Chapter 1.

Date: 1-27 Page:

Q. Define statistics. Write its advantage and disadvantages.

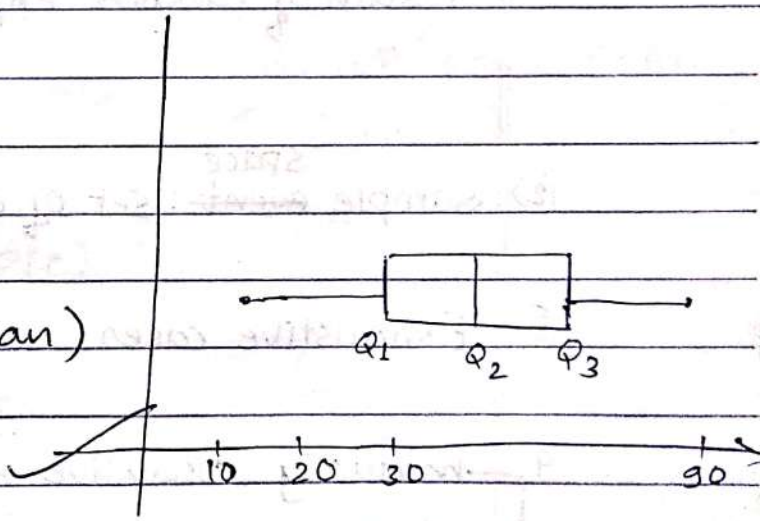
Q. Define and distinguish between diagrammic and graphical representation of data.

Q. Write the advantange and disadvantage of diagram and graph.

Q. Define box plot. Write its features. Take a individual frequency distribution of your own choice which consists at least 20 items and present the information with the help of box plot.

Box plot

1) minimum value
2) maximum value
3) First quartile
4) Second quartile (median)
5) Third quartile



Q. Define and distinguish between various measure of central tendency and measure of dispersion. Write the advantage and disadvantage of each

Date: Page:

Q. Probability

$$P(E) = \frac{n(E)}{n(S)} = \frac{m}{n} \quad \text{where } m \leq n$$

Some termoinologies

① Random experiment
→ result cannot be predicted.
eg: tossing a coin. [possible outcome is known]
result of random experiment: events or cases
Simple event & compound event

② Sample ~~event~~ space : set of all possible outcomes

③ Exhaustive cases: cardinal number of sample space.

4. mutually exclusive events:
and non mutually exclusive events → doesnot necess. ~~exp~~ exclude
→ happening of one event completly excludes other

5. Dependent and independent events
↳ second result depends upon the previous one.

Date: Page:

Addition theorem of probability

1. If A and B are non mutually exclusive events then,
$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

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

If A, B and C are non mutually exclusive then,
$$P(A \cup B \cup C) = P(A) + P(B) + P(C) - P(A \cap B) - P(B \cap C) - P(A \cap C) + P(A \cap B \cap C)$$

If mutually exclusive,
$$P(A \cup B \cup C) = P(A) + P(B) + P(C)$$

If A, B, and C are independent
$$P(A \cup B \cup C) = 1 - P(\overline{A \cup B \cup C})$$
$$= 1 - P(\bar{A} \cap \bar{B} \cap \bar{C})$$
$$= 1 - P(\bar{A}) \times P(\bar{B}) \times P(\bar{C})$$

It can be used upto n number of events also.

Date: Page:

Multiplication theorem of probability

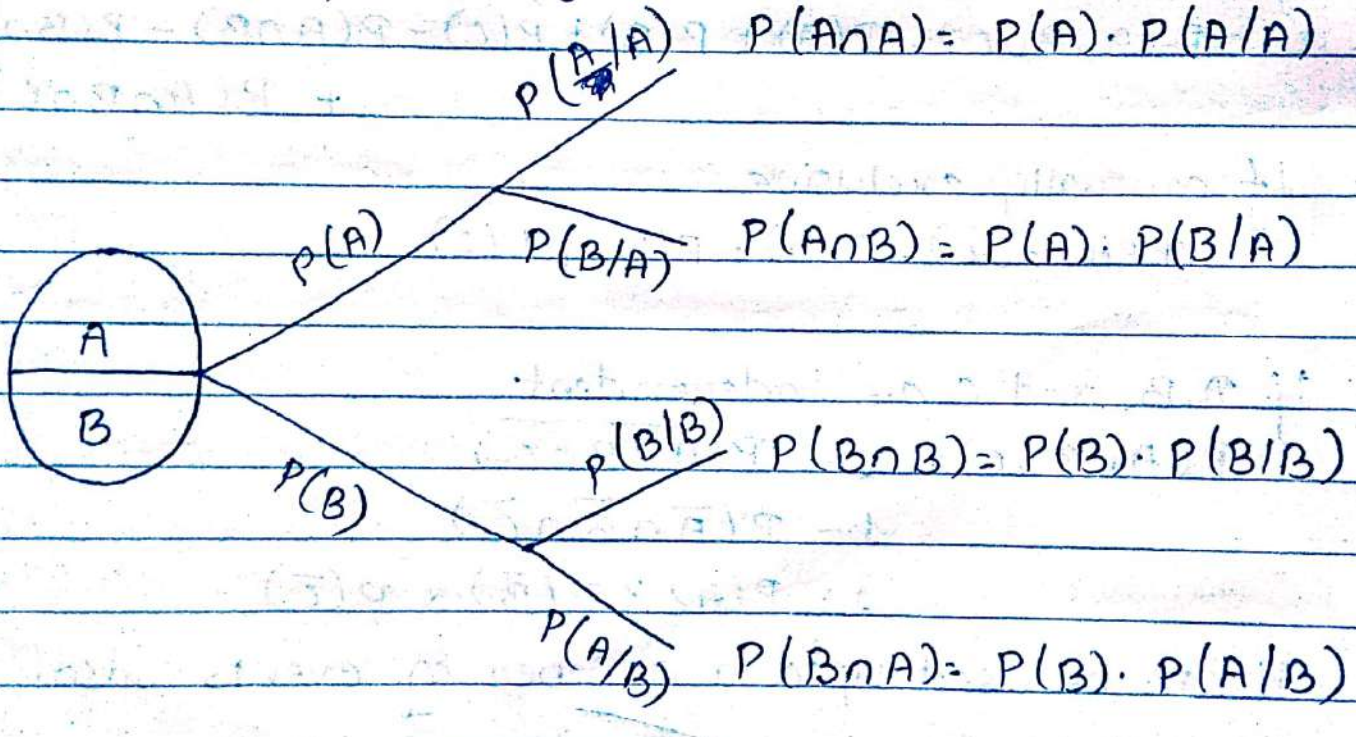
If $A_1, A_2 \ldots A_n$ are independent events then,

i. $P(A_1 \cap A_2) = P(A_1) \times P(A_2)$

ii. $P(A_1 \cap A_2 \cap A_3) = P(A_1) \times P(A_2) \times P(A_3)$

iii. $P(A_1 \cap A_2 \cap \ldots \cap A_n) = P(A_1) \times P(A_2) \times \ldots \ldots \ldots \times P(A_n)$

Conditional probability:



$$P(B \cap A) = P(B) \cdot P(A/B)$$

$$P(A/B) = \frac{P(B \cap A)}{P(B)}$$

If A and B are two different elements, then probability of happening of A when B is already happened is ~~defi~~ denoted by $P(A/B)$ is defined as the ratio of probability of simultaenously happening

Date: Page:

of A and B to the probability of happening of B, This is called conditional probability of A when B has already happened.

Mathematically, $P(A/B) = \frac{P(B \cap A)}{P(B)}$

Answers:

Date: Page:

Qno1: Statistics can be defined as the science which deals with collection, classification and tabulation of numerical facts as the basis for explaination, description ~~and classification~~ and comparison of phenomenon.

The advantages of statistics are:

a. design of experiment uses statistical technique to test and construct models of engineering components and techniques.

b. Quality control and process control use statistical techniques to ~~test and construct~~ manage conformance to specifications of manufacturing processes and their products

c. Time and methods engineering use statistics to study repeatitive operations in manufacturing in order to set standards and find optimum manufacturing procedures.

d. Reliability engineering uses statistics to measure the ability of the system to perform for its intended function and has tools for improving performance.

The limitations of statistics are:

a. As the statistics deals with the masses, it is not applicable for the single observation.

b. Statistics deals with the quantitative data only.

c. Statistical laws are only approximate and are not exact.

Due to this statistical conclusions are not universally true.

d. Statistics is liable to be misused.

e. Statistics is only a mean to draw conclusion about masses or population.

Qno2. A visual form of presentation of data in which facts are highlighted in the language of diagrams is known as diagramatic representation of data.

A visual form of presentation of data in which facts are highlighted in the graph is known as graphical representation of data.

Diagramatic representation	Graphical representation
i. In this representation, diagrams are constructed on plain paper.	i. In this representation, graphs are constructed on graph paper.
i. Diagrams may be of one, two or three dimensional.	ii. Graphs are generally of two dimensional.
ii. In diagrams, the numerical data are presented by bars, circles, cubes, rectangles etc.	iii. In graphs, data are presented in terms of points and lines.
iv. Presentation of frequency distribution in diagrams is not used.	iv. Presentation of frequency distribution and time series in graph is used.

Q. The advantages of diagram and graph are:

a. A clear picture of the variation in the values of a variable is much more easily obtained by diagram and graphs than the value given in table.

b. They can be easily understood by a ~~simple~~ common person.

c. They give delight to the eye and leave an ever lasting impression on the mind.

d. They give required information in less time without any mental strain.

e. They facilitates comparison of two or more sets of numerical data.

The disadvantages of graph are:

a. Graphical representation involves more time as it requires graph and figures.

b. Since, graphical representations are complex, there is each and every chance of error and mistake.

c. All may not be able to get the meaning of graphical representation

d. Graphical representation of reports are costly.

The disadvantages of diagram are:

a. Classified and tabulated data provides more information than diagrams.

b. Diagram offers a low level of precision of values.

c. Diagram donot allow the user to analyze the data further.

d. Diagram tends to potray only a limited number of characteristics.

Date: Page:

Q. Box plot: A box plot is a graphical representation of the data that displays five number summary of data set based on the minimum value $X_{smallest}$, lower quartile Q_1, upper quartile Q_3 and maximum value $X_{largest}$ of the data on a rectangular box aligned either horizontally or vertically.

Suppose the given data are:

11 15 23 ~~29~~ 29 19 22 21 20 15 25 17

Firstly, arrange the data in ascending order as:

11 15 15 17 19 20 21 22 23 25 29

Here, $Q_1 = 15$ $Q_2 = 20$ $Q_3 = 23$

The five number summaries are: 11, 15, 20, 23 and 29

Q. The difference between measure of central tendency and measure of dispersion are:

Measure of central tendency	Measure of dispersion
Measure of central tendency is a single value within the range of the data which represents a group of individuals values in a simple and consise manner.	Measure of dispersion is a descriptive statistical measure used to measure the variation or spread or scatterings in the data set.
It is also known as measure of location.	It is also known as measure of variation or measure of variability or measure of spread.

The various measure of central tendency are:

(1) mean

The advantage of using mean are:

a. mean is easy to understand and calculation is simple.
b. mean is rigidly defined and based on all the observation.
c. It is affected least by fluctuations of samplings.

The disadvantages of using mean are:

a. It is very much affected by extreme values.
b. It cannot be calculated by inspection and graphically
c. It cannot be used in open ended class.

2. median

a. It is rigidly defined
b. It is easy to understand and easy to calculate
c. It is a positional average and hence it is not affected by extreme values.

Disadvantage

a. Arrangement of data according to the magnitude is necessary.
b. It is not based on all observation.
c. It is not suitable for further mathematical treatment.

Date: Page:

3. Mode.

Advantages:

a. It can be obtained in open end classes.
b. It can be located by inspection or by graph
c. It is not affected by extreme values or set of observation.

Disadvantage.

a. It is ill defined ie. not rigidly defined.
b. It is not based on all the observation

The various measure of dispersion are:

1. Absolute measure
   a. Range
   b. Quartile deviation
   c. Standard deviation
   d. mean deviation or average deviation

2. Relative measure
   a. Coefficient of range
   b. coefficient of quartile deviation
   c. Coefficient of mean deviation
   d. Coefficient of variation.

Absolute measure

i. Range

Advantages:

- It is rigidly defined
- Only minimum time is required to know the variability with the help of range.

Disadvantages

- It is not based on all observation
- It cannot be calculated for the frequency distribution having open ended class.

ii. Quartile deviation

Advantages

- Since, it includes the lowest and highest 25% values it is not affected by extreme observations.
- It is better measure than range because it includes 50% values.

Disadvantages

- It is based on two positional values Q_1 and Q_3 and ignores the extreme 50% of the item.
- It is affected by fluctuating of samplings.

iii. Mean deviation

Advantages

- It is improved method than range and quartile deviation
- It is flexible because it can be calculated from any averages.

Disadvantages:

- In mean deviation, the algebric negative signs of the deviation are ignored which is mathematically unsound.
- When mode is ill defined, it is difficult to calculate MD taken from mode.

iv. Standard deviation

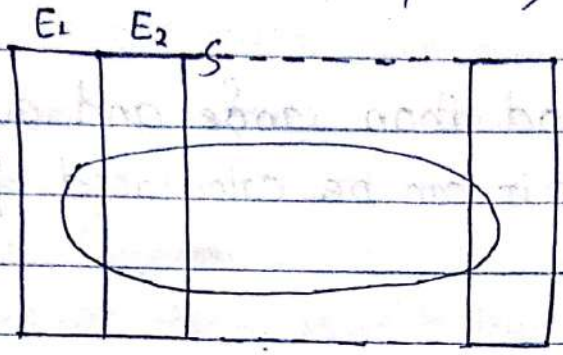
Advantages

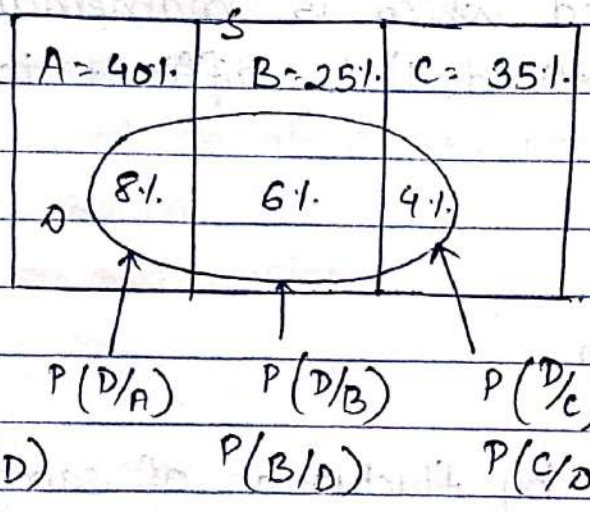
- It is least affected by fluctuation of sampling than other.
- It is suitable for further mathematical treatment.

Disadvantages

- It is not easy to calculate.
- It can't be calculated for open end classes.

VVI Baye's theorom: (State and prove)





Let $E_1, E_2 \dots E_n$ are n mutually exclusive ~~cases~~ and exhaustive cases such that $P(E_1) \neq 0$ for $n = 1, 2 \dots n$. If D is a set such that $D \subset \bigcup_{i=1}^{n} E_i$ and $P(D) \neq 0$ then the probability of happening of any event E_1 when D is already happened is denoted by $P(E_1/D)$ and given by

$$P(E_1/D) = \frac{P(E_1) \times P(D/E_1)}{\sum_{i=1}^{n} P(E_1) \times P(D/E_1)}$$

This is known as Baye's theorom for conditional probability.

Proof:

from figure ①

$$D = (E_1 \cap D) \cup (E_2 \cap D) \cup \cdots (E_n \cap D)$$

or, $P(D) = P\{(E_1 \cap D) \cup (E_2 \cap D) \cup \cdots (E_n \cap D)\}$

Since, $E_1 \cap D$, $E_2 \cap D \cdots$, $E_n \cap D$ are mutually exclusive so,

So,

$$P(D) = P(E_1 \cap D) + P(E_2 \cap D) + \cdots P(E_n \cap D)$$
$$= P(E_1) \cdot P(D/E_1) + P(E_2) \cdot P(D/E_2) + \cdots$$
$$\cdots P(E_n) \cdot P(D/E_n)$$

$$= \sum_{i=1}^{n} P(E_1) \cdot P(D/E_1)$$

We know, $P(D \cap E_1) = P(E_1 \cap D)$

or, $P(D) \cdot P(E_1/D) = P(E_1) \cdot P(D/E_1)$

$$P(E_1/D) = \frac{P(E_1) \cdot P(D/E_1)}{P(D)}$$

$$P(E_1/D) = \frac{P(E_1) \cdot P(D/E_1)}{\sum_{i=1}^{n} P(E_1) \cdot P(D/E_1)}$$

Hence, the theorem is proved.

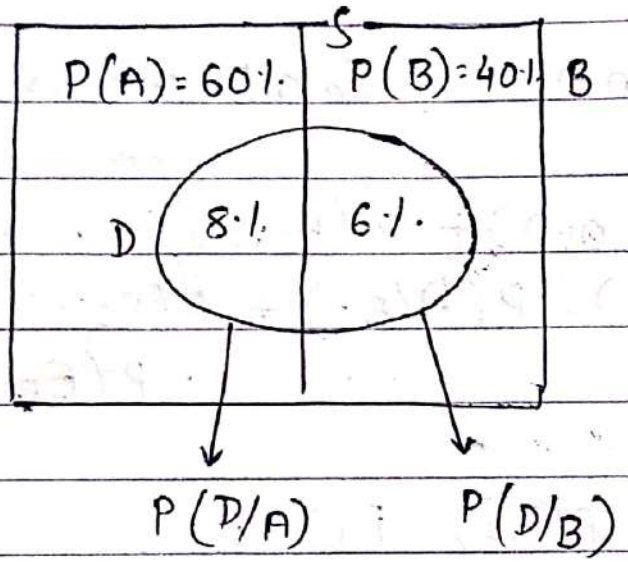
Date: Page:

Ex-2 (1-5 X 6,7,8,11,12,13) Class XII

14-18 related with WOP (VVI)

Qno20. A factory has two machines A and B .....

Solution,



Let A and B represent machines A and B respectively and D denotes respective outputs.

$P(A) = 60\% = 0.6$ $\quad$ $P(B) = 40\% = 0.4$

$P(D/A) = 8\% = 0.08$ $\quad$ $P(D/B) = 6\% = 0.06$

We have,

$$P(D) = P(A) \cdot P(D/A) + P(B) \cdot P(D/B)$$
$$= 0.6 \times 0.08 + 0.4 \times 0.06$$
$$= 0.048 + 0.024$$
$$= 0.072$$

Then,

$$P(B/D) = \frac{P(B) \cdot P(D/B)}{P(D)}$$

$$= \frac{0.4 \times 0.06}{0.072}$$

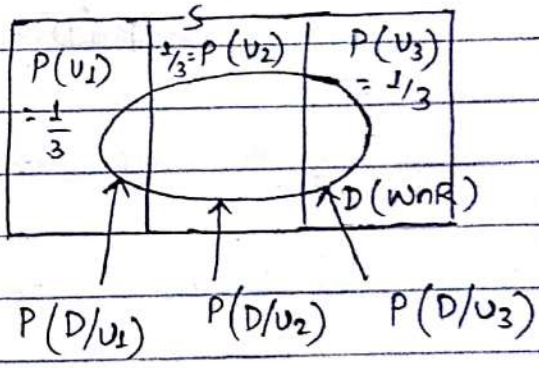
$$= \frac{0.024}{0.072} = \frac{1}{3}$$

Qno21. Solution,

Urn I $\quad 1W + 2B + 3R$

Urn II $\quad 2W + 1B + 1R$

Urn III $\quad 4W + 5B + 3R$



Now,

$P(U_1) = 1/3$

$P(D/U_1) = P(W_1 \cap R_2) + P$

$= P(W) \cdot P(R/W) + P(R) \cdot P(W/R)$

$= \frac{1}{6} \times \frac{3}{5} + \frac{3}{6} \times \frac{1}{5}$

$= \frac{1}{5}$

$P(U_2) = 1/3$

Again,

$P(D/U_2) = P(W) \cdot P(R/W) + P(R) \cdot P(W/R)$

$= \frac{2}{4} \times \frac{1}{3} + \frac{1}{4} \times \frac{2}{3}$

$= \frac{1}{3}$

Again,

$$P(U_3) = \frac{1}{3}$$

$$P(D/U_3) = P(W) \cdot P(R/W) + P(R) \cdot P(W/R)$$

$$= \frac{4}{12} \times \frac{3}{11} + \frac{3}{12} \times \frac{4}{11}$$

$$= \frac{2}{11}$$

We know,

$$P(D) = P(U_1) \cdot P(D/U_1) + P(U_2) \cdot P(D/U_2) + P(U_3) \cdot P(D/U_3)$$

$$= \frac{1}{15} + \frac{1}{9} + \frac{2}{33}$$

$$= \frac{118}{495}$$

(i) $$P(U_1/D) = \frac{P(U_1) \cdot P(D/U_1)}{P(D)}$$

$$= \frac{1/3 \times 1/5}{\frac{118}{495}} = \frac{33}{118}$$

(ii) $$P(U_2/D) = \frac{P(U_2) \cdot P(D/U_2)}{P(D)}$$

$$= \frac{1/3 \times 1/3}{\frac{118}{495}} = \frac{55}{118}$$

(iii) $P(U_3/D) = \frac{P(U_3) \cdot P(D/U_3)}{P(D)}$

$= \frac{1/3 \times 2/11}{\frac{118}{495}} = \frac{15}{118/2} = \frac{15}{59}$

ODD: The odd in the favour of any event E is defined as the ratio of number of favourable cases to the number of unfavourable cases. If O(E) denotes odd in the favour of any event E, a denotes the number of favourable cases and b denotes the number of unfavourable cases then, $O(E) = \frac{a}{b}$

Similarly, odd against E is denoted by $O(\bar{E})$

$O(\bar{E}) = \frac{b}{a}$

$= \frac{\text{No. of unfavourable cases}}{\text{no. of favourable cases}}$

From this we can obtain the probability of E as:

$P(E) = \frac{a}{a+b}$

Date: Page:

Qno9: Solution,

Let $O(\bar{A})$ = odd against solving a problem by A

$= 8:6 = \frac{8}{6}$

Then, $P(A) = \frac{6}{6+8} = \frac{6}{14}$

$= \frac{3}{7}$

Also, $O(B)$ = odd in the favour of solving the problem by B

$= 14:10$

$= \frac{14}{10} = \frac{7}{5}$

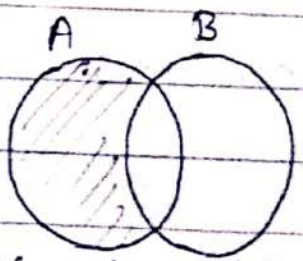
Then, $P(B) = \frac{14}{14+10}$

$= \frac{14}{24} = \frac{7}{12}$

(i) $P(A \cap B) = P(A) \times P(B)$

$= \frac{3}{7} \times \frac{7}{12}$

$= \frac{1}{4}$



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

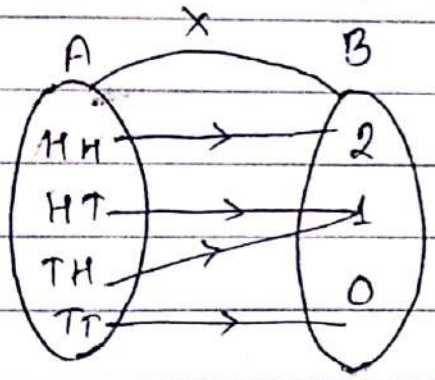
(ii) $P(A-B) = P(A) - P(A \cap B)$

$= \frac{3}{7} - \frac{1}{4} = \frac{5}{28}$

Random variable:

A random variable is a real valued function whose domain is the sample space of the random experiment and range is the set of the real numbers. Generally, random variable is represented by capital letters of english alphabets X, Y, Z and its value by corresponding small letters of an english alphabets x, y, z.

eg: When a coin is tossed twice and if the number of getting head is assumed as a random variable then it is a real valued function whose domain = {HH, HT, TH, T and range = {2, 1, 0}



X = no. of getting head.

Probability mass function (Pmf):

Let X is a discrete random variable which taken the values $x_1, x_2 \cdots x_n$, then a function denoted by $P(X = x_i) = P(x_i) = P_i$ (for i = 1, 2 ... n) is said to be probability mass function if

i. $P(x = x_i) \geq 0$ for each $i = 1, 2 \cdots n$

ii. $\sum_{i=1}^{n} P(x = x_i) = 1$

Date: Page:

Mathematical expectation:

Let X is a discrete random variable which takes the values $x_1, x_2 \cdots x_n$ with corresponding probabilities $P(x_1), P(x_2) \cdots P(x_n)$ then mathematical expectation of x is denoted by $E(x)$ and defined by

$$E(x) = \sum_{i=1}^{n} x_i P(x = x_i)$$

It is also called expectational value or mean value.

Properties:

(i) For any arbitary constant a,

$$E(a) = a$$

(ii) $E(ax+b) = aE(x)+b$

(iii) $E(x_1 + x_2 + \cdots x_n) = E(x_1) + E(x_2) + \cdots E(x_n)$

(iv) If $X_1, X_2 \cdots X_n$ are independent

$$E(x_1, x_2 \cdots x_n) = E(x_1), E(x_2) \cdots E(x_n)$$

(v) $E(x^n) = \sum x^n p(x)$

But

$$E\left(\frac{1}{x}\right) \neq \frac{1}{E(x)}$$

$$E(x^n) \neq$$

Date Page

Variance:

Let x is a discrete ~~variable~~ random variable having its mathematical expectation $E(x)$ and probability mass function $P(X=x)$ then variance of x is denoted by $V(x)$ or σ_x^2 and given by a relation,

$$V(x) = E(x-\mu)^2 \text{ when } \mu = E(x)$$

$$E(x^2 - 2x\mu + \mu^2)$$

$$= E(x^2) - E(2\mu x) + E(\mu^2)$$

$$= \Sigma x^2 P(x) - 2\mu E(x) + \mu^2$$

$$\Rightarrow V(x) = \Sigma x^2 P(x) - \mu^2$$

Date: Page:

Cumulative

Let X is a discrete random variable having its probability mass function (Pmf) $P(X = x)$ then cumulative distribution function of X is denoted by $F(x)$ and given by $F(x) = P(X \leq x)$

OR $F(x_i) = \sum_{J=1}^{i} P(x_j)$

Q. Solution, (1) $f(x) = x/5$

probability distribution

The distribution of A set of values of random variable and corresponding probabilities in the form of table or graph or mapping diagram etc is called

Discrete probability distribution

(i) Binomial

A discrete random variable x with parameters n and P is said to follow binomial probability distribution if its probability mass function (pmf) is denoted by $P(X=x)$ or $b(x; n, r)$ and given by a relation

$$b(x; n, p) = \begin{cases} c(n,x)\, p^x q^{n-x} & \text{for } x = 0, 1, 2 \cdots n \\ 0 & \text{otherwise} \end{cases}$$

where, n = number of trials

p = probability of success

1. mean $(u) = E(x) = np$
2. variance $(\sigma^2) = V(x) = npq$

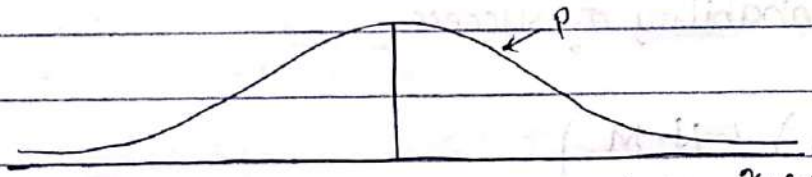
Conditions of using Binomials

1. When the number of trials are fixed and finite.
2. When every experiment consists only two possible outcomes success and failure.
3. When the successive experiment are performed under identical and independent condition.
4. When the probability of success is constant throughout the experiment.
5. When the probability of sucess/failure is neither too large nor too small.

Date: ... Page:

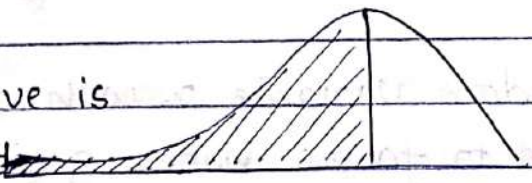
Properties of Binomial distribution

1. Binomial distribution can be determined completly if the parameters n and p are given.
2. Its mean and variance are:
   Mean (μ) = $E(x) = np$ mean > variance.
   variance (σ^2) = $V(x) = npq$
3. Its mean is always greater than variance
4. If $p = q = 0.5$, then the curve of the distribution is symmetrical.



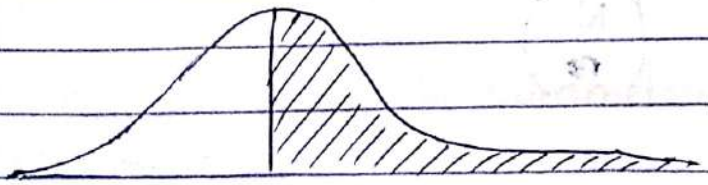
mean = median = mode

5. $P > 0.5$
   then the curve is negatively skewed



skewed to the left / negatively skewed

6. If $P < 0.5$, then the curve is positively skewed.



skewed to the right / positively skewed.

7. Its probability mass function (pmf) is

$$b(x; n, p) = \begin{cases} c(n,x)\, p^x q^{n-x}, & \text{for } x = 0, 1, 2 \cdots \\ 0 & \text{otherwise} \end{cases}$$

II. Hypergeometric distribution
(dependent event)
(population finite) = N

M = number of success
(N − M) = number of failure
x = probability of success

$$\frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

For discrete random variable x with parameters n, M and N is said to follow hyper geometric distribution if its probability mass function (pmf) is denoted by: $P(X = x)$ or $h(x; n, M, N)$ and

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

where,
N = Population size
M = number of success
n = number of trial
$x = 0, 1, 2, \cdots n$

Mean and variance

(i) mean $(\mu) = E(x) = n\frac{M}{N}$

(ii) variance $(\sigma^2) = V(x) = \frac{N-n}{N-1} n \frac{M}{N}\left(1 - \frac{M}{N}\right)$

where, $\frac{N-n}{N-1}$ = finite population correction factor

$\frac{N-n}{N-1} \to 1$ as $N \to \infty$

i.e. $\frac{N-n}{N-1} = \frac{N\left(1-\frac{n}{N}\right)}{N\left(1-\frac{1}{N}\right)} = \frac{1-n/\infty}{1-1/\infty} = \frac{1-0}{1-0} = 1$

Conditions of using hypergeometric distribution

1. When the items are not replaced after each experiment.
2. When the whole population (N) is divided into two character number of success (M) and number of failure (N-M)
3. When the parameters N, M and n are given.

Characteristics of hypergeometric distribution.

1. It depends upon the parameters N, M and n.
2. Its mean and variance are:

mean $(\mu) = E(x) = n\frac{M}{N}$

variance $(\sigma^2) = V(x) = \frac{N-n}{N-1} n \frac{M}{N}\left(1 - \frac{M}{N}\right)$

3. Items are not replaced after each experiment
4. Successive trials are not performed under identical condition.
5. The probability of success is not constant throughout the experiment.
6. The population size is decreased after each trial.
7. Every experiment consists only two possible outcomes; success and failure.
8. Its probability mass function is

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

Distinguish betⁿ binomial and hypergeometric distribution.

Date: Page:

Binomial	Hypergeometric
1. It depends upon the parameters n and p.	1. It depends upon the parameters M, N and n.
2. Items are replaced after each experiment.	2. Items are not replaced after each experiment.
3. Successive trials are performed under identical and independent conditions.	3. Successive trials are not performed under identical conditions.
4. Probability of success is constant throughout the experiment.	4. Probability of success is not constant throughout the experiment.
5. The ~~pre~~ population size remains same throughout the experiment.	5. The population size is decreased after each experiment
6. Its probability mass function is: $b(x; n, p) = \begin{cases} C(n,x)\, p^x \cdot q^{n-x} & \text{for } 0,1,2 \cdots n \\ = 0 & \text{otherwise} \end{cases}$	6. Its pmf is: $h(x; n, M, N) = \dfrac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$

Date: Page:

Similarities betn binomial and hypergeometric

i In both distributions;

a. every experiment consists only two possible outcomes; success and failure.

b. the probability of success / failure is neither too large nor too small.

Binomial approximation to Hypergeometric distribution

The hyper geometric distribution approximate into binomial distribution under the following conditions.

(i) When the population size is infinitely large.
i.e. $N \to \infty$

(ii) When the ratio M/N approximately equal to probability of success.

Date: ______ Page: ______

Exercise:

Q. During one stage in the manufacture ......

Solution,

p = probability of circuit chips receive a thick enough coating

= 70% = 0.7

q = 1 - p = 0.3

Number of chip (n) = 15

(i) At least 12 will have thick enough coatings

$x \geq 12$

Now,

$$P(x \geq 12) = 1 - P(x < 12)$$

$$= 1 - P(x \leq 11)$$

$$= 1 - \sum_{x=0}^{11} c(n, x)\, p^x q^{n-x}$$

$$= 1 - \sum_{x=0}^{11} c(15, x)\, (0.7)^x (0.3)^{15-x}$$

$$= 1 - 0.7031$$

$$= 0.2969$$

(ii) At most ;

$x \leq 6$

Now, $P(X \leq 6) = \sum_{n=0}^{6} c(n, x)\, p^x q^{n-x}$

$$= \sum_{n=0}^{6} c(15, x)\, (0.7)^x (0.3)^{15-x}$$

$$= 0.0152$$

Poisson probability distribution:

A discrete random variable X with single parameter mean (λ) is said to follow poisson probability distribution if its probability mass function (pmf) is denoted by $P(X=x)$ or $P(x;\lambda)$ and given by a relation as:

$$P(x;\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^x}{x!}, & \text{for } x = 0,1,2,3\ldots \\ 0 & \text{otherwise.} \end{cases}$$

Where, λ = mean of the distribution

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n = 2.7182$$

mean and variance

(i) mean (μ) = $E(x) = \lambda$

(ii) variance (σ^2) = $V(x) = \lambda$

Conditions of using Poisson probability distribution.

1) When the number of trials is infinitely large and probability of success is ~~0~~ nearly equal to 0.

2. When any event occurs rarely and uniformly within a certain interval of time.

3. When every experiment consists only two possible outcomes ie. success or failure.

Date: ________ Page: ________

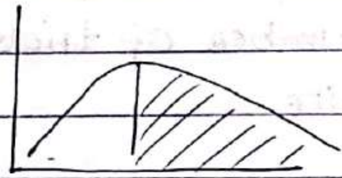
4. When sequential experiment are performed under identical and independent conditions.

5. Properties of PPD.

1. This distribution is depend upon single parameter (mean (λ).

2. This is the only one discrete probability distribution whose mean is always equal to the variance.

3. Being $P < 0.5$, it is always positively skewed.



4. Sum of two or more than two poisson distribution is again a poisson distribution.

5. The probability of success is constant throughout the experiment.

6. The number of trials is infinitely large and probability of success is nearly zero.

7. Sequential experiment are performed under identical and independent conditions.

8. Its pmf is $P(x;\lambda) = \begin{cases} \frac{e^{-\lambda}\lambda^{x}}{x} & \text{for } x = 0,1,2,3\cdots \\ = 0 & \text{otherwise} \end{cases}$

Dissimilarities between binomial and Poisson's distribution

Binomial	Poisson
1. It depends upon the parameters n and p.	It depends upon the parameters λ.
2. Its mean is always greater than variance.	Its mean and variance are equal.
3. The number of trials are fixed and finite	The no. of trials are infinitely large.
4. The probability of sucess/failure is neither too large nor too small.	The probability of succes is nearly zero and probability of failure is nearly 1.
5. The random variable takes the values such that 0, 1, 2 ... n.	The random variable takes the values such that it has no upper boundary.
6. Its pmf is $b(x; n, p) = \begin{cases} C(n,x)\, p^x q^{n-x} \\ \text{for } x = 0, 1, 2 \cdots \\ 0 \text{ otherwise} \end{cases}$	Its pmf is: $P(x; \lambda) = \begin{cases} \frac{e^{-\lambda} \lambda^x}{x} \text{ for } 0, 1, 2 \cdots \\ 0 \text{ otherwise} \end{cases}$

Date: Page:

Similarities:

1. In both distributions,
   (i) sequential experiments are performed under identical and independent condition.

   (ii) The probability of success is constant throughout the experiment.

   (iii) Every experiment consists only two possible outcomes success or failure.

Poisson approximation to binomial distribution.

The binomial distribution approximate into Poisson distribution under the following condition.

(i) When the number of trial are infinitely large i.e. $n \to \infty$

(ii) When the probability of success is nearly equal to zero ie. $p \to 0$

(iii) When mean $(\mu) = \lambda = np =$ finite.

Date: Page:

Negative Binomial Distribution.

A discrete random variable X with parameter r and p is said to follow negative binomial distribution if its probability mass function (pmf) is denoted by $P(X=x)$ or $nb(x; r, p)$ and given by a relation as

$$nb(x; r, p) = \begin{cases} \binom{x+r-1}{r-1} p^x q^x & \text{for } x = 0, 1, 2, \ldots r \\ 0 & \text{otherwise} \end{cases}$$

where,

r = no. of success

p = probability of success

q = 1 − p = " " failure

mean and variance

$$\text{mean } (\mu) = \frac{rq}{p}$$

$$\text{variance } (\sigma^2) = \frac{rq}{p^2}$$

Conditions of using Negative Binomial Distribution

(i) When the experiment is continued until the desirable number of success is occured.

(ii) When the no. of trial is a random variable, and no. of success is a given parameter.

(iii) When sequential experiments are performed, under identical and independent conditions

Date: Page:

(iv) When every experiment consists only two possible outcomes success or failure.

Properties of Negative Binomial Distribution

(i) It depends upon the parameters r and p.

(ii) Its variance is always greater than mean.

Its mean and variance are,

mean $(\mu) = rq/p$

and variance $(\sigma^2) = rq/p^2$

(iii) Its last trial is always success.

(iv) Its probability of success is never zero.

(v) The number of trial is a random variable and no. of success is a given parameter.

(vi) The probability of success is constant throughout the experiment.

(vii) Sequential exp. are performed under identical and independent conditions.

(viii) Its pmf is $nb(x; r, p) = \begin{cases} \binom{x+r-1}{r-1} p^x q^x & \text{for } x = 0, 1, 2 \cdots \\ 0 & \text{otherwise} \end{cases}$

Dissimilarities betn Negative binomial & binomial

Binomial	Negative Binomial
1. It depends upon the parameters n and p.	It depends upon the parameters r and p.
2. Its mean is always greater than variance.	Its variance is always greater than mean.
3. No. of trial is a given parameter and no. of success is a random variable.	No. of trial is a random variable and no. of sucess is a given parameter.
4. The last trial is either success or failure.	The last trial is always success.
5. Its pmf is: $b(x;n,p) = \begin{cases} C(n,x)p^x q^{n-x} & \text{for } 0,1,2\cdots n \\ 0 & \text{otherwise} \end{cases}$	Its pmf is: $nb(x;r,p) = \begin{cases} \binom{x+r-1}{r-1} p^x q^x & \text{for } 0,1,2\cdots n \\ 0 & \text{otherwise} \end{cases}$

Date: Page:

Similarities betn Neg. binomial and binomial.

1. In both distributions

(i) sequential exp. are performed under identical and independent conditions.

(ii) every experiment consists only two possible outcomes success or failure.

(iii) the probability of success / failure is neither too large nor too small.

(iv) the probability of success constant throughout the experiment.

Exercise: 3

29 Solution,

$N = 20$

Number of car chargers $(N) = 20$

Number of defective chargers $(M) = 5$

Sample size $(n) = 10$

$P(x = 2) = ?$

We have,

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \frac{\binom{5}{2}\binom{15}{8}}{\binom{20}{10}}$$

$$= \frac{C(5,2) \times C(15,2)}{C(20,10)}$$

$=$

(ii) $N = 100$, $M = 25$

a. By hypergeometric distribution,

$n = 10$; $x = 2$

We have,

$$h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

$$= \frac{\binom{25}{2}\binom{75}{8}}{\binom{100}{10}} = \frac{c(25,2) \times c(75,8)}{c(100,10)}$$

$$= 0.2923$$

b. By binomial distribution

$$p = \frac{M}{N} = \frac{25}{100} = 0.25$$

$$q = 1 - p = 0.75$$

$$n = 10$$

Then,

$$b(x; n, p) = c(n, x) p^x q^{n-x}$$

$$= c(10, 2) (0.25)^2 (0.75)^{10-2}$$

$$= 0.2815$$

35) Solution,

average call per minute $(\lambda) = 0.6$

a. At least one call in a minute

$$x \geq 1$$

Now,

$$P(x \leq 1) = 1 - P(x < 1)$$

$$= 1 - P(x = 0)$$

$$= 1 - \frac{e^{-\lambda} \lambda^x}{x!} = 1 - \frac{e^{-0.6} (0.6)^0}{0!}$$

$$= 1 - \frac{1}{e^{0.6}} = 0.4511$$

b. At least 3 calls in a four minute interval.

$x \geq 3$

Now, length of the interval $(d) = 4$ min

$$P(x \geq 3) = 1 - P(x < 3)$$

$$= 1 - P(x \leq 2)$$

$$= 1 - \sum_{x=0}^{2} \frac{e^{-\lambda d} (\lambda d)^x}{x!}$$

$$= 1 - \sum_{x=0}^{2} \frac{e^{-2.4} (2.4)^x}{x!}$$

$$= 0.430$$

46. Solution,

Probability of success $(p) = \frac{1}{50} = 0.02$

Sample size $(n) = 25$

Average lens to be defective $(\lambda) = np$

$= 25 \times 0.02$

$= 0.5$

$x \leq 1$

Now,

$$P(x \leq 1) = \sum_{x=0}^{1} \frac{e^{-\lambda} \lambda^x}{x!}$$

$$= \sum_{x=0}^{1} \frac{e^{-0.5}(0.5)^x}{x!}$$

$$= 0.9098$$

Required number of blade $= N \times P(x \leq 1)$

$= 10,000 \times 0.9098$

$= 9098$

New book.

59. Solution,

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

probability of success $(p) = \frac{6}{8} = \frac{3}{4}$

probability of failure $(q) = 1 - p$

$= 1 - 3/4 = 1/4$

Number of success $(r) = 1$

$n < 4$

$x + 1 < 4$ where, x is the number of failure

$x < 3$

We have,

$P(x < 3) = P(x \leq 2)$

$$= \sum_{x=0}^{2} \binom{x+r-1}{r-1} p^r q^x$$

$$= \sum_{x=0}^{2} \binom{x+1-1}{1-1} (0.75)^1 (0.25)^x$$

$$= \sum_{x=0}^{2} 1 \times 0.75 \times (0.25)^x$$

$$= 0.75 \sum_{x=0}^{2} (0.25)^x = \frac{63}{64}$$

$$= 0.9843$$

2. A student has taken five answers multiple choice question orally. He continue to ~~*~~ give the answer until he gets 3 correct answers. What is the probability that he will achieve this ~~not more than~~ at most 7 attempts.

Solution,

probability of success, $p = 1/5$

probability of failure, $q = 1 - p = \frac{4}{5}$

no. of success $= r = 3$

$n \leq 7$

$x + r \leq 7$ where x = number of failure

$x + 3 \leq 7$

$x \leq 4$

Date: Page:

Now,

$$P(x \leq 4) = \sum_{x=0}^{4} \binom{x+r-1}{r-1} p^r q^x$$

$$= \sum_{x=0}^{4} \binom{x+3-1}{3-1} \left(\frac{1}{5}\right)^3 \left(\frac{4}{5}\right)^x$$

$$= \frac{1}{125} \sum_{x=0}^{4} \binom{x+2}{2} \left(\frac{4}{5}\right)^x$$

$$= 0.148$$

Date: Page:

Continuous random variable.

A random variable is said to be continuous if it takes all possible values within a given interval of time. for eg: amount of rainfall, electric voltage, height, weight, temperature etc are continuous random variables.

Dissimilarities betn continuous and discrete random variables.

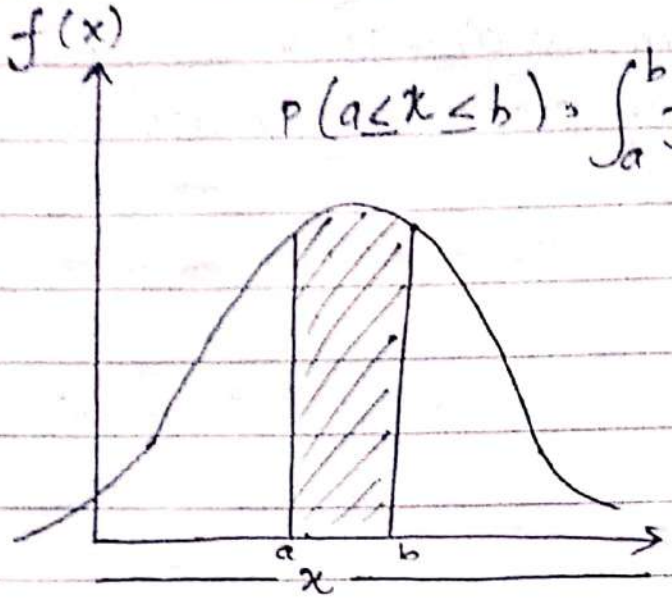
Discrete	Continuous
(1) discrete random variables are counted. for eg: the number of people queue in a hospital, number of cars parked in a parking station, the no. of stds in class etc.	Continuous random variables are measured. for eg: the amount of rainfall, electric voltage, height, weight, temperature etc.
(2) It takes only integral values.	It takes any real numbers within a certain range of interval.
(3) To study the ~~properties~~ probability of discrete random variable, we use mathematical tool ie. summation.	To study the probability of continuous random variables, we use calculus of mathematics. int, diff..

4) The probability of discrete random variable is represented given by probability mass function.

The probability of continuous random variable is the area enclosed by probability density function with x axis within a range or interval.

Date: Page:

Probability density function



$P(a \leq x \leq b) = \int_a^b f(x)\,dx$

(i) $f(x) \geq 0$

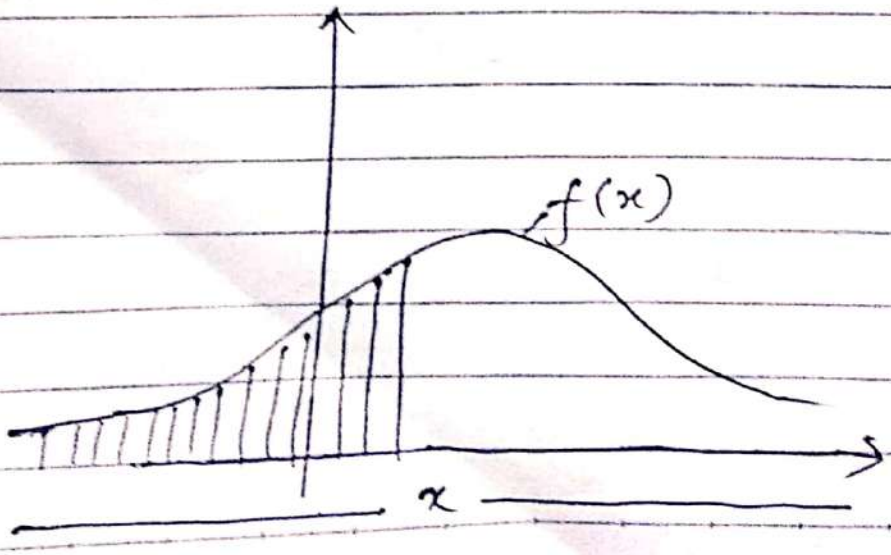
(ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Let x is a continuous random variable then a function $f(x)$ such that for $a \leq x \leq b$;

$P(a \leq x \leq b) = \int_a^b f(x)\,dx$ is said to be probability density function if it satisfies the following properties

(i) $f(x) > 0$ (ii) $\int_{-\infty}^{\infty} f(x)\,dx = 1$

Cumulative distribution function



Date: Page:

$$F(x) = P(X \leq x)$$
$$= \int_{-\infty}^{x} f(x)\,dx$$

Let X is a continuous random variable having probability density function (p.d.f) $f(x)$ then a new function denoted by $F(x)$ is said to be cumulative distribution function of X ~~is~~ if $F(x) = P(X \leq x) = \int_{-\infty}^{x} f(x)\,dx$

Properties:

(i) $F(-\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 0$

(ii) $F(\infty) = \int_{-\infty}^{\infty} f(x)\,dx = 1$

(iii) $0 \leq F(x) \leq 1$

(iv) $F'(x) = f(x)$

(v) $P(a \leq x \leq b) = P(a < x < b) = P(a \leq x < b) = P(a < x \leq b)$

Mean and variance

Let x is a continuous random variable having its probability density function (pdf) is $f(x)$ then mean (μ) and variance (σ^2) of x defined as follows:

i) mean $(\mu) = E(x) = \int_{-\infty}^{\infty} x f(x)\, dx$

(ii) variance $(\sigma^2) = V(x) = \int_{-\infty}^{\infty} x^2 f(x)\, dx - \mu^2$

Exercise: ~~3~~ (Page 211 1Q. Sure)

2. Solution,

The pdf of a random variable is given by:

$$f(x) = \begin{cases} kx^3 & \text{for } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$k = ?$

We know

$$\int_{-\infty}^{\infty} f(x)\, dx = 1$$

or, $$\int_{-\infty}^{0} f(x)\, dx + \int_{0}^{1} f(x)\, dx + \int_{1}^{\infty} f(x)\, dx = 1$$

$$0 + \int_{0}^{1} kx^3\, dx + 0 = 1$$

Date: Page:

$$\text{or,} \quad \left[\frac{kx^4}{4}\right]_0^1 = 1$$

$$\text{or,} \quad \frac{k}{4} = 1 \quad \therefore k = 4$$

$$P\left(\frac{1}{4} \leq x < \frac{3}{4}\right) = \int_{1/4}^{3/4} f(x)\,dx$$

$$= \int_{1/4}^{3/4} kx^3\,dx$$

$$= \left[\frac{kx^4}{4}\right]_{1/4}^{3/4}$$

$$= \left[\frac{3}{4}\right]^4 - \left[\frac{1}{4}\right]^4$$

$$= \frac{5}{16}$$

c) $P\left(x > \frac{2}{3}\right) = \int_{2/3}^{\infty} f(x)\,dx$

$= \int_{2/3}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$

$= \int_{2/3}^{1} kx^3\,dx$ = solve

mean $(\mu) = \int_{-\infty}^{\infty} x f(x)\,dx$

$= \int_{-\infty}^{0} x f(x)\,dx + \int_{0}^{1} x f(x)\,dx + \int_{1}^{\infty} x f(x)\,dx$

$= 0 + \int_{0}^{1} x f(x)\,dx + 0$

$= \int_{0}^{1} x \cdot kx^3$

$= \int_{0}^{1} kx^4 \quad : k\left[\frac{x^5}{5}\right]_0^1$

$= 4 \times \frac{1^5}{5}$

$= 4/5$

Date: Page:

Variance $(\sigma^2) = \int_{-\infty}^{\infty} x^2 f(x)\,dx - \mu^2$

$= \int_{-\infty}^{\infty} x^2 \cdot kx^3\,dx - \mu^2$

$= \int_{-\infty}^{0} kx^5\,dx + \int_{0}^{1} kx^5\,dx + \int_{1}^{\infty} kx^5\,dx - \mu^2$

$= k\left[\frac{x^6}{6}\right]_0^1 - \frac{16}{25}$

$= \frac{4}{6} - \frac{16}{25}$

$= 2/75$

Qno5. Solution,

Given function is:

$$f(x) = \begin{cases} 0 & \text{for } x < -a \\ \frac{1}{2a}(x+1) & \text{for } -a \le x \le a \\ 0 & \text{for } x > a \end{cases}$$

Now,

$$f(x) = F'(x) = \begin{cases} 0 & \text{for } x < -a \\ 1/2a & \text{for } -a \le x \le a \\ 0 & \text{for } x > a \end{cases}$$

Here,

$f(x)$ is exists. Now $f(x)$ should satisfy the properties of pdf to be $F(x)$ a distribution function.

(i) Clearly, $f(x) \ge 0$ for all x

(ii) $$\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{-a} f(x)\,dx + \int_{-a}^{a} f(x)\,dx + \int_{a}^{\infty} f(x)\,dx$$

$$= 0 + \int_{-a}^{a} \frac{1}{2a}\,dx + 0$$

$$= 1$$

Qno8.

The pdf of a rv is:

$$f(x) = \begin{cases} e^{-x} & \text{for } 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$

① mean $(\mu) = \int_{-\infty}^{\infty} x f(x) dx$

$$= \int_{-\infty}^{0} x f(x) dx + \int_{0}^{\infty} x f(x) dx$$

$$= 0 + \int_{0}^{\infty} x e^{-x} dx \qquad \left[\int_{0}^{\infty} e^{-x} x^{n-1} dx = \Gamma x\right]$$

$$= \int_{0}^{\infty} e^{-x} x^{2-1} dx$$

$$= \Gamma 2$$

$$= (2-1)! = 1$$

Variance $(\sigma^2) = \int_{0}^{\infty} x^2 e^{-x} dx - \mu^2$

$$= \Gamma(3) - 1$$

$$= (2-1)! - 1 = 1$$

Date: Page:

Q → $f(x) = \begin{cases} e^{-x^2/2} & \text{for } 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$

$$\text{mean}(\mu) = \int_{-\infty}^{\infty} x f(x)\, dx$$

$$= \int_{-\infty}^{0} x f(x)\, dx + \int_{0}^{\infty} x f(x)\, dx + \int_{\infty}^{\infty}$$

$$= 0 + \int_{0}^{\infty} x e^{-x^2/2}\, dx$$

Variance,

Let $-\frac{x^2}{2} = t$

$dt = -x \cdot dx$

When $x \to 0$, $t \to 0$

When $x \to \infty$, $t \to \infty$

$$= -\int_{0}^{\infty} e^t\, dt$$

$$= -e^t \Big|_0^{\infty} \qquad = -e^{\infty} + 1$$

$$= 1$$

Date: Page:

no12. Solution,

$$f(x) = \begin{cases} 6x(1-x) & \text{for } 0 \le x \le 1 \\ 0 & \text{otherwise} \end{cases}$$

a. pdf

Clearly, $f(x) \ge 0$ for all x

(ii) $$\int_{-\infty}^{\infty} f(x)\,dx = \int_{\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{\infty} f(x)\,dx$$

$$= \int_0^1 6x(1-x)\,dx$$

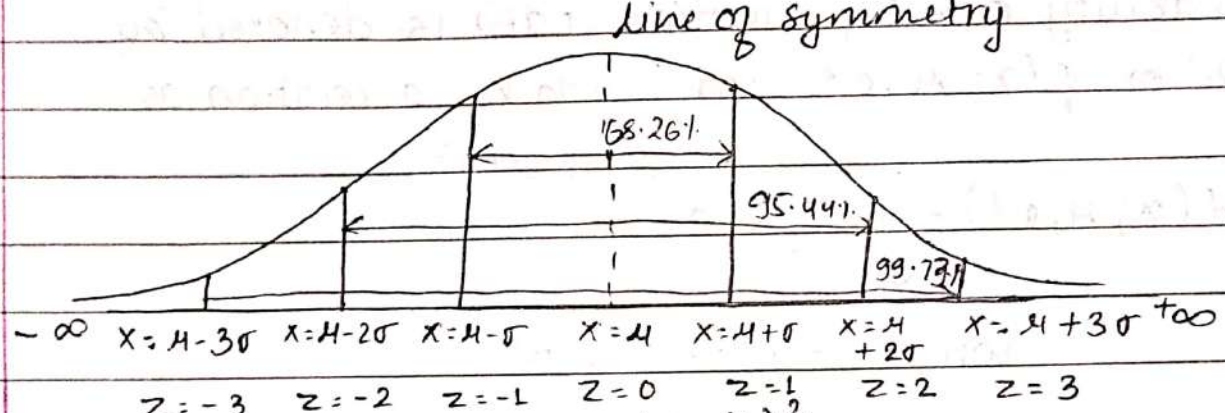
$$= \int_0^1 6x - 6x^2\,dx$$

$$= \left[\frac{6x^2}{2} - \frac{6x^3}{3}\right]_0^1 = 3 - 2 = 1$$

So, it is pdf.

MAYUR
Date
Page

Normal Probability Distribution:



$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

If $z = \frac{x-\mu}{\sigma}$, $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$

mean = 0 (μ)
Standard deviation = 1
→ standard normal distribution

A continuous random variable X with parameters mean (μ) and ~~variable~~ variance (σ^2) is said to follow normal probability distribution if it's probability density function (pdf) is denoted by $f(x)$ or $f(x; \mu, \sigma^2)$ and given by a relation as:

$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

where, $-\infty < \mu < \infty$

$-\infty < x < \infty$

$\sigma^2 \geq 0$

Standard Normal Distribution:

In normal distribution, if we put $z = \frac{x-\mu}{\sigma}$, then z follows normal distribution with mean (μ) = 0 and variance (σ^2) = 1, the distribution so obtained is known as standard normal distribution. The pdf of standard normal distribution is:

$$f(z; 0, 1) = \frac{1}{\sqrt{2\pi}} e^{-\frac{1}{2}z^2}$$

where,

$-\infty < z < \infty$

Properties of Normal Probability Distribution

1. The normal probability curve is bell shaped and symmetrical about the line $x = \mu$.

2. The mean, median and mode of this distribution are coincident. ie. mean = median = mode

3. Its probability density function is:
$$f(x; \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

4. Having only one single peak, it is unimodel distribution.
(only one mode)

5. Being symmetrical in nature, quartiles are equidistant from the median.
$$\text{i.e. } Q_3 - Md = Md - Q_1$$
$$Md = Q_1 + Q_3 / 2$$

6. Since the curve changes its direction at points $x = \mu \pm \sigma$ so, this is the point of inflection.

7. Its mean deviation from mean is $MD = \frac{4}{5}\sigma$ and
quartile deviation $QD = \frac{2}{3}\sigma$

8. X axis is the asymptote of normal probability curve

9. If $X_1, X_2 \cdots X_n$ are normal variables with corresponding means $\mu_1, \mu_2 \cdots \mu_n$ and variances $\sigma_1^2, \sigma_2^2 \cdots \sigma_n^2$ then its linear combination:

$$X = a_1X_1 + a_2X_2 + \cdots a_nX_n$$

is also a normal variable with mean

$$\mu = a_1\mu_1 + a_2\mu_2 + \cdots a_n\mu_n$$

and variance $\sigma^2 = a_1^2\sigma_1^2 + a_2^2\sigma_2^2 + \cdots a_n^2\sigma_n^2$

where, $a_1, a_2 \cdots a_n$ are arbitrary constants

10. Area Properties

(i) It covers 68.26% area within the range

$$X = \mu \pm \sigma$$

$$P(\mu - \sigma < X < \mu + \sigma) = 68.26\% = 0.6826$$

(ii) It covers 95.44% area within the range

$$X = \mu \pm 2\sigma$$

$$P(\mu - 2\sigma < X < \mu + 2\sigma) = 95.44\% = 0.9544$$

(iii) It covers 99.75% area within the range

$$X = \mu \pm 3\sigma$$

$$P(\mu - 3\sigma < X < \mu + 3\sigma) = 99.75\% = 0.9975$$

MAYUR
Date
Page

Normal approximation to Binomial distribution.

The binomial distribution approximate into Normal Probability distribution under the following two conditions.

(1) When the number of trials are infinitely large.
i.e. $n \to \infty$

(2) When the probability of success ~~is nearly~~ / failure is neither too large nor too small.

Normal approximation to Poisson distribution.

The Poisson probability distribution approximate into Normal probability distribution if the mean of the distribution is infinitely large. i.e. $\lambda \to \infty$

Importance of Normal Probability Distribution

1. Most of the discrete probability distributions like: Binomial, Poisson, can be approximated by Normal Distribution.

2. Most of the sampling distributions of actual sample test like: Z, t, F, χ^2 (chi square) follows Normal Probability Distribution. when $n \to \infty$

3. Most of the hypothesis testing like: Z, t, F, etc based on the assumption that their parent ~~assumption~~ population is taken from normal.

4. When $n \to \infty$, the central limit theorem (CLT) follows normal distribution.

5. Most of the probability distributions which are not normal can be made normal by simple transformation.

6. It is used in Statistical quality control in industry to set its quality limit.

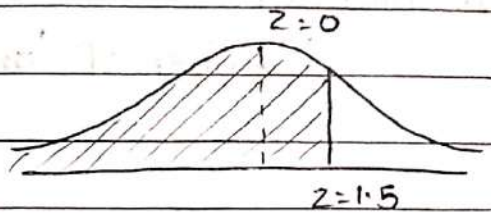
MAYUR
Date
Page

Exercise 4:

Normal distribution

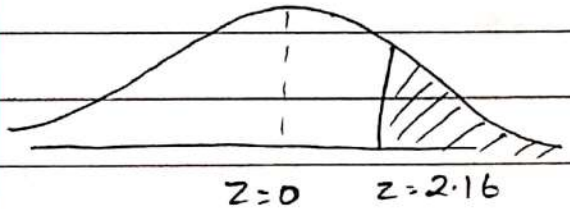
1.

a. less than 1.50



i. $P(z < 1.50) = F(1.50)$
$= 0.9332$ (from table)

ii. greater than 2.16

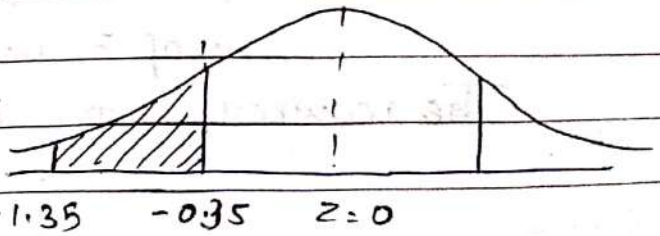


$P(z > 2.16) = 1 - P(z < 2.16)$
$= 1 - F(2.16)$
$= 1 - 0.9846$
$= 0.0154$

c. between -1.35 and -0.35

$P(-1.35 < z < -0.35) =$

$P(z < -0.35) - P(-1.35)$
$= F(-0.35) - F(-1.35)$
$= 0.3632 - 0.0885$
$= 0.2847$



Exercise: 4

Qno 21. Solution,

Number of lamps to be filled up $(N) = 10,000$

Average life $(\mu) = 1850$ hrs

S.D $(\sigma) = 200$ hrs

a) More than 2000 hours

$x > 2000 = x$

Then,

$$z = \frac{x - \mu}{\sigma}$$

$$= \frac{2000 - 1850}{200}$$

$$= 150/200 = 0.75$$

Now,

$$P(x > 2000) = P(z > 0.75)$$

$$= 1 - P(z \leq 0.75)$$

$$= 1 - 0.7734$$

$$= 0.2266$$

no. of

The required lamp $= N \times P(x > 2000)$

$$= 10000 \times 0.2266$$

$$= 2266$$

(b) Do yourself.

(c) Between 1540 and 1800 hours

$x_1 = 1540 < X < 1800$

Then,

$z_1 = \frac{x_1 - \mu}{\sigma}$, and $z_2 = \frac{x_2 - \mu}{\sigma}$

$= \frac{1540 - 1850}{2000}$ $= \frac{1800 - 1850}{2000}$

$= -1.55$ $= -0.25$

Now,

$P(1540 < X < 1800)$

$= P(-1.55 < Z < -0.25)$

$= F(-0.25) - F(-1.55)$

$= 0.3407$

Therefore, Required lamp $= N \times P(1540 < X < 1800)$

$= 10,000 \times 0.3407$

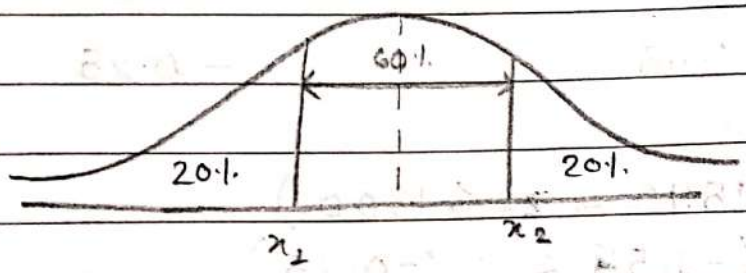
$= 3407$

no 25. Solution,

Average income (μ) = Rs 2000

S.D (σ) = Rs 200

a) Range of income of middle 60% employees



If x_1 is the lowest income and x_2 is the highest income of middle 60% employees.

If z_1 and z_2 are corresponding SNVs of x_1 and x_2 then

$$P(z < z_1) = 20\% = 0.2$$

from table

$z_1 = -0.84$ 0.2005.

$$\frac{x_1 - \mu}{\sigma} = -0.84$$

$$\frac{x_1 - 2000}{200} = -0.84$$

$$\therefore x_1 = -0.84 \times 200 + 2000$$

$$\boxed{x_1 = \text{Rs } 1832}$$

Also,

~~$P(z_1 < z < z_2$~~

$P(z > z_2) = 20\%$.

$1 - P(z < z_2) = 0.2$

$P(z < z_2) = +0.8$

from table,

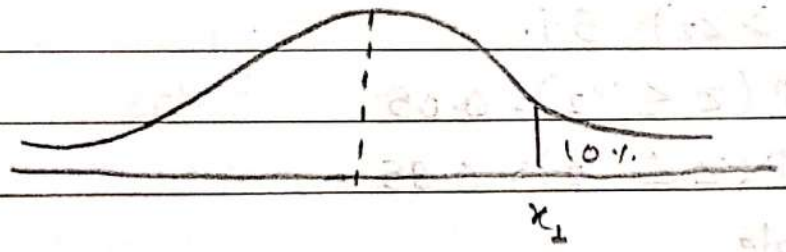
$z_2 = 0.84$

or, $\frac{x_2 - \mu}{\sigma} = 0.84$

or, $x_2 = 0.84 \times 200 + 2000$

$\therefore x_2 = 2168$

b. Lowest income of ~~richest~~ lowest 10% employees



$P(z > z_1) = 10\% = 0.1$

$1 - P(z \leq z_1) = 0.1$

$P(z \leq z_1) = 0.9$

from table,

$z = 1.28$

$\frac{x - \mu}{\sigma} = 1.28$

$x_1 = 1.28 \times 200 + 2000$

$= 2256$

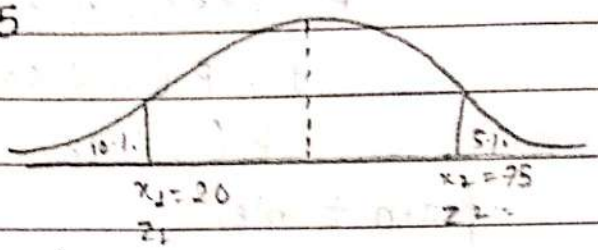
Qno39. Solution,

Let $x_1 = 20$ and $x_2 = 75$

then,

$P(z < z_1) = 10\%$

$= 0.1$



from table,

$z_1 = -1.28$

or, $\frac{x_1 - \mu}{\sigma} = -1.28$

or, $\frac{20 - \mu}{\sigma} = -1.28$

$\mu = 20 + 1.28\sigma$

Also,

$P(z > z_2) = 5\%$

$1 - P(z \leq z_2) = 0.05$

$P(z \leq z_2) = 0.95$

from table,

$z_2 = 1.645$

$\frac{x_2 - \mu}{\sigma} = 1.645$

or, $\frac{75 - \mu}{\sigma} = 1.645$

or, $75 - \mu = 1.645\sigma$

$\mu = 75 - 1.645\sigma$ ----(ii)

Equating ① and ⑪

$$20 + 1.28\sigma = 75 - 1.645\sigma$$

$$\text{or, } (1.28 + 1.645)\sigma = 75 - 20$$

$$\sigma = \frac{55}{1.28 + 1.645}$$

$$\boxed{\sigma = 18.8}$$

SAMPLING DISTRIBUTION

Sampling: The process of selection of a fraction of population from large group to know about its particular characteristics and making investigation about it and getting conclusion about the whole group is known as sampling. The fraction of population ~~from~~ which is selected from large group is called sample and the large group from which a fraction of population is selected is known as population or universe.

examples① the process of taking a ~~series of~~ syringe of blood from whole body of patient to diagonosis a particular disease and making conclusion about whole body is an eg. of sampling.

② the process of tasting a spoon full curry from the whole bucket to know about quantity of salt and making conclusion of whole bucket. is also an eg. of sampling.

Importance of sampling:

1. To study about population of infinite nature.
2. To study about the population of destructive nature.
3. To study about the population of complicated nature.
4. To get ~~to~~ quick result ~~from~~ by using limited resources.

5

Q.# Write the importance of sampling in the field of engineering.

Census:

The process of studying the every members of the population about its particular characteristics is known as census. In other words, a complete enumeration of a population is known as census. This method of study is more expensive and time consuming but it gives accurate result than sampling if ~~sampling and non-sampling errors are minimized~~ the man power and resources are expert and accurate.

Usually this is better to do to study about the populations having two members, a population of historical nature, ~~to~~ identify the machinary fault etc.

Distinguish betn census and sampling.

Parameters	Statistics
1. The statistical measures which are used to describe the characteristics of a population is known as parameters.	1. The statistical measure which are used to describe the characteristics of a sample are known as statistics.
2. Parameters are constant for a given population.	2. Statistics are different for different samples of a given population.
3. Some examples of the parameters are: Population mean = μ Population variance = σ^2 " proportion = p " correlation coefficient = ρ	3. Some examples of statistics are: sample mean = $\bar{x}$ " variance = s^2 " proportion = $\hat{p}$ " correlation coefficient = r

MAYUR
Date
Page

Parameters and statistics

↓ Population
mean = μ
variance: σ^2
proportion: p
Correlation coeff: ρ

↳ whole popn is divided into different samples.

sample: mean = $\bar{x}$
variance: s^2
proportion: $\hat{p}$
correlation coefficient: r

(proportion) → $p = \frac{10}{50}$ → those who like football / → total no. of population

Sample mean and sample variance

Let $x_1, x_2 \cdots x_n$ are n random sample taken from identically and independently distributed population. then, sample mean (mean of samples) and sample variance are denoted by $\bar{x}$ and s^2 respectively and defined by:

$$\bar{X} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n}$$

and

$$s^2 = \frac{\Sigma (x - \bar{x})^2}{n-1}$$

Sampling distribution:

$P = \{2, 4, 5, 6\}$

$N = 4$

$n = 2$

Possible samples without replacement

(2,4) (2,5), (2,6), (4,5) (4,6) (5,6)

with replacement (2,2) (2,4), (2,5), (2,6)

sample	means ($\bar{x}$)	$p(\bar{x})$	prop. of odd no.
(2,4)	3	1/6	0
(2,5)	3.5	1/6	1/2
(2,6)	4	1/6	0
(4,5)	4.5	1/6	1/2
(4,6)	5	1/6	0
(5,6)	5.5	1/6	1/2

$\Sigma \bar{x} =$

$\mu_{\bar{x}} =$ mean of $\Sigma \bar{x}$

$= \frac{\Sigma \bar{x}}{k}$ where $k = 6$

↳ mean of the sampling distribution of sample mean.

$\Sigma p(\bar{x})$

$\mu_p = \frac{\Sigma \hat{p}}{k}$

mean of sampling distribution of sample proportion.

$(\bar{x} - \mu_{\bar{x}})^2$

add all these.

Variance : without square root

Standard deviation $\longrightarrow \sqrt{\Sigma (\bar{x} - \mu_{\bar{x}})^2 / k}$ ↗ standard error of sampling mean

* Define sampling distribution of the sample mean. Illustrate it with example.

Ans:

The probability distribution of the sample means is known as sampling distribution of sample means. Consider a population consists N members from which 'n' samples are choosen without replacement (or with replacement). The possible number of samples so formed are $C(N,n)$ for without replacement and N^n for with replacement.

Let these samples have means $\overline{x_1}, \overline{x_2}, \overline{x_3} \cdots : \overline{x_k}$

Each samples means possesses definite probabilities.

Hence, the ~~probality~~ pr probability distribution of these sample means is known as sampling distribution of sample means.

* Define sampling distribution of the sample proportion and Illustrate it with an example.

Standard error:

The standard deviation of sampling distribution of a sample statistics is known as standard error of that statistics.

for eg: The standard deviation of sampling distribution of sample mean and sample proportion are standard error of sample mean and sample proportion respectively.

Mathematically,

$$S.E(\bar{x}) = \sigma_{\bar{x}} = \sqrt{\frac{\sum(\bar{x} - \mu_{\bar{x}})^2}{K}}$$

$$SE(\hat{p}) = \sigma_{\hat{p}} = \sqrt{\frac{\sum(\hat{p} - \bar{\hat{p}})^2}{K}}$$

where, $K = C(N, n)$ for without replacement

$K = N^n$ for with replacement.

MAYUR
Date
Page

Mean of sampling distribution of a sample mean.

Consider a population consists N members from which n samples are choosen with / without replacement The possible no. of the samples for with replacement is N^n and for without replacement is $((N,n)$.
Let these samples have means $\overline{x_1}, \overline{x_2}, \overline{x_3}, \ldots \overline{x_k}$.
Then, ~~the sampli~~ mean of ~~sampling distri~~ these means is known as mean of sampling distribution of sample means. It is denoted by '$\mu_{\bar{x}}$' or '$\overline{\bar{x}}$' and defined by $\mu_{\bar{x}} = \frac{\sum_{i=1}^{k} \overline{x}}{k}$

Mean of sampling distribution of sample proportion.

Standard error of sample mean (σ^2 known)

Case I: (finite & with replacement)

When the population is very large or samples are drawn with replacements.

Let $x_1, x_2, x_3 \cdots x_n$ are n random samples taken from identically and independently distributed population with mean (μ) and variance (σ^2). If $\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots x_n}{n}$

Then,

$$\mu_{\bar{x}} = E(\bar{x}) = \frac{E(x_1 + x_2 + \cdots x_n)}{n}$$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + \cdots E(x_n)\}$$

$$\because [E(ax) = a E(x)]$$

$$= \frac{1}{n}\{\mu + \mu + \mu \cdots \mu\}$$

$$= \frac{n\mu}{n}$$

$$\boxed{\mu_{\bar{x}} = \mu}$$

$$S.E(\bar{x}) = \sigma_{\bar{x}} = \sqrt{V(\bar{x})}$$

$$= \sqrt{\frac{V(x_1 + x_2 + \cdots x_n)}{n}}$$

$$= \sqrt{\frac{1}{n^2}\{V(x_1) + V(x_2) + \cdots V(x_n)}$$

$$= \sqrt{\frac{1}{n^2}(\sigma^2 + \sigma^2 + \cdots \sigma^2)}$$

$$= \sqrt{\frac{n\sigma^2}{n^2}} = \frac{\sigma}{\sqrt{n}}$$

Case II:

When the population is finite and samples are drawn without replacements.

Consider a population consists N members from which n samples choosen without replacement then

$$SE(\bar{x}) = \sqrt{\frac{N-n}{N-1}} \frac{\sigma}{\sqrt{n}}$$

where $\frac{N-n}{N-1} \to 1$ as $N \to \infty$

Exercise: 6

Qno 4. Given population = {2, 4, 6, 8, 10}

Population (N) = 5

Sample size (n) = 2

(a) Possible samples: without replacement

{(2,4), (2,6), (2,8), (2,10), (4,6), (4,8), (4,10), (6,8), (6,10), (8,10)}

(b) Mean and variance

$$\text{Pop}^n \text{ mean } (\mu) = \frac{\Sigma X}{N} = \frac{2+4+6+8+10}{5} = 6$$

Pop^n variance (σ^2)

X	$(X-\mu)^2$
2	$(2-6)^2 = 16$
4	$(4-6)^2 = 4$
6	$(6-6)^2 = 0$
8	$(8-6)^2 = 4$
10	$(10-6)^2 = 16$
	$\Sigma (X-\mu)^2 = 40$

We know, $\sigma^2 = \frac{\Sigma (X-\mu)^2}{N} = \frac{40}{5} = 8$

(c) Show that mean of sampling distribution of the sample mean is equal to the population mean.

Population sample	Sample means ($\bar{x}$)
(2, 4)	3
(2, 6)	4
(2, 8)	5
(2, 10)	6
(4, 6)	5
(4, 8)	6
(4, 10)	7
(6, 8)	7
(6, 10)	8
(8, 10)	9

$\Sigma \bar{x} = 60$

mean $(\mu_{\bar{x}}) = \frac{\Sigma \bar{x}}{k} = \frac{60}{10} = 6 = \text{pop}^{n}$ mean

Central limit theorem (CLT);

Let $X_1, X_2, X_3, \ldots X_n$ are n random samples taken from identically and independently

Theory of estimation:

Estimation:-

The process of guessing or predicting the value of true population parameter by past experience or its corresponding sample statistics with certain level of confidence.

The sample statistics which is used to predict the value of unknown population parameter is known as estimator. for eg: sample mean, sample variance and sample proportion are estimators of population mean, popⁿ variance and popⁿ proportion respectively.

Characteristics of good estimator

(i) Unbiasness

If the value of an estimator is approximately equal to the value of its corresponding population parameter then such estimator is said to be unbiased estimator of its unknown popⁿ parameter.

eg: mean of sampling distribution of sampling mean is unbiased estimator of its population mean.

ii. Consistency:

If the value of an estimator is more and more close with the value of its unknown population parameter ~~then~~ when the sample size is increased gradually. then that estimator is said to be consistent estimator of its population parameter.

eg: sample mean is consistent estimator of its population mean.

iii. Effeciency:

If the variance of one estimator is less than than the variance of another estimator then former estimator ~~later~~ is said to be efficient estimator than later.

for eg: sample mean is efficient estimator than sample median.

iv. Sufficiency:

If the max. possible information are used while obtaining the value of an estimator. Then, such estimator is said to be sufficient estimator of its population parameter.

for eg: while obtaining sample mean, all sample values are used. So, it is sufficient estimator of its population mean.

Types of estimation

a. Point estimation b. Interval estimation.

a. Point estimation:

The process of estimating the value of unknown population parameter with the help of its sample statistics by a single numerical value is known as point estimation.

for eg: sample mean and sample proportion are point estimators of population mean and population proportion respectively.

b. Interval estimation:

The process of estimating the value of its true popn parameter within a range of interval by certain level of confidence is known as interval estimation.

for eg: the daily checkout of the books from WRC library 460 – 500.

Formula of interval estimation

① for single population parameter

(a) Two tail: $C.I = \hat{\theta} \pm Z_{\alpha/2} \, SE(\hat{\theta})$

b. One tail $C.I = \hat{\theta} \pm Z_{\alpha} \, S.E.(\hat{\theta})$

(ii) for difference of parameters

a. for two tail

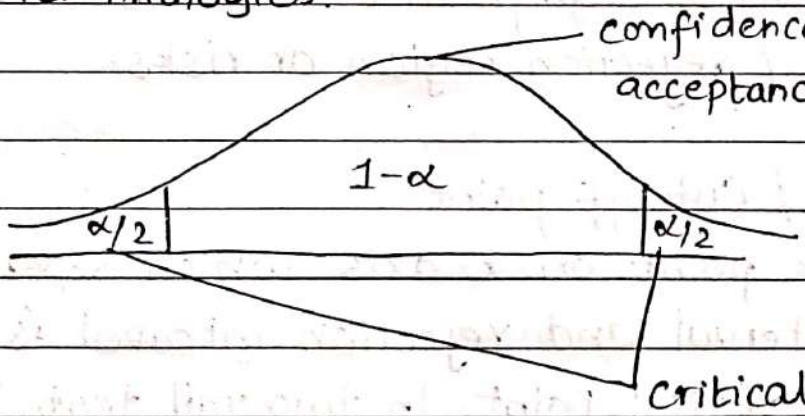
$$C.I = (\hat{\theta}_1 - \hat{\theta}_2) \pm Z_{\alpha/2}\, SE\,(\hat{\theta}_1 - \hat{\theta}_2)$$

b. for one tail

$$CI = |\theta_1 - \theta_2| \pm Z_{\alpha}\, SE\,(\hat{\theta}_1 - \hat{\theta}_2)$$

For small sample less than 30, we use t instead of z with corresponding degree of freedom

Some terminalogies:



Confidence region

The region or the probability in which true population parameter is expected to lie is called confidence region.
The interval on x axis of confidence region is called confidence interval.
It is denoted by $1-\alpha$.

Level of significance

The region or the probability in which true population parameter is unexpected to lie, is known as level of significance. It is also called critical region / rejection region or risks.

Critical point / Cut off point

The point on x axis which seperate the confidence interval and rejection interval is called ~~cutt o~~ critical point. In two tail test, it lies on both sides of probability curve and in one tail test, it lies only one side of the probability curve.

II. Standard error for some parameters

A. Concerning single sample mean

1. When the population is very large or sample are drawn with replacement

$$S.E(\bar{x}) = \frac{\sigma}{\sqrt{n}}$$

If σ is not given,

$$S.E(\bar{x}) = \frac{s}{\sqrt{n}}$$

where, $s = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}}$

B. Concerning single proportion:

1. When the population is very large or samples are drawn with replacement

$$SE(\hat{p}) = \sqrt{\frac{pq}{n}} \quad \text{where, } p = \frac{X}{N}, \; q = 1-p$$

X = number which carry specific character of the popⁿ.

If p is not given,

$$SE(\hat{p}) = \sqrt{\frac{\hat{p}\hat{q}}{n}} \quad \text{where, } \hat{p} = \frac{x}{n} \quad \hat{q} = 1-\hat{p}$$

Concerning difference of ~~means~~ proportion

When the population is very large or samples are drawn with replacement.

$$S.E(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{P_1 a_1}{n_1} + \frac{P_2 a_2}{n_2}}$$

where, $P_1 = \frac{X_1}{N_1}$ $\quad P_2 = \frac{X_2}{N_2}$

$q_1 = 1 - P_1$ $\quad q_2 = 1 - P_2$

(ii) If $P_1 = P_2 = P$ then,

$$SE(\hat{P}_1 - \hat{P}_2) = \sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}$$

iii. If P_1 and P_2 are not given and assumed to be unequal.

$$SE(\hat{P}_1 - \hat{P}_2) = \sqrt{\frac{\hat{P}_1 \hat{q}_1}{n_1} + \frac{\hat{P}_2 \hat{q}_2}{n_2}}$$

where, $\hat{P}_1 = \frac{x_1}{n_1}$

and $\hat{P}_2 = \frac{x_2}{n_2}$

Exercise: 7

Q. Solution,

$n = 39$

$\bar{x} = 22\,mg$

$s = 4\,mg$

(a) $(1-\alpha)100\% = 95\%$

$\alpha = 5\% = 0.05$

$CI: \bar{x} \pm Z_{\alpha/2} \cdot SE(\bar{x})$

Critical value $= Z_{\alpha/2} = Z_{0.025} = 1.96$ (from table)

(ignore − sign to get critical value)

Standard error

$$SE(\bar{x}) = \frac{s}{\sqrt{n}}$$

$$= \frac{4}{\sqrt{39}}$$

We know,

$$CI = \bar{x} \pm Z_{\alpha/2} \cdot SE(\bar{x})$$

$$= 22 \pm 1.96 \times \frac{4}{\sqrt{39}}$$

$$= (L, U)$$

$$= (20.745, 23.255)$$

t मा infinity मान

$n < 30$, t herni
$n > 30$, z hereni h

3. Solution,

$n = 16$

$\bar{x} = 3.42$ mg

$s = 0.68$ mg

$(1-\alpha)\ 100\% = 99\%$

$\alpha = 1\%$

$= 0.01$

Critical value,

$t(\alpha/2, n-1) = t(0.005, 15)$

$= 2.947$ (from table)

Standard error

$$S.E(\bar{x}) = \frac{1}{\sqrt{n}} = \frac{0.68}{4} = 0.17$$

We know,

$CI = \bar{x} \pm t(\alpha/2, n-1)\ SE(\bar{x})$

$= 3.42 \pm 2.947 \times 0.17$

$= 3.42 \pm 0.5001$

$(L, U) = (2.92, 3.92)$

Date: Page:

7. Solution,

$n_1 = 80$, $n_2 = 35$

$\overline{x_1} = 50$, $\overline{x_2} = 45$

$S_1^2 = 16$ and $S_2^2 = 25$

The samples are drawn from two normal populations $N_1(\mu_1, \sigma^2)$ and $N_2(\mu_2, \sigma^2)$

$(1-\alpha)\,100\% = 95\%$ $\quad \alpha = 0.05$ $\quad = 5\%$

Critical value,

$Z_{\alpha/2} = Z_{0.025} = 1.96$

Standard error

$$SE(\overline{x_1} - \overline{x_2}) = S_p \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

where, $S_p = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}}$

$$SE(\overline{x_1} - \overline{x_2}) = \sqrt{\frac{(n_1-1)S_1^2 + (n_2-1)S_2^2}{n_1+n_2-2}} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}$$

We know, $= 0.876$

$$CI = | \overline{x_1} - \overline{x_2} | \pm Z_{\alpha/2}\, SE(\overline{x_1} - \overline{x_2})$$

$$= (50-45) \pm 1.96 \times 0.876$$

$$= 5 \pm 1.717$$

$$(L, U) = (3.28,\ 6.717)$$

Date: Page:

8 Soln:

(CI for dependent samples)

Matched pairs or pair samples:

If Two samples taken from the same population or different populations are pair wise dependent then such samples are said to be match pairs or pair sample. for eg: marks of the students before and after tution, sugar level of the patient before and after medication, ~~cells of~~ sells of the artides before and after advertisement etc.

Before (x)	After (y)	$d = y - x$	$(d-\bar{d})^2$ ~~$(y-\bar{d})^2$~~
10	12	2	
15	17	2	
9	8	-1	
3	5	2	
7	6	-1	
12	11	-1	
16	18	2	
17	20	3	
4	3	-1	
		$\Sigma d = 7$	$\Sigma (d-\bar{d})^2 =$

23.46

Date: Page:

We know,

$$\bar{d} = \frac{\Sigma d}{n} = \frac{7}{9} = 0.778$$

$$Sd = \sqrt{\frac{\Sigma(d-\bar{d})^2}{n-1}} = 1.7007$$

$(1-\alpha) \times 100\% = 95\%$

$\alpha = 5\% = 0.05$

Critical value

$t(\alpha/2, n-1) = t(0.025, 8)$

$= 2.306$

We know,

$$CI = \bar{x} \pm t_{(\alpha/2, n-1)} * \frac{Sd}{\sqrt{n}}$$

$$= 0.77 \pm 2.306 * \frac{1.7007}{\sqrt{9}}$$

$$= 0.77 \pm 1.307$$

$(UL, LL) = (2.077, -0.537)$

Date: Page:

Proportion:

① Solution,

$n = 100$

$\hat{p} = 55\% = 0.55$

$\hat{q} = 1 - \hat{p} = 0.45$

$(1-\alpha)\,100\% = 95\%$

$\alpha = 5\% = 0.05$

Critical value,

$Z_{\alpha/2} = Z_{0.025} = 1.96$

Standard error

$$S.E(\hat{p}) = \sqrt{\frac{\hat{p}\cdot\hat{q}}{n}}$$

$$= \sqrt{\frac{0.55 \times 0.45}{100}}$$

$$= 0.049$$

We know,

$$CI = \hat{p} \pm Z_{\alpha/2}\, S.E(\hat{p})$$

$$= 0.55 \pm 1.96 \times 0.049$$

$$= 0.55 \pm 0.097$$

$$(U, L) = (0.64, 0.453)$$

Date: Page:

7. Solution,

$x_1 = 2400 \quad n_1 = 5000$

$x_2 = 1500 \quad n_2 = 2500$

Now,

$$\hat{p}_1 = \frac{x_1}{n_1} = \frac{2400}{5000} = 0.48$$

$$\hat{q}_1 = 1 - \hat{p}_1 = 1 - 0.48 = 0.52$$

$$\hat{p}_2 = \frac{x_2}{n_2} = \frac{1500}{2500} = 0.6$$

$$\hat{q}_2 = 1 - \hat{p}_2 = 1 - 0.6 = 0.4$$

$(1-\alpha) \times 100\% = 95\%$

$\alpha = 5\% = 0.05$

Critical value,

$Z_{\alpha/2} = Z_{0.025}$

$= 1.96$

Standard error

$$SE.(\hat{p}_1 - \hat{p}_2) = \sqrt{\frac{\hat{p}_1\hat{q}_1}{n_1} + \frac{\hat{p}_2\hat{q}_2}{n_2}}$$

$$= \sqrt{\frac{0.48 \times 0.52}{5000} + \frac{0.6 \times 0.4}{2500}}$$

$= 0.012$

Date: Page:

We know,

$$CI = |\hat{p}_1 - \hat{p}_2| \pm Z_{\alpha/2} \; S.E(\hat{p}_1 - \hat{p}_2)$$

$$= |0.48 - 0.6| \pm 1.96 \times 0.012$$

$$= 0.12 \pm 0.02352$$

$$\therefore (L, U) = (0.09648, 0.14352)$$

Hypothesis testing

Hypothesis: A hypothesis is an assumption or claim made about true population parameter with certain level of confidence and its truthness or falsity will be determined after testing it through number of systematic steps.

Types of hypothesis

(i) Null hypothesis (H_0)

(ii) Alternative hypothesis (H_1)

Date: Page:

(i) Null Hypothesis (H_0)

The assumption of no significance difference between true population parameter and ~~plain~~ claim value about that population parameter is known as null hypothesis. It is denoted by H_0 and defined by one of the following ways.

$H_0 : \theta = \theta_0$ (common)

OR

$H_0 : \theta \leq \theta_0$ OR $H_0 : \theta \geq \theta_0$

where, θ is true popⁿ parameter

and θ_0 is ~~plai~~ claim value about that population parameter.

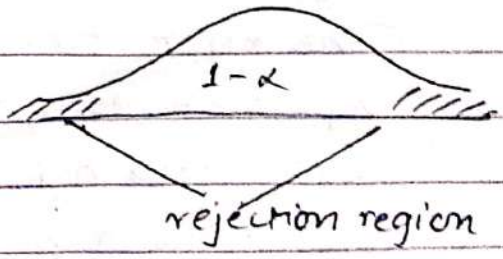
(ii) Alternative hypothesis (H_1)

The assumption of significance difference between ~~true~~ popⁿ parameter and ~~plain valu~~ claim value about it is known as alternative hypothesis.

In other words, a complementory hypothesis of null hypothesis is known as alternative hypothesis.

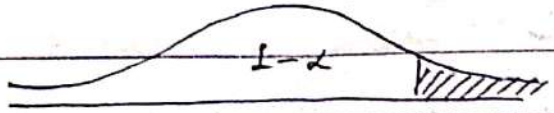
It is denoted by H_1 and defined by one of the following ways

$H_1 : \theta \neq \theta_0$ (Two tail)



OR

$H_1 : \theta > \theta_0$ (right ~~side~~ tail)



$H_1 : \theta < \theta_0$ (left ~~side~~ tail)

Date: Page:

Types of decision:

① Right decisions
i) Accepting Ho when it is true.
ii) Rejecting Ho when it is false.

② Wrong decisions
i) Rejecting Ho when it is true
(Type I error)
ii) Accepting Ho when it is false
(Type II error)

Type I error:

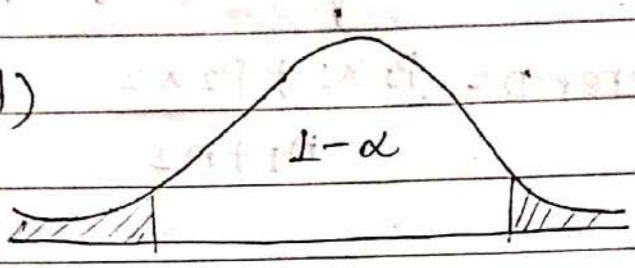
The error committed in rejecting the true null hypothesis is known as type I error. In this type of the error, null hypothesis is rejected even it is true. It is denoted by α. Since, the producers have to bear this kind of risks so it is also known as producers risks.

Type II error.

The error committed in accepting the false null hypothesis is known as type II error. In this type of error, null hypothesis is accepted even it is false. It is denoted by β (Greek letter). Since, the consumers have to bear this kind of risks. so, it is also known as consumers risks.

Tails of test in hypothesis testing:

① Two tail test
(equal / unequal, unbiased)



If the rejection region lies on the both side of the probability curve while testing hypothesis, such type of test is called two tail test. This type of test can be identified if the problem of a statement consists keywords: as much as, ~~equal~~ exactly, exactly equal to, no significance, no longer than, no shorter than, unbaised, change, difference etc. Then, we need to understand that it belongs to two tail test. In this type of test, hypothesis can be defined as follows

Null hypothesis

$H_0 : \theta = \theta_0$

Alternative hypothesis

$H_1 : \theta \neq \theta_0$

Right tail test

If the rejection region lies only right side of probability curve while testing hypothesis, then such type of test is called right tail test. In this type of test, the problem of the statement consists keywords: greater than, increase, more than, superior than, improve than, gained etc. Then, we need to understand, it belongs to right tail test. In this type of test, hypothesis can be defined as follows:

(i) Null hypothesis

$H_0: \theta = \theta_0$

OR, $H_0: \theta \leq \theta_0$

(ii) Alternative

$H_1: \theta > \theta_0$ (Right)

Left tail test.

If the rejection region lies only left side of the probability curve while testing hypothesis then such type of test is called left tail test. In this type of test, the problem of a statement consists keywords: less than, decrease, reduce, inferior than, below, smaller than, shorter than etc. Then, we need to understand, it belongs to left tail test. In this type of test, hypothesis can be defined as follows:

(i) Null hypothesis

$H_0: \theta = \theta_0$

$H_0: \theta \geq \theta_0$

Alternative hypothesis

$H_1: \theta < \theta_0$ (Right)

Imp

Systematic steps of testing hypothesis

(1) Concerning mean
(2) Concerning ~~population~~ proportion.
Concerning single parameter

Step I: (Formulation of hypothesis)

Null hypo: $H_0: \theta = \theta_0$ → claimed value
↳ true popⁿ parameter

OR, $H_0: \theta \geq \theta_0$

OR $H_0: \theta \leq \theta_0$

Alternative hypothesis:

$H_1: \theta \neq \theta_0$ (Two tail)

OR,

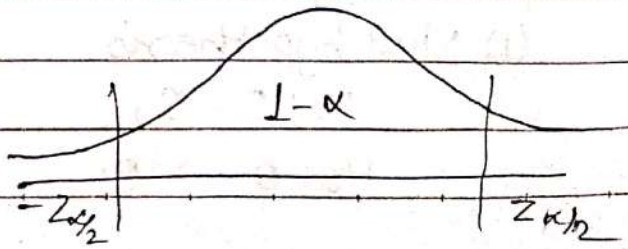
$H_1: \theta > \theta_0$ (Right tail)

OR,

$H_1: \theta < \theta_0$ (Left tail)

Step II: (Choosing level of significance)
Choose $\alpha = 5\%$ unless it is stated

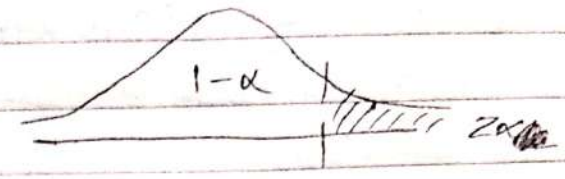
Step III: (Finding critical value/ Tabulated value)
two tail:
$Z_{tab} = Z_{\alpha/2}$



MAYUR
Date
Page

One tail:

$Z_{tab} = Z_\alpha$



for small sample $n < 30$ we use t instead of 30.

Step IV: (Calculated value)

Test statistics under $H_0: \theta = \theta_0$

$$Z_{cal} \text{ or } t_{cal} = \frac{\hat{\theta} - \theta_0}{SE(\hat{\theta})}$$

Step V: (Decision)

or, $|Z_{cal} \text{ or } t_{cal}| < Z_{tab} \text{ or } t_{tab}$

accept H_0 and reject H_1

Read yourself !!

Exercise: 8 (Page 388)

Qno2. Solution,

$\mu_0 = 150.2$ ~~150.2~~ 115.2 (claimed value)

$n = 49$ (for sample)

$\bar{x} = 117.4$

$\sigma = 8.4$

$\alpha = 0.05$

Step I:

$H_0 =$ $\mu = 115.2$

$H_1 : \mu > 115.2$ (right tail)

Step II:

$\alpha = 0.05$

Step III: (critical value)

$Z_{tab} = Z_\alpha$

$= Z_{0.05}$

$= 1.645$

Step IV:

Test statistics under ~~the~~ $H_0 : \mu = 115.2$

$$Z_{cal} = \frac{\bar{x} - \mu_0}{SE(\bar{x})} = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$$

$$= \frac{117.4 - 115.2}{8.4/\sqrt{49}}$$

$|Z_{cal}| = 1.83 > 1.645$

$|Z_{cal}| > |Z_{tab}|$

Null hypothesis is rejected and alternative hypothesis is accepted.

Step VI Conclusion:

True mean miles/gallen is greater than 115.2

Q. Z test for difference

Solution,

$n_1 = 40$ $n_2 = 50$

$\overline{x_1} = 74$ $\overline{x_2} = 78$

$S_1 = 8$ $S_2 = 7$

$\alpha = 0.05$

Step I:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$

Step II:

$\alpha = 0.05$

Step III (critical value)

$Z_{tab} = Z_{\alpha/2}$

$= Z_{0.025}$

$= 1.96$

Step IV:

Test statistics under $H_0: \mu_1 = \mu_2$

$$Z_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{SE(\bar{x}_1 - \bar{x}_2)}$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

~~Step V :~~

$$|Z_{cal}| = 2.48$$

$$|Z_{cal}| > Z_{tab}$$

H_0 is rejected and H_1 is accepted.

Step V: Conclusion:

There is significance difference beth the two values of performances of different classes.

test in single mean

Q. Solution,

$n = 16$

$\bar{x} = 41.5$ inches

$\Sigma (x - \bar{x})^2 = 135$ sq. inches

$\mu_0 = 43.5$ inches

$(1-\alpha)100 = 95\%$

$\alpha = 5\% = 0.05$

Step I:

$H_0 : \mu = 43.5$

$H_1 : \mu \neq 43.5$

Step II:

$\alpha = 0.05$

Step III

Critical value

$Z_{tab} = Z_{\alpha/2}$

$= Z_{0.025}$

$= 1.96$

Step IV:

Test statistics under $H_0: \mu = 43.5$

Step 5. $t_{tab} = t_{\alpha/2}\ n-1$

$= t_{0.025}\ 15$

$= 2.131$

$$S = \sqrt{\frac{\Sigma(x-\bar{x})^2}{n-1}} = \sqrt{\frac{135}{15}} = 3$$

$$t_{cal} = \frac{\bar{x} - \mu_0}{S/\sqrt{n}}$$

$$= \frac{41.5 - 43.5}{3/\sqrt{16}}$$

$$= -2.68$$

$|t_{cal}| = 2.68$

$t_{cal} > t_{tab}$

H_0 is rejected and H_1 is accepted

Conclusion;

MAYUR
Date
Page

Exercise: 8

t-test for difference.

1. Soln.

D_1	D_2	$(D_1 - \overline{D_1})^2$	$(D_2 - \overline{D_1})^2$
8	10	1	0.5625
12	8	9	7.5625
13	12	16	1.5625
9	15	0	18.0625
3	6	36	22.56
8	11	1	0.0625
10	12	1	1.5625
9	12	0	1.5625
$\Sigma D_1 = 72$	$\Sigma D_2 = 86$	64	53.4975

We know,

$$\overline{D_1} = \frac{\Sigma D_1}{n} = \frac{72}{8} = 9$$

$$\overline{D_2} = \frac{\Sigma D_2}{n} = \frac{86}{8} = 10.75$$

$$S_{D_1} = \sqrt{\frac{\Sigma (D_1 - \overline{D_1})^2}{n-1}} = \sqrt{\frac{64}{8}} = 3.02$$

$$S_{D_2} = \sqrt{\frac{\Sigma (D_2 - \overline{D_2})^2}{n-1}} = 2.7645$$

Step 1;

$\mu D_2 \leq \mu D_1$

$\mu D_2 > \mu D_1$

Step II;

$\alpha = 5.1 = 0.05$

Step III,

$$\nu : \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{\left(\frac{s_1^2}{n_1}\right)^2}{n_1 - 1} + \frac{\left(\frac{s_2^2}{n_2}\right)^2}{n_2 - 1}} = 13.889 \approx 14$$

$t_{tab} : t(\alpha, \nu) = t(0.05, 4) = 1.761$

Step IV: Test statistics under

$H_0 : \mu_1 = \mu_2$

$$t_{cal} = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = 1.21$$

The relationship betn a dependent variable and one or more than one independent variable is known as correlation.

foreg: the relationship betn demand and supply, quantity and cost, the production of crops and quality of seeds.

Correlation coefficient:

The numerical measurement of strengths of relationship betn a dependent variable and one or more than one independent variables is known as correlation coefficient. There are three types of correlation coefficient 1) Simple correlation coefficient
2) Partial " "
3) Multiple " "

1. Simple correlation coefficient

The numerical measurement of strength of relationship betn a dependent and an independent variable is known as simple correlation coefficient.

If X is dependent variable and Y is an independent variable then the simple correlation coefficient betn X and Y is denoted by r_{xy} and given by following formula:

$$r_{xy} = \frac{Cor(x,y)}{\sqrt{var\,x}\ \sqrt{var\,y}}$$

where, $$Cor(x,y) = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{n-1}$$

Date: Page:

→ If two independents variables are kept constant, it is said to be partial correlation coefficient of second order and soon.

Coefficient of partial determination.

The square of partial correlation coefficient is known as coefficient of partial determination.

$$r_{12.3}^2 = \frac{(r_{12} - r_{23} \cdot r_{31})^2}{(1 - r_{13})^2 (1 - r_{23})^2}$$

It is used to interpret the value of partial correlation coefficient.

If $r_{12.3} = 0.9$, then,

$$r_{12.3}^2 = 0.81 = 81\%.$$

means 81% of variation of x_1 is explained by x_2 by keeping x_3 constant.

Multiple correlation coefficient.

The numerical measurement of strength of relationship between a dependent variable and two or more than two independent variables by keeping the effect of all independent variables together.

They can be obtained by solving eqs :

$$\Sigma y = na + b\Sigma x + c\Sigma z$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz$$

Properties of regression coefficient.

The simple correlation coefficient is geometric mean of two regression coefficients.

Both regression coefficients have ~~sio~~ same sign.

If one regression coefficient is greater than unity, then another regression coefficient must be less than unity.

If It is independent of change of origin but not of scale.

The arithmetic mean of two regression coefficient is not less than correlation coefficient.

The mean ~~is~~ always passes through the line of regression.

If $r=0$ then two lines of regression are parallel.

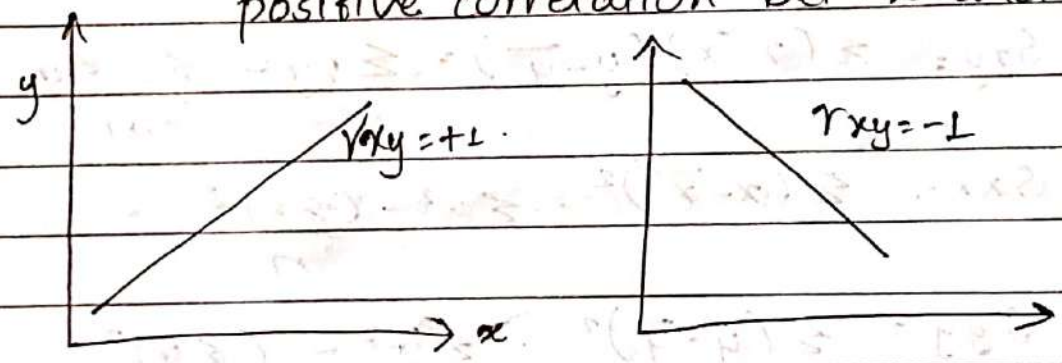
Distinguish between correlation and regression.

Date: Page:

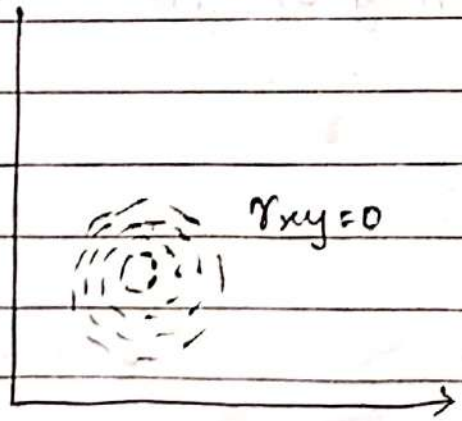
Properties of simple correlation coefficient.

1. The dependent and independent variables are interchangable. $r_{xy} = r_{yx}$
2. It is unitless.
3. It is independent of change of scale.
4. It's value always lies beth -1 and +1.
5. It is the geometric mean of two simple regression coefficient. $r_{xy} = \sqrt{b_{xy}\, b_{yx}}$

If $r_{xy} = +1$, there is perfectly positive correlation beth x and y.



If $r_{xy} = -1$, there is perfectly negative correlation between x and y.



If $r_{xy} = 0$, there is no correlation between x and y.

Data: Page:

a and
How to find b?

Method I: 'Direct method'

$$b = byx = \frac{n\Sigma xy - \Sigma x \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$a = \bar{y} - b\bar{x} \quad \because y = \frac{\Sigma y}{n}, \quad \bar{x} = \frac{\Sigma x}{n}.$$

Method II: solving its normals

$$\Sigma y = na + b\Sigma x$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2$$

x on y

$$x = a + by$$

Similar as y on x.

Multiple regression:

The relationship between a dependent variable and two or more than two independent variables in which the value of dependent variable is predicted with the help of independent variables is known as multiple regression. Regression line of y on x and z.

$$y = a + bx + cz$$

How to find a, b and c?

Date:

$$\text{Var}(x) = \frac{\Sigma (x-\bar{x})^2}{n-1}$$

$$\text{var}(y) = \frac{\Sigma (y-\bar{y})^2}{n-1}$$

$$r_{xy} = \frac{\Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\Sigma (x-\bar{x})^2 \Sigma (y-\bar{y})^2}}$$ (central moment ~~morf~~ formula)

$$r_{xy} = \frac{S_{xy}}{\sqrt{S_{xx}}\sqrt{S_{yy}}}$$

where,

$$S_{xy} = \Sigma (x-\bar{x})(y-\bar{y}) = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$S_{xx} = \Sigma (x-\bar{x})^2 = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$S_{yy} = \Sigma (y-\bar{y})^2 = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

4. $$r_{xy} = \frac{n \Sigma xy - \Sigma x \Sigma y}{\sqrt{n \Sigma x^2 - (\Sigma x)^2}\sqrt{n \Sigma y^2 - (\Sigma y)^2}}$$

2>

Date: Page:

$$Se = \sqrt{\frac{Syy - (Sxy)^2 - Sxx}{n-2}}$$

$= 8.874$

$(1-\alpha)\ 100\% = 95\%$

$\alpha = 5\% = 0.05$

$t(\alpha/2, n-2) = t(0.025, 2) = 4.303$

$$CI = b \pm t(\alpha/2, n-2)\ se \sqrt{\frac{1}{Sxx}}$$

$= 9.15 \pm 4.303 \times 8.874 \times 0.0447$

$= 9.15 \pm 1.707$

$= (7.45, 10.857)$

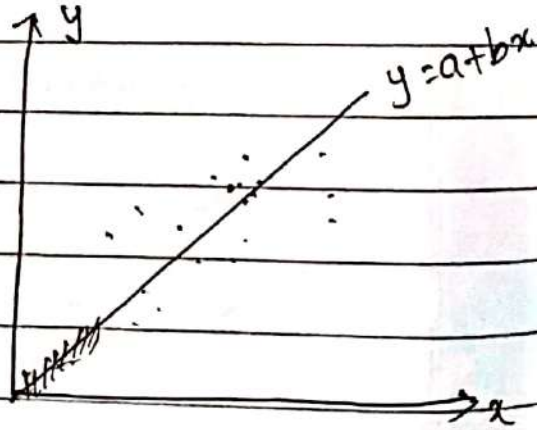
Date: Page:

Regression:

The relationship between a dependent variable and one or more than one independent variables in which value of dependent variable is predicted with the help of independent variables is known as regression. The correlation analyse the strength of relationship betn the dependent and independent variables. On the other hand, regression analyse the nature of relationship betn the dependent and independent variables.

Types of regression:

1) Simple
2) Multiple



4. Simple regression line:

y on x

$y = a + bx$

a = y intercept

b = b_{yx} = slope or regression coefficient of y on x.

They can be obtained by solving eqs:

$$\Sigma y = na + b\Sigma x + c\Sigma z$$

$$\Sigma xy = a\Sigma x + b\Sigma x^2 + c\Sigma xz$$

Properties of regression coefficient.

The simple correlation coefficient is geometric mean of two regression coefficients.

Both regression coefficients have ~~sio~~ same sign.

If one regression coefficient is greater than unity, then another regression coefficient must be less than unity.

If It is independent of change of origin but not of scale.

The arithmetic mean of two regression coefficient is not less than correlation coefficient.

The mean ~~is~~ always passes through the line of regression.

If $r=0$ then two lines of regression are parallel.

Distinguish between correlation and regression.

Date: Page:

x	y	x^2	y^2	xy
30	160	900		4800
40	240	1600		9600
50	330	2500		16500
60	435	3600		2100
		$\Sigma x^2 =$		

Let the line of regression is: $y = a + bx$

$$b = \frac{n\Sigma xy - \Sigma x \cdot \Sigma y}{n\Sigma x^2 - (\Sigma x)^2}$$

$$\bar{x} = \frac{\Sigma x}{n} = \frac{180}{4} = 45$$

$$\bar{y} = \frac{\Sigma y}{n} = \frac{1165}{4} = 29$$

$$a = \bar{y} - b\bar{x}$$

Substituting the value of $\bar{y}, b, \bar{x}$

$a =$

CI for slope

$$\int xy = \Sigma xy - \frac{\Sigma x \Sigma y}{n}$$

$$\int xx = \Sigma x^2 - \frac{(\Sigma x)^2}{n}$$

$$\int yy = \Sigma y^2 - \frac{(\Sigma y)^2}{n}$$

If X_1 is a dependent variable and X_2 and X_3 are independent variables then the multiple correlation coefficient between X_1, X_2 and X_3 is denoted by $R_{1.23}$ and given by the relation as:

$$R_{1.23} = \sqrt{\frac{r_{12}^2 + r_{13}^2 - 2 r_{12} r_{13} r_{23}}{1 - r_{23}^2}}$$

Properties:

→ Its value lies between 0 to 1.

→ The position of the subscripts after the dots can be interchanged: $R_{1.23} = R_{1.32}$

→ $R_{1.23} = 0$ If and only if $R_{12} = 0$ and $R_{13} = 0$

→ $R_{1.23} \geq r_{12}$, r_{23} and r_{13}

Coefficient of multiple determination.

- The square of multiple correlation coefficient is known as coefficient of multiple determination. It is used to interpret the value of multiple correlation coefficient.

If $R_{1.23} = 0.8$ then $R^2_{1.23} = 0.64$ means 64% of variation in dependent ~~variation~~ variable X_1 is explained by independent variables X_2 and X_3 together.

Theorem.

Prove that: $-1 \leq r_{xy} \leq 1$

Proof:

Consider the sum of square as:

$$\Sigma \left(\frac{x-\bar{x}}{s_x} \pm \frac{y-\bar{y}}{s_y} \right)^2 \geq 0$$

$$\text{or, } \Sigma \left\{ \frac{(x-\bar{x})^2}{s_x^2} \pm 2 \cdot \frac{(x-\bar{x})}{s_x} \cdot \frac{(y-\bar{y})}{s_y} + \frac{(y-\bar{y})^2}{s_y^2} \right\} \geq 0$$

$$\text{or, } \frac{\Sigma (x-\bar{x})^2}{s_x^2} \pm \frac{2 \Sigma (x-\bar{x})(y-\bar{y})}{s_x \cdot s_y} + \frac{\Sigma (y-\bar{y})^2}{s_y^2} \geq 0 \quad \text{— (1)}$$

Since,

$$s_x^2 = \frac{\Sigma (x-\bar{x})^2}{n-1} \Rightarrow \Sigma (x-\bar{x})^2 = (n-1) s_x^2$$

$$s_y^2 = \frac{\Sigma (y-\bar{y})^2}{n-1} \Rightarrow \Sigma (y-\bar{y})^2 = (n-1) s_y^2$$

Then, from (1)

$$\frac{(n-1) s_x^2}{s_x^2} \pm \frac{2 \Sigma (x-\bar{x})(y-\bar{y})}{\sqrt{\frac{\Sigma (x-\bar{x})^2}{n-1}} \sqrt{\frac{\Sigma (y-\bar{y})^2}{n-1}}} + \frac{(n-1) s_y^2}{s_y^2} = 0$$

$$\text{or, } 2(n-1) \pm 2(n-1) r_{xy} \geq 0$$

$$\text{or, } 2(n-1)(1 \pm r_{xy}) \geq 0$$