frac{\bar{x}-\mu}{s/\sqrt{n}}$ **~** t_{n-1}

Table 12 Summary - Selecting confidence interval for means

Variance	Sample size	95% confidence interval
σ^2 known	Large or small	$\bar{x} \pm 1.96 \frac{\sigma}{\sqrt{n}}$
σ^2 unknown	Large	$\bar{x} \pm 1.96 \frac{s}{\sqrt{n}}$
σ^2 unknown	Small	$\bar{x} \pm 2.54 \times t_{\alpha,n-1} \times \frac{s}{\sqrt{n}}$

Eg. Given the following GPA for 6 students: 2.80, 3.20, 3.75, 3.10, 2.95, 3.40. Calculate a 95% confidence interval for the population mean GPA.

Ans: 3.2 $\pm 2.57 \times \frac{0.339}{\sqrt{6}}$ = [2.84 – 3.56] (based on case 3)

If we assume normal the CI = 3.2 $\pm$ 1.96$\times \frac{0.339}{\sqrt{6}}$ = [2.93 – 3.47]

Note: You can try using Minitab software

10. POINT ESTIMATOR AND CONFIDENCE INTERVAL FOR PROPORTION

Suppose we desire to estimate of *p*, the proportion of an event in a population based on a sample size of *n*. Let *X* = number of successes in a sample. Then the estimator for the proportion of success(*p*, say) can be obtained by $\hat{p} = \frac{\mathrm{x}}{\mathrm{n}}$

$X \sim B(n, p) \rightarrow X \sim N(np, npq)$ where $q = 1-p$

$E(\frac{x}{n}) = \frac{1}{n}E(x) = \frac{np}{n} = p$ Thus $\hat{\mathrm{p}}$ is an unbiased estimator for p.

$V(\frac{x}{n}) = \frac{1}{n^2}V(x) = npq/n^2 = \frac{pq}{n}$

Assuming $Y = X/n \sim N(p, p\times(1-p)/n)$ CI for the proportion is given by

$$\hat{p} \pm \lambda_{\alpha/2} SD(\hat{p}) = \hat{p} \pm \lambda_{\alpha/2} \times \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}.$$

Example: Out of random sample of 100 boxes coming from a particular machine, 82 were non defective. Construct the 99% CI estimate of the proportion of non defectives.

Estimator for non defective = $\hat{p} = \frac{82}{100}$

Required CI = $0.82 \pm 2.576 \times \sqrt{\frac{0.82(1-0.82)}{100}}$ = [72.1, 91.9]

Example. The Ceylon Daily News reported that a poll in Jaffna 46% of the population was in the favour of the present paddy prices with a margin error of ±3%. How many people were questioned?

Margin error = $\pm \lambda_{\alpha/2} \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$,

Thus $1.96^2 \times \frac{0.46 \times (1-0.46)}{n} = 0.03^2$ $n = 1060$

Example: A sample pole of 100 voters chosen at random from all voters in a given district indicated that 55% of them were in favor of a particular candidate. Find the (a) 95% and (b) 99% confidence limits for the proportion of all the voters in favor of this candidate. How larger sample of voters should we take in order to be 99% confident that the candidate will be elected.

Home Exercises (Tutorial) – Applying Concepts

1. Using the company records of last 500 working days the Manager of the company has summarized the number of cars sold per day as follow.

Number of cars sold/day	Frequency
0	40
1	100
2	142
3	66
4	36
5	30
6	26
7	20
8	16
9	14
10	8
11	2

(a) Construct the pdf and cdf of the number of cars sold (say, Y).

(b) Compute the expected value of Y and SD

(c) Find the number of cars sold less than or equals to 50%?

2. Prove the followings.

If X and Y are two continuous rvs then

(a) $E(X+Y)=E(X)+E(Y)$

(b) $\text{Cov}(aX+bY, cU+dV)=ac\text{Cov}(X, U) + ad\text{Cov}(X, V)+ bc\text{Cov}(Y, U) + bd\text{Cov}(Y,V)$

(c) $V(Y) = pq$ if the rv Y takes the value 1 and 0 with probabilities p and q respectively.

3. The following data are the estimated market value (in Rs. 100,000) of 50 companies.

26.8	8.6	6.5	30.6	15.4	18.0	7.6	21.5
11.0	10.2	28.3	15.5	31.4	23.4	4.3	20.2
33.5	7.9	11.2	1.0	11.7	18.5	6.8	22.3
12.9	29.8	1.3	14.1	29.7	18.7	6.7	31.4
30.4	20.6	5.2	37.8	13.4	18.3	27.1	32.7
6.1	.9	9.6	35.0	17.1	1.9	1.2	16.6
31.1	16.1						

(a) Determine the mean, standard deviation and the median of the market values and interpret.

(b) Using the empirical rule about 95% of the values would occur between what values.

(c) Determine the coefficient of variation and interpret.

(d) Estimate Q1 and Q3 values and interpret.

(e) Draw a Box plot and write brief report of the variability of data.

4. Consider the following joint distribution of X and Y.

X	Y				Total
	-2	-1	4	3	
1	0.1	0.2	0.0	0.3	0.6
2	0.2	0.1	0.1	0.0	0.4
Total	0.3	0.3	0.1	0.3	

Find $E(X)$, $E(Y)$, Cov(X,Y) and ρ_{XY} .

3. The portfolio expected return and portfolio risk of two asset investments X and Y is given by $E(P) = wE(X) + (1-w)E(Y)$ and $D(P)=\sqrt{w^2V(X)+(1-w)^2V(Y)+2w(1-w)Cov(X,Y)}$. In two investments X and Y $E(X)$=Rs. 50, $E(Y)$=Rs. 100, $V(X)$=9000, $V(Y)$=15,000 and Cov (X,Y)=7500. The weight assigned to investment X of portfolio asset is 0.4. Compute portfolio expected mean and risk.