SIMATS
School of Engineering

Probability and Statistics

Science & Humanities

Saveetha Institute of Medical And Technical Sciences,Chennai.

UBA09 – PROBABILITY AND STATISTICS

INDEX

UBA09 - PROBABILITY AND STATISTICS

UNIT - I RANDOM VARIABLE

1. Probability and Conditional Probability (1)
2. Discrete Random variable - Mathematical Expectation and Moment Generating Function (MGF) (2)
3. Conditional Random Variable - Mathematical Expectation and Moment Generating Function (MGF) (3)

UNIT II - DISCRETE DISTRIBUTION

1. Binomial Distribution - Mean, Variance & MGF (4)
2. Poisson Distribution - Mean, Variance & MGF (5)
3. Geometric Distribution - Mean, Variance & MGF (6)

UNIT III - CONTINUOUS DISTRIBUTION

1. Uniform Distribution - Mean, Variance & MGF (7)
2. Exponential Distribution - Mean, Variance & MGF (8)
3. Normal (Gaussian) Distribution - Area under the normal curve and Probability. (9)

UNIT IV - LARGE SAMPLE (TEST OF SIGNIFICANCE)

1. Single Mean. (10)
2. Difference of Means. (11)
3. Single Proportion. (12)
4. Difference of Proportions. (13)

UNIT V - SMALL SAMPLE (TEST OF SIGNIFICANCE)

t - Test

1. Single Mean (15)
2. Difference of Means. (16)

F - Test

3. Ratio of Variances (17)

Chi - Square Test

4. Goodness of Fit (18)
5. Independence of Attributes. (19)

Typed in A3

Trail and Event :

* The performance of a random experiment is called Trail.
* The outcomes is called an Event

Example : Throwing of a coin is a trail and getting H or T is an event.

Sample Space :

The totality of the possible outcomes of a random experiment is called Sample space and it is denoted by S.

Example: Tossing two coins simultaneously then $S = \{HH, HT, TH, TT\}$

Mutually Exclusive Event :

Two events A & B are said to be mutually exclusive events or disjoint events if $A \cap B$ is the null set.

Example: When a coin is tossed getting

Exhaustive Events :

A set of events is said to be exhaustive if no event outside this set occurs and atleast one of these events must happen as a result of an experiment

Example : If a coin is tossing either the head or tail turns up, there is no other probability.

PROBABILITY

Probability : Let S be the sample space and A be an event associated with a random experiment. Let n(S) & n(A) be the number of elements of S & A respectively.

ie., $P(A) = \dfrac{n(A)}{n(S)}$.

Axioms of Probability :

(i) $0 \le P(E) \le 1$ (ii) $P(S) = 1$

(iii) If A & B are mutually exclusive events

$P(A \cup B) = P(A) + P(B)$

Theorems :-

(i) $P(\phi) = 0$ (ii) $P(\bar{A}) = 1 - P(A)$

(iii) Addition theorem : If A & B are any two events are not disjoint, then

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

Independent Event : If the happening of an event A does not depend on the happening of event B then they are called independent event ie., $P(A \cap B) = P(A) \cdot P(B)$

Conditional Probability : The conditional Probability of A given B is

$$P(A/B) = \frac{P(A \cap B)}{P(B)}, \text{ if } P(B) \neq 0$$

Multiplication Rule :

$$P(A \cap B) = \begin{cases} P(B) \cdot P(A/B), & \text{if } P(B) \neq 0 \\ P(A) \cdot P(B/A), & \text{if } P(A) \neq 0 \end{cases}$$

Problems :-

① If $P(A) = 0.35$, $P(B) = 0.73$, $P(A \cap B) = 0.14$. Find $P(A \cup B)$

Solution:-

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

$= 0.35 + 0.73 - 0.14 = 0.94$

② A card is drawn at random from a well-shuffled deck of 52 cards. Find the Probability of drawing a queen or a king.

Solution:-

Given $n(A) = 4$, $n(B) = 4$

$P(A \text{ or } B) = P(A \cup B) = P(A) + P(B)$

$= \frac{4}{52} + \frac{4}{52} = \frac{2}{13}$

③ Two cards are drawn from a pack of 52 cards in Succession. Find the Probability that both are kings, when (i) the first drawn card is replaced (ii) the card is not replaced.

Solution:-

Given $n(A) = 4$, $n(B) = 4$, $n(S) = 52$

(i) When card is replaced :

A & B are independent, A will not affect the prob. of occurence of B.

$\therefore P(A \cap B) = P(A) \cdot P(B) = \frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$

(ii) When card is not replaced ie., A & B are not independent

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

$= \frac{4}{52} \cdot \frac{3}{51}$

$= \frac{1}{221}$

$P(A \cap B) = \frac{1}{221}$

Random Variable: A real valued function over the sample space.

Discrete R.V: $x = 0, 1, 2, 3, \ldots$

(i) $P(x_i) \geq 0, \forall i$

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

Problem 1: A R.V "X" has the folg. probability function.

X :	0	1	2	3	4
P(X) :	k	3k	5k	7k	10k

Find (i) k, (ii) $P(X \geq 3)$ (iii) $P(0 < X < 4)$

(iv) CDF.

Solution: (i) $\sum P(x_i) = 1$

$k + 3k + 5k + 7k + 10k = 1$

$\Rightarrow 25k = 1 \Rightarrow \boxed{k = \frac{1}{25}}$

(ii) $P(X \geq 3) = P(X=3) + P(X=4)$

$= 7k + 9k = 16k.$

$\boxed{P(X \geq 3) = \frac{16}{25}}$

(iii) $P(0 < X < 4) = P(1) + P(2) + P(3)$

$= 15k = \frac{15}{25} = \frac{3}{5}$

$\boxed{P(0 < X < 4) = 3/5}$

(iv) CDF

X	0	1	2	3	4
P(X)	$\frac{1}{25}$	$\frac{3}{25}$	$\frac{5}{25}$	$\frac{7}{25}$	$9/25$
F(X)	$\frac{1}{25}$	$\frac{4}{25}$	$\frac{9}{25}$	$\frac{16}{25}$	1

Discrete Random Variable

Commulative Distribution function:

* $F(x) = \sum_{x_k \leq x} P(x_k)$.

Mathematical Expectation.

* $E(X) = \sum x P(X)$
* $Var(X) = E(X^2) - (E(X))^2$

Properties of Expectation

* $E(c) = c$, c is a Constant
* $E(aX + b) = aE(X) + b$.
* $E(X + Y) = E(X) + E(Y)$
* $E(XY) = E(X)E(Y)$, If X & Y are Independent.

Properties of Variance

* $Var(aX) = a^2 Var(X)$
* $Var(X + c) = Var(X)$
* $Var(X + Y) = Var(X) + Var(Y)$

If X & Y are independent

Problem 2: Let X be the number that turns up when a dice is thrown. Find Mean and Variance of X.

Solution: X is a Discrete R.V

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

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

Mean $= E(X) = \sum x p(x)$

$= (1 \times \frac{1}{6}) + (2 \times 1/6) + (3 \times \frac{1}{6}) + (4 \times 1/6) + (5 \times 1/6) + (6 \times 1/6)$

$= \frac{7}{2}$

$E(X^2) = \sum x^2 P(x)$

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

$= \frac{91}{6}$

$Var(X) = E(X^2) - (E(X))^2$

$= \frac{91}{6} - \left(\frac{7}{2}\right)^2 = \frac{35}{12}$

Moment Generating function

$M_X(t) = E(e^{tx}) = \sum e^{tx} p(x)$

Problem 3: A perfect coin is tossed twice. If X denotes the no. of heads that appear, find MGF of X, also find Mean and Variance.

Solution:-

x	0	1	2
P(x)	1/4	1/2	1/4
e^{tx}	1	e^t	e^{2t}

$M_X(t) = E(e^{xt}) = \frac{1}{4} + e^t(1/2) + e^{2t}(1/4)$

$= \frac{1}{4}(1 + e^t)^2$

$M_X'(t) = \frac{1}{2}(1 + e^t)e^t \Rightarrow \mu_1' = M_X'(0) = 1$

$M_X''(t) = \frac{1}{2}[e^t(1 + e^t) + e^{2t}] \Rightarrow \mu_2' = M_X''(0) = \frac{3}{2}$

$\therefore$ Mean $= \mu_1' = 1$

$Var(X) = \mu_2' - (\mu_1')^2 = \frac{3}{2} - 1 = \frac{1}{2}$

CONTINUOUS RANDOM VARIABLE

PROBABILITY DENSITY FUNCTION:

* $P(x \in [a,b]) = \int_a^b f(x)\,dx$
* $f(x) \geq 0$
* $\int_{-\infty}^{\infty} f(x)\,dx = 1$

PROBLEM:-

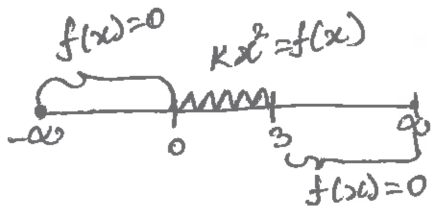
Find K, if $f(x) = \begin{cases} Kx^2, & 0<x<3 \\ 0, & \text{otherwise} \end{cases}$

is a pdf and compute.

(i) $P(1<X<2)$; (ii) $P(X<2)$; (iii) $P(X\geq 2)$; (iv) $P(X=2)$.

SOLUTION:-

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



$\int_{-\infty}^{\infty} f(x)\,dx = \int_0^3 f(x)\,dx = 1$

$\boxed{K = \frac{1}{9}}$

i) $P(1<X<2) = \int_1^2 f(x)\,dx = \boxed{\frac{7}{27}}$

ii) $P(X<2) = \int_{-\infty}^{2} f(x)\,dx = \boxed{\frac{8}{27}}$

iii) $P(X\geq 2) = 1 - P(X<2)$

$= 1 - \frac{8}{27} = \boxed{\frac{19}{27}}$

iv) $P(X=2) = \boxed{0}$

Cummulative Distribution fn

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

MATHEMATICAL EXPECTATIONS:

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

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

$Var(X) = E(x^2) - (E(x))^2$.

PROBLEM:-

Find the mean and variance of X given.

$f(x) = \begin{cases} x; & 0 \leq x \leq 1 \\ 2-x; & 1 \leq x \leq 2 \\ 0; & \text{otherwise.} \end{cases}$

SOLUTION:-

Mean $= E(X) = \int_{-\infty}^{\infty} x f(x)\,dx$

$= \int_0^1 x \cdot x\,dx + \int_1^2 x(2-x)\,dx$

$= \left[\frac{x^3}{3}\right]_0^1 + \left[\frac{2x^2}{2} - \frac{x^3}{3}\right]_1^2$

$= \left[\frac{1}{3} - 0\right] + \left[4 - \frac{8}{3} - 1 + \frac{1}{3}\right] = 1$

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

$= \int_0^1 x^2 \cdot x\,dx + \int_1^2 x^2(2-x)\,dx$

$= \left[\frac{x^4}{4}\right]_0^1 + \left[\frac{2x^3}{3} - \frac{x^4}{4}\right]_1^2$

$= \left[\frac{1}{4} - 0\right] + \left[\frac{16}{3} - \frac{16}{4} - \frac{2}{3} + \frac{1}{4}\right] = \frac{7}{6}$

$\therefore Var(X) = E(X^2) - [E(X)]^2$

$= \frac{7}{6} - (1)^2 = \frac{1}{6}$.

MOMENT GENERATING FUNCTION:

$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f_X(x)\,dx$.

PROBLEM: 1. The number of hours of satisfactory operations that a certain brand of TV set will give is a random variable with pdf

$f(x) = 500e^{-500x}$, $x>0$
$= 0$, $x \leq 0$

Find the m.g.f of X, mean and variance of X.

$M_x(t) = E(e^{xt}) = \int_0^{\infty} e^{xt}\, 500e^{-500x}\,dx$

$= 500\int_0^{\infty} e^{(t-500)x}\,dx$

$= 500\left[\frac{e^{(t-500)x}}{t-500}\right]_0^{\infty} = \frac{500}{500-t}$

$M'_x(t) = \frac{500}{(500-t)^2}$; $M'_x(0) = \frac{1}{500}$

$M''_x(t) = \frac{500\times 2}{(500-t)^3}$; $M''_x(0) = \frac{2}{500^2}$

$\therefore$ Mean $= \mu_1' = M'_x(0) = \frac{1}{500}$

Variance $= \mu_2' - \mu_1'^2 = \frac{2}{500^2} - \left[\frac{1}{500}\right]^2 = \frac{1}{500^2}$.

PROBLEM:

2. Obtain the M.G.F of the RV X having

pdf $f(x) = \begin{cases} x, & 0 \leq x \leq 1 \\ 2-x, & 1 \leq x < 2 \\ 0, & \text{otherwise} \end{cases}$

$M_X(t) = E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx = \int_0^1 x e^{tx}\,dx + \int_1^2 (2-x)e^{tx}\,dx$

$= \left[x\frac{e^{tx}}{t} - \frac{e^{tx}}{t^2}\right]_0^1 + \left[(2-x)\frac{e^{tx}}{t} - (-1)\frac{e^{tx}}{t^2}\right]_1^2$.

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

$= M_X(t) = \frac{1}{t^2}[e^t - 1]^2$.

The probability of x successes in 'n' trails is

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

$$x = 0, 1, 2, \cdots n$$

Moment Generating Function:

$$M_x(t) = (q + pe^t)^n$$

$$E(x) = \text{Mean} = np$$

$$E(x^2) = n^2p^2 + npq$$

$$Var(x) = npq$$

$$S.D = \sqrt{npq}$$

Problems:

1. Find the binomial distribution if mean = 4 and variance = 3.

Solution:

$$E(x) = 4 = np$$

$$Var(x) = 3 = npq$$

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

$$p = 1 - q = \frac{1}{4}, \quad np = 4 \Rightarrow n = 16$$

$$p(x) = 16C_x \left(\tfrac{1}{4}\right)^x \left(\tfrac{3}{4}\right)^{16-x}$$

$$x = 0, 1, 2, \cdots 16.$$

BINOMIAL DISTRIBUTION

2. $M_x(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$ for a R.v. X. Find $E(x)$, $Var(x)$ and $p(x=2)$.

Solution:

$$M_x(t) = (q + pe^t)^n = \left(\tfrac{1}{4} + \tfrac{3}{4}e^t\right)^5$$

$$\Rightarrow q = \tfrac{1}{4}, \quad p = \tfrac{3}{4} \quad n = 5$$

$$E(x) = np = \frac{15}{4}$$

$$Var(x) = npq = \frac{15}{16}$$

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

$$p(x=2) = 5C_2 \left(\tfrac{3}{4}\right)^2 \left(\tfrac{1}{4}\right)^3$$

$$= 0.0879$$

3. In a large consignment of electric bulbs 10% are defective. A random sample of 20 is taken for inspection. Find the probability that (i) All are good bulbs, (ii) Atmost there are 3 defective bulbs, (iii) Exactly there are three defective bulbs, (iv) 2 are defective

Solution: Here $p = \frac{10}{100} = 0.1$, $q = 1 - p = 0.9$, $n = 20$

(i) P(all are good bulbs)

$= p(\text{none are defective}) = p(0)$

$= nC_0 p^0 q^{n-0}$

$= 20C_0 (0.1)^0 (0.9)^{20}$

$= 0.1216$

(ii) p(at most there are 3 defective bulbs) $= p(x \le 3)$

$= p(0) + p(1) + p(2) + p(3)$

$= 20C_0 (0.1)^0 (0.9)^{20} + 20C_1 (0.1)(0.90)^{19} + 20C_2 (0.1)^2 (0.9)^{18} + 20C_3 (0.1)^3 (0.9)^{17}$

$= 0.1215 + 0.27 + 0.285 + 0.19$

$= 0.8666$

(iii) p(exactly 3 defective bulbs)

$= p(3)$

$= nC_3 p^3 q^{n-3}$

$= 20C_3 (0.1)^3 (0.9)^{17}$

$= 0.19$

(iv) p[exactly 2 defective bulbs)

$= p(2)$

$= nC_2 p^2 q^{n-2}$

$= 20C_2 (0.1)^2 (0.9)^{18}$

$= 0.285$

POISSON DISTRIBUTION

POISSON DISTRIBUTION

Probability Mass function

$$P[X=x] = P(x) = \begin{cases} \dfrac{e^{-\lambda}\lambda^{x}}{x!}, & x = 0,1,2\ldots \lambda > 0 \\ 0 & \text{otherwise} \end{cases}$$

λ is the parameter of possion distribution

Mean $= \lambda$	VARIANCE $= \lambda$	S.D $= \sqrt{\lambda}$

EXAMPLE

* Number of defective items produced in the factory.
* Number of deaths due to rare disease.
* Number of mistake committed by a typist per page.

PROBLEM-1

Write down the probability mass function of the poisson distribution, which is approximately equivalent to B(100, 0.02)

SOLUTION: GIVEN $n = 100$
$p = 0.02$

$\lambda = np = 100 \times 0.02 = 2$

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

$$= \frac{e^{-2}\,2^{x}}{x!}$$

When $x = 0, 1, 2, \ldots\ldots$

Problem. 2 The number of typing mistakes that a typist make on the given page has a poisson distribution with a mean of 3 mistakes. What is the probability that she makes

(i) Exactly 7 mistakes

(ii) Fewer than 4 mistakes

(iii) No mistake on a given Page

Given mean $= \lambda = 3$

$$P[X=x] = \frac{e^{-\lambda}\lambda^{3}}{3!}$$

(i) P[Exactly 7 mistakes]

$$P[X=7] = \frac{e^{-3}(3)^{7}}{7!}$$

$= 0.0216$

(ii) P[Fewer than 4 mistakes]

$= P[X<4]$

$= P[X=0] + P[X=1] + P[X=2] + P[X=3]$

$$= \frac{e^{-3}(3)^0}{0!} + \frac{e^{-3}(3)^1}{1!} + \frac{e^{-3}(3)^2}{2!} + \frac{e^{-3}(3)^3}{3!}$$

$$= e^{-3} + 3e^{-3} + \frac{9e^{-3}}{2} + \frac{9e^{-3}}{2}$$

$= 13e^{-3}$

$= 13 \times 0.0498$

$= 0.6474$

(iii)

P[No mistake on a given Page]

$= P[X=0]$

$$= \frac{e^{-3}(3)^0}{0!}$$

$= e^{-3} = 0.0498$

GEOMETRIC DISTRIBUTION

Definition:

PMF of X

$P(X=x) = q^{x-1}p; \quad x = 1, 2, \ldots$

Mean $E(X) = 1/p$

Variance $var(X) = \frac{q}{p^2}$

MGF: $M_X(t) = \frac{pe^t}{1-qe^t}$.

Another form of G.D is

$P(X=x) = q^x p$

$x = 0, 1, 2, \ldots$

Memoryless property (Ageless Property)

$P(X > m+n / X > m) = P(X > n)$

Problems:

1. If the probability that a target is destroyed on any shot is 0.5, then find the probability that it would be destroyed on 6^{th} attempt?

Solution:-

Given $p = 0.5$, $q = 1-p = 0.5$

$P(X=x) = q^{x-1} \cdot p$

$P(X=6) = q^{6-1} p = (0.5)^5 \cdot (0.5)$

$= \frac{1}{26}$

2. The prob. that an applicant for a driver's licence will pass the road test on any given trial is 0.8, find

(i) The probl: that pass the test on 4^{th} trial

(ii) The probl: that pass the test in fewer than 4 trials?

Solution:

Given $p = 0.8$, $q = 0.2$

$P(X=x) = q^{x-1} \cdot p = P(X=2) = (0.2)^{x-1}(0.8)$

(i) $P(X=4) = (0.2)^{4-1}(0.8)$

$= (0.2)^3(0.8) = 0.0064$

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

$= [1 + 0.2 + (0.2)^2][0.8]$

$= \boxed{0.992}$

3. A die is cast until 6 appears. What is the probability that it must be cast more than five times?

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

$P(X=x) = q^x p$

$q = \frac{5}{6}, \; p = \frac{1}{6} \quad \therefore P(X=x) = \left(\frac{5}{6}\right)^x \cdot \frac{1}{6}$

$P(X>5) = 1 - P(X \leq 5)$

$= 1 - \{P(0) + P(1) + P(2) + P(3) + P(4) + P(5)\}$

$= 1 - \left\{\frac{1}{6} + \frac{5}{6} \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^2 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^3 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^4 \cdot \frac{1}{6} + \left(\frac{5}{6}\right)^5 \cdot \frac{1}{6}\right\}$

$= 1 - \frac{1}{6}\left\{1 + \frac{5}{6} + \left(\frac{5}{6}\right)^2 + \left(\frac{5}{6}\right)^3 + \left(\frac{5}{6}\right)^4 + \left(\frac{5}{6}\right)^5\right\}$

$= 1 - \frac{1}{6}(3.586) = 1 - 0.598 = \boxed{0.402}$

4. The probability that an applicant for a driver's licence will pass the road test on any given trial is 0.7. Find the probability that he will pass the test (i) on the third trial (ii) before the fifth trial.

Solution: Let X denote the number of trials required for pass then X follows geometric distribution with probability function

$P(X=x) = q^{x-1} p, \quad x = 1, 2, \ldots$

$p = 0.7$, $q = 1-p = 0.3$

(i) $P(X=3) = (0.3)^2(0.7) = 0.063$

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

$= p + pq + pq^2 + pq^3$

$= 0.7 + (0.3)(0.7) + (0.7)(0.3)^2 + (0.7)(0.3)^3$

$= 0.7 + 0.21 + 0.063 + 0.019 = 0.992$

UNIFORM DISTRIBUTION

Definition: A random variable 'X' is said to have continuous uniform distribution, if its p.d.f is given by $f(x) = \begin{cases} \frac{1}{b-a}, & a<x<b \\ 0, & \text{otherwise} \end{cases}$ where a & b are two parameters.

M.G.F

$$M_X(t) = E[e^{tx}]$$
$$= \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$
$$= \frac{1}{b-a}\int_a^b e^{tx}\,dx$$

$$M_X(t) = \frac{e^{bt}-e^{at}}{b-a}$$

Mean

$$\mu_r' = \int_a^b x^r f(x)\,dx$$
$$\mu_1' = \int_a^b x f(x)\,dx$$
$$= \frac{1}{b-a}\int_a^b x\,dx$$
$$= \frac{b^2-a^2}{2(b-a)}$$
$$= \frac{(b+a)(b-a)}{2(b-a)}$$

$$\text{Mean} = \frac{b+a}{2}$$

Problems:

① If X is uniformly distributed over (0,10), find the probability that (i) $X<2$ (ii) $X>8$ (iii) $3<X<9$

Solution: $X \sim U.D(0,10)$

$$f(x) = \frac{1}{b-a}, \quad a<x<b$$
$$= \begin{cases} \frac{1}{10}, & 0<x<10 \\ 0, & \text{otherwise} \end{cases}$$

(i) $P(X<2) = \int_0^2 f(x)\,dx = \int_0^2 \frac{1}{10}\,dx$
$$= \frac{1}{10}(x)_0^2 = \frac{1}{5}$$

(ii) $P(X>8) = \int_8^{10} \frac{1}{10}\,dx = \frac{1}{5}$

(iii) $P(3<X<9) = \int_3^9 f(x)\,dx$
$$= \int_3^9 \frac{1}{10}\,dx = \frac{3}{5}$$

② A random variable 'X' has a uniform distribution over (-3,3). Compute (i) $P(|X|<2)$ (ii) $P(|X-2|<2)$ (iii) Find k, $P(X>k) = \frac{1}{3}$

Solution: $X \sim U.D(-3,3)$

$$f(x) = \frac{1}{2a}, \quad -a<x<a$$
$$= \begin{cases} \frac{1}{6}, & -3<x<3 \\ 0, & \text{o.w} \end{cases}$$

(i) $P(|X|<2) = P(-2<X<2)$
$$= \int_{-2}^{2} \frac{1}{6}\,dx = \frac{4}{6} = \frac{2}{3}$$

(ii) $P(|X-2|<2)$
$$= P(-2<(X-2)<2)$$
$$= P(0<x<4) = \int_0^3 \frac{1}{6}\,dx$$
$$= \frac{1}{2}$$

(iii) Given $P(X>k) = \frac{1}{3}$

$$\Rightarrow \int_k^3 f(x)\,dx = \frac{1}{3}$$
$$\Rightarrow \frac{1}{6}(3-k) = \frac{1}{3}$$
$$\Rightarrow 3-k = 2$$
$$\Rightarrow \boxed{k=1}$$

③ Subway trains on a certain line run every half an hour between mid night and six in the morning. What is the prob. that a man entering the station at a random time during this period will have to wait atleast twenty minutes.

Solution:

Given $f(x) = \begin{cases} \frac{1}{30}, & 0<x<30 \\ 0, & \text{otherwise} \end{cases}$

P[a man is waiting for atleast 20 minutes]

$$= P[X \geq 20]$$
$$= P[X \geq 20] = \int_{20}^{30} f(x)\,dx$$
$$= \int_{20}^{30} \frac{1}{30}\,dx$$
$$= \frac{1}{30}(30-20)$$
$$= \frac{1}{3}$$

$$\boxed{\therefore P[X \geq 20] = \frac{1}{3}}$$

Probability density function is

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & \lambda > 0,\ x > 0 \\ 0, & \text{otherwise} \end{cases}$$

Mean $= E(X) = 1/\lambda$

Variance $= 1/\lambda^2$

Standard deviation $= \sqrt{\text{Variance}} = 1/\lambda$

Moment Generating Function : (M.G.F)

$$M_X(t) = E[X(t)] = \int_{-\infty}^{\infty} e^{tx} f(x)\,dx$$

$$= \int_{-\infty}^{\infty} e^{tx} \lambda e^{-\lambda x}\,dx = \lambda \int_{-\infty}^{\infty} e^{-(\lambda - t)x}\,dx$$

$$= \lambda \left[\frac{e^{-(\lambda - t)x}}{-(\lambda - t)} \right]_{-\infty}^{\infty} = \frac{\lambda}{\lambda - t}, \quad \lambda > t$$

Problems:

① The time in hrs required to repair a machine is exponential distribution with parameter $\lambda = 1/2$, Find the probability (a) Exceeds 2 seconds (b) Exceeds 5 seconds.

Solution:

Given $\lambda = \frac{1}{2}$. P.d.f is $f(x) = \frac{1}{2} e^{-x/2}, x > 0$

(a) $P(X > 2) = \int_2^{\infty} \frac{1}{2} e^{-x/2}\,dx = \frac{1}{2}\left[\frac{e^{-x/2}}{-1/2}\right]_2^{\infty} = e^{-1}$

(b) $P(X > 5) = \int_5^{\infty} \frac{1}{2} \cdot e^{-x/2}\,dx = e^{-5/2} = e^{-2.5}$

EXPONENTIAL DISTRIBUTION

② The time required to repair a machine is exponentially distributed with parameter $\lambda = 1/2$. What is the probability that a repair takes atleast 10 hours given that its duration exceeds 9 hours.

Solution :-

Given $\lambda = 2$, $f(x) = \frac{1}{2} e^{-x/2}, \quad x > 0$

$$P[X > 10 / X > 9] = P[X > 9 + 1 / X > 9] = P[X > 1]$$

$$= e^{-\frac{1}{2}} = 0.6065$$

Since, $P[X > m+n / X > m] = P[X > n] = e^{-\lambda n}$

③ Suppose the duration 'X' in minutes of long distance calls from your home follows exponential law with p.d.f

$$f(x) = \begin{cases} \frac{1}{5} e^{-x/5}, & x > 0 \\ 0 & \text{o.w} \end{cases}$$

Find (i) $P(X > 5)$ (ii) $P(3 \le X \le 6)$ (iii) Mean of X and variance of X

Solution:

Given $\lambda = \frac{1}{5}$

(i) $P(X > 5) = \int_5^{\infty} f(x)\,dx = \int_5^{\infty} \frac{1}{5} e^{-x/5}\,dx$

$$= \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_5^{\infty} = e^{-1}$$

(ii) $P(3 \le X \le 6) = \int_3^6 \frac{1}{5} e^{-x/5}\,dx = \frac{1}{5}\left[\frac{e^{-x/5}}{-1/5}\right]_3^6$

$$= e^{-3/5} - e^{-6/5}$$

Memoryless Property:

$$P(X > m+n / X > m) = P(X > n)$$

Proof:

$$P(X > n) = \int_n^{\infty} \lambda e^{-\lambda x}\,dx = \left[\frac{e^{-\lambda x}}{-\lambda}\right]_n^{\infty} \lambda$$

$$= -(0 - e^{-n\lambda}) = e^{-\lambda n}$$

So, $P[X > m+n / X > m] = \dfrac{P[X > m+n]}{P[X > m]}$

$$= \frac{e^{-\lambda(m+n)}}{e^{-\lambda m}} = \frac{e^{-\lambda m} \cdot e^{-\lambda n}}{e^{-\lambda m}}$$

$$= e^{-\lambda n} = P(X > n)$$

Hence Proved.

(iii) Mean $= \frac{1}{\lambda} = \frac{1}{1/5} = 5$

Variance $= \frac{1}{\lambda^2} = \frac{1}{(1/25)} = 25$

④ Suppose that the no. of kilometers that a car can run before it wears out is exponentially distributed with an avg. value of 12,000 kms. If a person desires to go on a tour covering a distance of 3000 kms, what is the probability that the person will be able to complete the tour without replacing the battery?

Solution:

Given mean $= 12000$ $\therefore \lambda = \frac{1}{12{,}000}$

$\therefore P[X > t + 3000 / X > t] = P[X > 3000]$

$$= \int_{3000}^{\infty} f(x)\,dx = \int_{3000}^{\infty} \frac{1}{12000} e^{-\frac{x}{12000}}\,dx$$

$$= \frac{1}{12000}\left[\frac{e^{-x/12000}}{-1/12000}\right]_{3000}^{\infty} = e^{-1/4}$$

$$= 0.7788$$

NORMAL DISTRIBUTION

Normal Distribution :-

X - Continuous Random variable, follows Normal with μ - mean; σ^2 - variance

p.d.f $f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$

$-\infty < x < \infty$, $\sigma > 0$

$-\infty < \mu < \infty$

$\Downarrow$

Standard Normal Curve

$P(z_1 < z < z_2)$

$= \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} dz$

$= \phi(z)$

$\Downarrow$

n^{th} - Central Moments

$\boxed{M_x(t) = e^{\mu t + t^2\sigma^2/2}}$

$E(e^{t(x-\mu)}) = e^{t^2\sigma^2/2}$

$\boxed{\mu_n - \text{Coefft. of } \frac{t^n}{n!}}$

$\mu_1 = \text{Coefft of } \frac{t}{1!} = 0$

$\mu_2 = \text{Coefft of } t^2/2! = \sigma^2$

$\mu_3 = \text{Coefft of } t^3/3! = 0$

$\mu_4 = \text{Coefft of } t^4/4! = 3\sigma^4$

Mean = 0 ; Var = σ^2

Example :-

Normal distribution with mean $\mu = 20$ & S.D = $\sigma = 10$, Find $P(15 \le x \le 40)$

Solution :-

$\mu = 20$, $\sigma = 10$

$z = \frac{x-\mu}{\sigma} = \frac{x-20}{10}$

$\downarrow$

When $x = 15 \Rightarrow z = -0.5$

$x = 40 \Rightarrow z = 2$

$\downarrow$

$P(15 \le x \le 40) = P(-0.5 \le z \le 2)$

$= P(-0.5 \le z \le 0)$

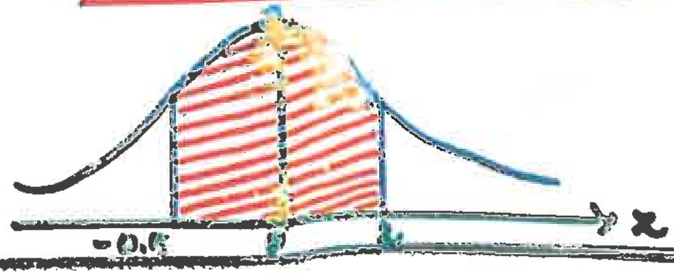
$+ P(0 \le z \le 2)$

$= 0.1915 + 0.4772$

$= 0.6687$

Solution :-

$P(15 \le x \le 40) = 0.6687$



Example :-

Mean height of Soldiers - 68.22 inch with variance 10.8 inch. How many soldiers of 1000 would be expected to be over 6 feet tall.

Solution

$\mu = 68.22$, $\sigma^2 = 10.8$, $\sigma = 3.286$

$\downarrow$

$z = \frac{x-\mu}{\sigma} = \frac{x - 68.22}{3.286}$

P(height of Soldiers)

$= P(x > 6)$ (in feet) $= P(x > 72)$ (In Inches)

$\downarrow$

$z = \frac{72 - 68.22}{3.286}$

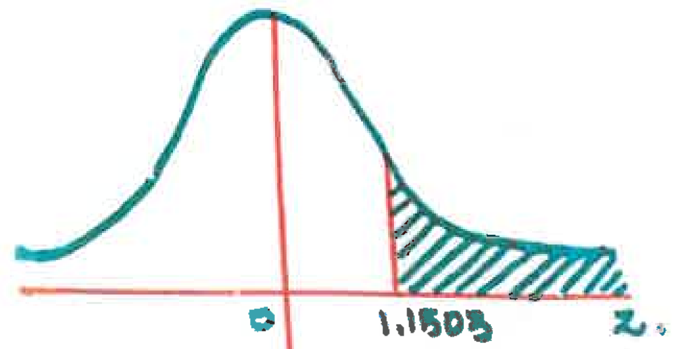
$z = 1.1503$

$\downarrow$

For 1000 Soldiers ;

0.1251×1000

= 125 Soldiers.



LARGE SAMPLE : MEAN:

Step-1 : $\boxed{\begin{array}{l} H_0: \bar{x} = \mu \\ H_1: \bar{x} \neq \mu \end{array}}$

(or)

$\boxed{\begin{array}{l} \bar{x} > \mu \\ \bar{x} < \mu \end{array}}$

Step-2 : $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$

Step-3 : Find tab(z_α)

where α is level of significance.

Step-4 : $\boxed{\text{If cal}(z) < \text{tab}(z_\alpha)}$

then we accept H0

otherwise reject H0.

Step-5 : $\boxed{\text{CONCLUSION :}}$

As for the Given problem:

PROBLEMS:

1) A sample of 100 students is Taken from a large population. The Mean height of the students in this sample is 160 cm. Can it be reasonably regarded that in the population, the mean height is 165 cm, and The S.D is 10cm? [Test if At 1% level of significance]

GIVEN : Sample mean $\bar{x} = 160$

Population mean $\mu = 165$ S.D $= \sigma = 10$, $n = 100$

STEP-1 : $\boxed{\begin{array}{l} H_0: \bar{x} = \mu \\ H_1: \bar{x} \neq \mu \end{array}}$ (two-tailed test)

Step-2 : $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}} = \dfrac{160 - 165}{10/\sqrt{100}} = -5$

$\therefore \boxed{|z| = |-5| = 5}$

Step-3 : $\boxed{z_{1\%} = z_{0.01} = 2.58}$

Step-4 : $\boxed{\text{cal}(z) > \text{tab}(z_\alpha)}$

Step-5 : $\boxed{\text{CONCLUSION}}$: $\boxed{\text{REJECT H0}}$.

2) The Mean breaking strength of the cables supplied by a manufacture -re is 1800 with a S.D of 100. By a New Technique in the Manufacturing Process. If is claimed that the Breaking strength of the cable has increased. In order to test the claim a sample of 50 cables is tested and it is found that the mean breaking strength is 1850. Can we support the claim at 1% level of significance.

GIVEN : Sample mean $\bar{x} = 1850$

Sample size $n = 50$ population mean $\mu = 1800$ and S.D $= \sigma = 100$

Step-1 : $\boxed{H_0: \bar{x} - \mu \quad H_1: \bar{x} > \mu}$

Step-2 : $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}} = \dfrac{1850 - 1800}{100/\sqrt{50}} = 3.54}$

Step-3 : $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}}$

Step-4 : $\boxed{\text{cal}(z) > \text{tab}(z_{0.01})}$

Step-5 : $\boxed{\text{REJECT H0}}$

3) An automatic Machine fills in tea in sealed tins with mean weight of tea 1 kg and S.D. 1 gm. A random sample of 50 tins was examined and it was found that their mean weight was 999.50gms. Is the machine working properly?

Step-1 : $\boxed{\begin{array}{l} H_0: \mu = 1\,kg \\ H_1: \mu \neq 1\,kg \end{array}}$

Step-2 : $\boxed{z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}} = \dfrac{999.50 - 1000}{1/\sqrt{50}} = -3.54$

Step-3 : $\boxed{z_{1\%} = z_{0.01} = 2.58}$

Step-4 : $\boxed{\text{cal}(z) > \text{tab}(z_\alpha)}$

Step-5 : $\boxed{\text{REJECTED H0}}$

LARGE SAMPLE - DIFFERENCE OF MEANS

i) If the samples are drawn from the same population (i.e) $[\sigma_1 = \sigma_2 = \sigma]$ then,

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

ii) If σ_1 & σ_2 are not known then

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

iii) If the samples drawn from two normal population with same S.D then

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

Problem: 1 In a random sample of size 500, the mean is found to be 20. If another independent sample of size 400, the mean is 15, Could the sample have been drawn from the same population with S.D is 4? [At 1% level of significance]

Gn: $\bar{x}_1 = 20$ $n_1 = 500$ $\sigma = 4$
$\bar{x}_2 = 15$ $n_2 = 400$

STEP 1: $H_0 : \bar{x}_1 = \bar{x}_2$
$H_1 : \bar{x}_1 \neq \bar{x}_2$ [TWO TAILED TEST]

STEP 2: $\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}} = \frac{20-15}{4\sqrt{1/500 + 1/400}} = 18.6$

$|Z| = 18.6$

STEP 3: $tab(Z_{1\%}) = tab(Z_{0.01}) = 2.58$

STEP 4: $Z_{cal} > Z_{tab}(1\%)$

Reject H_0

Problem: 2 A simple sample of heights of 6400 Englishmen has a mean of 170 cm & a S.D of 6.4 cm, while a simple sample of heights of 1600 Americans has a mean of 172 cm & S.D of 6.3 cm. Do the Data indicate that Americans are, on the average taller than Englishmen?

Gn: $n_1 = 6400$ $\bar{x}_1 = 170$ $s_1 = 6.4$
$n_2 = 1600$ $\bar{x}_2 = 172$ $s_2 = 6.3$

STEP 1: $H_0 : x_1 = x_2$ (or) $\mu_1 = \mu_2$ [LEFT TAILED]
$H_1 : x_1 < x_2$ (or) $\mu_1 < \mu_2$

STEP 2: $\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}} \Rightarrow \frac{170-172}{\sqrt{\frac{(6.4)^2}{6400} + \frac{(6.3)^2}{1600}}}$

$|Z| = |-11.32| = 11.32$

STEP 3: $tab(Z_\alpha) = tab(Z_{5\%}) = 1.645$

STEP 4: $Z_{cal} > Z_{tab}(5\%)$

Reject H_0

Problem: 3 Test the significance of the difference b/w the means of the samples, drawn from the normal populations with the same S.D from the following data.

	Size	Mean	S.D
Sample 1	100	61	4
Sample 2	200	63	6

STEP 1: $H_0 = \bar{x}_1 = \bar{x}_2$ (or) $\mu_1 = \mu_2$
$H_1 = \bar{x}_1 \neq \bar{x}_2$ (or) $\mu_1 \neq \mu_2$

STEP 2: $\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_2} + \frac{s_2^2}{n_1}}}} = \frac{61-63}{\sqrt{\frac{16}{200} + \frac{36}{100}}}$

$|Z| = 3.02$

STEP 3: $tab(Z_\alpha) = tab(Z_{5\%}) = 1.96$

STEP 4: $Z_{cal} > Z_{tab}(5\%)$

Reject H_0

LARGE SAMPLE TEST

SINGLE PROPORTION

STEP 1:
$H_0 : p = P$
$H_1 : p \neq P$ or $(p < P)$ or $(p > P)$

STEP 2:
$z = p - P / \sqrt{\frac{PQ}{n}}$
$Q = 1 - P$

STEP 3:
Find, $Z_{tab} = Z_\alpha$
α is level of Significance

STEP 4:
$|Z_{cal}| < |Z_{tab}|$ we accept H_0 otherwise reject H_0

STEP 5:
conclusion

TABLE VALUE FOR Z-TEST

TEST\LOS	1% (0.01)	5% (0.05)	10% (0.1)
Two tailed	2.58	1.96	1.645
Right tailed	2.33	1.645	1.28
Left tailed	-2.33	-1.645	-1.28

PROBLEM 1: The fatality rate of typhoid patients believed to be 17.26%. In a certain year 640 patients suffering from typhoid were treated in a hospital and only 63 patients died. Can you consider the hospital efficient.

$p = 63/640 = 0.984$, $P = 0.1726$
$Q = 1 - P = 0.8274$

STEP 1 $H_0 : p = P$, $H_1 : p < P$ — ONE TAILED LEFT

STEP 2 $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.0984 - 0.1726}{\sqrt{\frac{0.1726 \times 0.8274}{640}}}$

$z = -4.96$
$|z| = 4.96$

STEP 3 $Z_{1\%} = \cancel{Z_{0.01}} = -2.33$
$|Z_\alpha| = 2.33$

STEP 4 $|Z_{cal}| > |Z_{\alpha tab}|$

REJECT H_0

PROBLEM 2: A Salesman in a departmental store claims that atmost 60 percent of the shoppers entering the store leaves without making a purchase. A random sample of 50 shoppers showed that 35 of them left without making a purchase. Are these sample results consistent with the claim of the sales man? use a level of significance of 0.05

$p = 35/50 = 0.7$ $P = 60\% = 0.6$
$Q = 1 - P = 0.4$

STEP 1 $H_0 : P = P$, $H_1 : p > P$ — RIGHT TAILED TEST

STEP 2 $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{0.7 - 0.6}{\sqrt{\frac{0.6 \times 0.4}{50}}} = 1.443$

$|z| = 1.443$

STEP 3 $Z_{tab} = 1.645$ — TWO TAIL TEST

STEP 4 $|Z_{cal}| < |Z_{tab}|$

ACCEPT H_0

PROBLEM 3: Experience has shown that 20% of a manufactured product is of top quality. In one day's production of 400 articles only 50 are of top quality. Show that either the production of the day chosen was not a representative sample or the hypothesis of 20% was wrong. Based on the particular day's production

$P = 20\% = 1/5$ $p = 50/400 = 1/8$
$Q = 1 - P = 4/5$

STEP 1 $H_0 : P = P$, $H_1 : p \neq P$ — TWO TAILED TEST

STEP 2 $z = \frac{p - P}{\sqrt{\frac{PQ}{n}}} = \frac{1/8 - 1/5}{\sqrt{\frac{1/5 \times 4/5}{400}}} = -3.75$

$|z| = 3.75$

STEP 3: $Z_{tab}^{0.05} = 1.96$

STEP 4 $|Z_{cal}| > |Z_{\alpha tab}|$

REJECT H_0

DIFFERENCE OF PROPORTIONS

STEP 1

$H_0 : P_1 = P_2$

$H_1 : P_1 \neq P_2 / P_1 > P_2 / P_1 < P_2$

STEP 2

$$Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$$

Where $P = \frac{P_1 n_1 + P_2 n_2}{n_1 + n_2}$

STEP 3

To find Z_α,

α = level of significance

STEP 4 If $|Z| < |Z_\alpha|$ then we accept H_0

Otherwise reject H_0

Conclusion - As for the given problem.

PROBLEM: 1 In a large City A 20% of a random sample of 900 school boys had a slight physical defect. In another large city B, 18.5% of a random sample of 1600 school boys had the same defect. Is the difference b/w the proportions significant?

Gn: $P_1 = 0.2$ $n_1 = 900$

$P_2 = 0.185$ $n_2 = 1600$

step 1: $H_0 : P_1 = P_2$ [TWO TAILED]

$H_1 : P_1 \neq P_2$

step 2: $Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

Whr $P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = \frac{180+296}{900+1600}$

$= 0.1904$

$Q = 1 - P = 0.8096$

$$Z = \frac{0.2 - 0.185}{\sqrt{0.1904 \times 0.8096\left(\frac{1}{900}+\frac{1}{1600}\right)}}$$

$Z_{cal} = 0.92$

step: 3 $Z_{tab} = Z_{0.05} = 1.96$

step: 4 $Z_{(cal)} < Z_{(tab)}$

Accept H_0

PROBLEM: 2 15.5% of a random sample of 1600 undergraduates were smokers, whereas 20% of a random sample of 900 postgraduates were smokers in a state. Can we conclude that less number of undergraduate are smokers than the postgraduates?

Gn: $P_1 = 15.5\%$ $n_1 = 1600$

$P_2 = 20\%$ $n_2 = 900$

step 1: $H_0 : P_1 = P_2$ [LEFT TAILED]

$H_1 : P_1 < P_2$

step 2: $Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = 0.1712$

$Q = 1 - P = 0.8288$

$$Z = \frac{0.155 - 0.2}{\sqrt{0.1712 \times 0.8288\left(\frac{1}{1600}+\frac{1}{900}\right)}}$$

$Z_{cal} = 2.87$

step 3: $Z_{tab} = Z_{0.05} = 1.645$

step 4: $Z_{cal} > Z_{tab}$

Reject H_0

PROBLEM: 3 Before an increase in exercise duty on tea, 800 people out of a sample of 1000 were customers of tea. After the increase in duty, 800 people were consumed of tea in a sample of 1200 persons. Find whether there is significant decrease in the consumption of tea after the increase in duty at 1% level of significance?

step 1: $H_0 : P_1 = P_2$ [Right tailed]

$H_1 : P_1 > P_2$

step 2: $Z = \frac{P_1 - P_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$

$P = \frac{n_1 P_1 + n_2 P_2}{n_1 + n_2} = 0.7273$

$Q = 1 - P = 0.2727$

$$Z = \frac{0.8 - 0.67}{\sqrt{0.7273 \times 0.2727\left(\frac{1}{1000}+\frac{1}{1200}\right)}}$$

$Z_{cal} = 6.82$

step 3: $Z_{tab} = Z_{0.01} = 2.33$

step 4: $Z_{cal} > Z_{tab}$

Reject H_0

1. $H_0 = \bar{x} = \mu$

$H_1 = \bar{x} \neq \mu$ (or) $\bar{x} > \mu$, $\bar{x} < \mu$

2. $t = \dfrac{\bar{x}-\mu}{S/\sqrt{n-1}}$, S is the SD of the given Sample.

3. $t_{tab} = t_{\alpha,v} = t_{5\%}$ n-1 df

4. If $t_{cal} < t_{tab}$ then we accept Ho otherwise we reject.

5. Conclusion.

Problems:

1. The mean life time of a Sample of 25 bulbs is found as 1550 hour with S.D of 120 hours. The company manufacturing the bulbs is 1600 hours. Is the claim acceptable at 5% level of Significance?

Solution: Sample mean : $\bar{x} = 1550$

Sample S.D : $S = 120$

Sample size : $n = 25$

$H_0 : \bar{x} = \mu$, $H_1 : \bar{x} < \mu$

$t = \dfrac{\bar{x}-\mu}{S/\sqrt{n-1}} = \dfrac{1550-1600}{120/\sqrt{24}}$

$|t| = 2.04$, $V = n-1 = 24$

t-Test for Single Mean

tab ($t_{5\%}$) for one-tailed test

= tab ($t_{10\%}$) for two-tailed test

(ie) tab 10% at 24 = 1.71

Cal (t) > tab (t)

Reject Ho.

2. A certain injection admitted to each of 12 patterns resulted in the following increases of blood pressure 5, 2, 8, -1, 3, 0, 6, -2, 1, 5, 0, 4 can it be concluded that the injection will be general, accompanied by increase in Bp

Solution: $\bar{x} = \Sigma x/n = 31/12 = 2.58$

$S^2 = \Sigma x^2/n - \bar{x}^2 = 8.76$, $S = 2.91$

$H_0 : \bar{x} = \mu$ $\quad t = \dfrac{\bar{x}-\mu}{S/\sqrt{n-1}} = \dfrac{2.58-0}{2.91/\sqrt{11}}$

$H_1 : \bar{x} > \mu$,

$|t| = 2.89$

tab ($t_{5\%}$) for right-tailed test

= tab ($t_{10\%}$) for two-tailed test

(ie) (tab (t)) at 10% level of Signifi-cance

for V = 11 ⇒ 1.80

cal (t) > tab (t)

Reject Ho.

3. A machine is designed to produce insulating washers for electrical devices of Average thickness of 0.025cm. A random Sample of 10 washers was found to have an average thickness of 0.024 cm with a S.D of 0.002cm. Test the significance of the deviation. Value of t for 9 degrees of freedom at 5% level is 2.262.

Solution: $\mu = 0.025$cm, $n = 10$, $\bar{x} = 0.024$cm, $S = 0.002$

$n = 10 < 30$, the Sample is Small. We can apply t-test for testing the mean.

$H_0 : \mu = 0.025$, $H_1 : \mu \neq 0.025$

The test Statistic is

$t = \dfrac{\bar{x}-\mu}{S/\sqrt{n-1}} = \dfrac{0.024-0.025}{0.002/\sqrt{9}} = -1.5$

$|t| = 1.5$

ndf $= n-1 = 10-1 = 9$

Table value of t for 9df is 5% level is 2.262.

Ho is accepted at 5% level since the Calculated value of |t| is less than the table value.

∴ The deviation is not Significant.

t - TEST OF SIGNIFICANCE FOR THE DIFFERENCE BETWEEN TWO MEANS

PROCEDURE

STEP 1:

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 \neq \bar{x}_2$ (or)

$\bar{x}_1 < \bar{x}_2$ (or) $\bar{x}_1 > \bar{x}_2$

STEP 2:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

Where, $\sigma^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

STEP 3: $t_{tab} = t_{(\alpha, v)}$

where $\alpha = 5\%$ / $v = n_1 + n_2 - 2$

STEP 4:

If $t_{cal} < t_{tab}$ then we accept H_0 otherwise reject H_0.

STEP 5:

Conclusion as per the problem

PROBLEM 1:

Sample of two types of electric bulbs were tested for length of life and the following data

	size	Mean	S.D
Sample 1	8	1234 hrs	36 hrs
Sample 2	7	1036 hrs	40 hrs

Is the difference in the means sufficient to warrant that type-I bulb superior to type-II

Gn: $\bar{x}_1 = 1234$ $s_1 = 36$ $n_1 = 8$

$\bar{x}_2 = 1036$ $s_2 = 40$ $n_2 = 7$

STEP 1

$H_0 : \bar{x}_1 = \bar{x}_2$

$H_1 : \bar{x}_1 > \bar{x}_2$

STEP 2

$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$, $\sigma^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

$$\sigma^2 = \frac{8(36)^2 + 7(40)^2}{13} = \frac{21568}{13} = 1659.07$$

$$t = \frac{1234 - 1036}{40.73\sqrt{1/8 + 1/7}} = \frac{198}{21.082} = 9.392$$

STEP 3 $t_{tab} = t_{(\alpha, v)} = t_{(10, 13)} = 1.77$

STEP 4 $t_{cal} > t_{tab}$ **Reject H_0**

PROBLEM 2:

Two horses A & B were tested according to time (in seconds) to run a particular race with the following results.

A	28	30	32	33	33	29	34
B	29	30	30	24	27	29	

Test whether horse A is running faster than B at 5% level.

Gn: $n_1 = 7$ $n_2 = 6$

STEP 1:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 > \mu_2$

STEP 2:

$t = \dfrac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{1/n_1 + 1/n_2}}$ $\quad \sigma^2 = \dfrac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2}$

$\bar{x}_1 = A + \dfrac{\Sigma d_1}{n_1} = 33 - 12/7$

$= 31.29$

$d_1 = x_1 - A$

$d_2 = x_2 - B$

$\bar{x}_2 = B + \dfrac{\Sigma d_2}{n_2} = 30 - 1/6 = 28.17$

$$\sigma^2 = \frac{(7 \times 4.50) + (6 \times 4.48)}{11} = 5.31$$

$$t = \frac{31.29 - 28.17}{2.30\sqrt{1/7 + 1/6}} = 2.49$$

STEP 3: $t_{tab} = t_{(\alpha, v)} = t_{(5, 11)} = 1.796$

STEP 4: $t_{cal} > t_{tab}$ **Reject H_0**

Hence B runs faster than A

SMALL SAMPLE TEST

F-test for Significance

DIFFERENCE b/w TWO POPULATION VARIANCES:

STEP 1:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

STEP 2:

$$F_{cal} = \frac{\sigma_1^2}{\sigma_2^2} \text{ if } \sigma_1^2 > \sigma_2^2$$

(OR)

$$F_{cal} = \frac{\sigma_2^2}{\sigma_1^2} \text{ if } \sigma_2^2 > \sigma_1^2$$

where σ_1^2 & σ_2^2 are population variances and n_1 & n_2 are sample sizes.

STEP 3:

$$F_{tab} = F_\alpha(v_1, v_2) \text{ if } s_1^2 > s_2^2$$

(OR)

$$F_{tab} = F_\alpha(v_2, v_1) \text{ if } s_2^2 > s_1^2$$

where $v_1 = n_1 - 1$ $\quad v_2 = n_2 - 1$

Step 4: If $F_{cal} > F_{tab}$ then accept Ho otherwise we reject Ho

---

PROBLEM 1: A sample of size 13 gave an estimated population variance of 3.0, while another sample of size 15 gave an estimated population variance of 2.5. Could both samples be from populations with same variances?

Given: $n_1 = 13$ $\quad \sigma_1^2 = 3.0$

$n_2 = 15$ $\quad \sigma_2^2 = 2.5$

Step 1:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

Step 2:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{3}{2.5} = 1.2$$

Step 3:

$v_1 = n_1 - 1 = 13 - 1 = 12$

$v_2 = n_2 - 1 = 15 - 1 = 14$

$$F_{tab} = F_{5\%(12,14)} = 2.53$$

Step 4:

$F_{cal} < F_{tab}$

Accept Ho

Step 5: Both samples are came from same variances

PROBLEM 2: Two samples of size 9 and 8 gave the sum of squares of deviations from their respective means equal to 160 and 91 respectively. Can they be regarded as drawn from the same normal population?

Given: $n_1 = 9$ $\quad \Sigma(x-\bar{x})^2 = 160$

$n_2 = 8$ $\quad \Sigma(y-\bar{y})^2 = 91$

$$\sigma_1^2 = \frac{n_1 s_1^2}{n_1 - 1} = \frac{160}{8} = 20$$

$$\sigma_2^2 = \frac{n_2 s_2^2}{n_2 - 1} = \frac{91}{7} = 13$$

Step 1:

$H_0 : \sigma_1^2 = \sigma_2^2$

$H_1 : \sigma_1^2 \neq \sigma_2^2$

Step 2:

$$F = \frac{\sigma_1^2}{\sigma_2^2} = \frac{20}{13} = 1.54$$

Step 3:

$$F_{tab} = F_{5\%(8,7)} = 3.73$$

Step 4: cal(F) < tab(F)

Accept Ho

Step 5: Both samples came from same normal population.

χ^2 - TEST

Goodness Of Fit

Test Procedure

Step 1 :- H_0 : Null Hypothesis
H_1 : Alternate Hypothesis

Step 2 :- Calculate theoretical frequency

Step 3 :- Test statistic

$$\chi^2 = \Sigma \left(\frac{(O_i - E_i)^2}{E_i} \right)$$

Step 4 :- Degrees of freedom = n - 1

Step 5 :- Compute $\chi^2_{(tab)}$ at α %

Step 6 :- Compare cal(χ^2) & Tab(χ^2)

Step 7 :- If cal(χ^2) < tab(χ^2) accept H_0 otherwise reject H_0

Step 8 :- Draw the conclusion from cal(χ^2) & tab(χ^2)

Problem 1 :-

5 coins are tossed 256 times whose observed frequency is as follows. Examine the goodness of fit.

No of Heads	0	1	2	3	4	5
Frequency	5	35	75	84	45	12

Solution :-

Step 1 :- H_0 : Binomial is a good fit
H_1 : Binomial is not a good fit

Step 2 :- $\alpha = 5\%$ df = 6 - 1 = 5

Step 3 :- Theoretical frequencies are $N(q+p)^n$

$= 256\left(\frac{1}{2}+\frac{1}{2}\right)^5$

$= 256[{}^5C_0 + {}^5C_1 + {}^5C_2 + {}^5C_3 + {}^5C_4 + {}^5C_5]$

$= 8(1+5+10+10+5+1)$

Theoretical frequencies are :- 8, 40, 80, 80, 40, 8

Step 4 :-

O_i :	5	35	75	84	45	12
E_i :	8	40	80	80	40	8
$\frac{(O_i-E_i)^2}{E_i}$:	1.25	0.625	0.312	0.2	0.63	2

Step 5 :- χ^2(tab) = 11.07

Step 6 :- cal(χ^2) < tab(χ^2). Accept H_0

Problem 2 :-

The theory predicts that the proportion of beans in 4 given groups should be 9:3:3:1. In an examination with beans the no's in the 4 groups were 882, 313, 287 and 118. Does the experimental result support the theory.

Solution :-

Step 1 :- H_0 : 4 groups are in the ratio 9:3:3:1
H_1 : 4 groups are not in the ratio 9:3:3:1

Step 2 :-

O_i :	882	313	287	118	Total
E_i :	900	300	300	100	
$\frac{(O_i-E_i)^2}{E_i}$:	0.36	0.563	0.563	3.24	4.276

Step 3 :- cal$(\chi^2) = \frac{(O_i-E_i)^2}{E_i} = 4.726$

Step 4 :- tab(χ^2) = 7.81 at 5% level with df = 3

Step 5 :- cal(χ^2) < tab(χ^2)

Step 6 :- Accept H_0

Step 7 :- Hence the 4 groups in the ratio 9:3:3:1

χ^2 TEST - INDEPENDENCE OF ATTRIBUTES

PROBLEM NO: 1

Find if there is any association between extravagance in fathers and extravagance in sons from the following data.

	Extravagant Father	Miserly Father
Extravagant son	327	741
Miserly son	545	234

Determine the coefficient of association also.

SOLUTION: Here $a = 327$, $b = 741$, $c = 545$, $d = 234$

1. H_0: Namely that the extravagance in sons and father are not significant
2. H_1: Significant
3. $\alpha = 0.05$, $d.f = (r-1)(s-1) = (2-1)(2-1) = 1$
4. Table value of χ^2 : 3.841
5. The test statistic is $\chi^2 = \dfrac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$

i.e $\chi^2 = \dfrac{[(327)(234)-(741)(545)]^2 \times (327+741+545+234)}{(872)(975)(1068)(779)} = 279.77$

6. Conclusion: Here, Cal χ^2 > table χ^2 i.e, $279.77 > 3.841$

So, we reject H_0 at 5% level of significance

$\therefore$ There is dependence between the attributes.

7. Coefficient of attribute $= \dfrac{ad-bc}{ad+bc} = \dfrac{-327327}{480363} = -0.6814$

PROBLEM NO: 2

Two sample polls of votes for two candidates A and B for a public office are taken one from among residents of rural areas. The results are given below. Examine whether the nature of the area is related to voting preference in the election.

Area / Votes for	A	B	Total
Rural	620	380	1000
Urban	550	450	1000
Total	1170	830	2000

SOLUTION: Here, $a = 620$, $b = 380$, $c = 550$, $d = 450$

1. H_0: The nature area is independent of voting preference in the election
2. H_1: dependent
3. $\alpha = 0.05$, $d.f = (r-1)(s-1) = (2-1)(2-1) = 1$
4. Table value of χ^2 : 5.991
5. The test statistic is $\chi^2 = \dfrac{(ad-bc)^2(a+b+c+d)}{(a+b)(c+d)(a+c)(b+d)}$

i.e, $\chi^2 = \dfrac{(620 \times 450 - 380 \times 550)^2 (620+450+380+550)}{(620+380)(550+450)(620+550)(380+450)} = 10.09$

6. Conclusion: Here, Cal χ^2 > table χ^2

i.e, $10.09 > 5.991$

So, we reject H_0 at 5% level of significance

SIMATS
SCHOOL OF ENGINEERING

Engineer to Excel

Approved by AICTE | IET-UK Accreditation

Saveetha Nagar, Thandalam, Chennai - 602 105, TamilNadu, India