classmate

Date
Page

# Unit-1 Random variables & Probability distribution

Probablity (Introduction) ] Definition of probability was given by Pierre Siemen Laplace in 1795
T. Carden an Italian Physician and a mathematician wrote the first book on probability name book of games

1. Games of chance:-
probability has been used extensively in many areas such as biology, physics etc. It also used in forecast of weather, result of election, population, earthquakes, crop production etc.

2. Random experiment:-
An experiment is said to be random if its outcome cannot be predicted that is the outcome of an experiment does not obey any rule.
Eg:- 1. Tossing a coin is a random exp.
2. Throwing a die

3. Outcome:-
The result of a random experiment will be called as outcome.

4. Sample space:-
The set of all possible outcomes of an experiment are called a sample space
Ex:- 1. A coin is tossed neither head or tail occurs.
2. Sample space $S = \{H, T\}$

5. Event:-
Any subset E of a sample space is called an event
Ex:- When coin is tossed getting a head.

⇒ Probability (definition):-
The No. of favourable outcomes of an event by total No. of possible outcomes of the Sample space

$$P(E) = \frac{n(E)}{n(S)}$$

Q. A class consists of 6 girls and 10 boys if a committee of 3 is chosen at random from the class. Find the probability that

i). 3 boys are selected

ii). exactly 2 girls are selected

Sol: i). $n(S) = 16C_3$

$n(E) = 10C_3$

$P(E) = \frac{n(E)}{n(S)} = \frac{10C_3}{16C_3} = \frac{3}{14}$

ii). $n(E) = 6C_2 \times 10C_1$

$P(E) = \frac{n(E)}{n(S)}$

$= \frac{6C_2 \times 10C_1}{16C_3} = \frac{15}{56}$

6.5

- Random variable -

A real variable X whose values is determined by the outcome of a random experiment is called a random variable

Ex:- Consider of tossing a coin twice

$S = \{S_1, S_2, S_3, S_4\}$

where $S_1 = \{HH\}$

$S_2 = \{TH\}$

$S_3 = \{HT\}$

$S_4 = \{TT\}$

let $X(S)$ = The no. of heads

$X(S_1) = 2 \quad X(S_2) = 1 \quad X(S_3) = 1 \quad X(S_4) = 0$

* Types of random variables
1. Discrete random variable
2. Continuous random variable

1. Discrete random variable:-
A random variable X which can take finite number of discrete values in an interval of domain is called a discrete random variable
Ex:- A random variable denoting number of students in a class variable

2. Continuous random
A random variable X which can take value continuously that is which take all possible values in a given interval is called Continuous random variable
Ex:- height, age, weight of individuals

- Probability function of discrete random variables-
- For a discrete random variable X the real valued function $P(x)$ such that $P(X=x)$ i.e, $P(X=x)=P(x)$ then $P(x)$ is called probability function or probability density function.

* Properties of Probability function:-
- if $p(x)$ is probability function of random variable X then it possess following properties
1) $P(x) \geq 0 \ \forall x$
2) $\sum P(x) = 1 \ \forall x$
3) $P(x)$ cannot be negative for any value of $x$.

* Mean of a probability discrete distribution-
- It is denoted as a $\mu = \sum P_i x_i = E(x)$ where $E(x) \rightarrow$ Expectation

* Variants of a probability discrete distribution - denoted as $\sigma^2$ and It is defined as $\sigma^2 = \sum P_i x_i^2 - \mu^2$

classmate

Date ______
Page ______

* Standard Deviation - It is positive sq. root of variants. It is denoted as

$$\sigma = \sqrt{\sum P_i x_i^2 - \mu^2}$$

* Problems:

1. Two dice are thrown. Let $X$ assign to each point $(a,b)$ in $S$. The maximum of its number that is $X(a,b) = \max(a,b)$ find the prob. distrib. $X$ is a random variable with $X(S) = \{1,2,3,4,5,6\}$ Also find mean, variants, std. deviation

sol: ~~n(S)=36~~ The total no. of cases i.e., $n(S) = 36$

The maximum number could be for $X(S) = \{1,2,3,4,5,6\}$ (3,1)

i.e., $X(a,b) = \max(a,b)$ (1,3) (2,2)

The no. 1 will appear maximum in one case only i.e., $(1,1)$ (3,3) (2,3)

$\therefore P(X=1) = P(1) = \frac{1}{36}$

The no. 2 will occur maximum in 3 cases i.e., $(1,2), (2,1), (2,2)$

$\therefore P(X=2) = P(2) = \frac{3}{36}$

$||^{ly}$ $P(3) = \frac{5}{36}$ $[(1,3), (3,1), (2,3), (3,2), (3,3)]$

$P(4) = \frac{7}{36}$ $[(4,1), (1,4), (2,4), (4,2), (3,4), (4,3), (4,4)]$

$P(5) = \frac{9}{36}$

$P(6) = \frac{11}{36}$

The P.D

| $X=x$ | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|
| $P(X=x) = P(x)$ | $1/36$ | $3/36$ | $5/36$ | $7/36$ | $9/36$ | $11/36$ |

$\mu = \sum P_i x_i$

$= \frac{1}{36} + \frac{2\times 3}{36} + \frac{3\times 5}{36} + \frac{4\times 7}{36} + \frac{5\times 9}{36} + \frac{6\times 11}{36}$

$= \frac{161}{36} = 4.47$

classmate

Date

Page

$\sigma^2 = \sum P_i x_i^2 - \mu^2$

$= \left[1^2 \times \frac{1}{36} + 2^2 \times \frac{3}{36} + 3^2 \times \frac{5}{36} + 4^2 \times \frac{7}{36} + 5^2 \times \frac{9}{36} + 6^2 \times \frac{11}{36}\right] - [4.47]^2$

$= \frac{791}{36} - (4.47)^2$

$= 21.972 - (4.47)^2$

$= 1.973$

$SD = \sqrt{\text{variance}} = \sqrt{1.973} = 1.404.$

2. A random variable X has the Probability funcn.

| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 |
|---|---|---|---|---|---|---|---|---|
| $P(x)$ | 0 | $k$ | $2k$ | $2k$ | $3k$ | $k^2$ | $2k^2$ | $7k^2+k$ |

i) find k

ii) Evaluate $P(X<6)$, $P(X\geq 6)$, $P(0<X<5)$, $P(0\leq X\leq 4)$

iii) if $P(X\leq t) > 1/2$

find minimum value of $t$

iv) determine the distribution funcn of X

v) mean, variance, S.D.

i) $\sum P(x) = 1$

$k + 2k + 2k + 3k + k^2 + 2k^2 + 7k^2 + k = 1$

$10k^2 + 9k - 1 = 0$

$\boxed{k = 0.1}$

ii) 1. $P(X<6) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4) + P(X=5)$

$= 0 + k + 2k + 2k + 3k + k^2$

$= 0.81$

2. $P(X\geq 6) = \cancel{P(X=0)+P}\ 1 - P(X<6)$

$= 1 - 0.81$

$= 0.19$

classmate
Date
Page

3. $P(0<x<5) = P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= k + 2k + 2k + 3k$

$= 0.8$

4. $P(0 \le x \le 4) = P(X=0) + P(X=1) + P(X=2) + P(X=3) + P(X=4)$

$= 0 + k + 2k + 2k + 3k$

$= 0.8$

v) $\text{Mean}(\mu) = \Sigma P_i x_i$

$= 1\times k + 2\times 2k + 3\times 2k + 4\times 3k + 5\times k^2 + 6\times 2k^2 + 7(7k^2+k)$

$= 30k + 66k^2$

$= 3.66$

$\text{Variance}(\sigma^2) = \Sigma P_i x_i^2 - \mu^2$

$= 1^2\times k + 2^2\times 2k + 3^2\times 2k + 4^2\times 3k + 5^2\times k^2 + 6^2\times 2k^2 + 7^2(7k^2+k) - (3.66)^2$

$= k + 8k + 18k + 48k + 25k^2 + 72k^2 + 343k^2 + 49k - (3.66)^2$

$= 124k + 440k^2 - (3.66)^2$

$= 3.404$

$S.D = \sqrt{\text{variance}} = \sqrt{3.404} = 1.843$

iii) $P(X \le t) > \frac{1}{2}$

$P(X \le 0) = 0 < \frac{1}{2}$

$P(X \le 1) = P(X=0) + P(X=1)$

$= k = 0.1 < \frac{1}{2}$

$P(X \le 2) = P(X=0) + P(X=1) + P(X=2)$

$= 0 + k + 2k = 3k = 0.3 < \frac{1}{2}$

$P(X \le 3) = P(X=0) + P(X=1) + P(X=2) + P(X=3)$

$= 0 + k + 2k + 2k = 5k = 0.5 = \frac{1}{2}$

$P(X \le 4) = 0.8 > \frac{1}{2}$

$P(X \le 5) = 0.81 > \frac{1}{2}$

| $x$ | $F(x) = P(X \le x)$ |
|---|---|
| 0 | 0 |
| 1 | 0.1 |
| 2 | 0.3 |
| 3 | 0.5 |
| 4 | 0.8 |
| 5 | 0.81 |
| 6 | 0.83 |
| 7 | 1 |

iv) $t = 4$

Q. A Sample of 4 items is selected from a box containing 12 items of which 5 are defective. find the expected number E of a defective item.

Sol. Let X denotes the no. of defective items among 4 items drawn from 12 items

$\therefore X = \{0, 1, 2, 3, 4\}$

The good items are 7

bad items are 5

$n(S) = {}^{12}C_4$

$P(X=0) = P(\text{No defective}) = \frac{{}^{7}C_4}{{}^{12}C_4} = 0.07$

$P(X=1) = P(1 \text{ def., } 3 \text{ Good}) = \frac{{}^{5}C_1 \cdot {}^{7}C_3}{{}^{12}C_4} = 0.35$

$P(X=2) = P(2 \text{ def., } 2 \text{ Good}) = \frac{{}^{5}C_2 \cdot {}^{7}C_2}{{}^{12}C_4} = 0.42$

$P(X=3) = P(3 \text{ def, } 1 \text{ Good}) = \frac{{}^{5}C_3 \cdot {}^{7}C_1}{{}^{12}C_4} = 0.14$

$P(X=4) = P(\text{all def}) = \frac{{}^{5}C_4}{{}^{12}C_4} = 0.01$

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $P(x)$ | 0.07 | 0.35 | 0.42 | 0.14 | 0.01 |

$\mu = \sum p_i x_i$

$= 0 \times 0.07 + 1 \times 0.35 + 2 \times 0.42 + 3 \times 0.14 + 4 \times 0.01$

$= 1.65$

9:- Continuous Probability distribution -

when a random variable $X$ takes every value in an Interval it gives rise to continuous distribution of $X$.

- properties of probability of density funcn $f(x)$ -

1. $f(x) \geq 0 \; \forall x$

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

3. The probability $P(E)$ is defined as $\int_E f(x)\,dx$

=> mean(expectation) - The mean of a distribution $\mu = E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$.

if $X$ is defined from $(a,b)$ then $\mu = E(x) = \int_a^b x f(x)\,dx$

=> Median - median is the point which divides the entire distribution of two equal parts.

i.e, $\int_a^M f(x)\,dx = \int_M^b f(x)\,dx = \frac{1}{2}$

solving $M$ we get the median

=> Mode - Mode is a value of $x$ for which $f(x)$ is maximum that is mode is given by $f'(x) = 0$ & $f''(x) < 0 \; \forall \; a < x < b$.

=> Variance - It is defined as $\sigma^2 = \int_a^b x^2 f(x)\,dx - \mu^2$

Q. The prob. density func$^n$ of a continuous random variable is given by $f(x) = ce^{-|x|}$ $\forall -\infty < x < \infty$ S.T $c = \frac{1}{2}$ and also find mean, variance of the distribution. Find the probability that variable lies b/w 0 & 4

sol. (i) We know that

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

$$\Rightarrow \int_{-\infty}^{\infty} ce^{-|x|}\,dx = 1$$

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

$$\Rightarrow 2c\left[\frac{e^{-x}}{-1}\right]_0^{\infty} = 1$$

$$\Rightarrow -2c[e^{-\infty} - e^{0}] = 1$$

$$\Rightarrow -2c[0-1] = 1$$

$$\Rightarrow c = \frac{1}{2}$$

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

$$= \frac{1}{2}\int_{-\infty}^{\infty} x e^{-|x|}\,dx$$

$$= 0$$

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

$$= \frac{1}{2}\int_{-\infty}^{\infty} x^2 e^{-|x|}\,dx$$

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

$\Rightarrow \left[x^2\left[\frac{e^{-x}}{-1}\right] - 2x\frac{e^{-x}}{1} + (2)\frac{e^{-x}}{-1}\right]_0^{\infty}$

$\frac{1}{e^{\infty}}$

$\frac{1}{\infty} = 0$

$\Rightarrow \left[-x^2e^{-x} - 2xe^{-x} - 2e^{-x}\right]_0^{\infty}$

$\Rightarrow \left[-\infty^2 e^{-\infty} - 2(\infty)e^{-\infty} - e^{-\infty} + 0 + 0 + 2\right]$

$\Rightarrow 2$

$P(E) = \int_0^4 f(x)\,dx$

$= \frac{1}{2}\int_0^4 e^{-|x|}dx$

$= \frac{1}{2}\left[\frac{e^{-x}}{-1}\right]_0^4$

$= \frac{-1}{2}\left[e^{-4} - 1\right]$

$= \frac{1}{2}\left[1 - e^{-4}\right]$

$= 0.4908.$

$-c\left[e^{-x}\right]_{-\infty}^{\infty} = 1$

$-c\left[e^{-\infty} - e^{\infty}\right] = 1$

$-c\left[0 - \right.$

$\frac{1}{e^{\infty}}$

$\frac{1}{\infty}$

$\Rightarrow 0$

Q. Probability density function of a random variable $X$

$$f(x) = \begin{cases} \frac{1}{2}\sin x & \text{for } 0 \le x \le \pi \\ 0 & \text{else where} \end{cases}$$

find mean, median, mode and also find the probability b/wn $0$ & $\frac{\pi}{2}$

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

$$= \underbrace{\int_{-\infty}^{0} x f(x) dx}_{\substack{\text{odd functn} \\ \Rightarrow 0}} + \int_{0}^{\pi} x f(x) dx + \int_{\pi}^{\infty} x f(x) dx$$

$$= \int_{0}^{\pi} x f(x) dx$$

$$= \frac{1}{2}\int_{0}^{\pi} x \sin x \, dx$$

$$= \frac{1}{2}\left[x(-\cos x) - (1)(-\sin x)\right]_{0}^{\pi}$$

$$\Rightarrow \frac{1}{2}\left[\pi + 0 - 0 - 0\right] \Rightarrow \frac{\pi}{2}$$

$$P(E) = \int_{0}^{\pi/2} f(x) dx$$

$$= \frac{1}{2}\int_{0}^{\pi/2} \sin x \, dx$$

$$= \frac{1}{2}\left[-\cos x\right]_{0}^{\frac{\pi}{2}}$$

$$\Rightarrow \frac{1}{2}\left[-0 + 1\right] \Rightarrow \frac{1}{2}$$

Median:- $\int_a^M f(x)dx = \int_M^b f(x)dx = \frac{1}{2}$

$\Rightarrow \frac{1}{2}\int_0^M \sin x\,dx = \frac{1}{2}\int_M^\pi \sin x\,dx = \frac{1}{2}$

$\Rightarrow \frac{1}{2}\int_0^M \sin x\,dx = \frac{1}{2} \Rightarrow \int_0^M \sin x\,dx = 1$

$\Rightarrow [-\cos x]_0^M = 1$

$\Rightarrow -\cos M + \cos 0 = 1$

$\Rightarrow \cos M = 0$

$M = \cancel{\frac{n\pi}{2}}\ (2n+1)\frac{\pi}{2}$

& $M = \frac{\pi}{2}$

Mode - We know that

$f'(x) = 0$ & $f''(x) < 0 \rightarrow -\frac{1}{2}\sin x$

here $f(x) = \frac{1}{2}\sin x$

$\Rightarrow \frac{1}{2}\cos x = 0$

$\cos x = 0$

$x = \cos^{-1}(0)$

$x = \pi/2$

$= -\frac{1}{2}\sin\left(\frac{\pi}{2}\right)$

$= -\frac{1}{2} < 0$

$\therefore x = \pi/2$ is the mode for $f(x)$

classmate
Date
Page

Q. let X denote the sum of 2 numbers that appear when a pair of fair die is tossed find distribution funcn, mean, variance

Q. A Continuous random variable X has distribn funcn

$$F(x) = \begin{cases} 0, & \text{if } x \le 1 \\ k(x-1)^4 & \text{if } 1 < x \le 3 \\ 1 & \text{if } x > 3 \end{cases}$$

determine i) $f(x)$ ii). $k$ iii). $\mu$ iv). $\sigma^2$

1 sol. $X : \{\cancel{0}, \cancel{1}, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12\}$

$P(X=2) = \frac{1}{36}$ (1,1)

$P(X=3) = \frac{2}{36}$

$P(X=4) = \frac{3}{36}$ (3,1), (1,3), (2,2)

$P(X=5) = \frac{4}{36}$ (1,4), (2,3), (3,2), (4,1)

$P(X=6) = \frac{5}{36}$ (1,5), (2,4), (3,3), (4,2), (5,1)

$P(X=7) = \frac{6}{36}$ (1,6), (2,5), (3,4), (4,3),

$P(X=8) = \frac{5}{36}$ (2,6), (3,5), (4,4), (5,3)

$P(X=9) = \frac{4}{36}$

$P(X=10) = \frac{3}{36}$

$P(X=11) = \frac{2}{36}$

$P(X=12) = \frac{1}{36}$.

| $x$ | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| $P(x)$ | $1/36$ | $2/36$ | $3/36$ | $\frac{4}{36}$ | $5/36$ | $6/36$ | $5/36$ | $4/36$ | $3/36$ | $2/36$ | $1/36$ |

$\mu = \sum P_i x_i$

$= \frac{2}{36} + \frac{6}{36} + \frac{12}{36} + \frac{20}{36} + \frac{30}{36} + \frac{42}{36} + \frac{40}{36} + \frac{36}{36} + \frac{30}{36} + \frac{22}{36} + \frac{12}{36}$

$= \frac{252}{36}$

$\mu = 7$

$\sigma^2 = \sum P_i x_i^2 - \mu^2$

$= 2^2 \times \frac{1}{36} + 3^2 \times \frac{2}{36} + 4^2 \times \frac{3}{36} + 5^2 \times \frac{4}{36} + 6^2 \times \frac{5}{36} + 7^2 \times \frac{6}{36} + 8^2 \times \frac{5}{36} + 9^2 \times \frac{4}{36}$

$+ 10^2 \times \frac{3}{36} + 11^2 \times \frac{2}{36} + 12^2 \times \frac{1}{36}$

$= \frac{4}{36} + \frac{18}{36} + \frac{48}{36} + \frac{100}{36} + \frac{180}{36} + \frac{294}{36} + \frac{320}{36}$

classmate
Date
Page

2 sol. $f(x) = \frac{d}{dx} F(x)$

i) $f(x) = \begin{cases} 0, & \text{if } x \le 1 \\ 4k(x-1)^3, & 1 < x \le 3 \\ 0 & x > 3 \end{cases}$

ii). $\int_1^3 4k(x-1)^3 dx = 1$

$\Rightarrow 4k\left[\frac{(x-1)^4}{4}\right]_1^3 = 1$

$\Rightarrow \cancel{4}k\left[\frac{16}{\cancel{4}}\right] = 1$

$\Rightarrow k = \frac{1}{16}$

iii). $f(x) = \begin{cases} 0, & \text{if } x \le 1 \\ \frac{1}{4}(x-1)^3 & 1 < x \le 3 \\ 0 & x > 3 \end{cases}$

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

$= \int_{-\infty}^{1} 0\, dx + \int_1^3 x\frac{1}{4}(x-1)^3 dx + \int_3^{\infty} 0\, dx$

$= \cancel{\frac{1}{4}\left[\frac{(x-1)^4}{4}\right]_1^3} \Rightarrow \frac{1}{4}\left[x\frac{(x-1)^4}{4} - (1)\frac{(x-1)^5}{20}\right]_1^3$

$= \cancel{\frac{1}{4}\left[\frac{16}{4}\right]} \Rightarrow \frac{1}{4}\left[\cancel{3}\cdot\frac{16}{\cancel{4}} - \frac{32}{20}\right] \Rightarrow \frac{1}{4}\left[12 - 1.6\right]$

$\Rightarrow \mu = 2.6$

$\mu = 1$

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

$= \int_1^3 x^2\frac{1}{4}(x-1)^3 dx - (2.6)^2$

$= \frac{1}{4}\int_1^3 x^2(x-1)^3 dx - (2.6)^2$

$$= \frac{1}{4}\left[x^2\frac{(x-1)^4}{4} - 2x\frac{(x-1)^5}{20} + \frac{2(x-1)^6}{120}\right]_1^3 - (2.6)^2$$

$$= \frac{1}{4}\left[9\times\frac{16}{4} - 6\times\frac{32}{20} + 2\times\frac{64}{120} - 0\right] - (2.6)^2$$

$$\Rightarrow \frac{1}{4}\left[36 - \frac{48}{5} + \frac{16}{15}\right] - (2.6)^2$$

$\Rightarrow 5.866 \Rightarrow 0.106$

Binomial Distribution:- (B.D)

A random variable X has binomial distribution if it assumes only non-negative values and its prob. function is given by

$$P(r) = P(X=r) = {}^nc_r\, p^r q^{n-r}$$

Ex: ① The number of defective bolts in a box containing n bolts

② The number of postgraduate in a group of n men.

③ *Conditions of B.D:-

1. Trials are repeated under identical condition for a fixed numbers of numbers say n times.
2. There are only 2 possible outcomes
   Ex:- success, failure
3. The probability of success in each trial remains constant and does not change from trade to trade
4. Trials are independent that is probability of an event in any trade is not affected by the result of other ~~traders~~ trials.
   Ex:- Tossing a coin, Birth of baby, auditing a bill

* Mean of a binomial distribution:
It is defined as $\mu = E(x) = np$

* Variance$(\sigma^2) = npq$

$q = 1 - p$

$\therefore p + q = 1$

* Standard deviation of B.D $= \sqrt{npq}$

Q. 10 coins are thrown simultaneously. Find the probability of getting atleast
① 7 heads
② 6 heads

Given $n = 10$

$p$ = Probability of getting a head = $\frac{1}{2}$

$q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2}$

(i) 7 heads

here $r = 7$

$P(X = r) = {}^nC_r \cdot p^r q^{n-r}$

$P(X \geq 7) = P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$= {}^{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right) + {}^{10}C_{10} \left(\frac{1}{2}\right)^{10}$

$= 0.171$

(ii) 6 heads

here $r = 6$

$P(X = r) = {}^nC_r \cdot p^r \cdot q^{n-r}$

$P(X \geq 6) = P(X=6) + P(X=7) + P(X=8) + P(X=9) + P(X=10)$

$= {}^{10}C_6 \left(\frac{1}{2}\right)^6 \left(\frac{1}{2}\right)^4 + {}^{10}C_7 \left(\frac{1}{2}\right)^7 \left(\frac{1}{2}\right)^3 + {}^{10}C_8 \left(\frac{1}{2}\right)^8 \left(\frac{1}{2}\right)^2 + {}^{10}C_9 \left(\frac{1}{2}\right)^9 \left(\frac{1}{2}\right) + {}^{10}C_{10} \left(\frac{1}{2}\right)^{10}$

$= 0.3769$

2. out of 800 families with 5 children each how many would you expect to have

a) 3 boys

b) 5 girls

c) either 2 or 3 boys

d) atleast one boy Assume equal probability for boys and girls

sol: let the number of boys in each family = $x$

$p$ = prob. of each boy = $\frac{1}{2}$

$q$ = prob. of each girl = $\frac{1}{2}$ $\quad 1-p = 1-\frac{1}{2} = \frac{1}{2}$

$n$ = no. of children = 5

We know Binomial Distribution

$$P(X=x) = P(x) = {}^nC_x p^x q^{n-x}$$

a) $P(X=3) = {}^5C_3 \left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2$

$= 0.3125$

Thus for 800 families the probability of no. of families having 3 boys

$\Rightarrow 800 \times 0.3125$

$\Rightarrow 250$

b) $P(X=5) = P(5 \text{ girls})$ or $P(X=0)$ i.e., $P(0 \text{ boys})$

$$P(X=5) = P(5) = {}^5C_5 \left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^{5-5} = {}^5C_5 \left(\frac{1}{2}\right)^5\left(\frac{1}{2}\right)^0$$

$= 0.03125$

Thus for 800 families = 25

c) $P(X=2 \text{ or } 3) = P(X=2) + P(X=3)$

$= {}^5C_2\left(\frac{1}{2}\right)^2\left(\frac{1}{2}\right)^3 + {}^5C_3\left(\frac{1}{2}\right)^3\left(\frac{1}{2}\right)^2$

$= {}^5C_2\left(\frac{1}{2}\right)^5 + {}^5C_3\left(\frac{1}{2}\right)^5$

$= 0.625$

Thus for 800 families the probability of number of families having either 2 & 3 boys $= 0.625 \times 800$

$= 500$

d. $P(X \geq 1) = 1 - P(X=0)$

$= 1 - P(X=0)$

$= 1 - {}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^5$

$= 0.9687$

Thus for 800 families $= 0.9687 \times 800$

$= 775$

3. Out of 800 families with 4 children how many families would be expected to have

a) 2 boys and 2 girls
b) Atleast one boy
c) No girl
d) Atmost two girls

Sol. let the No. of boys in each family $= x$

$p =$ probability of each boy $= \frac{1}{2}$

$q =$ probability of each girl $= \frac{1}{2}$

$n =$ number of children $= 4$

$P(X=r) = P(r) = {}^nC_r \, p^r q^{n-r}$

a) $P(X=2) + P(X=2) = 2P(X=2)$

$= 2 \cdot {}^4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^2$

$\Rightarrow 0.75$

Thus for 800 families Prob. $= 0.75 \times 800 = 600$

b) $P(X \geq 1) = 1 - P(X<0)$

$= 1 - P(X=0)$

$= 1 - {}^4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^4$

$= 0.9375$

Thus for 800 families prob. $= 0.9375 \times 800 = 750$

c) $P(\text{No girl}) = P(X=0)$ or $P(X=4)$ i.e., $P(4 \text{ boys})$

$= {}^4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^0 = 4\left(\frac{1}{2}\right)^4 = 0.25$

Thus for 800 families prob. $= 0.25 \times 800$

$= 200$

4. The mean and variance of a binomial distribution are 4 and $4/3$. Find $P(X \geq 1)$

Sol. Given mean and variance are

$np = 4$ — ①

$npq = \frac{4}{3}$ — ②

Solving ② ÷ ①

$\frac{npq}{np} = \frac{\frac{4}{3}}{4}$

$q = \frac{1}{3}$ ; $p = 1 - q$

$= 1 - \frac{1}{3}$

$p = \frac{2}{3}$

from ①

$np = 4 \Rightarrow n\left(\frac{2}{3}\right) = 4 \Rightarrow n = 6$

$P(X \geq 1) = 1 - P(X < 1)$

$= 1 - P(X=0)$

$= 1 - {}^6C_0 \left(\frac{2}{3}\right)^0 \left(\frac{1}{3}\right)^6$

$= 0.99$

5. In 8 throws of a dice the numbers 5 or 6 is considered as success. Find the mean and standard deviation of the numbers of success.

Sol. Given $n = 8$

$P$ = prob. of 5 or 6 as a success $= \frac{2}{6} = \frac{1}{3}$

$q = 1 - p = 1 - \frac{1}{3} = \frac{2}{3}$

Mean $(\mu) = np$

$= 8\left(\frac{1}{3}\right) = 2.66$

variance $= npq$

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

$= 1.77$

Standard deviation $= \sqrt{npq} = \sqrt{1.77}$

$= 1.330.$

6. A binomial distribution to the following frequency distribution.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x$ | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
| $f$ | 13 | 25 | 52 | 58 | 32 | 16 | 4 |

Sol. $n = 6$ ; $N = \Sigma f_i = 200$

Mean $= \dfrac{\Sigma f_i x_i}{\Sigma f_i} = 2.67$

The mean of binomial distribution $= np$

$2.67 = 6(P)$

$P = 0.44$

$q = 1 - P = 0.56$

We know that

$(P+q)^n = P^n + {}^nC_1 P^{n-1} q + {}^nC_2 P^{n-2} q^2 + \cdots + {}^nC_r P^{n-r} q^r$

To fit a binomial distribution table is

$N(q+P)^n = 200[(0.56 + 0.44)^6]$

$\Rightarrow 200[(0.56)^6 + 6C_1 (0.56)^5 (0.44) + 6C_2 (0.56)^4 (0.44)^2 + 6C_3 (0.56)^3 (0.44)^3 + 6C_4 (0.56)^2 (0.44)^4 + 6C_5 (0.56)^1 (0.44)^5 + 6C_6 (0.44)^6]$

$\Rightarrow 200[0.030 + 0.1453 + 0.285 + 0.299 + 0.1763 + 0.055 + 0.0072]$

$\Rightarrow \cancel{199.36} \Rightarrow [6 + 29.06 + 57 + 59.8 + 35.2 + 11 + 1.44]$

The binomial distribution table is

| X | 0 | 1 | 2 | 3 | 4 | 5 | 6 |
|---|---|---|---|---|---|---|---|
| Frequency | 13 | 25 | 52 | 58 | 32 | 16 | 4 |
| Theoretical values | 6 | 29 | 57 | 60 | 35 | 11 | 2 |

* Poisson Distribution-

It is a rare distribution of rare events i.e., the events whose prob. of is very small but the no. of trials which could lead to the occurence of the events very large

- A random variable 'x' is said to follow a poisson distribution if it assumes only non-negative values and the function is given by

$$P(x,\lambda) = P(X=x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \; \forall x=0$$

where $\lambda>0$ is called the parameter of a distribution.

Ex:- 1) Consider the number of persons born blind per year in a large city.
2) The number of Telephone calls per minute at a switch board.

* Conditions of Poisson distribution-

-

- The occurence are rare
- The number of trials is large
- The probability of success is very small
- $np=\lambda$ is finite

=> Mean of a poisson distribution-

$$\mu = E(x) = \lambda \quad [\because np=\lambda]$$

Variance :- It is defined as mean of the distribution

$$V(x)=\lambda$$

hence, the standard deviation is $\sqrt{\lambda}$

* Recurrence relation for poisson distribution is

$$P(x+1) = \left[\frac{\lambda}{x+1}\right] P(x)$$

or

$$P(x) = \frac{\lambda}{x} P(x-1)$$

Problems

1. A car hire firm has 2 cars which it hires out day by day the number of demand for a car each day is distributed with a mean 1.5. Calculate:

(i) on which there is no demand

(ii) $P(x>2)$

sol. Given mean $= \lambda = 1.5$

W.K.T

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!}$$

(i) $P(X=0) = P(0) = \frac{e^{-1.5}(1.5)^{0}}{0!}$

$= 0.2231$

(ii) $P(x>2) = 1 - P(X \leq 2)$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left[0.2231 + \frac{e^{-1.5}(1.5)^{1}}{1!} + \frac{e^{-1.5}(1.5)^{2}}{2!}\right]$

$= 0.1911$

2. If the prob. that an individual suffers a bad reaction from a certain infection is 0.001. Determine the probability that out of 2000 individuals.

(i) exactly 3

(ii) More than 2 individuals

(iii) None

(iv) More than one individual suffers a bad reaction

sol. Given $P = 0.001$, $n = 2000$

$\lambda = np$

$= (0.001)(2000)$

$\lambda = 2$

(i) $P(X=3) = \frac{e^{-2} 2^3}{3!}$

$= 0.1804$

(ii) $P(X>2) = 1 - P(X \leq 2)$

$= 1 - [P(X=0) + P(X=1) + P(X=2)]$

$= 1 - \left[\frac{e^{-2}2^0}{0!} + \frac{e^{-2}2}{1!} + \frac{e^{-2}2^2}{2!}\right]$

$= 1 - \left[e^{-2} + 2e^{-2} + \frac{4e^{-2}}{2}\right]$

$= 1 - [5e^{-2}]$

$P(X>2) = 0.3233$

(iii) $P(X=0) = \frac{e^{-2}2^0}{0!}$

$= 0.1353$

(iv) $P(X>1) = 1 - P(X \leq 1)$

$= 1 - [P(X=0) + P(X=1)]$

$= 1 - \left[0.1353 + \frac{e^{-2}2^1}{1!}\right]$

$= 1 - [0.1353 + 0.2706]$

$= 0.5941.$

3. In a random variable has a poisson distribution such that $P(1) = P(2)$ find

(i) mean

(ii) $P(4)$

(iii) $P(X \geq 1)$

(iv) $P(1 < X < 4)$

Sol. Given, $P(1) = P(2)$

(i) W.k.T $P(x) = \frac{e^{-\lambda}\lambda^x}{x!}$ i.e., $\frac{e^{-\lambda}\lambda^1}{1!} = \frac{e^{-\lambda}\lambda^2}{2!}$

$\lambda = 2$

(ii) $P(4) = P(X=4)$

$= \frac{e^{-2}2^4}{4!}$

$P(4) = 0.0902$

(iii) $P(X \geq 1) = 1 - P(X < 1)$

$= 1 - P(X=0)$

$= 1 - \left[\frac{e^{-2}2^0}{0!}\right]$

$= 1 - [e^{-2}]$

$P(X \geq 1) = 0.8646$

(iv) $P(1 < X < 4) = P(X=2) + P(X=3)$

$= \frac{e^{-2}2^2}{2!} + \frac{e^{-2}2^3}{3!}$

$= e^{-2}.2 + \frac{e^{-2}.\cancel{8}^4}{\cancel{6}3}$

$P(1 < X < 4) = 0.4511.$

5. Fit the Poisson distribution for the following data

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $f(x)$ | 109 | 65 | 22 | 3 | 1 |

Sol. Let $N = \Sigma f_i$

$= 200$

$$\text{Mean} = \frac{\Sigma f_i x_i}{\Sigma f_i} = 0.61$$

$\lambda = 0.61$

To fit a P.D is $N \cdot P(x)$

$200 \, P(X=x)$ here $x = 0, 1, 2, 3, 4$

when $x = 0$

$$\Rightarrow NP(0) = 200 \frac{e^{-0.61} (0.61)^0}{0!} = 108.6$$

when $x = 1$

$$NP(1) = 200 \frac{e^{-0.61} (0.61)^1}{1!} = 66.28$$

when $x = 2$

$$NP(2) = 200 \frac{e^{-0.61} (0.61)^2}{2!} = 20.21$$

when $x = 3$

$$NP(3) = 200 \frac{e^{-0.61} \cdot (0.61)^3}{3!}$$

$= 4.11$

when $x = 4$

$$NP(4) = 200 \frac{e^{-0.61} (0.61)^4}{4!} = 0.62$$

| $x$ | 0 | 1 | 2 | 3 | 4 |
|---|---|---|---|---|---|
| $f(x)$ | 109 | 65 | 22 | 3 | 1 |
| Expected | 109 | 66 | 20 | 4 | 1 |

Q. A random variable X has the following probability funcn

| $x$ | $-3$ | $-2$ | $-1$ | $0$ | $1$ | $2$ | $3$ |
|---|---|---|---|---|---|---|---|
| $P(x)$ | $k$ | $0.1$ | $k$ | $0.2$ | $2k$ | $0.4$ | $2k$ |

Find i) k

ii). Mean

iii). Variance

Q. A function is defined by $f(x) = \begin{cases} 0, & x < 2 \\ \frac{1}{18}(2x+3); & 2 \le x \le 4 \\ 0, & x > 4 \end{cases}$

S.T $f(x)$ is a density function or not and also find for the interval $2 \le x \le 3$

Q. If 10% of rivets produced by a machine are defective. Find the prob. that out of 5 rivets chosen at random

i) None will be defective

ii). One will be defective

iii). atmost two will be defective

Q. 20% of items produced from a factory are defective. Find the prob. that sample of 5 chosen at random (i) 1

(ii) None

(iii) $P(1 < X < 4)$

Q. It is found that 2% of the tools produced by certain machines are defective. What is the prob. that in a shipment of 400 such tools

i) 3% or more

ii) 2% or less will be defective

Q. If 2 cards are drawn from a pack of 52 cards which are diamonds use Poisson distrib. to find Prob. of getting two diamonds atleast 3 times in 51 consecutive trials of 2 cards drawn each trial.

- The distributions binomial and poisson are discrete distributions whereas normal distribution is continuous distribution.
- The random variable $X$ is said to have normal distribution if its density function is defined $z = \frac{x-\mu}{\sigma}$ where $\mu \rightarrow$ mean, $\sigma \rightarrow$ S.D of $x$

* To find the probability density of normal curve :-

- The probability that normal variate $X$ with a mean ($\mu$), S.D ($\sigma$) lies b/wn two specific values $x_1, x_2$

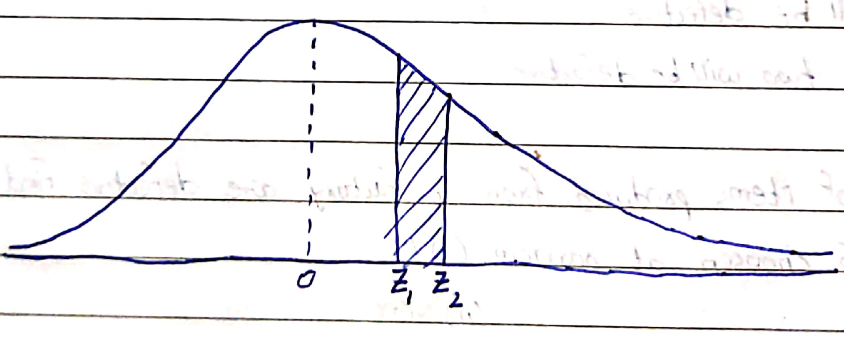
Step-I - The change of scale $z = \frac{x-\mu}{\sigma}$ and find $Z_1, Z_2$ corresponding to the values of $x_1$ and $x_2$

Step 2 -

1. To find $P(x_1 \le X \le x_2) = P(Z_1 \le Z \le Z_2)$

case (i) - If both $Z_1$ and $Z_2$ are positive or both are negative then

$P(x_1 \le X \le x_2) = |A(Z_2) - A(Z_1)|$



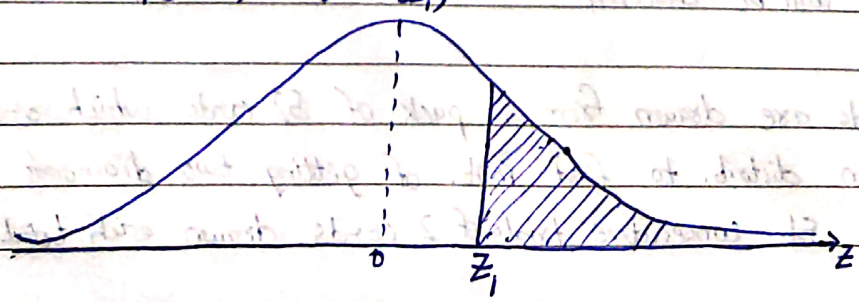
Case (ii) - If both $Z_1 < 0$, $Z_2 > 0$ or viceversa then

$P(x_1 \le X \le x_2) = |A(Z_2) + A(Z_1)|$

2. To find $P(Z > Z_1)$

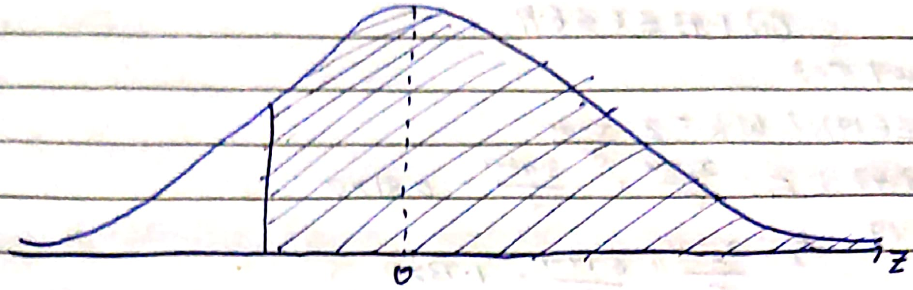
case (i) - If $Z_1 > 0$ then

$P(Z > Z_1) = 0.5 - A(Z_1)$



case (ii) - if $Z_1 < 0$ then probability

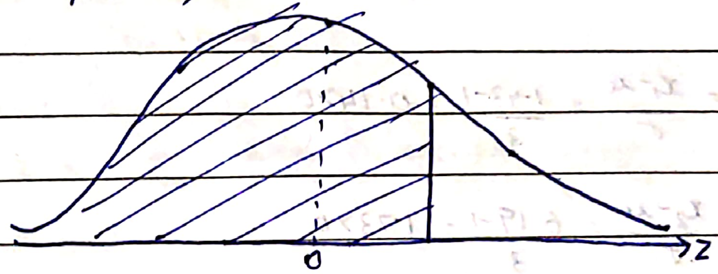
$P(Z > Z_1) = 0.5 + A(Z_1)$



3. To find $P(Z < Z_1)$

case (i) - if $Z_1 > 0$ then

$P(Z < Z_1) = 0.5 + A(Z_1)$



case (ii) If $Z_1 < 0$ then

$P(Z < Z_1) = 0.5 - A(Z_1)$

Q: for a normally distributed variate with a mean = 1 and S.D = 3 find the probabilities that (i) $3.43 \le x \le 6.19$

(ii) $1.43 \le x \le 6.19$

Given that $\mu = 1$ $\sigma = 3$

(i) $P(3.43 \le x \le 6.19) = ?$ W.K.T $Z = \frac{x-\mu}{\sigma}$

here $x_1 = 3.43$ ; $Z_1 = \frac{x_1-\mu}{\sigma} = \frac{3.43-1}{3} = 0.81 > 0$

$x_2 = 6.19$ ; $Z_2 = \frac{x_2-\mu}{\sigma} = \frac{6.19-1}{3} = 1.73 > 0$

$P(3.43 \le x \le 6.19) = |A(Z_2) - A(Z_1)|$

$= |A(1.73) - A(0.81)|$

$= |0.4582 - 0.2910|$

$= 0.1672$

(ii) $P(1.43 \le x \le 6.19) = ?$

here $x_1 = 1.43$, $Z_1 = \frac{x_1-\mu}{\sigma} = \frac{1.43-1}{3} = 0.14 > 0$

$x_2 = 6.19$ ; $Z_2 = \frac{x_2-\mu}{\sigma} = \frac{6.19-1}{3} = 1.73 > 0$

$P(1.43 \le x \le 6.19) = |A(Z_2) - A(Z_1)|$

$= |A(1.73) - A(0.14)|$

$= |0.4582 - 0.0948|$

$= 0.4025$

classmate
Date
Page

Q. If $X$ is a normal variant with $\mu = 30$, $\overset{(S.D)}{\sigma} = 5$ find

(i) $P(26 \le x \le 40)$

(ii) $P(x \ge 45)$

(i) W.k.T $Z = \frac{x-k}{\sigma}$

here $x_1 = 26$ ; $Z_1 = \frac{x_1 - \mu}{\sigma} = -0.8 < 0$

$x_2 = 40$ ; $Z_2 = \frac{x_2 - \mu}{\sigma} = 2 > 0$

$P(26 \le x \le 40) = |A(Z_2) + A(Z_1)|$

$= 0.7653$

(ii) $P(x \ge 45) = ?$

here $x = 45$ $Z = \frac{x-\mu}{\sigma} = \frac{45-30}{5} = 3 > 0$

$P(x \ge 45) = 0.5 - A(Z)$

$= 0.5 - A(3)$

$= 0.5 - 0.4987$

$= 0.0013$

Q. find the mean and standard deviation of a normal distribution in which 7% of items are under 35 and 89% are under 63

sol. $\mu = ?$, $\sigma = ?$

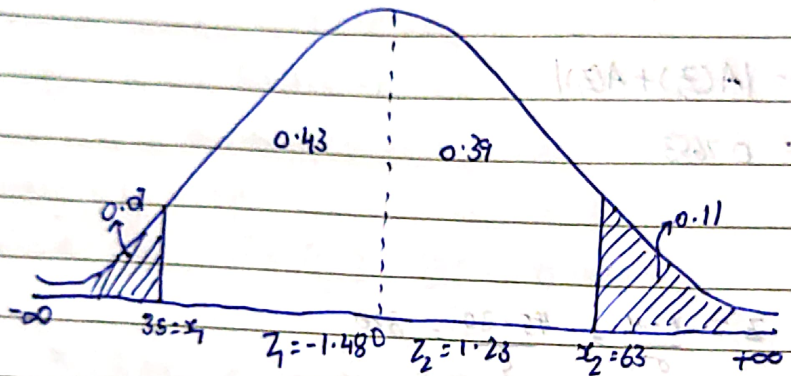
$P(X < 35) = 0.07$

$P(X < 63) = 0.89$

$\Rightarrow P(X > 63) = 1 - P(X < 63)$

$= 1 - 0.89$

$= 0.11$




when $x_1 = 35$, $z_1 = -1.48$ [from the table]

$x_2 = 63$, $z_2 = 1.23$ [from the table]

$\Rightarrow$ when $x_1 = 35$

$$z_1 = \frac{x_1 - \mu}{\sigma}$$

$$-1.48 = \frac{35 - \mu}{\sigma}$$

$$\mu + 1.48\sigma = 35 \text{ — } ①$$

when $x_2 = 63$

$$z_2 = \frac{x_2 - \mu}{\sigma}$$

$$1.23 = \frac{63 - \mu}{\sigma}$$

$$\mu + 1.23\sigma = 63 \text{ — } ②$$

By solving ① & ②  $\mu = 50$

$\sigma = 10$

Q. ~~In~~ With the mass of 300 students are normally distributed 68 kgs and S.D 3 kg how many students have the mass (i) greater than 72 kg
(ii) $\leq 64$ kgs
(iii) b/w 65 and 71 kg

sol. $\mu = 68, \sigma = 3$

(i) $P(X>72)$

$x_1 = 72; \; Z_1 = \frac{x_1 - \mu}{\sigma}$

$= \frac{72-68}{3}$

$= 1.33 > 0$

$P(X>72) = 0.5 - A(Z_1)$

$= 0.5 - A(1.33)$

$= 0.0918$

(ii) $P(X \leq 64)$

$x_1 = 64, \; Z_1 = -1.33 < 0$

$P(X \leq 64) = 0.5 - A(Z_1)$

$= 0.0918$ $[\because A(-Z) = A(Z)]$

for 300 student $= 300 \times 0.0918$

$= 27.5$

$\cong 28$ $(Z_1 < 0; Z_2 > 0)$

(iii) $P(65 < X < 71) \Rightarrow A(Z_1) + A(Z_2) \Rightarrow A(1) + A(1)$

$x_1 = 65; \; Z_1 = \frac{x - \mu}{\sigma}$

$= \frac{65-68}{3}$

$Z_1 \Rightarrow -1$

$x_2 = 71; \; Z_2 = \frac{x - \mu}{\sigma}$

$= \frac{71-68}{3}$

$= \frac{3}{3}$

$Z_2 = 1$

$\Rightarrow 0.3413 + 0.3413$

$\Rightarrow 0.6826$

for 300 students $= 300 \times 0.6826$

$= 204.78$

$= 205$

* Moments - moments are used to describe the various characteristics of a frequency distribution. Kurtosis enables us to study the shape of the top of
* the curve
* skewness - is a measure refers to the extent of symmetry as symmetry distribution.

→ Moments is defined as arithmetic mean of various power of deviation of items from their means (assume & natural) will give the required power of moments of the distribution.

- If the deviations of the items are taken from the arithmatic mean is known as central moments
- The moments are denoted by '$\mu$'

* central moments (or) det moments about actual mean:

1. central moments for individual series - let $\bar{X}$ be the mean of the individual series

let $x$ be the deviation of $X$ from its mean

$$X = d = x - \bar{x}$$

let $N$ be the total number of items or observation

$$\mu_1 = \frac{\Sigma d}{N}$$

$$\mu_2 = \frac{\Sigma d^2}{N}$$

$$\mu_3 = \frac{\Sigma d^3}{N} \quad ||^{ly} \quad \mu_r = \frac{\Sigma d^r}{N}$$

2. Central moments for frequency distribution:-

let $n$ be the observations of $x_1, x_2, -----, x_n$ occuring with the frequencies $f_1, f_2, ------ f_n$ with A.M of a frequency is defined as

$$\bar{X} = \frac{\Sigma f_i x_i}{N} \quad \text{where } N = \sum_{i=1} f_i$$

- The deviation is $d$ = r taken as $d = x - \bar{x}$ and the moments are defined as

$$\mu_1 = \frac{\Sigma fd}{N} \qquad \mu_3 = \frac{\Sigma fd^3}{N}$$

$$\mu_2 = \frac{\Sigma fd^2}{N} \qquad ||^{ly} \quad \mu_r = \frac{\Sigma fd^r}{N}$$

* Properties of

1. The first moment about mean is always zero i.e, $\mu_1 = 0$
2. The second moment about mean measures varianle i.e., $\mu_2 = \sigma^2$ (or) $\sigma = \pm\sqrt{\mu_2}$
3. The third moment about mean measures skewness

   case 1 : if $\mu_3 > 0$ the distribution is positively skewed

   case 2 : if $\mu_3 < 0$ the distribution negatively skewed

   case 3 : if $\mu_3 = 0$ the distribution symmetrical
4. The fourth moment about mean measures the kurtosis

* Raw moments about the origin

- Moments about the arbitrary origin are called raw moments are denoted by $(\mu_r')$
- The moments are defined as $\mu_1' = \frac{\Sigma fd}{N}$ $\quad \mu_3' = \frac{\Sigma fd^3}{N}$

$$\mu_2' = \frac{\Sigma fd^2}{N} \qquad \text{IIly } \mu_r' = \frac{\Sigma fd^r}{N}$$

The moments about the origin are defined as

$$\mu_1 = 0$$

$$\mu_2 = \mu_2' - (\mu_1')^2$$

$$\mu_3 = \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3$$

$$\mu_4 = \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4$$

efined as

$^2 - 3(\mu_1')^4$

Q. find the first four moments for the set of numbers 2, 4, 6, 8

sol. Given $x = 2, 4, 6, 8$ $\quad N = 4$

$\bar{x} = \frac{\Sigma x}{N} = \frac{20}{4} = 5$

| $x$ | $d = x - \bar{x}$ | $d^2$ | $d^3$ | $d^4$ |
|---|---|---|---|---|
| 2 | -3 | 9 | -27 | 81 |
| 4 | -1 | 1 | -1 | 1 |
| 6 | 1 | 1 | 1 | 1 |
| 8 | 3 | 9 | 27 | 81 |
| | $\Sigma d = 0$ | $\Sigma d^2 = 20$ | $\Sigma d^3 = 0$ | $\Sigma d^4 = 164$ |

$\mu_1 = \frac{\Sigma d}{N} = 0$

$\mu_2 = \frac{\Sigma d^2}{N} = 5$

$\mu_3 = \frac{\Sigma d^3}{N} = 0$

$\mu_4 = \frac{\Sigma d^4}{N} = 41.$

Q. Calculate the first 4 moments for the following distribution and also calculate $\beta_1$ and $\beta_2$

| $x$ | $f$ | $fx$ | $d = x - \bar{x}$ | $fd^2$ | $fd^3$ | $fd^4$ | $fd$ |
|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | -4 | 16 | | | |
| 1 | 8 | 8 | -3 | 72 | | | |
| 2 | 28 | 56 | -2 | 112 | | | |
| 3 | 56 | 168 | -1 | 56 | | | |
| 4 | 70 | 280 | 0 | 0 | | | |
| 5 | 56 | 280 | 1 | 56 | | | |
| 6 | 28 | 168 | 2 | 112 | | | |
| 7 | 8 | 56 | 3 | 72 | | | |
| 8 | 1 | 8 | 4 | 16 | | | |
| | | $\Sigma fx = 1024$ | $\Sigma d = 0$ | | | | |

$\because$ $N = \Sigma f = 256$

$$\bar{x} = \frac{\Sigma f_i x_i}{N} = 3.9 \cong 4$$

$\mu_1 = 0$

$$\mu_2 = \frac{\Sigma f d^2}{N}$$

$$\mu_3 = \frac{\Sigma f d^3}{N}$$

$$\mu_4 = \frac{\Sigma f d^4}{N}$$

$$\beta_1 = \frac{\mu_3^2}{\mu_2^3}$$

$$\beta_2 = \frac{\mu_4}{\mu_2^2}$$

# Correlation, Regretion and sampling Distribution

**Correlation (Introduction):** For study the characteristic of only one variable like marks, weights, heights, prices, ages, sales etc.

* This type of analysis is called univariate analysis.
* If there exists some relationship between two variables then the statistical analysis of such data is called bivariate analysis.

* Correlation refers to the relationship of two or more variable. There exists a relationship between the height of th a father and a son. The study of the relation is called Correlation. It measures the closeness of the relationship between the variable.

**Definition:** Correlation is a statistical analysis which measures and analyze the degree or extent to which two variables fluctuate with reference to each other.

* The Correlation express the relationship or interdependent of two sets of variable upon each other. one variable may be called the subject (independent) and the other relative (dependent).
* A distribution involving two variables is known as bivariate distribution. If these two variable vary such that change in one variable effects the change in other variable and the variables are said to be Correlated

26-10-2023

Ex: There exist some relation between height and weight of person.

2) price of commitee and its demand.

Note:

(i) The degree of relationship between the variable under consideration is measured through the correlation analysis.

(ii) The measures of correlation is called as Coefficient of Correlation or Correlation index.

Types of Correlation:

* Correlation are classified into many types.

(i) positive and negative

(ii) simple and multiple

(iii) Total and partial

(i) positive and negative : If two variables tends to move together in the same direction that is an increase in the value of one variable is accompanied by an increase in another variable or vice versa is called a "positive Correlation".

* If two variables tends to move together in opposite direction so that an increase or decrease in the value

26-10-2023

of one variable by ~~a degrees~~ is accompanied by a degrees of another variable is called "negative correlation"

(ii) simple and multiple : When we study only two variables i.e. so relationship is called "simple".

* when we study more than two variables simultaneously that relationship is called "multiple".

(iii) Total and partial : The study of two variables excluding some other variables is known as partial.

* In total Correlation all the facts are considered taken into account.

linear and Non-linear Correlation: If the ration of change between two variables is uniform then there will be a linear correlation between them.

* The amount of change in one variable does not bear a constant ration of the amount of change in other variable is known as Non-linear Corelation.

Methods of studying Correlation : There are two methods

1) Graphical method

2) Mathematical method.

Mathematical methods : There are two types

(i) Karl person coefficient of Correlation

30-10-2023

(ii) Spearman's rank Correlation

(i) **Coefficient of Correlation**: This method is used for measuring the magnitude of linear relationship between two variables. It is denoted by '$r$' and it is defined as

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left\{\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right\}\left\{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right\}}}$$

where
$$\begin{cases} X = x - \bar{x}, & \bar{x} = \frac{\Sigma x}{N} \\ Y = y - \bar{y}, & \bar{y} = \frac{\Sigma y}{N} \end{cases}$$
deviation

* $r$ lies between $\pm 1$

**Problems:**

1, Calculate the Coefficient of Correlation for the following data

~~x = 12, 9, 8~~

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $X$ | 12 | 9 | 8 | 10 | 11 | 13 | 7 |
| $Y$ | 14 | 8 | 6 | 9 | 11 | 12 | 3 |

30-10-2023

Sol:

| $X$ | $Y$ | $X^2$ | $Y^2$ | $XY$ |
|---|---|---|---|---|
| 12 | 14 | 144 | 196 | 168 |
| 9 | 8 | 81 | 64 | 72 |
| 8 | 6 | 64 | 36 | 48 |
| 10 | 9 | 100 | 81 | 90 |
| 11 | 11 | 121 | 121 | 121 |
| 13 | 12 | 169 | 144 | 156 |
| 7 | 3 | 49 | 9 | 21 |
| $\Sigma X = 70$ | $\Sigma Y = 63$ | $\Sigma X^2 = 728$ | $\Sigma Y^2 = 651$ | $\Sigma XY = 676$ |

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left\{\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right\}\left\{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right\}}}$$

$$r = \frac{676 - \frac{(70 \times 63)}{7}}{\sqrt{\left\{728 - \frac{(\cancel{728}\ 70)^2}{7}\right\}\left\{651 - \frac{(\cancel{651}\ 63)^2}{7}\right\}}}$$

$$r = 0.9485$$

3-10-2023

2, Find the Coefficient of correlation between heights and weights given below

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| Height (in inches) x | 57 | 59 | 62 | 63 | 64 | 65 | 55 | 58 |
| weight y | 113 | 117 | 126 | 126 | 130 | 129 | 111 | 116 |

Sol:

| x | $x = X - \bar{X}$ | y | $Y = y - \bar{y}$ | $x^2$ | $Y^2$ | $xY$ |
|---|---|---|---|---|---|---|
| 57 | -3 | 113 | ~~113~~ -7 | 9 | 49 | 21 |
| 59 | -1 | 117 | ~~117~~ -3 | 1 | 9 | 3 |
| 62 | 2 | 126 | ~~126~~ 6 | 4 | 36 | 12 |
| 63 | 3 | 126 | ~~126~~ 6 | 9 | 36 | 18 |
| 64 | 4 | 130 | ~~130~~ 10 | 16 | 100 | 40 |
| 65 | 5 | 129 | ~~129~~ 9 | 25 | 81 | 45 |
| 55 | -5 | 111 | ~~111~~ -9 | 25 | 81 | 45 |
| 58 | -2 | 116 | ~~116~~ -4 | 4 | 16 | 8 |
| 57 | -3 | 112 | 112 -8 | 9 | 64 | 24 |
| $\Sigma X = 540$ | $\Sigma x = 0$ | $\Sigma y = 1080$ | $\Sigma Y = 0$ | $\Sigma x^2 = 102$ | $\Sigma Y^2 = 472$ | $\Sigma xY = 216$ |

$$\bar{X} = \frac{\Sigma x}{N} = \frac{540}{9} = 60$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{1080}{9} = 120$$

$$r = \frac{\Sigma xY - \frac{\Sigma x \Sigma Y}{N}}{\sqrt{\left\{\Sigma x^2 - \frac{(\Sigma x)^2}{N}\right\}\left\{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right\}}}$$

30-10-2023

$$r = \frac{216 - 0}{\sqrt{\{102 - 0\}\{472 - 0\}}}$$

$$r = 0.98$$

3, Calculate Coefficient of Correlation for the following data

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 28 | 41 | 40 | 38 | 35 | 33 | 40 | 32 | 36 | 33 |
| y | 23 | 34 | 33 | 34 | 30 | 26 | 28 | 31 | 36 | 38 |

Sol:

| X | $x = X - \bar{x}$ | y | $Y = y - \bar{y}$ | $x^2$ | $Y^2$ | $xY$ |
|---|---|---|---|---|---|---|
| 28 | -8 | 23 | -8 | 64 | 64 | 64 |
| 41 | 5 | 34 | 3 | 25 | 9 | 15 |
| 40 | 4 | 33 | 2 | 16 | 4 | 8 |
| 38 | 2 | 34 | 3 | 4 | 9 | 6 |
| 35 | -1 | 30 | -1 | 1 | 1 | 1 |
| 33 | -3 | 26 | -5 | 9 | 25 | 15 |
| 40 | +4 | 28 | -3 | 16 | 9 | -12 |
| 32 | -4 | 31 | 0 | 16 | 0 | 0 |
| 36 | 0 | 36 | 5 | 0 | 25 | 0 |
| 33 | -3 | 38 | 7 | 9 | 49 | -21 |
| $\Sigma X = 356$ | $\Sigma x = -4$ | $\Sigma y = 313$ | $\Sigma Y = 3$ | $\Sigma x^2 = 160$ | $\Sigma Y^2 = 195$ | $\Sigma xY = 76$ |

$$\bar{X} = \frac{\Sigma x}{N} = \frac{356}{10} = 35.6 \cong 36$$

$$\bar{Y} = \frac{\Sigma y}{N} = \frac{313}{10} = 31.3 \cong 31$$

31-10-2019

$$r = \frac{\Sigma XY - \frac{\Sigma X \Sigma Y}{N}}{\sqrt{\left\{\Sigma X^2 - \frac{(\Sigma X)^2}{N}\right\}\left\{\Sigma Y^2 - \frac{(\Sigma Y)^2}{N}\right\}}}$$

$$r = \frac{76 - \frac{(-4)(3)}{10}}{\sqrt{\left\{160 - \frac{(\cancel{3}4)^2}{10}\right\}\left\{195 - \frac{(3)^2}{10}\right\}}}$$

$$r = 0.44$$

**Rank correlation (or) non-repeated ranks**: This method is based on rank and is used in dealing with qualitative characteristics such as intelligence, beauty, morality etc.

* It is based on the ranks given to the observation.
* Rank correlation is applicable only for the individual observation.
* It is denoted as $\rho$ and it is defined as

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

~~where~~ where $D$ = sum of squares of difference between two ranks

$N$ = number of paired observation

31-10-2019

**problem**:

1, The following are the ranks obtained by 10 students in 2 subjects

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| statistical values (s) | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 |
| mathematical (m) values | 2 | 4 | 1 | 5 | 3 | 9 | 7 | 10 | 6 | 8 |

sol:

| $x$ | $y$ | $D = x-y$ | $D^2$ |
|---|---|---|---|
| 1 | 2 | -1 | 1 |
| 2 | 4 | -2 | 4 |
| 3 | 1 | 2 | 4 |
| 4 | 5 | -1 | 1 |
| 5 | 3 | 2 | 4 |
| 6 | 9 | -3 | 9 |
| 7 | 7 | 0 | 0 |
| 8 | 10 | -2 | 4 |
| 9 | 6 | 3 | 9 |
| 10 | 8 | 2 | 4 |
| | | | $\Sigma D^2 = 40$ |

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$\rho = 1 - \frac{6 \times 40}{10(10^2-1)}$$

$$\rho = 0.75$$

31-10-2013

2, A random sample of 5 college students are selected and their grades in mathematics and statistic values are found to be

| | | | | | |
|---|---|---|---|---|---|
| M | 85 | 60 | 73 | 40 | 90 |
| S | 93 | 75 | 65 | 50 | 80 |

Sol:-

| x | X | y | Y | D = X − Y | $D^2$ |
|---|---|---|---|---|---|
| 85 | 2 | 93 | 1 | 1 | 1 |
| 60 | 4 | 75 | 3 | 1 | 1 |
| 73 | 3 | 65 | 4 | −1 | 1 |
| 40 | 5 | 50 | 5 | 0 | 0 |
| 90 | 1 | 80 | 2 | −1 | 1 |
| | | | | | $\Sigma D^2 = 4$ |

$$\rho = 1 - \frac{6 \Sigma D^2}{N(N^2-1)}$$

$$\rho = 1 - \frac{6(4)}{5(5^2-1)}$$

$$\rho = 0.8$$

31-10-2013

**Equal or Repeated ranks**: If any two or more persons are bracketed equally in any classification or if there is more than one item with the same value in the series then we will apply repeated rank correlation.

* The common rank is the average of the ranks which these items would have assumed if they work different from each other and the next item will get the rank next to ranks already assumed and it is defined as

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \frac{1}{12}(m^3-m) + \cdots\right]}{N(N^2-1)}$$

where m = number of items repeated

**Problem:**

1, From the following data calculate the rank correlation coefficient after making adjustment for tied ranks.

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| x | 48 | 33 | 40 | 9 | 16 | 16 | 65 | 24 | 16 | 57 |
| y | 13 | 13 | 24 | 6 | 15 | 4 | 20 | 9 | 6 | 19 |

01-12-23

Sol:

| x | X | y | Y | D = X − Y | $D^2$ |
|---|---|---|---|---|---|
| 48 | 8 | 13 | 5.5 | 2.5 | 6.25 |
| 33 | 6 | 13 | 5.5 | 0.5 | 0.25 |
| 40 | 7 | 24 | 10 | −3 | 9 |
| 9 | 1 | 6 | 2.5 | −1.5 | 2.25 |
| 16 | 3 | 15 | 7 | −4 | 16 |
| 16 | 3 | 4 | 1 | 2 | 4 |
| 65 | 10 | 20 | 9 | 1 | 1 |
| 24 | 5 | 9 | 4 | 1 | 1 |
| 16 | 3 | 6 | 2.5 | 0.5 | 0.25 |
| 57 | 9 | 19 | 8 | 1 | 1 |
| | | | | | $\Sigma D^2 = 41$ |

Here $m = 3, 2, 2$

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

$$\rho = 1 - \frac{6\left[41 + \frac{1}{12}(3^3 - 3) + \frac{1}{12}(2^3 - 2) + \frac{1}{12}(2^3 - 2)\right]}{10(10^2 - 1)}$$

$$\rho = 0.73$$

01-12-23

2) A sample of 12 fathers and their elder sons gave the following data

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| x | 65 | 63 | 67 | 64 | 68 | 62 | 70 | 66 | 68 | 67 | 69 | 71 |
| y | 68 | 66 | 68 | 65 | 69 | 66 | 68 | 65 | 71 | 67 | 68 | 70 |

Sol:

| x | X | y | Y | D = X − Y | $D^2$ |
|---|---|---|---|---|---|
| 65 | 4 | 68 | 7.5 | −3.5 | 12.25 |
| 63 | 2 | 66 | 3.5 | −1.5 | 2.25 |
| 67 | 7 | 68 | 7.5 | −0.5 | 0.25 |
| 64 | 3 | 65 | 1.5 | 1.5 | 2.25 |
| 68 | ~~4~~ 7.5 | 69 | 10 | −2.5 | 6.25 |
| 62 | 1 | 66 | 3.5 | −2.5 | 6.25 |
| 70 | 10 | 68 | 7.5 | 2.5 | 6.25 |
| 66 | 5 | 65 | 1.5 | 3.5 | 12.25 |
| 68 | ~~4~~ 7.5 | 71 | 12 | −4.5 | 20.25 |
| 67 | 6 | 67 | 5 | 1 | 1 |
| 69 | 9 | 68 | 7.5 | 1.5 | 2.25 |
| 71 | 11 | 70 | 11 | 0 | 0 |
| | | | | | $\Sigma D^2 = 71.5$ |

Here $m =$

$$\rho = 1 - \frac{6\left[\Sigma D^2 + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m) + \frac{1}{12}(m^3 - m)\right]}{N(N^2 - 1)}$$

**Regression** (2M) The study of Correlation measures the direction and the strength of relationship between two variables.

* In Correlation we can estimate the value of other variable when the value of one variable is given.
* But in regression we can estimate the value of one variable with the value of other variable which is known.
* The statistical method which helps us to estimate the unknown value of one variable from the known value of the related variable is called Regression.

**Methods of studying Regression:**

(i) 1) Graphical method

2) Algebraic method

**Regression line**: A regression line is a straight line fitted to the data by the method of least squares.

* There are always two regression lines constructed for the relationship between two variables x and y.

**Regression equation for a straight line equation of y on x:**

$Y = a + bX$

The normal equation

$\Sigma Y = Na + b\Sigma X$

$\Sigma XY = a\Sigma X + b\Sigma X^2$

**Regression equation for a straight line equation of x on y:**

$X = a + bY$

The normal equation

$\Sigma X = Na + b\Sigma Y$

$\Sigma XY = a\Sigma X + b\Sigma Y^2$

**Deviation taken from arithmetic mean:**

1) Regression equation of x on y

* The equation is $X - \bar{X} = r\frac{\sigma x}{\sigma y}(Y - \bar{Y})$

where $r\frac{\sigma x}{\sigma y} = b_{xy} = \frac{\Sigma XY}{\Sigma Y^2}$

$X = x - \bar{x}$ $\quad \bar{x} = \frac{\Sigma x}{N}, \quad \bar{y} = \frac{\Sigma y}{N}$

$Y = y - \bar{y}$

2) Regression equation of Y on X

$Y - \bar{Y} = r\frac{\sigma y}{\sigma x}(X - \bar{X})$

01-11-2013

where $r\frac{\sigma_y}{\sigma_x} = b_{yx} = \frac{\Sigma xy}{\Sigma x^2}$

$x = x - \bar{x}$ $\qquad \bar{x} = \frac{\Sigma x}{N}, \ \bar{y} = \frac{\Sigma y}{N}$

$y = y - \bar{y}$

Problem :

1, determine the equation of a straight line which fits best data

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| x | 10 | 12 | 13 | 16 | 17 | 20 | 25 |
| y | 10 | 22 | 24 | 27 | 29 | 33 | 37 |

Sol:

| x | y | $x^2$ | $y^2$ | xy |
|---|---|---|---|---|
| 10 | 10 | 100 | 100 | 100 |
| 12 | 22 | 144 | 484 | 264 |
| 13 | 24 | 169 | 576 | 312 |
| 16 | 27 | 256 | 729 | ~~384~~ 432 |
| 17 | 29 | 289 | 841 | 493 |
| 20 | 33 | 400 | 1089 | 660 |
| 25 | 37 | 625 | 1369 | 925 |
| $\Sigma x = 113$ | $\Sigma y = 182$ | $\Sigma x^2 = 1983$ | $\Sigma y^2 = 5188$ | $\Sigma xy = 3186$ |

The straight line equation for y on x is

$y = a + bx$ —— (1)

The normal equations are

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

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

$182 = 7a + 113b$

$3186 = 113a + 1983b$

$a = 0.79, \quad b = 1.56$

From (1)

$$\boxed{y = 0.79 + 1.56x}$$

2, A panel of 2 judges P and Q graded 7 dramatic performance by independently awarding marks as follows

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| Marks of P | 46 | 42 | 44 | 40 | 43 | 41 | 45 |
| marks of Q | 40 | 38 | 36 | 35 | 39 | 37 | 41 |

The 8th performance which judge Q would not attend was awarded 37 marks by judge P. If judge Q has also been present how many marks would be expected who have been awarded by him for the 8th performance.

04-11-2023

Sol:

| $x$ | $X = x - \bar{x}$ | $y$ | $Y = y - \bar{Y}$ | $XY$ | $X^2$ | $Y^2$ |
|---|---|---|---|---|---|---|
| 46 | 3 | 40 | 2 | 6 | 9 | 4 |
| 42 | -1 | 38 | 0 | 0 | 1 | 0 |
| 44 | 1 | 36 | -2 | -2 | 1 | 4 |
| 40 | -3 | 35 | -3 | 9 | 9 | 9 |
| 43 | 0 | 39 | 1 | 0 | 0 | 1 |
| 41 | -2 | 37 | -1 | 2 | 4 | 1 |
| 45 | 2 | 41 | 3 | 6 | 4 | 9 |
| $\Sigma x = 301$ | $\Sigma X = 0$ | $\Sigma y = 266$ | $\Sigma Y = 0$ | $\Sigma XY = 21$ | $\Sigma X^2 = 28$ | $\Sigma Y^2 = 28$ |

$\bar{x} = \frac{\Sigma x}{N}$ $\qquad$ $\bar{Y} = \frac{\Sigma y}{N}$

$\bar{x} = \frac{301}{7} = 43$ $\qquad$ $\bar{Y} = \frac{266}{7} = 38$

Regression equation of $Y$ on $x$.

$$Y - \bar{Y} = r \frac{\sigma y}{\sigma x} (x - \bar{x}) \text{ --- } (1)$$

$$r \frac{\sigma y}{\sigma x} = \frac{\Sigma XY}{\Sigma X^2} = \frac{21}{28} = 0.75$$

From (1)

$$Y - 38 = 0.75 (x - 43)$$

$$Y = 0.75x - 32.25 + 38$$

$$Y = 0.75x + 5.75$$

When $x = 37 \Rightarrow Y = 0.75(37) + 5.75$

$Y = 33.5$

04-11-2013

Deviation taken from assumed frequency if the actual mean is a fraction this method is used.

* Regression equation of $Y$ on $X$.

$$Y - \bar{Y} = \frac{r \sigma y}{\sigma x} (X - \bar{X})$$

where $X - \bar{Y} = r \frac{\sigma y}{\sigma x} (x - \bar{x})$

**Problems:**

1) Price index of cotton and wool are given below for the 12 months of a year obtained the regression equation of lines between the two index

| | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| X | 78 | 77 | 85 | 88 | 87 | 82 | 81 | 77 | 76 | 83 | 97 | 93 |
| Y | 84 | 82 | 82 | 85 | 89 | 90 | 88 | 92 | 83 | 89 | 98 | 99 |

Sol:

| $x$ | $dx = x - A$ ($A$ = 84) | $y$ | $dy = y - A$ ($A$ = 88) | $dx^2$ | $dy^2$ | $dx\,dy$ |
|---|---|---|---|---|---|---|
| 78 | −6 | 84 | −4 | 36 | 16 | 24 |
| 77 | −7 | 82 | −6 | 49 | 36 | 42 |
| 85 | +1 | 82 | −6 | 1 | 36 | −6 |
| 88 | 4 | 85 | −3 | 16 | 9 | −12 |
| 87 | 3 | 89 | 1 | 9 | 1 | 6 |
| 82 | −2 | 90 | 2 | 4 | 4 | |
| 81 | −3 | 88 | 0 | 9 | | 6 |
| 77 | −7 | 92 | 4 | 49 | 16 | |
| 76 | −8 | 83 | −5 | 64 | 25 | |
| 83 | −1 | 89 | −1 | 1 | 1 | |
| 97 | 13 | 98 | 10 | 169 | 100 | |
| 93 | 9 | 99 | 11 | 81 | 121 | |
| $\Sigma x$ = | | | | | | |

Regression equation of x on Y.

$$x - \bar{x} = \frac{r \sigma x}{\sigma y} (y - \bar{y}) \quad \text{--- (1)}$$

where $$\frac{r \sigma x}{\sigma y} = \frac{\Sigma dx\, dy - \frac{\Sigma dx\, \Sigma dy}{N}}{\Sigma dy^2 - \frac{\Sigma (dy)^2}{N}}$$

$$\frac{r \sigma x}{\sigma y} =$$

## Sampling distribution:

**Population**: It is the aggregate or totality or statistical data forming a subject of investigation.

Ex: Heights of India or Nationalized banks in India.

Note: The number of observation in the population is defined to be the size of the population.

* It is denoted by 'N'.
* It may be finite or infinite.

**Sampling**: Most of the time study of entire population may not be possible to carry out and hence a part alone is selected from the given population.

* A portion of the population which is examine with a view to determine a population characteristic is called sample.
* A sample is a subset of population and number of objects in the sample is called size of the sample and denoted by n.

Ex: Cars produced in India is the population and 'NANO' cars comes under sample.

**Classification of samples:**

* The samples are classified into two ways.

1) large sample 2) small sample

**Large sample**: If the size of the sample i.e if $n >= 30$ the sample is said to be large sample.

**Small sample**: If the size of the sample i.e if $n <= 30$ the sample is said to be small sample.

Note:

* The number of samples with replacement (infinite is $N^n$)
* The number of samples without replacement (finite) is $^NC_n$

**Simple mean**: If $x_1, x_2, \ldots x_n$ represent a random sample of size $n$ then the simple mean is defined as $\bar{x} = \frac{\Sigma x}{n}$

**Simple variance**: If $x_1, x_2 \ldots x_n$ represents a random sample of size $n$ then the simple variance is defined as

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

**Standard error**: The sampling distribution of a statistic is known as it's standard error and it is denoted by S.E.

$$S.E = \frac{\sigma}{\sqrt{n}}$$

**Central limit theorem**: If $\bar{x}$ is the mean of the sample and 'n' is the size of the sample drawn from a population mean with a mean '$\mu$' under standard deviation $\sigma$, then the standardised simple mean is

01-07-2019

defined as $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

**Correction Factor (C.F)**

$$CF = \frac{N-n}{N-1}$$

**Problem**

1. What is the value of Correction factor if $n=5$, $N=200$

Sol:

$$CF = \frac{N-n}{N-1}$$

$$= \frac{200-5}{200-1} = \frac{195}{199}$$

$$\boxed{CF = 0.979}$$

2. How many different samples of size 2 can be chosen from a finite population of size 25.

Sol: $N = 25$, $n = 2$.

Number of samples $N_{C_n} = {}^{25}C_2$

$= 300$ ways

3. A population consists of 5 numbers, 2, 3, 6, 8 and 11. Consider all possible samples of size 2 which can be drawn with replacement from this population. Find
(i) mean of the population
(ii) standard deviation of the population
(iii) mean of the sampling distribution of means
(iv) S.D of the sampling distribution

Sol: Given population are 2, 3, 6, 8, 11, $N=5$, $n=2$

(i) $\mu = \frac{2+3+6+8+11}{5}$

$= 6$

(ii) $\sigma^2 = \sum_{i=1}^{5} \frac{(x-\bar{x})^2}{N}$

$= \frac{(2-6)^2+(3-6)^2+(8-6)^2+(6-6)^2+(1-6)^2}{5}$

$= 10.8$

(ii) $\sigma = \sqrt{10.8} = 3.28$

(iii) The number of samples with replacement is

$N^n = 5^2 = 25$ ways

(2, 2) (2, 3) (2, 6) (2, 8) (2, 11) (3, 2) (3, 3) (3, 6) (3, 8)
(3, 11) (6, 2) (6, 3) (6, 6) (6, 8) (6, 11) (8, 2) (8, 3) (8, 6)
(8, 8) (8, 11) (11, 2) (11, 3) (11, 6) (11, 8) (11, 11)

The mean of sample are

| | | | | | |
|---|---|---|---|---|---|
| | | 4 | 5 | 6.5 | 2.5 |
| 2 | 2.5 | 4 | 5 | 6.5 | 2.5 |
| 3 | 4.5 | 5.5 | 7 | 4 | 4.5 |
| 6 | 7 | 8.5 | 5 | 5.5 | 7 |
| 8 | 9.5 | 6.5 | 7 | 8.5 | 9.5 | 11 |

The mean of the standard deviation of mean is

$$\bar{x} = \frac{2 + 2.5 + 4 + \cdots\cdots + 11}{25}$$

$$= 6$$

(iv) $$\sigma^2 = \frac{\sum_{i=1}^{5} (x - \bar{x})^2}{n}$$

$$= \frac{(2-6)^2 + (2.5-6)^2 + \cdots\cdots + (11-6)^2}{25}$$

$\sigma^2 = 5.5 \Rightarrow \sigma = \sqrt{5.5}$

$\sigma = 2.4$

2, A population consists of 5, 10, 14, 18, 13, 24. Consider the possible samples of size 2 which can be drawn without replacement. Find

(i) Mean

(ii) S.D of population

(iii) Mean of the sampling distribution of mean

(iv) S.D of the sampling distribution.

Given population are 5, 10, 14, 18, 13, 24, $N = 6$, $n = 2$

$$\mu = \frac{5 + 10 + 14 + 18 + 13 + 24}{6}$$

$\mu = 14$

(ii) $$\sigma^2 = \frac{\sum_{i=1}^{6} (x - \bar{x})^2}{N}$$

$$= \frac{(5-14)^2 + (10-14)^2 + (14-14)^2 + (18-14)^2 + (13-14)^2 + (24-14)^2}{6}$$

$\sigma^2 = 35.6$

$\sigma = \sqrt{35.6} = 5.97$

(iii) The number of samples without replacement is

$N_{C_n} = {}^6C_2 = 15$ ways

(5, 10) (5, 14) (5, 18) (5, 13) (5, 24) (10, 14) (10, 18)

(10, 13) (10, 24) (14, 18) (14, 13) (14, 24) (18, 13) (18, 24)

(13, 24)

The mean of sample are

7.5 9.5 11.5 9 14.5 12 14 11.5 17 16 13.5

18 15.5 21 18.5

$\bar{x} = 7$

The mean of the standard deviation of mean is

$$\bar{x} = \frac{7 + 7.5 + 11.5 + 9 + \cdots\cdots + 18.5}{15} = 14$$

07-11-2023

(iv) $\sigma^2 = \dfrac{\sum_{i=1}^{6} (x - \bar{x})^2}{n}$

$$\sigma^2 = \frac{(7.5-14)^2 + (9.5-14)^2 + (11.5-14)^2 + \cdots\cdots (18.5-14)^2}{15}$$

$\sigma^2 = 14.26 \Rightarrow \sigma = \sqrt{14.26}$

$\sigma = 3.77$

3, The variance of a population is 2. The size of the sample collected from the population is 169. What is the ~~S.D~~ S.E?

Sol: $\sigma^2 = 2 \Rightarrow \sigma = \sqrt{2} = 1.414$

$n = 169$

$S.E = \dfrac{\sigma}{\sqrt{n}}$

$= \dfrac{\sqrt{2}}{\sqrt{169}} = 0.108$

08-11-2023

4, A random sample of size 100 is taken from an infinite population having the mean $\mu = 76$ and the variance $\sigma^2 = 256$. What is the probability that $\bar{x}$ will be between 75 and 78?

Sol: $\mu = 76$, $\sigma^2 = 256$, $\sigma = 16$, $n = 100$, $\bar{x}_1 = 75$, $\bar{x}_2 = 78$

we know that

$$z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$$

when $\bar{x}_1 = 75$

$z_1 = \dfrac{\bar{x}_1 - \mu}{\sigma / \sqrt{n}}$

$z_1 = \dfrac{75 - 76}{16 / \sqrt{100}} = \dfrac{-1}{16/10}$

$z_1 = -0.625 < 0$

when $\bar{x}_2 = 78$

$z_2 = 1.25 > 0$

$P(75 < \bar{x} < 78) = |A(z_2) + A(z_1)|$

$= |A(1.25) + A(-0.625)|$

$= |A(1.25) + A(0.625)|$

$= |A(1.2 + 0.05) + A(0.6 + 0.02)|$

$= |0.3944 + 0.2324| = 0.6268$

5, A random sample of size 64 taken from a normal population with a mean $\mu = 51.4$ and $\sigma = 6.8$. What is the probability that the mean of the sample will be

(i) Exceed 52.9

(ii) Falls between 50.5 and 52.3

(iii) less than 50.6

Sol: $n = 64$, $\mu = 51.4$, $\sigma = 6.8$,

(i) $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$z_1 = \dfrac{52.9 - 51.4}{6.8/\sqrt{64}} = 1.76 > 0$

$P(z > z_1) = 0.5 - A(z_1)$

$= 0.5 - A(1.76)$

$= 0.5 - A(1.7 + 0.06)$

$= 0.5 - 0.4608$

$= 0.0392$

(ii) $z_1 = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

when $x_1 = 50.5$

$z_1 = \dfrac{50.5 - 51.4}{6.8/\sqrt{64}}$

$z_1 = -1.05 < 0$

$z_2 =$

when $x_2 = 52.3$

$z_2 = \dfrac{52.3 - 51.4}{6.8/\sqrt{64}}$

$z_2 = 1.05 > 0$

$P(50.5 < z < 52.3) = |A(z_2) + A(z_1)|$

$= |A(1.05) + A(-1.05)|$

$= |0.3531 + 0.3531|$

$= 0.7062$

(iii) $z_1 = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$= \dfrac{50.6 - 51.4}{6.8/8}$

$= -0.94$

$= 0.5 - A(z_1)$

$= |0.5 - 0.3264|$

$= 0.1736$

6, what is the effect on standard error if sample is taken from infinite population of sample size is increased from 400 to 900

Sol: $n_1 = 400$ , $n_2 = 900$

$$SE = \frac{\sigma}{\sqrt{n}}$$

$$SE_1 = \frac{\sigma}{\sqrt{n_1}} = \frac{\sigma}{\sqrt{400}} = \frac{\sigma}{20}$$

$$SE_2 = \frac{\sigma}{\sqrt{n_2}} = \frac{\sigma}{\sqrt{900}} = \frac{\sigma}{30}$$

$$SE_1 = \frac{3}{2} SE_2 .$$

**Estimation :**

**Estimate** : To find an unknown population parameter or judgement on a statement is made which is an estimate.

**Estimator** : The method or rule to determine an unknown population parameter is called estimator.

* The estimate can be done in 2 ways

1) Point estimation
2) Interval estimation

**Maximum error of estimate** : The maximum error of estimate is $E_{max} = Z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$

**Sample size** (when mean is given):

$$n = \left[\frac{Z_{\alpha/2}\,\sigma}{E_{max}}\right]^2$$

**Sample size** (when proportion is given):

$$n = \left[\frac{Z_{\alpha/2}}{E_{max}}\right]^2 PQ$$

where $P$ = success of the proportion
$Q$ = Failure of the proportion

* Maximum error $E_{max} = Z_{\alpha/2}\sqrt{\frac{PQ}{n}}$

**Confidence interval estimate of parameters:**

* In an interval estimation of the population parameter ($\theta$), if we can find two quantities $t_1$ and $t_2$ based on a sample observation drawn from the population such that the unknown parameter $\theta$ is included in the interval $[t_1, t_2]$ it in a specified percentage of cases then this interval is called a confidence interval for the parameter $\theta$.

## Confidence limits:

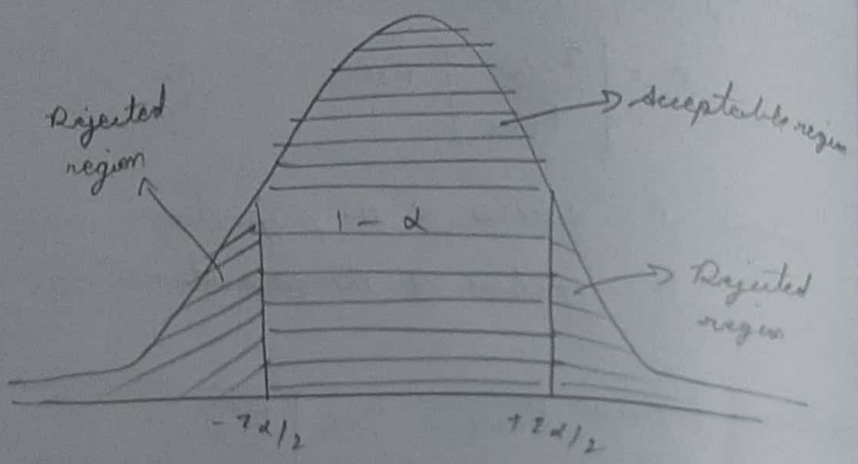
1) 95% Confidence limits are 1.96 i.e., $Z_{\alpha/2} = 1.96$

2) 99% Confidence limits are 2.58 i.e., $Z_{\alpha/2} = 2.57$

3) 98% Confidence limits are 2.33 i.e., $Z_{\alpha/2} = 2.33$

4) 90% Confidence limits are 1.64 i.e., $Z_{\alpha/2} = 1.64$

$$E_{max} = Z_{\alpha/2}\sqrt{\frac{pq}{n}}$$




## Confidence Intervals:

$$C.\ Interval = \left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$$

## Problems:

1) In a study of automobile insurance a random sample of 80 bodies body repair cost had a mean of rupees 472.36 And a standard deviation of rupees 62.35. If $\bar{x}$ is used as the point estimate to the true average repair cost with what confidence we can assert that the maximum error does not exceed rupees 10.

Sol:

$\bar{x} = 472.36$, $\sigma = 62.35$, $n = 80$, $E_{max} = 10$.

Confidence intervals (C.I) = ?

$$E_{max} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$$

$$10 = Z_{\alpha/2}\frac{62.35}{\sqrt{80}}$$

$$Z_{\alpha/2} = \cancel{62}\ \frac{\sqrt{80}\times 10}{62.35}$$

$$Z_{\alpha/2} = 1.43$$

$$\frac{\alpha}{2} = 0.4236 \quad [\because \text{From the normal distribution table}]$$

$$\alpha = 0.8472$$

$\therefore$ The ~~C.L~~ Confidence levels C.L [i.e. $1-\alpha$] = 84.72%

2. What is the size of the smallest sample required to estimate an unknown proportion within a maximum error 0.06 with atleast 95% Confidence.

Sol: $E_{max} = 0.06$

$n = ?$

$Z_{\alpha/2} = 1.96$ [$\because$ 95%]

$P = \frac{1}{2}$ [$\because$ If proportion is not given then it is $\frac{1}{2}$]

$Q = \frac{1}{2}$

$n = \left[\frac{Z_{\alpha/2}}{E_{max}}\right]^2 PQ$

$n = \left[\frac{1.96}{0.06}\right]^2 \left(\frac{1}{2}\right)\left(\frac{1}{2}\right)$

$n = 266.77$

$n = 267$ [only sample size should be a whole number]

3. Assuming that $\sigma = 20$, how large a random sample be taken to assert with the probability 0.95 that the sample mean will not differ from the true mean by 3 point max error.

Sol: $n = ?$, $\sigma = 20$, $E_{max} = 3$, $Z_{\alpha/2} = 1.96$

$n = \left[\frac{Z_{\alpha/2}\sigma}{E_{max}}\right]^2$

$n = \left[\frac{1.96 \times 20}{3}\right]^2 = 170.73$

$n = 171$

4. A random sample of size 100 has a standard deviation 5. What can you say about the maximum errors with 95% Confidence.

Sol: $E_{max} = ?$, $Z_{\alpha/2} = 1.96$ [$\because$ 95%]

$n = 100, \sigma = 5$

$E_{max} = Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}$

$= 1.96 \times \frac{5}{\sqrt{100}}$

$= 0.98$

5. A random sample of size 81 was taken whose variance is 20.25 and the mean is 32. Construct 98%. Confidence interval.

Sol: $n = 81$, $\bar{x} = 32$, $\sigma^2 = 20.25 \Rightarrow \sigma = 4.5$

$Z_{\alpha/2} = 2.33$

$C.I = \left(\bar{x} - Z_{\alpha/2}\frac{\sigma}{\sqrt{n}},\ \bar{x} + Z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right)$

$= \left(32 - \left(2.33 \times \frac{4.5}{\sqrt{81}}\right),\ 32 + \left(2.33 \times \frac{4.5}{\sqrt{81}}\right)\right)$

$= [30.85,\ 33.13]$

11-11-2019

6) Find the mean values of x and y and Correlation Coefficient of Correlation from the following regression equation

$2y - x = 50$, $3y - 2x = 10$

?

Sol: Given regression lines of y on x are

$2y - x = 50$ —①, $3y - 2x = 10$ —②

$x = 130$, $y = 90$

$\bar{x} = 130$, $\bar{y} = 90$

Rewrite the equation ① and ②

From ① $\Rightarrow$ $y = \frac{x}{2} + 25$

From ② $\Rightarrow$ $x = \frac{3}{2}y - 5$

$r\frac{\sigma y}{\sigma x} = \frac{1}{2}$

$r\frac{\sigma x}{\sigma y} = \frac{3}{2}$

$r^2 = \frac{3}{4}$

$r = 0.86$

## Testing of Hypothesis I

**Hypothesis:** There are many problem in which rather than estimating the value of a parameter we need to decide whether to accept or reject a statement about the parameter

* This statement is called hypothesis and the decision making process about the hypothesis is called testing of hypothesis

* A drug chemist is to decide whether a newly drug is really effective in curing a disease.

* A quality control manager is to determine whether the process is working properly.

* There are two types of hypothesis

1) Null hypothesis ($H_0$): A null hypothesis is the hypothesis which assist that there is no significance to difference between the statistics and the population parameter and whatever observed, difference is there merely due to fluctuation in a sampling from the sample population

2) Alternative hypothesis ($H_1$): Any hypothesis which contradicts the null hypothesis is called Alternative Hypothesis.

* It is denoted by $H_1$ and it is defined as

a) $H_1 : \mu \neq \mu_0$

b) $H_1 : \mu > \mu_0$ – Right tailed } one-tailed

c) $H_1 : \mu < \mu_0$ – left tailed

14-11-2019

**Level of significance:** The level of significance is denoted by $\alpha$ is the confidence with which we reject or accept the null hypothesis ($H_0$).

* The level of significance is generally specified by some certain levels.

$\alpha = 5\%$ (95% confidence)

$\alpha = 10\%$ (90% confidence)

$\alpha = 1\%$ (99% confidence)

**Note:** If the level of significance is not mentioned then by default it is considered as 5%.

**Errors of sampling:**

The object in sampling theory is to draw valid inference about the population parameter on the basis of the sample result.

* In a practice we decide to accept or reject after examining a sample.

* There are two types of errors:

1) **Type I error:** Reject $H_0$ when it is true i.e. if the null hypothesis $H_0$ is true but it is rejected by the test procedure then the error made is called Type I error.

2) **Type II error:** Accept $H_0$ when it is wrong i.e. accept $H_0$ when $H_1$ is true if the $H_0$ is false but it is accepted by the test procedures, then the error committed is called Type II error.

**Critical values:**

| | Level of significance | | |
|---|---|---|---|
| | 1% (0.01) | 5% (0.05) | 10% (0.1) |
| Two tailed Test | $\lvert Z_2 \rvert = 2.58$ | $Z_2 = 1.96$ | $Z_2 = 1.64$ |
| Right tailed Test | $Z_2 = 2.33$ | $Z_2 = 1.64$ | $Z_2 = 1.28$ |
| Left tailed Test | $Z_2 = -2.33$ | $Z_2 = -1.64$ | $Z_2 = -1.28$ |

**Procedure for testing of Hypothesis:**

**Step-1:** Null hypothesis ($H_0$): Define or setup a null hypothesis ($H_0$) taking into consideration, the nature of the problem and the data involved.

**Step 2:** Alternative hypothesis ($H_1$): Setup the alternative hypothesis so that we could decide whether we should use one tailed or two-tailed test.

step-3: level of significance ($\alpha$): select the appropriate level of significance ($\alpha$) usually we chose 5%. level of significance.

step-4: Test of statistics (z test): compute the test of statistics under the null of hypothesis

step-5: Conclusion: i) if $|z| < z_\alpha$, $H_0$ is accepted

(ii) if $|z| > z_\alpha$, $H_0$ is rejected

$|z|$ = Calculated value

$|z_\alpha|$ = Tabular value.

Test of significance for large sample (when sample mean is not given):

step-1: $H_0$

step-2: $H_1$

step-3: $\alpha$

step-4: Test of statistics $z = \frac{x - \mu}{\sigma}$

step-5: Conclusion.

problems:

1) If a coin is tossed 960 times and returned head 183 times. Test the hypothesis that the coin is unbiased.

sol: Given, $n = 960$ ($n > 30$ large sample)

$x = 183$

$p = \frac{1}{2}$, $q = \frac{1}{2}$

$\mu = np = 960\left(\frac{1}{2}\right) = 480$

$\sigma = \sqrt{npq}$

$= \sqrt{960\left(\frac{1}{4}\right)}$

$= 15.49$

$\alpha = 5\%$

step I: Null Hypothesis $H_0$: Coin is unbiased

step II: Alternative Hypothesis $H_1$: Coin is biased

step III: level of significance - $\alpha = 5\%$ at $z_\alpha = 1.96$

step IV: Test of statistics $z = \frac{x - \mu}{\sigma} = \frac{180 - 480}{15.49}$

$= -19.17$

step V: Conclusion: $|z| > z_\alpha$ $H_0$ is Rejected

14-11-24

2, A dice is tossed 960 times and it falls with 5 upwards 184 times. Is the dice unbiased at the level of significance 1%.

Sol: Given, $n = 960$ ($n > 30$ large sample)

$x = 184$

$p = \frac{1}{6}, q = \frac{5}{6}$

$p + q = 1$

$\frac{1}{6} + q = 1$

$q = 1 - \frac{1}{6}$

$q = \frac{5}{6}$

$\mu = np = \cancel{960}^{160}\left(\frac{1}{\cancel{6}}\right) = 160$

$\sigma = \sqrt{npq} = \sqrt{\cancel{160 \times \frac{5}{6}}}$

$\sigma = \sqrt{\cancel{160}^{80} \times \frac{5}{\cancel{6}_3}} = \sqrt{\frac{400}{3}}$

$\sigma = 11.54$

$\alpha = 1\%$

step-I: $H_0$ : dice is unbiased

step-II: $H_1$ : dice is biased

step-III: $\alpha = 1\%$

Testing of Hypothesis for a single mean (large samples):

step-I: Null hypothesis $H_0 : \mu = \mu_0$ (Two tailed)

step-II: Alternative hypothesis $H_1 : \mu > \mu_0$ (Right tailed)

step-III: level of significance ($\alpha$) : 5%. 10%. 1%.

By default 5% is used when $\alpha$ is not given.

step-IV: Testing of statistic $z = \frac{\bar{x} - \mu}{\sigma / \sqrt{n}}$

step-V: Conclusion: 1) If $|z| < z_\alpha$, $H_0$ is accepted

2) If $|z| > z_\alpha$, $H_0$ is rejected

Problem:

1, According to the norms established for a mechanical aptitude test persons who are 18 years old have an average height of 73.2 with a standard deviation of 18.6. If 40 randomly selected person of that age average is 76.7 test the hypothesis as $\mu = 73.2$ and again its alternative hypothesis $\mu > 73.2$ at the level of significance 1%.

Sol: Given, $n = 40$, $\bar{x} = 76.7$, $\mu = 73.2$, $\alpha = 1\%$.

1) $H_0 : \mu = \mu_0 (= 76.7)$

2) $H_1 : \mu > \mu_0 (= 76.7)$ [Right tailed test]

3) $\alpha = 1\%$ i.e. $z_\alpha = 2.33$

18-11-2013

4) $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{76.7 - 73.2}{8.1/\sqrt{40}} = 2.5$

5) $|z| > z_\alpha$

$\therefore$ $H_0$ is rejected

2, A sample of 64 students have a mean weight of 70 kgs. Can this be regarded as a sample mean from a population mean with weight 56 kgs and a standard deviation 26 kgs.

Sol: Given, $n = 64$, $\bar{x} = 70$, $\mu = 56$, $\alpha = 5\%$, $\sigma = 28$

1) $H_0 : \mu = \mu_0 = 70$

2) $H_1 : \mu \neq \mu_0$ [Two tailed test]

3) $\alpha = 5\%$ i.e. $z_\alpha = 1.96$

4) $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = 4$

5) $|z| > z_\alpha$

$\therefore$ $H_0$ is rejected.

3, An ambulance service claim that it takes an average less than 10 min to reach its destination in emergency class. A sample of 36 calls has a mean of 11 mins and the variance of 16 mins. Test the level of significance at 5%.

Sol: Given, $n = 36$, $\bar{x} = 11$, $\mu = 10$, $\alpha = 5$, $\sigma^2 = 16$, $\sigma = 4$

1) $H_0 : \mu = \mu_0$

2) $H_1 : \mu < \mu_0$ [left tailed test]

3) $\alpha = 5\%$ i.e. $z_\alpha = -1.64$

4) $z = \frac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \frac{11 - 10}{4/\sqrt{36}} = 1.5$

5) $|z| < z_\alpha$

$\therefore$ $H_0$ is ~~rejected~~ accepted.