PROBABILITY
&
QUEUEING THEORY

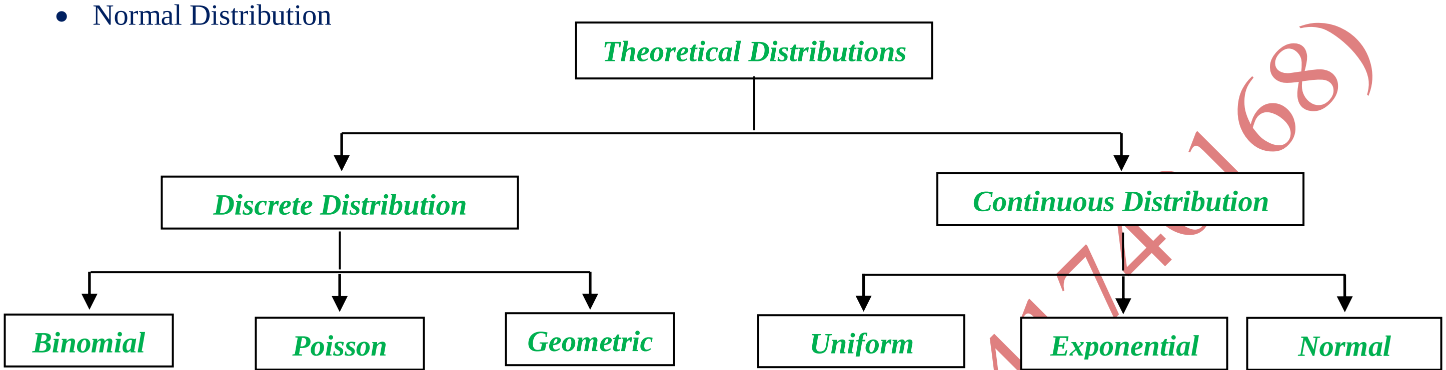
(As per SRM INSTITUTE OF SCIENCE AND TECHNOLOGY Syllabus)

Dr. P. GODHANDARAMAN
M.Sc., M.A., M.B.A., Ph.D.
Assistant Professor (Senior Grade)
Department of Mathematics
SRM IST, Kattankulathur – 603 203
Email : godhanda@gmail.com
Mobile : 9941740168

UNIT – II - THEORETICAL DISTRIBUTIONS

Syllabus

- Binomial Distribution
- Poisson Distribution
- Geometric Distribution
- Uniform Distribution
- Exponential Distribution
- Normal Distribution



DISCRETE DISTRIBUTION				
Discrete	**Probability Mass Function (p.m.f.)**	**Moment Generating Function (m.g.f.) $M_X(t)$**	**Mean $E(X)$**	**Variance $V(X)$**
Binomial	$P(X=x)=nC_x p^x q^{n-x},\ \ x=0,1,\dots n$	$(q+pe^t)^n$	np	npq
Poisson	$P(X=x)=\dfrac{e^{-\lambda}\,\lambda^x}{x!},\ \ x=0,1,\dots\infty$	$e^{\lambda(e^t-1)}$	λ	λ
Geometric	$P(X=x)=p\,q^{x-1},\ \ x=1,\dots\infty$	$\dfrac{p\,e^t}{1-q\,e^t}$	$\dfrac{1}{p}$	$\dfrac{q}{p^2}$

CONTINOUS DISTRIBUTION				
Continuous	**Probability Density Function (p.d.f.)**	**Moment Generating Function (m.g.f.) $M_X(t)$**	**Mean $E(X)$**	**Variance $V(X)$**
Uniform	$f(x)=\dfrac{1}{b-a},\quad a<x<b$	$\dfrac{e^{bt}-e^{at}}{t(b-a)}$	$\dfrac{b+a}{2}$	$\dfrac{(b-a)^2}{12}$
Exponential	$f(x)=\lambda\,e^{-\lambda x},\quad x\ge 0$	$\dfrac{\lambda}{\lambda-t}$	$\dfrac{1}{\lambda}$	$\dfrac{1}{\lambda^2}$
Normal	$f(x)=\frac{1}{\sigma\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$, $-\infty<x<\infty,-\infty<\mu<\infty,\sigma>0$	$e^{\left(\mu t+\frac{t^2\sigma^2}{2}\right)}$	μ	σ^2

BINOMIAL DISTRIBUTION

Bernoulli Distribution **:** A Bernoulli distribution is one having the following properties

(i) The experiment consists of n repeated trials.

(ii) Each trial results in an outcome that may be classified under two mutually exclusive categories as a success or as a failure.

(iii) The probability of success denoted by p, remains constant from trial to trial.

(iv) The repeated trials are independent.

Binomial Distribution**:**

A Bernoulli trial can result in a success with probability p and a failure with probability $q=1-p$. Then the probability distribution of the binomial random variable X, the number of successes in n independent trials

$$\boldsymbol{P(X=x)=nC_x p^x q^{n-x},\ \ x=0,1,\dots n}$$

The quantities n & p are called the parameters of binomial distribution.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Areas of Application:

1. Quality control measures and Sampling processes in industries to classify items as defective or non defective.
2. Medical applications as success or failure of a surgery, cure or no cure of a patient.

Moment Generating Function (m.g.f.) in Binomial Distribution

$$M_X(t) = \sum_{x=0}^{n} e^{tx}\, p(x) = \sum_{x=0}^{n} e^{tx}\, nC_x p^x q^{n-x} = \sum_{x=0}^{n} nC_x (pe^t)^x q^{n-x}$$
$$= nC_0 (pe^t)^0 q^{n-0} + nC_1 (pe^t)^1 q^{n-1} + nC_2 (pe^t)^2 q^{n-2} + \cdots + nC_n (pe^t)^n q^{n-n}$$
$$= q^n + nC_1 (pe^t)^1 q^{n-1} + nC_2 (pe^t)^2 q^{n-2} + \cdots + (pe^t)^n = \boldsymbol{(q + pe^t)^n}$$

Mean and Variance using Moment Generating Function in Binomial Distribution

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[\frac{d}{dt}(q + pe^t)^n\right]_{t=0} = [n(q + pe^t)^{n-1}\, pe^t]_{t=0} = n(q + pe^0)^{n-1}\, pe^0 = \boldsymbol{np}$$
$$E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = \left[\frac{d}{dt} np\, (q + pe^t)^{n-1}\, e^t\right]_{t=0} = np[(n-1)(q + pe^t)^{n-2}\, pe^t + (q + pe^t)^{n-1} e^t]_{t=0}$$
$$= np[(n-1)(q + pe^0)^{n-2}\, pe^0 + (q + pe^0)^{n-1} e^0] = np[(n-1)p + 1] = \boldsymbol{n^2p^2 - np^2 + np}$$
$$V(X) = E(X^2) - [E(X)]^2 = n^2p^2 - np^2 + np - n^2p^2 = -np^2 + np = np(1-q) = \boldsymbol{npq}$$

Problems in Binomial Distribution

1. ***Four coins are tossed simultaneously. What is the probability of getting 2 heads and at least 2 heads?***

Solution : $n = 4,\ p = \frac{1}{2},\quad q = 1 - p = 1 - \frac{1}{2} = \frac{1}{2},\ P(X = x) = nC_x p^x q^{n-x},\ x = 0, 1, \ldots n.$

(i) $P(2\ heads) = P(X = 2) = 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} = \frac{3}{8}$

(ii) $P(atleast\ 2\ heads) = P(X \geq 2) = P(X = 2) + P(X = 3) + P(X = 4)$

$$= 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} + 4C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{4-3} + 4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4} = \frac{3}{8} + \frac{1}{4} + \frac{1}{16} = \frac{11}{16}$$

2. ***The probability of a bomb hitting a target is $\frac{1}{5}$. Two bombs are enough to destroy a bridge. If six bombs are aimed at the bridge, find the probability that the bridge is destroyed.***

Solution : $n = 6,\ p = \frac{1}{5},\quad q = 1 - p = \frac{4}{5},\ P(X = x) = 6C_x \left(\frac{1}{5}\right)^x \left(\frac{4}{5}\right)^{6-x}$

$$P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)] = 1 - 6C_0 \left(\frac{1}{5}\right)^0 \left(\frac{4}{5}\right)^{6-0} - 6C_1 \left(\frac{1}{5}\right)^1 \left(\frac{4}{5}\right)^{6-1} = 0.3446$$

3. ***The probability that a patient recovers from a disease is 0.3. If 18 people are affected from this disease, What is the probability that (i) At least 10 survive (ii) Exactly 6 survive (iii) 4 to 7 survive***

Solution: $n = 18,\ p = 0.3,\quad q = 1 - p = 1 - 0.3 = 0.7,\ P(X = x) = 18C_x (0.3)^x (0.7)^{18-x}$

(i) $P(X \geq 10) = 1 - P(X < 10) = 1 - [P(0) + P(1) + \cdots + P(9)] = 1 - 0.9790 = 0.021$

(ii) $P(X = 6) = 18C_6 (0.3)^6 (0.7)^{18-6} = 0.1873$

(iii) $P(4\ to\ 7\ survice) = P(X = 4) + P(X = 5) + P(X = 6) + P(X = 7)$

$$= 18C_4 (0.3)^4 (0.7)^{18-4} + 18C_5 (0.3)^5 (0.7)^{18-5} + 18C_6 (0.3)^6 (0.7)^{18-6} + 18C_7 (0.3)^7 (0.7)^{18-7} = 0.6947$$

4. ***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 (ii) At most there are 3 defective (iii) Exactly there are 3 defective bulbs.***

Solution : $n = 20,\ p = 10\% = \frac{10}{100} = 0.1,\quad q = 1 - p = 1 - 0.1 = 0.9,\ P(X = x) = 20C_x (0.1)^x (0.9)^{20-x}$

(i) $P(X = 0) = 20C_0 (0.1)^0 (0.9)^{20-0} = 0.1216$

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

$$= 0.1216 + 20C_1 (0.1)^1 (0.9)^{20-1} + 20C_2 (0.1)^2 (0.9)^{20-2} + 20C_3 (0.1)^3 (0.9)^{20-3} = 0.8671$$

(iii) $P(X = 3) = 20C_3 (0.1)^3 (0.9)^{20-3} = 0.1901$

5. ***It is known that the probability of an item produced by a certain machine will be defective is 5%. If the produced items are sent to the market in packets of 20, find the number of packets containing (i) At least 2 defective items (ii) At most 2 defective items in a consignment of 1000 packets using Binomial distribution.***

Solution : $n = 20,\ p = 5\% = \frac{5}{100} = 0.05,\quad q = 1 - p = 0.95,\ P(X = x) = 20C_x (0.05)^x (0.95)^{20-x}$

(i) $P(X \geq 2) = 1 - P(X < 2) = 1 - [P(X = 0) + P(X = 1)]$

$$= 1 - 20C_0 (0.05)^0 (0.95)^{20-0} - 20C_1 (0.05)^1 (0.95)^{20-1} = 1 - 0.3585 - 0.3774 = 0.2641$$

$N\, P(X \geq 2) = 1000 \times 0.2641 = 264$

(ii) $P(X \leq 2) = P(X = 0) + P(X = 1) + P(X = 2)$

$$= 20C_0 (0.05)^0 (0.95)^{20-0} + 20C_1 (0.05)^1 (0.95)^{20-1} + 20C_2 (0.05)^2 (0.95)^{20-2} = 0.9246$$

$N\, P(X \leq 2) = 1000 \times 0.9246 = 925$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

6. ***A and B play a game in which their chances of winning are in the ratio 3 : 2. Find A's chance of winning at least three games out of the five games played.***

Solution : $n = 5,\ p = \frac{3}{5},\quad q = 1 - p = \frac{2}{5},\ P(X = x) = 5C_x \left(\frac{3}{5}\right)^x \left(\frac{2}{5}\right)^{5-x}$

$$P(X \geq 3) = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$$
$$= 1 - 5C_0 \left(\frac{3}{5}\right)^0 \left(\frac{2}{5}\right)^{5-0} - 5C_1 \left(\frac{3}{5}\right)^1 \left(\frac{2}{5}\right)^{5-1} - 5C_2 \left(\frac{3}{5}\right)^2 \left(\frac{2}{5}\right)^{5-2}$$
$$= 1 - 0.01024 - 0.0768 - 0.2304 = 0.68$$

7. ***A multiple choice test consists of 8 questions with 3 answers to each question (of which only one is correct). A student answers each question by rolling a balanced die and checking the first answer if he gets 1 or 2, the second answer if he gets 3 or 4 and the third answer if he gets 5 or 6. To get a distinction, the student must secure at least 75% correct answers. If there is no negative making, what is the probability that the student secures a distinction?***

Solution : Since there are three answers to each question, out of which only one is correct, the probability of getting an answer to a question correctly is given by $n = 8,\ p = \frac{1}{3},\quad q = 1 - p = \frac{2}{3},\ P(X = x) = 8C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{8-x}$

The required probability of securing a distinction (i.e. of getting correct answers to al least 6 out of the 8 questions)

$$P(X = 6) + P(X = 7) + P(X = 8) = 8C_6 \left(\frac{1}{3}\right)^6 \left(\frac{2}{3}\right)^{8-6} + 8C_7 \left(\frac{1}{3}\right)^7 \left(\frac{2}{3}\right)^{8-7} + 8C_8 \left(\frac{1}{3}\right)^8 \left(\frac{2}{3}\right)^{8-8} = 0.0197$$

8. ***If on the average rain falls on 10 days in every 30 days, obtain the probability that (i) rain will fall on at least 3 days of a given week (ii) first three days of a given week will be fine and the remaining 4 days wet.***

Solution : $n = 7,\ p = \frac{10}{30} = \frac{1}{3},\quad q = 1 - p = \frac{2}{3},\ P(X = x) = 7C_x \left(\frac{1}{3}\right)^x \left(\frac{2}{3}\right)^{7-x}$

(i) $P(X \geq 3) = 1 - P(X < 3) = 1 - [P(X = 0) + P(X = 1) + P(X = 2)]$
$$= 1 - 7C_0 \left(\frac{1}{3}\right)^0 \left(\frac{2}{3}\right)^{7-0} - 7C_1 \left(\frac{1}{3}\right)^1 \left(\frac{2}{3}\right)^{7-1} - 7C_2 \left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^{7-2}$$
$$= 1 - 0.0585 - 0.2048 - 0.3073 = 0.4293$$

(ii) Probability that the first 3 days are fine and the remaining days wet $= q^3\, p^4 = \left(\frac{2}{3}\right)^3 \left(\frac{1}{3}\right)^4 = 0.0037$

9. ***Out of 800 families with 4 children each, how many families would be expected to have (i) 2 boys and 2 girls (ii) At least 1 boy (iii) At most 2 girls (iv) Children of both genders. Assume equal prob. for boys and girls.***

Solution : Considering each child as a trial $n = 4$. Assuming that birth of a boy is a success, $p = \frac{1}{2},\quad q = \frac{1}{2}$.

Let X denote the no. of successes (boys). $P(X = x) = nC_x p^x q^{n-x},\ \ x = 0,1, \dots n$. $P(X = x) = 4C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{4-x}$

(i) $P(2 \text{ boys and } 2 \text{ girls}) = P(X = 2) = 4C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{4-2} = \frac{3}{8} = 0.375$

800 families having 2 boys and 2 girls $= N\, P(X = 2) = 800 \left(\frac{3}{8}\right) = 300$

(ii) $P(\text{At least 1 boy}) = P(X \geq 1) = 1 - P(X = 0) = 1 - 4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} = 1 - \frac{1}{16} = \frac{15}{16} = 0.9375$

800 families having At least 1 boy $= N\, P(X \geq 1) = 800 \left(\frac{15}{16}\right) = 750$

(iii) $P(\text{At most 2 girls }) = P(\text{exactly } 0\text{ , } 1 \text{ or } 2 \text{ girls}) = P(4) + P(3) + P(2) = 1 - [P(0) + P(1)]$
$$= 1 - 4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} - 4C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{4-1} = \frac{11}{16} = 0.6875$$

800 families having At most 2 girls $= 800 \left(\frac{11}{16}\right) = 550$

(iv) $P(\text{Children of both sexes }) = 1 - P(\text{children of the same sex}) = 1 - [P(\text{all are boys}) + P(\text{all are girls})]$
$$= 1 - P(4) - P(0) = 1 - 4C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{4-4} - 4C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{4-0} = \frac{7}{8} = 0.875$$

800 families having Children of both sexes $= 800 \left(\frac{7}{8}\right) = 700$

10. ***For a binomial distribution the mean is 6 and variance is 2. Find the distribution and find $P(X = 1)$.***

Solution : Mean $= np = 6$, Variance $= npq = 2$, $\frac{npq}{np} = \frac{2}{6} \Rightarrow \boldsymbol{q} = \frac{\mathbf{1}}{\mathbf{3}},\quad p = 1 - q = 1 - \frac{1}{3} = \frac{\mathbf{2}}{\mathbf{3}}$

$np = 6 \ \Rightarrow n = \frac{6}{p} \Rightarrow n = \frac{6}{\left(\frac{2}{3}\right)} = \frac{18}{2} = 9,\ \boldsymbol{n = 9},\ \ \boldsymbol{p} = \frac{\mathbf{2}}{\mathbf{3}},\ \ \boldsymbol{q} = \frac{\mathbf{1}}{\mathbf{3}},\quad P(X = 1) = 9C_1 \left(\frac{2}{3}\right)^1 \left(\frac{1}{3}\right)^{9-1} = 0.0012$

11. ***A Binomial variable X satisfies the relation $9P(X = 4) = P(X = 2)$ when $n = 6$. Find the parameter p.***

Solution : $\quad n = 6,\quad P(X = x) = nC_x p^x q^{n-x},\ \ x = 0,1, \dots n$

$$9P(X = 4) = P(X = 2) \ \Rightarrow\ 9 \times 6C_4 p^4 q^{6-4} = 6C_2 p^2 q^{6-2} \ \Rightarrow\ 9 \times 6C_4 p^4 q^2 = 6C_2 p^2 q^4$$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

$$9 \times 6C_4 p^2 = 6C_2 q^2 \Rightarrow 135p^2 = 15q^2 \Rightarrow 9p^2 - q^2 = 0 \Rightarrow 9p^2 - (1-p)^2 = 0 \Rightarrow 8p^2 + 2p - 1 = 0$$

$$p = -\frac{1}{2}\ (or)\ \frac{1}{4}, \qquad \boldsymbol{p = \frac{1}{4}}\ (\because \boldsymbol{p\ cannot\ be\ negative})$$

12. *A discrete RV X has moment generating function $M_X(t) = \left(\frac{1}{4} + \frac{3}{4}e^t\right)^5$. Find $E(X), Var(X)$ and $P(X = 2)$.*

Solution : $M_X(t) = (q + pe^t)^n$, $p = \frac{3}{4}$, $q = \frac{1}{4}$, $n = 5$, $E(X) = np = \frac{15}{4}$, $V(X) = npq = \frac{15}{16}$

$$P(X = 2) = 5C_2 \left(\frac{3}{4}\right)^2 \left(\frac{1}{4}\right)^{5-2} = \frac{45}{512} = 0.8654$$

13. *Fit a binomial distribution for the following data. Find the parameters of the distribution.*

x	***0***	***1***	***2***	***3***	***4***	***5***	***6***	***Total***
f	***5***	***18***	***28***	***12***	***7***	***6***	***4***	***80***

Solution : Fitting a binomial distribution means assuming that the given distribution is approximately binomial and hence finding the probability mass function and the finding the theoretical frequencies.

To find the binomial frequency distribution $N(q + p)^n$, which fits the given data, we require N, n *and* p.

We assume N = total frequency = 80 and n = no. of trials = 6 from the given data.

x	0	1	2	3	4	5	6	Total
f	5	18	28	12	7	6	4	80
fx	0	18	56	36	28	30	24	192

$$\bar{x} = \frac{\sum fx}{\sum f} = \frac{192}{80} = 2.4\ , np = 2.4 \Rightarrow 6p = 2.4\ ,\ p = 0.4,\ \ q = 1 - p = 0.6$$

Theoretical frequencies are given by N $P(X = x) = N\, nC_x p^x q^{n-x}$, $x = 0,1, \ldots n$

$80\ P(X = 0) = 80 \times 6C_0\ (0.4)^0 (0.6)^{6-0} = 3.73$, $\quad 80\ P(X = 1) = 80 \times 6C_1\ (0.4)^1 (0.6)^{6-1} = 14.93$

$80\ P(X = 2) = 80 \times 6C_2\ (0.4)^2 (0.6)^{6-2} = 24.88$, $80\ P(X = 3) = 80 \times 6C_3\ (0.4)^3 (0.6)^{6-3} = 22.12$

$80\ P(X = 4) = 80 \times 6C_4\ (0.4)^4 (0.6)^{6-4} = 11.06$, $80\ P(X = 5) = 80 \times 6C_5\ (0.4)^5 (0.6)^{6-5} = 2.95$

$80\ P(X = 6) = 80 \times 6C_6\ (0.4)^6 (0.6)^{6-6} = 0.33$, Converting these values into whole numbers consistent with the condition that the total frequency is 80, the corresponding binomial frequency distribution is as follows

X	0	1	2	3	4	5	6	*Total*
Theoretical f	4	15	25	22	11	3	0	*80*

POISSON DISTRIBUTION

Definition : The probability distribution of the Poisson random variable X, representing the number of outcomes occurring in a given time interval or specified region represented as t, is $\boldsymbol{P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}}$, $\boldsymbol{x = 0, 1, \ldots \infty}$**.** where λ is the average number of outcomes per unit time or region.

Poisson distribution as Limiting Form of Binomial Distribution

Poisson distribution is a limiting case of binomial distribution under the following conditions

(i) n, the number of trials is indefinitely large, i.e., $n \to \infty$.

(ii) p, the constant probability of success in each trial is very small, i.e., $p \to 0$.

(iii) $\lambda = np$ is finite or $p = \frac{\lambda}{n}$ and $q = 1 - \frac{\lambda}{n}$, where λ is a positive real number.

Areas of Application:

1. The number of misprints on a page of a book.
2. The number of deaths due to accidents in a month on national highway 47.
3. The number of break downs of a printing machine in a day.
4. The number of vacancies occurring during a year in a particular department.

Moment Generating Function (m.g.f.) in Poisson Distribution

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\, p(x) = \sum_{x=0}^{\infty} e^{tx} \frac{e^{-\lambda}\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} e^{tx} \frac{\lambda^x}{x!} = e^{-\lambda} \sum_{x=0}^{\infty} \frac{(\lambda e^t)^x}{x!} \qquad \left(\because \boldsymbol{e^x = 1 + \frac{x}{1!} + \frac{x^2}{2!} + \cdots}\right)$$

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

Mean and Variance Using Moment Generating Function in Poisson Distribution

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[\frac{d}{dt} e^{\lambda(e^t-1)}\right]_{t=0} = \left[e^{\lambda(e^t-1)}\, \lambda e^t\right]_{t=0} = e^{\lambda(e^0-1)}\, \lambda e^0 = \boldsymbol{\lambda} \qquad (\because \boldsymbol{e^0 = 1})$$

$$E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\ \lambda e^t\, e^{\lambda(e^t-1)}\right]_{t=0} = \lambda\left[e^t\, e^{\lambda(e^t-1)}\, \lambda e^t + e^t\, e^{\lambda(e^t-1)}\right]_{t=0} = \boldsymbol{\lambda(\lambda + 1)}$$

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

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Problems in Poisson distribution

1. ***On an average a typist makes 2 mistakes per page. What is the probability that she will make (i) No errors (ii) 4 or more errors on a particular page?***

Solution : $\lambda = 2$, Let X represent the number of errors on a page. $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0,1, \dots \infty$

(i) $P(X = 0) = \frac{e^{-2}\, 2^0}{0!} = e^{-2} = 0.1353$

(ii) $P(X \geq 4) = 1 - P(X < 4) = 1 - [P(0) + P(1) + P(2) + P(3)] = 1 - \left(\frac{e^{-2}\, 2^0}{0!} + \frac{e^{-2}\, 2^1}{1!} + \frac{e^{-2}\, 2^2}{2!} + \frac{e^{-2}\, 2^3}{3!}\right)$

$= 1 - e^{-2}\left(1 + \frac{2^1}{1!} + \frac{2^2}{2!} + \frac{2^3}{3!}\right) = 0.1434$

2. ***The number of monthly breakdown of a computer is a random variable having a Poisson distribution with mean equal to 1.8. Find the probability that this computer will function for a month (i) Without a breakdown (ii) With only 1 breakdown (iii) With at least 1 breakdown***

Solution : $\lambda = 1.8$, Let X denote the number of breakdown of the computer in a month.

(i) $P(X = 0) = \frac{e^{-1.8}\, (1.8)^0}{0!} = e^{-1.8} = 0.1653$ (ii) $P(X = 1) = \frac{e^{-1.8}\, (1.8)^1}{1!} = e^{-1.8}(1.8) = 0.2975$

(ii) $P(X \geq 1) = 1 - P(X = 0) = 0.8347$

3. ***A travel company has two cars for hiring. The demand for a car on each day is distributed as Poisson variate, with mean 1.5. Calculate the proportion of days on which (i) neither cars were used (ii) some demand is refused.***

Solution : $\lambda = 1.5$, Let X denote the number of demands for cars.

(i) $P(X = 0) = \frac{e^{-1.5}\, (1.5)^0}{0!} = e^{-1.5} = 0.2231$

(ii) $P(X > 2) = 1 - P(X \leq 2) = 1 - [P(0) + P(1) + P(2)] = 1 - \left[\frac{e^{-1.5}\, (1.5)^0}{0!} + \frac{e^{-1.5}\, (1.5)^1}{1!} + \frac{e^{-1.5}\, (1.5)^2}{2!}\right] = 0.191$

4. ***In a book of 520 pages, 390 typographical errors occur. Assuming Poisson law for the number of errors per page, find the probability that a random sample of 5 pages will contain no error.***

Solution : $\lambda = \frac{390}{520} = 0.75$, $P(1\, page\ contains\ no\ error) = P(X = 0) = \frac{e^{-0.75}\, (0.75)^0}{0!} = e^{-0.75}$

$P(5\, pages\ contains\ no\ error) = (e^{-0.75})^5 = 0.0235$

5. ***It is known that the probability of an item produced by a certain machine will be defective is 5%. If the produced items are sent to the market in packets of 20, find the number of packets containing (i) At least 2 defective items (ii) At most 2 defective items in a consignment of 1000 packets using Poisson distribution***

Solution : $p = 5\% = 0.05$, $q = 0.95$, $n = 20$, $\lambda = np = 20 \times \frac{5}{100} = 1$, $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0,1, \dots \infty$

(i) $P(X \geq 2) = 1 - P(X < 2) = 1 - \left[\frac{e^{-1}\, (1)^0}{0!} + \frac{e^{-1}\, (1)^1}{1!}\right] = 1 - [e^{-1} + e^{-1}] = 0.2642$

$NP(X \geq 2) = 1000 \times 0.2642 = 264$

(ii) $P(X \leq 2) = P(0) + P(1) + P(2) = \frac{e^{-1}\, (1)^0}{0!} + \frac{e^{-1}\, (1)^1}{1!} + \frac{e^{-1}\, (1)^2}{2!} = e^{-1} + e^{-1} + \frac{e^{-1}}{2} = 0.9197$

$N\, P(X \leq 2) = 1000 \times 0.9197 = 920$

6. ***A manufacturer of cotter pins knows that 5% of his product is defective. If he sells cotter pins in boxes of 100 and guarantees that not more than 10 pins will be defective. What is approximate probability that a box will fail to meet the guaranteed quality?***

Solution: $p = 5\% = 0.05$, $q = 0.95$, $n = 100$, $\lambda = np = 100 \times \frac{5}{100} = 5$, $P(X = x) = \frac{e^{-\lambda} \lambda^x}{x!}$, $x = 0,1, \dots \infty$

$P(X > 10) = 1 - P(X \leq 10)$

$= 1 - [P(0) + P(1) + P(2) + P(3) + P(4) + P(5) + P(6) + P(7) + P(8) + P(9) + P(10)]$

$= 1 - e^{-5}\left[\frac{5^0}{0!} + \frac{5^1}{1!} + \frac{5^2}{2!} + \frac{5^3}{3!} + \frac{5^4}{4!} + \frac{5^5}{5!} + \frac{5^6}{6!} + \frac{5^7}{7!} + \frac{5^8}{8!} + \frac{5^9}{9!} + \frac{5^{10}}{10!}\right] = 0.014$

7. ***Let X be a RV following Poisson distribution such that $P(X = 2) = 9P(X = 4) + 90P(X = 6)$. Find the mean and standard deviation of X.***

Solution : $P(X = 2) = 9P(X = 4) + 90P(X = 6) \Rightarrow \frac{e^{-\lambda} \lambda^2}{2!} = 9\frac{e^{-\lambda} \lambda^4}{4!} + 90\frac{e^{-\lambda} \lambda^6}{6!}$, *Dividing by* $e^{-\lambda}\ \lambda^2$

$\frac{1}{2!} = \frac{9\,\lambda^2}{4!} + \frac{90\ \lambda^4}{6!} \Rightarrow \frac{1}{2} = \frac{3\,\lambda^2}{8} + \frac{\lambda^4}{8} \Rightarrow \frac{\lambda^4}{8} + \frac{3\,\lambda^2}{8} - \frac{1}{2} = 0$

$\lambda^4 + 3\,\lambda^2 - 4 = 0 \Rightarrow \lambda = 1, -4$, **Mean =** $\boldsymbol{\lambda = 1}$ **, Variance =** $\boldsymbol{\lambda = 1}$ **, S.D. =** $\boldsymbol{\sqrt{Variance} = 1}$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

8. ***Fit a Poisson distribution for the following data:***

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400

Finding the probability mass function and then finding the theoretical frequencies.

Solution :

x	0	1	2	3	4	5	Total
f	142	156	69	27	5	1	400
fx	0	156	138	81	20	5	400

$\lambda = \frac{\sum fx}{\sum f} = \frac{400}{400} = 1,\ N = 400,$ Theoretical frequencies are given by $N\,P(X = x) = \frac{N\,e^{-\lambda}\,\lambda^x}{x!} = \frac{400\,e^{-1}\,1^x}{x!}$

$400\,P(X=0) = \frac{400\,e^{-1}\,1^0}{0!} = 147.15,\ 400\,P(X=1) = \frac{400\,e^{-1}\,1^1}{1!} = 147.15$

$400\,P(X=2) = \frac{400\,e^{-1}\,1^2}{2!} = 73.58\,,\ \ 400\,P(X=3) = \frac{400\,e^{-1}\,1^3}{3!} = 24.53$

$400\,P(X=4) = \frac{400\,e^{-1}\,1^4}{4!} = 6.13\,,\ \ 400\,P(X=5) = \frac{400\,e^{-1}\,1^5}{5!} = 1.23$

x	0	1	2	3	4	5	Total
Theoretical f	147	147	74	25	6	1	400

GEOMETRIC DISTRIBUTION

Definition: If repeated independent trials can result in a success with probability p and a failure with probability $q = 1 - p$, then the probability distribution of the random variable X, the number of trials on which the first success occurs, is $\boldsymbol{P(X = x) = p\,q^{x-1},\ \ x = 1, 2, \ldots \infty}$ **.**

Application: Geometric distribution has important application in queueing theory, related to the number of units which are being served or waiting to be served at any given time.

Moment Generating Function (M.G.F.) in Geometric Distribution

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\,p(x) = \sum_{x=1}^{\infty} e^{tx}\,p\,q^{x-1} = \frac{p}{q}\sum_{x=1}^{\infty}(q\,e^t)^x = \frac{p}{q}[(q\,e^t)^1 + (q\,e^t)^2 + (q\,e^t)^3 + \cdots]$$

$$= \frac{p}{q}(q\,e^t)[1 + (q\,e^t)^1 + (q\,e^t)^2 + \cdots] = pe^t[1 - q\,e^t]^{-1} = \boldsymbol{\frac{pe^t}{1 - q\,e^t}} \qquad [\because (\mathbf{1 - x})^{-1} = \mathbf{1 + x + x^2 + \cdots}]$$

Mean and Variance using Moment Generating Function in Geometric Distribution

$$E(X) = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{pe^t}{1-q\,e^t}\right)\right]_{t=0} = \left[\frac{(1-q\,e^t)pe^t - pe^t(-q\,e^t)}{(1-q\,e^t)^2}\right]_{t=0} \qquad \left[\because \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{vu' - uv'}{v^2}\right]$$

$$= \left[\frac{pe^t - p\,q\,e^{2t} + p\,q\,e^{2t}}{(1-q\,e^t)^2}\right]_{t=0} = \left[\frac{pe^t}{(1-q\,e^t)^2}\right]_{t=0} = \frac{pe^0}{(1-q\,e^0)^2} = \frac{p}{p^2} = \boldsymbol{\frac{1}{p}} \qquad (\because \boldsymbol{p + q = 1})$$

$$E(X^2) = \left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\frac{pe^t}{(1-q\,e^t)^2}\right]_{t=0} = \left[\frac{(1-q\,e^t)^2pe^t - pe^t2(1-q\,e^t)^1(-q\,e^t)}{(1-q\,e^t)^4}\right]_{t=0} \qquad \left[\because \frac{d}{dx}(\boldsymbol{u\,v}) = \boldsymbol{u\,v' + v\,u'}\right]$$

$$= \left[\frac{(1-q\,e^0)^2pe^0 - pe^02(1-q\,e^0)^1(-q\,e^0)}{(1-q\,e^0)^4}\right] = \frac{p^3 + 2p^2q}{p^4} = p^2\left(\frac{p+2q}{p^4}\right) = \boldsymbol{\frac{1+q}{p^2}}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{1+q}{p^2} - \frac{1}{p^2} = \frac{1+q-1}{p^2} = \boldsymbol{\frac{q}{p^2}}$$

ANOTHER FORM OF GEOMETRIC DISTRIBUTION

Definition: If X denotes the number of failure before the first success, then $\boldsymbol{P(X = x) = p\,q^x,\ \ x = 0, 1, 2, \ldots \infty}$ **.**

Moment Generating Function (M.G.F.) in Geometric Distribution

$$M_X(t) = \sum_{x=0}^{\infty} e^{tx}\,p(x) = p\sum_{x=0}^{\infty}(q\,e^t)^x = p[1 + (q\,e^t)^1 + (q\,e^t)^2 + (q\,e^t)^3 + \cdots] = p[1 - q\,e^t]^{-1} = \boldsymbol{\frac{p}{1-q\,e^t}}$$

Mean and Variance using Moment Generating Function in Geometric Distribution

$$E(X) = \left[\frac{d}{dt}M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{p}{1-q\,e^t}\right)\right]_{t=0} = \left[p\,\frac{d}{dt}(1 - q\,e^t)^{-1}\right]_{t=0} = [p(-1)(1-q\,e^t)^{-1-1}(-q\,e^t)]_{t=0}$$

$$= [pqe^t\,(1 - q\,e^t)^{-2}]_{t=0} = pqe^0\,(1 - q\,e^0)^{-2} = \boldsymbol{\frac{q}{p}} \qquad (\because \boldsymbol{p + q = 1})$$

$$E(X^2) = \left[\frac{d^2}{dt^2}M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\frac{p\,q\,e^t}{(1-q\,e^t)^2}\right]_{t=0} = \left[\frac{(1-q\,e^t)^2p\,q\,e^t - p\,q\,e^t2(1-q\,e^t)^1(-q\,e^t)}{(1-q\,e^t)^4}\right]_{t=0} \qquad \left[\because \frac{d}{dx}\left(\frac{u}{v}\right) = \frac{vu' - uv'}{v^2}\right]$$

$$E(X^2) = \left[\frac{(1-q\,e^0)^2p\,q\,e^0 - p\,q\,e^02(1-q\,e^0)^1(-q\,e^0)}{(1-q\,e^0)^4}\right] = \frac{p^3q + 2p^2q^2}{p^4} = p^2q\left(\frac{p+2q}{p^4}\right) = \boldsymbol{\frac{q^2+q}{p^2}}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{q^2+q}{p^2} - \frac{q^2}{p^2} = \frac{q^2+q-q^2}{p^2} = \boldsymbol{\frac{q}{p^2}}$$

Memory less Property of Geometric Distribution

If X is a RV with geormetic distribution, then X lacks memory, in the sense that $P(X > s + t/X > s) = P(X > t)$.

Proof : $P(X = x) = p\, q^{x-1},\ \ x = 1, \dots \infty$

$$P(X > s + t/X > s) = \frac{P(X>s+t \cap X>s)}{P(X>s)} = \frac{P(X>s+t)}{P(X>s)}$$

$$P(X > k) = \sum_{x=k+1}^{\infty} p\, q^{x-1} = p\, q^{k+1-1} + p\, q^{k+2-1} + p\, q^{k+3-1} + \cdots = p\, q^{k} + p\, q^{k+1} + p\, q^{k+2} + \cdots$$

$$= p\, q^{k}(1 + q^1 + q^2 + \cdots) = p\, q^{k}(1-q)^{-1} = \frac{p\, q^k}{p} = \boldsymbol{q^k}$$

Hence $P(X > s + t) = q^{s+t}$ and $P(X > s) = q^s$; $P(X > s + t/X > s) = \frac{q^{s+t}}{q^s} = \frac{q^s\, q^t}{q^s} = q^t = \boldsymbol{P(X > t)}$

Problems in Geometric Distribution

1. ***If the probability that an applicant for a drivers license will pass the road test on any given trial is 0.8. What is the probability that he will finally pass the test (i) On the fourth trial (ii) In less than 4 trials?***

Solution : $p = 0.8,\ q = 1 - p = 0.2,\ P(X = x) = p\, q^{x-1},\ \ x = 1, \dots \infty$

(i) $P(X = 4) = (0.8)\,(0.2)^{4-1} = 0.0064$

(ii) $P(X < 4) = P(1) + P(2) + P(3) = (0.8)\,[(0.2)^{1-1} + (0.2)^{2-1} + (0.2)^{3-1}] = 0.992$

2. ***A typist types 2 letters erroneously for every 100 letters. What is the probability that the 10th letter typed is the 1st erroneous letter?***

Solution : $p = \frac{2}{100} = 0.02,\ q = 1 - p = 0.98,\ P(X = 10) = (0.02)\,(0.98)^{10-1} = 0.0167$

3. ***A die is tossed until 6 appears. What is the probability that it must be tossed more than 5 times?***

Solution : $p = \frac{1}{6},\ q = 1 - p = \frac{5}{6},\ P(X = x) = p\, q^{x-1},\ \ x = 1, \dots \infty$

$P(X > 5) = 1 - P(X \le 5) = 1 - [P(1) + P(2) + P(3) + P(4) + P(5)]$

$$= 1 - \frac{1}{6}\left[\left(\frac{5}{6}\right)^{1-1} + \left(\frac{5}{6}\right)^{2-1} + \left(\frac{5}{6}\right)^{3-1} + \left(\frac{5}{6}\right)^{4-1} + \left(\frac{5}{6}\right)^{5-1}\right] = 0.4019$$

4. ***A trainee soldier shoots a target in an independent fashion. If the probability that the target is shot on any one shot is 0.8. (i) What is the probability that the target would be first hit at the 6th attempt? (ii) What is the probability that it takes less than 5 shots?***

Solution : $p = 0.8,\ q = 1 - p = 0.2,$ (i) $P(X = 6) = (0.8)\,(0.2)^{6-1} = 0.00026$

(ii) $P(X < 5) = P(1) + P(2) + P(3) + P(4) = (0.8)\,[(0.2)^{1-1} + (0.2)^{2-1} + (0.2)^{3-1} + (0.2)^{4-1}] = 0.9984$

5. ***The probability that a candidate can pass in an exam is 0.6. (i) What is the probability that he pass in the 3rd trial (ii) What is the probability that he pass before the 3rd trial?***

Solution : : $p = 0.6,\ q = 1 - p = 0.4$ (i) $P(X = 3) = (0.6)\,(0.4)^{3-1} = 0.096$

(ii) $P(X < 3) = P(1) + P(2) = (0.6)\,[(0.4)^{1-1} + (0.4)^{2-1}] = 0.84$

6. ***A discrete RV X has moment generating function $M_X(t) = (5 - 4\,e^t)^{-1}$ find $P(X = 5 \text{ or } 6)$.***

Solution : $P(X = x) = p\, q^x,\ \ x = 0, 1, \dots \infty,\qquad M_X(t) = p(1 - q\, e^t)^{-1}$

$$M_X(t) = (5 - 4\, e^t)^{-1} = \frac{1}{5}\left(1 - \frac{4}{5}\, e^t\right)^{-1},\ p = \frac{1}{5},\ q = \frac{4}{5}$$

$$P(X = 5 \text{ or } 6) = P(x = 5) + P(X = 6) = \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^5 + \left(\frac{1}{5}\right)\left(\frac{4}{5}\right)^6 = 0.118$$

7. ***A coin is tossed until the first head occurs. Assuming that the tosses are independent and the probability of a head occurring is p, find the value of p so that the probability that an odd number of tosses is required is equal to 0.6. Can you find a value of p so that the probability is 0.5 that an odd number of tosses is required?***

Solution : $P(X = x) = p\, q^{x-1},\ \ x = 1, \dots \infty$

$$P(X = odd\ number) = P(1) + P(3) + P(5) + \cdots = p\,(q^{1-1} + q^{3-1} + q^{5-1} + \cdots)$$

$$= p(1 + q^2 + q^4 + \dots) = p[1 + (q^2) + (q^2)^2 + \dots] = p(1 - q^2)^{-1} = \frac{p}{1-q^2} = \frac{p}{(1-q)(1+q)} = \frac{p}{p(1+q)} = \frac{1}{1+q}$$

Now $\frac{1}{1+q} = 0.6 \Rightarrow \frac{1}{2-p} = 0.6 \Rightarrow 0.6(2 - p) = 1 \Rightarrow 0.6p = 0.2 \Rightarrow p = \frac{1}{3}$

Now $\frac{1}{1+q} = 0.5 \Rightarrow \frac{1}{2-p} = 0.5 \Rightarrow 0.5(2 - p) = 1 \Rightarrow 0.5p = 0 \Rightarrow p = 0$

Though we get $p = 0$ it is meaningless. Hence the value of p cannot be found out.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

UNIFORM DISTRIBUTION

Definition : A continuous RV X with parameters a and b is uniform, if it has the p.d.f is $\boldsymbol{f(x)} = \begin{cases} \frac{1}{b-a}, & a < x < b \\ 0, & otherwise \end{cases}$

Moment Generating Function (M.G.F.) in Negative Binomial Distribution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx} f(x)\, dx = \int_a^b e^{tx} \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b e^{tx}\, dx = \frac{1}{b-a}\left[\frac{e^{tx}}{t}\right]_a^b = \boldsymbol{\frac{e^{bt}-e^{at}}{t(b-a)}}$$

Mean and Variance using Moment Generating Function in Negative Binomial Distribution

$$E(X) = \int_{-\infty}^{\infty} x\, f(x)\, dx = \int_a^b x \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x\, dx = \frac{1}{b-a}\left[\frac{x^2}{2}\right]_a^b = \frac{b^2-a^2}{2(b-a)} = \frac{(b-a)(b+a)}{2(b-a)} = \boldsymbol{\frac{b+a}{2}}$$

$$E(X^2) = \int_{-\infty}^{\infty} x^2 f(x)\, dx = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{1}{b-a}\int_a^b x^2\, dx = \frac{1}{b-a}\left[\frac{x^3}{3}\right]_a^b = \frac{b^3-a^3}{3(b-a)} = \frac{(b-a)(b^2+ab+a^2)}{3(b-a)} = \boldsymbol{\frac{b^2+ab+a^2}{3}}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{b^2+ab+a^2}{3} - \left(\frac{b+a}{2}\right)^2 = \frac{b^2+ab+a^2}{3} - \frac{b^2+2ab+a^2}{4} = \boldsymbol{\frac{(b-a)^2}{12}}$$

Problems in Uniform Distribution

1. ***If X is uniformly distributed over $(0, 10)$, find (i) $P(X < 2)$ (ii) $P(X > 8)$ (iii) $P(3 < X < 9)$***

Solution: $f(x) = \frac{1}{b-a},\ a < x < b,\quad f(x) = \frac{1}{10},\ 0 < x < 10$

(i) $P(X < 2) = \int_0^2 \left(\frac{1}{10}\right) dx = \frac{1}{10}[x]_0^2 = \frac{1}{5}$ (ii) $P(X > 8) = \int_8^{10} \left(\frac{1}{10}\right) dx = \frac{1}{10}[x]_8^{10} = \frac{1}{5}$

(iii) $P(3 < X < 9) = \int_3^9 \left(\frac{1}{10}\right) dx = \frac{1}{10}[x]_3^9 = \frac{3}{5}$

2. ***A random variable X has a uniform distribution over $(-3, 3)$.***

(i) $P(X < 2)$ (ii) $P(|X| < 2)$ (iii) $P(|X-2| < 2)$ (iv) Find K for which $P(X > K) = \frac{1}{3}$

Solution: $f(x) = \frac{1}{b-a},\ a < x < b,\quad f(x) = \frac{1}{6},\ -3 < x < 3$

(i) $P(X < 2) = \int_{-3}^2 \left(\frac{1}{6}\right) dx = \frac{1}{6}[x]_{-3}^2 = \frac{5}{6}$ (ii) $P(|X| < 2) = \int_{-2}^2 \left(\frac{1}{6}\right) dx = \frac{1}{6}[x]_{-2}^2 = \frac{2}{3}$

(iii) $P(|X-2| < 2) = P(-2 < (x-2) < 2) = P(0 < x < 4) = \int_0^3 \left(\frac{1}{6}\right) dx = \frac{1}{6}[x]_0^3 = \frac{1}{2}$

(iv) $P(X > k) = \frac{1}{3} \Rightarrow \int_k^3 \left(\frac{1}{6}\right) dx = \frac{1}{3} \Rightarrow \frac{1}{6}[x]_k^3 = \frac{1}{3} \Rightarrow 3 - k = 2 \Rightarrow \boldsymbol{k = 1}$

3. ***Busses arrive at a specified stop at 15 min intervals starting at 7 am this is they arrive at 7, 7.15, 7.30, 7.45 and so on. If a passenger arrives at the stop at a random time that is uniformly distributed between 7 and 7.30 am find the probability that he waits (i) Less than 5 min for a bus. (ii) At least 12 min for a bus.***

Solution: $f(x) = \frac{1}{b-a},\ a < x < b,\quad f(x) = \frac{1}{30},\ 0 < x < 30$

(i) $P(<\ 5\ minutes) = P(10 < X < 15) + P(25 < X < 30) = \int_{10}^{15} \frac{1}{30}\, dx + \int_{25}^{30} \frac{1}{30} dx = \frac{1}{30}[x]_{10}^{15} + \frac{1}{30}[x]_{25}^{30} = \frac{1}{3}$

(ii) $P(\ Atleast\ 12\ min) = P(0 < X < 3) + P(15 < X < 18) = \int_0^3 \frac{1}{30}\, dx + \int_{15}^{18} \frac{1}{30} dx = \frac{1}{30}[x]_0^3 + \frac{1}{30}[x]_{15}^{18} = \frac{1}{5}$

4. ***Trains arrive at a station at 15 minutes intervals starting at 4 am. If a passenger arrive to the station at a time that is uniformly distributed between 9.00 and 9.30, find the probability that he has to wait for the train for (i) Less than 6 minutes (ii) More than 10 minutes.***

Solution: $f(x) = \frac{1}{b-a},\ a < x < b,\quad f(x) = \frac{1}{30},\ 0 < x < 30$

(i) $P(<\ 6\ minutes) = P(9 < X < 15) + P(24 < X < 30) = \int_9^{15} \frac{1}{30}\, dx + \int_{24}^{30} \frac{1}{30} dx = \frac{1}{30}[x]_9^{15} + \frac{1}{30}[x]_{24}^{30} = \frac{2}{5}$

(ii) $P(More\ than\ 10\ min) = P(0 < X < 5) + P(15 < X < 20) = \int_0^5 \frac{1}{30}\, dx + \int_{15}^{20} \frac{1}{30} dx = \frac{1}{30}[x]_0^5 + \frac{1}{30}[x]_{15}^{20} = \frac{1}{3}$

5. ***Electric Trains in a particular route run every half an hour between 12 midnight and 6 am. Find the probability that a passenger entering the station at any time during this period will have to wait at least twenty minutes.***

Solution: $f(x) = \frac{1}{b-a},\ a < x < b,\quad f(x) = \frac{1}{30},\ 0 < x < 30$

$P(\text{at least 20 minutes}) = P(0 < X < 10) = \int_0^{10} \frac{1}{30}\, dx = \frac{1}{30}[x]_0^{10} = \frac{1}{3}$

6. ***If the MGF of a uniform distribution for a random variable X is $\frac{1}{t}(e^{5t} - e^{4t})$, find $E(X)$.***

Solution: $M_X(t) = \frac{e^{bt}-e^{at}}{t(b-a)},\ b = 5,\ a = 4,\quad E(X) = \frac{b+a}{2} = \frac{5+4}{2} = 4.5$

EXPONENTIAL DISTRIBUTION

Definition: A continuous RV X defined in $(0, \infty)$ is said to follow an exponential distribution if the probability density function is $\boldsymbol{f(x) = \lambda\, e^{-\lambda x},\ x \geq 0}$.

Application: Exponential distribution is useful in queueing theory and reliability theory. Time to failure of a component and time between arrivals can be modeled using exponential distribution.

Moment Generating Function (M.G.F.) in Exponential Distribution

$$M_X(t) = \int_{-\infty}^{\infty} e^{tx}\, f(x)\, dx = \int_0^{\infty} e^{tx}\, \lambda\, e^{-\lambda x}\, dx = \lambda \int_0^{\infty} e^{-(\lambda - t)x}\, dx = \lambda \left[\frac{e^{-(\lambda-t)x}}{-(\lambda - t)}\right]_0^{\infty} = \boldsymbol{\frac{\lambda}{\lambda - t}}$$

Mean and Variance using Moment Generating Function in Exponential Distribution

$$E(X) = \left[\frac{d}{dt} M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\left(\frac{\lambda}{\lambda - t}\right)\right]_{t=0} = \lambda\left[\frac{d}{dt}(\lambda - t)^{-1}\right]_{t=0} = \lambda[(-1)(\lambda - t)^{-2}(-1)]_{t=0} = \boldsymbol{\frac{1}{\lambda}}$$

$$E(X^2) = \left[\frac{d^2}{dt^2} M_X(t)\right]_{t=0} = \left[\frac{d}{dt}\, \lambda\, (\lambda - t)^{-2}\right]_{t=0} = \lambda\, [(-2)(\lambda - t)^{-3}(-1)]_{t=0} = \boldsymbol{\frac{2}{\lambda^2}}$$

$$V(X) = E(X^2) - [E(X)]^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = \boldsymbol{\frac{1}{\lambda^2}}$$

Memory less Property of Exponential Distribution

Statement: ***If X is exponential distributed with parameter λ, then for any 2 positive integers s and t*** $\boldsymbol{P(X > s + t / X > s) = P(X > t)}$.

Proof : $P(X > s + t / X > s) = \frac{P(X > s+t\, \cap\, X > s)}{P(X > s)} = \frac{P(X > s+t)}{P(X > s)}, \quad f(x) = \lambda\, e^{-\lambda x}, \ x \geq 0$

$$P(X > k) = \int_k^{\infty} f(x)\, dx = \int_k^{\infty} \lambda\, e^{-\lambda x}\, dx = \lambda \left[\frac{e^{-\lambda x}}{-\lambda}\right]_k^{\infty} = \boldsymbol{e^{-\lambda k}}$$

Hence $P(X > s + t) = e^{-\lambda(s+t)}$ *and* $P(X > s) = e^{-\lambda s}$; $P(X > s + t / X > s) = \frac{e^{-\lambda(s+t)}}{e^{-\lambda s}} = e^{-\lambda t} = \boldsymbol{P(X} > t)$

Problems in Exponential Distribution

1. ***The time (in hours) required to repairs a machine is exponential, distributed with parameter $\lambda = \frac{1}{2}$. (i) What is the probability that the repair time exceeds 2 hours? (ii) What is the conditional prob. that a repair takes at least 10h given that its duration exceeds 9h?***

Solution: $f(x) = \lambda\, e^{-\lambda x}, \ x > 0, \quad (i)\ P(X > 2) = \int_2^{\infty} \frac{1}{2}\, e^{-\frac{x}{2}} dx = \frac{1}{2}\left[\frac{e^{-\frac{x}{2}}}{-\frac{1}{2}}\right]_2^{\infty} = \frac{e^{-\infty} - e^{-1}}{-1} = e^{-1} = 0.3679$

$(ii)\ P(X \geq 10 / X > 9) = P(X > 1) = \int_1^{\infty} f(x) dx = \int_1^{\infty} \frac{1}{2}\, e^{-\frac{x}{2}} dx = \frac{1}{2}\left[\frac{e^{-\frac{x}{2}}}{-\frac{1}{2}}\right]_1^{\infty} = \frac{e^{-\infty} - e^{-\frac{1}{2}}}{-1} = e^{-\frac{1}{2}} = 0.6065$

2. ***Suppose that during a rainy season in a tropical island, the length of the shower has an exponential distribution, with average 2 minutes. Find the probability that the shower will be there for more than three minutes. If the shower has already lasted for 2 minutes, what is the probability that it will last for at least one more minute.***

Solution: Average length $= \frac{1}{\lambda} = 2, \ \lambda = \frac{1}{2}$, $(i)\ P(X > 3) = \int_3^{\infty} \frac{1}{2}\, e^{-\frac{x}{2}} dx = \frac{1}{2}\left[\frac{e^{-\frac{x}{2}}}{-\frac{1}{2}}\right]_3^{\infty} = e^{-\frac{3}{2}} = 0.2231$

$(ii)\ P(X > 3 / X > 2) = P(X > 1) = \int_1^{\infty} f(x) dx = \int_1^{\infty} \frac{1}{2}\, e^{-\frac{x}{2}} dx = \frac{1}{2}\left[\frac{e^{-\frac{x}{2}}}{-\frac{1}{2}}\right]_1^{\infty} = \frac{e^{-\infty} - e^{-\frac{1}{2}}}{-1} = e^{-\frac{1}{2}} = 0.6065$

3. ***The mileage which car owners get with a certain kind of radial tire is a RV having an exponential distribution with mean 40,000 km. Find the prob. that one of these tires will last (i) Atleast 20,000 km (ii) At most 30,000 km***

Solution: *Mean* $= \frac{1}{\lambda} = 40{,}000\ km, \ \lambda = \frac{1}{40{,}000}, \quad f(x) = \lambda\, e^{-\lambda x}, \ x > 0, \quad f(x) = \frac{1}{40{,}000}\, e^{-\frac{x}{40{,}000}}, \ x > 0$

(i) $P(X \geq 20{,}000) = \int_{20{,}000}^{\infty} f(x) dx = \int_{20{,}000}^{\infty} \frac{1}{40{,}000}\, e^{-\frac{x}{40{,}000}} dx = \frac{1}{40{,}000}\left[\frac{e^{-\frac{x}{40{,}000}}}{-\frac{1}{40{,}000}}\right]_{20{,}000}^{\infty} = e^{-\frac{1}{2}} = 0.6065$

(ii) $P(X \leq 30{,}000) = \int_0^{30{,}000} f(x) dx = \int_0^{30{,}000} \frac{1}{40{,}000}\, e^{-\frac{x}{40{,}000}} dx = \frac{1}{40{,}000}\left[\frac{e^{-\frac{x}{40{,}000}}}{-\frac{1}{40{,}000}}\right]_0^{30{,}000} = 1 - e^{-\frac{3}{4}} = 0.527$

4. ***The length of time a person speaks over phone follows exponential distribution with mean 6. What is the probability that the person will talk for (i) More than 8 min (ii) Less than 4 min (iii) Between 4 and 8 min***

Solution: *Mean* $= \frac{1}{\lambda} = 6, \ \lambda = \frac{1}{6}, \ f(x) = \lambda\, e^{-\lambda x}, \ x > 0$ $(i)\ P(X > 8) = \int_8^{\infty} \frac{1}{6}\, e^{-\frac{x}{6}}\, dx = \frac{1}{6}\left[\frac{e^{-\frac{x}{6}}}{-\frac{1}{6}}\right]_8^{\infty} = e^{-\frac{4}{3}} = 0.2635$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

(ii) $P(X < 4) = \int_0^4 \frac{1}{6}\, e^{-\frac{x}{6}}\, dx = \frac{1}{6}\left[\frac{e^{-\frac{x}{6}}}{-\frac{1}{6}}\right]_0^4 = 1 - e^{-\frac{2}{3}} = 0.4865$

(iii) $P(4 \le X \le 8) = \int_4^8 \frac{1}{6}\, e^{-\frac{x}{6}}\, dx = \frac{1}{6}\left[\frac{e^{-\frac{x}{6}}}{-\frac{1}{6}}\right]_4^8 = e^{-\frac{2}{3}} - e^{-\frac{4}{3}} = 0.5134 - 0.2635 = 0.2499$

5. ***The amount of time that a watch can run without having to be reset is a random variable having exponential distribution, with mean 120 days. Find the prob. that such a watch will have to be reset in less than 24 days.***

Solution:, $Mean = \frac{1}{\lambda} = 120\ days,\ \lambda = \frac{1}{120},\ \ f(x) = \lambda\, e^{-\lambda x},\ x > 0,\ \ f(x) = \frac{1}{120}\, e^{-\frac{x}{120}},\ x > 0$

$P(X < 24) = \int_0^{24} f(x)dx = \int_0^{24} \frac{1}{120}\, e^{-\frac{x}{120}} dx = \frac{1}{120}\left[\frac{e^{-\frac{x}{120}}}{-\frac{1}{120}}\right]_0^{24} = 1 - e^{-\frac{1}{5}} = 0.1813$

6. ***The number of kilo meters that a car can run before its battery has to be replaced is exponentially distributed with an average of 10,000 kms. If the owner desires to take a tour consisting of 8000 kms, what is the probability that he will be able to complete is his tour with our replacing the battery?***

Solution:, $Mean = \frac{1}{\lambda} = 10{,}000,\ \lambda = \frac{1}{10{,}000},\ \ f(x) = \lambda\, e^{-\lambda x},\ x > 0,\ \ f(x) = \frac{1}{10{,}000}\, e^{-\frac{x}{10{,}000}},\ x > 0$

$P(X > 8000) = \int_{8000}^{\infty} f(x)dx = \int_{8000}^{\infty} \frac{1}{10{,}000}\, e^{-\frac{x}{10{,}000}} dx = \frac{1}{10{,}000}\left[\frac{e^{-\frac{x}{10{,}000}}}{-\frac{1}{10{,}000}}\right]_{8000}^{\infty} = e^{-\frac{4}{5}} = 0.4493$

7. ***In a construction site, 3 lorries unload materials per hour, on an average. What is the probability that the time between arrival of successive lorries will be (i) at least 30 minutes (ii) less than 10 minutes?***

Solution: $\lambda = 3,\ f(x) = \lambda\, e^{-\lambda x},\ x > 0,\ \ f(x) = 3\, e^{-3x},\ x > 0$

(i) Probability that the time between arrival of successive lorries equal to 30 minutes or $\frac{1}{2}$ hour

$P\left(X \ge \frac{1}{2}\right) = \int_{\frac{1}{2}}^{\infty} f(x)dx = \int_{\frac{1}{2}}^{\infty} 3\, e^{-3x} dx = 3\left[\frac{e^{-3x}}{-3}\right]_{\frac{1}{2}}^{\infty} = e^{-\frac{3}{2}} = 0.223$

(ii) Probability that the time between arrival of successive lorries equal to 10 minutes or $\frac{1}{6}$ hour

$P\left(X < \frac{1}{6}\right) = \int_0^{\frac{1}{6}} f(x)dx = \int_0^{\frac{1}{6}} 3\, e^{-3x} dx = 3\left[\frac{e^{-3x}}{-3}\right]_0^{\frac{1}{6}} = 1 - e^{-\frac{1}{2}} = 0.393$

NORMAL DISTRIBUTION

The Normal distribution was first described by De Moive in 1933 as the limiting form of Binomial distribution as the number of trials becomes infinite. This discovery came into limelight after its discovery by both Laplace and Gauss half a century later. So this distribution is also called Gaussion distribution.

Definition: A continuous RV X, with parameters μ and σ^2 is normal if it has a probability density function

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

Standard Normal distribution: If X is a RV following normal distribution with parameter μ and σ, then $z = \frac{X-\mu}{\sigma}$ is called a Standard Normal variate and the p.d.f. of the standard variate Z is given by $\boldsymbol{\varphi(z) = \frac{1}{\sqrt{2\,\pi}}\, e^{-\frac{z^2}{2}}}, -\infty < z < \infty$

Application: (i) The most important continuous probability distribution in the statistics field is Normal distribution. (ii) In nature like rainfall and meteorological studies (iii) In industry (iv) In error calculation of experiments (v) Statistical quality control (vi) Radar applications and in research.

Problems in Normal Distribution

1. ***If X is normally distributed and the mean X is 12 and the SD is 4. Find out the following***
(i) $\boldsymbol{P(X \ge 20)}$ (ii) $\boldsymbol{P(X \le 20)}$ (iii) $\boldsymbol{P(0 \le X \le 12)}$

Solution: $\mu = 12,\ \sigma = 4$

(i) $P(X \ge 20) = P\left(\frac{X-\mu}{\sigma} \ge \frac{20-\mu}{\sigma}\right) = P\left(Z \ge \frac{20-12}{4}\right) = P(Z \ge 2) = 0.5 - P(0 \le Z \le 2) = 0.5 - 0.4772 = 0.0228$

(ii) $P(X \le 20) = P\left(\frac{X-\mu}{\sigma} \le \frac{20-\mu}{\sigma}\right) = P\left(Z \le \frac{20-12}{4}\right) = P(Z \le 2) = 0.5 + P(0 \le Z \le 2) = 0.5 + 0.4772 = 0.9772$

(iii) $P(0 \le X \le 12) = P\left(\frac{0-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{12-\mu}{\sigma}\right) = P\left(\frac{-12}{4} \le Z \le \frac{12-12}{4}\right) = P(-3 \le Z \le 0) = P(0 \le Z \le 3) = 0.4987$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

2. ***If X is a normal variate with $\mu = 30$ and $\sigma = 5$. Find (i) $P(26 \le X \le 40)$ (ii) $P(X \ge 45)$ (iii) $P(|X-30| \ge 5)$.***

Solution: $\mu = 30,\ \sigma = 5$

(i) $P(26 \le X \le 40) = P\left(\frac{26-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{40-\mu}{\sigma}\right) = P\left(\frac{26-30}{5} \le Z \le \frac{40-30}{5}\right) = P(-0.8 \le Z \le 2)$

$= P(-0.8 \le Z \le 0) + P(0 \le Z \le 2) = P(0 \le Z \le 0.8) + P(0 \le Z \le 2) = 0.2881 + 0.4772 = 0.7653$

(ii) $P(X \ge 45) = P\left(\frac{X-\mu}{\sigma} \ge \frac{45-\mu}{\sigma}\right) = P\left(Z \ge \frac{45-30}{5}\right) = P(Z \ge 3) = 0.5 - P(0 \le Z \le 3) = 0.5 - 0.4987 = 0.0013$

(iii) $P(|X-30| \ge 5) = 1 - P(|X-30| \le 5) = 1 - P(-5 \le X - 30 \le 5) = 1 - P(25 \le X \le 35)$

$= 1 - P\left(\frac{25-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{35-\mu}{\sigma}\right) = 1 - P\left(\frac{25-30}{5} \le Z \le \frac{35-30}{5}\right) = 1 - P(-1 \le Z \le 1)$

$= 1 - [P(-1 \le Z \le 0) + P(0 \le Z \le 1)] = 1 - P(0 \le Z \le 1) - P(0 \le Z \le 1)$

$= 1 - 2\,P(0 \le Z \le 1) = 1 - 2(0.3413) = 0.3174$

3. ***The savings bank account of a customer showed an average balance of Rs. 150 and a S.D. of Rs. 50. Assuming that the account balances are normally distributed (i) What percentage of account is over Rs. 200? (ii) What percentage of account is between Rs. 120 & Rs. 170? (iii) What % of account is less than Rs. 75?***

Solution: $\mu = 150,\ \sigma = 50$

(i) $P(X \ge 200) = P\left(Z \ge \frac{200-150}{50}\right) = P(Z \ge 1) = 0.5 - P(0 \le Z \le 1) = 0.5 - 0.3413 = 0.1587$

Percentage of account is over Rs. 200 is 15.87%

(ii) $P(120 < X < 170) = P\left(\frac{120-150}{50} < Z < \frac{170-150}{50}\right) = P(-0.6 < Z < 0.4)$

$= P(0 < Z < 0.6) + P(0 < Z < 0.4) = 0.2257 + 0.1554 = 0.3811$

Percentage of account is between Rs. 120 & Rs. 170 is 38.11%

(iii) $P(X < 75) = P\left(Z < \frac{75-150}{50}\right) = P(Z < -1.5) = 0.5 - P(0 < Z < 1.5) = 0.5 - 0.4332 = 0.0668$

Percentage of account is less than Rs. 75 is 6.68%

4. ***In an engineering examination, a student is considered to have failed, secured second class, first class and distinction, according as he scores less than 45%, between 45% and 60%, between 60% and 75% and above 75% respectively. In a particular year 10% of the students failed in the examination and 5% of the students got distinction. Find the percentages of students who have got first class and second class.***

Solution: Let X follow the distribution N(μ, σ).

Given: $P(X < 45) = 0.10$ *and* $P(X > 75) = 0.05$

$P\left(-\infty < \frac{X-\mu}{\sigma} < \frac{45-\mu}{\sigma}\right) = 0.1$ *and* $P\left(\frac{75-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \infty\right) = 0.05$

$P\left(-\infty < Z < \frac{45-\mu}{\sigma}\right) = 0.1$ *and* $P\left(\frac{75-\mu}{\sigma} < Z < \infty\right) = 0.05$

$P\left(0 < Z < \frac{\mu-45}{\sigma}\right) = 0.4$ *and* $P\left(0 < Z < \frac{75-\mu}{\sigma}\right) = 0.45$

From the table, $\frac{\mu-45}{\sigma} = 1.28$ *and* $\frac{75-\mu}{\sigma} = 1.64$

$\mu - 1.28\sigma = 45$ (1) *and* $\mu + 1.64\sigma = 75$ (2)

Solving equations (1) and (2), $\mu = 58.15$ *and* $\sigma = 10.28$

$P(Students\ gets\ first\ class) = P(60 < X < 75) = P\left(\frac{60-\mu}{\sigma} < \frac{X-\mu}{\sigma} < \frac{75-\mu}{\sigma}\right) = P\left(\frac{60-58.15}{10.28} < Z < \frac{75-58.15}{10.28}\right)$

$= P(0.18 < Z < 1.64) = P(0 < Z < 1.64) - P(0 < Z < 0.18)$

$= 0.4495 - 0.0714 = 0.3781$

Percentage of students getting first class = 38

Now percentage of students getting second class = 100 – (students who have failed, got 1st class and got distinction)

Percentage of students getting second class $= 100 - (10 + 38 + 5) = 47.$

5. ***If the actual amount of instant coffee which a filling machine puts into '6 – ounce' jars is a RV having a normal distribution with S.D. is 0.05 ounce and if only 3% of the jars are to contain less than 6 ounce of coffee, what must be the mean fill of these jars?***

Solution: Let X be the actual amount of coffee put into the jars. Then X follows N(μ, σ), σ = 0.05

$P(X < 6) = 3\% = \frac{3}{100} = 0.03 \Rightarrow\ P\left(-\infty < Z < \frac{6-\mu}{0.05}\right) = 0.03\ \Rightarrow\ P\left(0 < Z < \frac{\mu-6}{0.05}\right) = 0.5 - 0.03 = 0.47$

From the table, $\frac{\mu-6}{0.05} = 1.808,\quad \boldsymbol{\mu = 6.0904\ ounces}$

6. ***In a newly constructed township, 2000 electric lamps are installed with an average life of 1000 burning hours and standard deviation of 200hours. Assuming the life of the lamps follows normal distribution, find (i)The no. of lamps expected to fail during the first 700 hrs. (ii) In what period of burning hours 10% of the lamps fail.***

Solution: $\mu = 1000,\ \sigma = 200$

(i) $P(X \le 700) = P\left(\frac{X-\mu}{\sigma} < \frac{700-1000}{200}\right) = P(Z < -1.5) = P(Z > 1.5) = 0.5 - P(0 < Z < 1.5)$

$= 0.5 - 0.4332 = 0.0668$ ***(∵From Normal Table)***

The no. of lamps that fail to burn in the first 700 hours $= 2000 \times 0.0668 = 133.6 \approx 134$

(ii) Let t be the period at which 10% of lamps fail.

$P(X \le t) = 0.1 \ \Rightarrow\ P\left(\frac{X-\mu}{\sigma} \le \frac{t-1000}{200}\right) = 0.1 \Rightarrow P\left(Z \ge \frac{1000-t}{200}\right) = 0.1$

$P\left(0 \le Z \le \frac{1000-t}{200}\right) = 0.5 - 0.1 = 0.4 \ \Rightarrow\ \frac{1000-t}{200} = 1.28 \ \Rightarrow$ $\boldsymbol{t = 744}$ ***(∵From Normal Table)***

7. ***In a normal distribution 31% of the items are under 45 and 8% are over 64. Find the mean and S.D.***

Solution: Let mean by μ and standard deviation σ

31% of the items are under 45 ***and*** ***8% are over 64***

$P(X < 45) = 31\%$ ***and*** $P(X > 64) = 8\%$

$P\left(\frac{X-\mu}{\sigma} < \frac{45-\mu}{\sigma}\right) = 0.31$ ***and*** $P\left(\frac{X-\mu}{\sigma} > \frac{64-\mu}{\sigma}\right) = 0.08$

$P\left(Z < \frac{45-\mu}{\sigma}\right) = 0.31$ ***and*** $P\left(Z > \frac{64-\mu}{\sigma}\right) = 0.08$

$P\left(0 < Z < \frac{\mu-45}{\sigma}\right) = 0.5 - 0.31 = 0.19$ ***and*** $P\left(0 < Z < \frac{64-\mu}{\sigma}\right) = 0.5 - 0.08 = 0.42$

$\frac{\mu-45}{\sigma} = 0.5$ ***and*** $\frac{64-\mu}{\sigma} = 1.4$

$\mu - 0.5\sigma = 45$ ***and*** $\mu + 1.4\sigma = 64$

Solving for μ and σ, we get $\boldsymbol{\mu = 50}$ and $\boldsymbol{\sigma = 10}$

8. ***In a distribution exactly normal 7% of the items are under 35 and 89% are under 63. What are the mean and S.D. of the distribution?***

Solution: Let mean by μ and standard deviation σ

7% of the items are under 35 ***and*** ***89% are over 63***

$P(X < 35) = 7\%$ ***and*** $P(X < 63) = 89\%$

$P\left(\frac{X-\mu}{\sigma} < \frac{35-\mu}{\sigma}\right) = 0.07$ ***and*** $P\left(\frac{X-\mu}{\sigma} < \frac{63-\mu}{\sigma}\right) = 0.89$

$P\left(Z < \frac{35-\mu}{\sigma}\right) = 0.07$ ***and*** $P\left(Z < \frac{63-\mu}{\sigma}\right) = 0.89$

$P\left(0 < Z < \frac{\mu-35}{\sigma}\right) = 0.5 - 0.07 = 0.43$ ***and*** $P\left(0 < Z < \frac{63-\mu}{\sigma}\right) = 0.89 - 0.5 = 0.39$

$\frac{\mu-35}{\sigma} = 1.48$ ***and*** $\frac{63-\mu}{\sigma} = 1.23$

$\mu - 1.48\sigma = 35$ ***and*** $\mu + 1.23\sigma = 63$

Solving for μ and σ, we get $\boldsymbol{\mu = 50.3}$ and $\boldsymbol{\sigma = 10.33}$

9. ***In a normal distribution of a large group of men 5% are under 60 in height and 40% are between 60 and 65. Find the mean height and S.D.***

Solution: Let mean by μ and standard deviation σ

5% of the items are under 60 ***and*** ***40% are between 60 and 65***

$P(X < 60) = 5\%$ ***and*** $P(60 < X < 65) = 40\% = 0.4$

$P\left(\frac{X-\mu}{\sigma} < \frac{60-\mu}{\sigma}\right) = 0.05$ ***and*** $P(X < 65) = P(X < 60) + P(60 < X < 65)$

$P\left(Z < \frac{60-\mu}{\sigma}\right) = 0.05$ ***and*** $P\left(\frac{X-\mu}{\sigma} < \frac{65-\mu}{\sigma}\right) = 0.05 + 0.4 = 0.45$

$P\left(0 < Z < \frac{\mu-60}{\sigma}\right) = 0.5 - 0.05 = 0.45$ ***and*** $P\left(0 < Z < \frac{\mu-65}{\sigma}\right) = 0.5 - 0.45 = 0.05$

$\frac{\mu-60}{\sigma} = 1.645$ ***and*** $\frac{\mu-65}{\sigma} = 0.13$

$\mu - 1.645\sigma = 60$ ***and*** $\mu + 0.13\sigma = 65$

Solving for μ and σ, we get $\boldsymbol{\mu = 65.42}$ and $\boldsymbol{\sigma = 3.29}$

10. ***If X is $N(3,4)$. Find k so that $P(|X-3| > k) = 0.05$.***

Solution: Let mean by μ and standard deviation σ. $\mu = 3,\ \sigma^2 = 4,\ \sigma = 2,\ Z = \frac{X-\mu}{\sigma} = \frac{X-3}{2}$

$P(|X-3|>k)=0.05 \Rightarrow P\left(\left|\frac{X-3}{2}\right|>\frac{k}{2}\right)=0.05 \Rightarrow P\left(|Z|>\frac{k}{2}\right)=0.05 \Rightarrow 2\,P\left(Z>\frac{k}{2}\right)=0.05$

$P\left(Z>\frac{k}{2}\right)=0.025 \Rightarrow 0.5-P\left(0<Z<\frac{k}{2}\right)=0.025 \Rightarrow P\left(0<Z<\frac{k}{2}\right)=0.475 \Rightarrow \frac{k}{2}=1.96 \Rightarrow \quad \boldsymbol{k=3.92}$

11. ***The marks obtained by a number of students in a certain subject are approximately normally distributed with mean 65 and standard deviation 5. If 3 students are selected at random from this group, what is the probability that at least 1 of them would have scored above 75?***

Solution: $\mu=65,\ \sigma=5$

$P(X>75)=P\left(\frac{X-\mu}{\sigma}>\frac{75-\mu}{\sigma}\right)=P\left(Z>\frac{75-65}{5}\right)=P(Z>2)=0.5-P(0<Z<2)=0.5-0.4772=0.0228$

$p=P(a\ student\ scores\ above\ 75)=0.0228\ ,\ q=1-p=0.9772\ and\ n=3$

$P(y)=nC_y p^y q^{n-y},\ \ y=0,1,\ldots n\ ,\ P(y)=3C_y(0.0228)^y(0.9772)^{3-y},\ \ y=0,1,\ldots n$

$P(Y\geq 1)=1-P(Y<1)=1-P(0)=1-3C_0(0.0228)^0(0.9772)^{3-0}=0.0667$

7. ***In an examination the marks obtained by the students in Maths, Physics and Chemistry are normally distributed about mean 50, 52, 48 and S.D. 15, 12, 16 respectively. Find the prob. of securing a total mark of 180 or above.***

Solution: Let X,Y,Z be the marks of respective subjects. The total marks $T=X+Y+Z$

$\mu=E(T)=E(X+Y+Z)=E(X)+E(Y)+E(Z)=50+52+48=150$

$\sigma^2=V(T)=V(X+Y+Z)=V(X)+V(Y)+V(Z)=15^2+12^2+16^2=225+144+256=625\ ,\quad \sigma=25$

$$P(T\geq 180)=P\left(Z\geq\frac{180-150}{25}\right)=P(Z\geq 1.2)=0.5-P(0\leq Z\leq 1.2)=0.5-0.3849=0.1151$$

8. ***The percentage X of a particular compound contained in a rocket fuel follows the distribution $N(33,\ 3)$, through the specification for X is that it should lie between 30 and 35. The manufacturer will get a net profit (per unit of the fuel) of Rs. 100, if $30<X<35$, Rs. 50, if $25<X\leq 30$ or $35\leq X<40$ and incur a loss of Rs. 60 per unit of the fuel otherwise. Find the expected profit of the manufacturer. If he wants to increase his expected profit by 50% by increasing the net profit on that category of the fuel that meets the specification, what should be the new net profit per unit of the fuel of this category?***

Solution: $N(\mu,\ \sigma),\ N(33,\ 3),\ \mu=33,\ \sigma=3$

$P(30<X<35)=P\left(\frac{30-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{35-\mu}{\sigma}\right)=P\left(\frac{30-33}{3}<Z<\frac{35-33}{2}\right)=P(-1<Z<0.67)$

$=P(-1<Z<0)+P(0<Z<0.67)=P(0<Z<1)+P(0<Z<0.67)$

$=0.3413+0.2486=0.5899$ ***(∵From Normal Table)***

$P(25<X\leq 30)=P\left(\frac{25-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{30-\mu}{\sigma}\right)=P\left(\frac{25-33}{3}<Z<\frac{30-33}{2}\right)=P(-2.67<Z<-1)$

$=P(1<Z<2.67)=P(0<Z<2.67)-P(0<Z<1)=0.4962-0.3413=0.1549$

$P(35\leq X<40)=P\left(\frac{35-\mu}{\sigma}<\frac{X-\mu}{\sigma}<\frac{40-\mu}{\sigma}\right)=P\left(\frac{35-33}{3}<Z<\frac{40-33}{2}\right)=P(0.67<Z<2.33)$

$=P(0<Z<2.33)-P(0<Z<0.67)=0.4901-0..2486=0.2415$

$P[(25<X\leq 30)\ or\ (35\leq X<40)]=P(25<X\leq 30)+P(35\leq X<40)=0.1549+0.2415=0.3964$

$P(X<25\ or\ X>40)=1-(0.5899+0.3964)=0.0137$

Profit / Unit	***Probability***
Rs. 100	0.5899
Rs. 50	0.3964
Rs. -60	0.0137

$E(Profit\ per\ unit)\ =\ Rs.(100\times 0.5899+50\times 0.3964-60\times 0.0137)=\ Rs.79$

Let the revised net profit per unit of the first category fuel be k.

$E(Revised\ Profit\ per\ unit)=Rs.(k\times 0.5899+50\times 0.3964-60\times 0.0137)=\ Rs.(0.5899k+18.998)$

$E(Revised\ Profit\ per\ unit)=Rs.79+Rs.39.5$

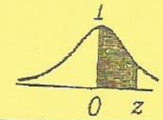
$0.5899k+18.998=118.5\Rightarrow k=\frac{118.5-18.998}{0.5899}=168.68\Rightarrow \boldsymbol{k=169}$

All the Best

Dr. P. GodhandaRaman, ***M.Sc., M.A., M.B.A., Ph.D.***
Assistant Professor (Senior Grade)
Department of Mathematics
SRM IST, Kattankulathur – 603 203
Email : godhanda@gmail.com
Mobile : 9941740168

Normal Distribution Table

Area under the Normal curve
from 0 to z



z	0	1	2	3	4	5	6	7	8	9
0.0	0.0000	0.0040	0.0080	0.0120	0.0160	0.0199	0.0239	0.0279	0.0319	0.0359
0.1	0.0398	0.0438	0.0478	0.0517	0.0557	0.0596	0.0636	0.0675	0.0714	0.0754
0.2	0.0793	0.0832	0.0871	0.0910	0.0948	0.0987	0.1026	0.1064	0.1103	0.1141
0.3	0.1179	0.1217	0.1255	0.1293	0.1331	0.1368	0.1406	0.1443	0.1480	0.1517
0.4	0.1554	0.1591	0.1628	0.1664	0.1700	0.1736	0.1772	0.1808	0.1844	0.1879
0.5	0.1915	0.1950	0.1985	0.2019	0.2054	0.2088	0.2123	0.2157	0.2190	0.2224
0.6	0.2258	0.2291	0.2324	0.2357	0.2389	0.2422	0.2454	0.2486	0.2518	0.2549
0.7	0.2580	0.2612	0.2642	0.2673	0.2704	0.2734	0.2764	0.2794	0.2823	0.2852
0.8	0.2881	0.2910	0.2939	0.2967	0.2996	0.3023	0.3051	0.3078	0.3106	0.3133
0.9	0.3159	0.3186	0.3212	0.3238	0.3264	0.3289	0.3315	0.3340	0.3365	0.3389
1.0	0.3413	0.3438	0.3461	0.3485	0.3508	0.3531	0.3554	0.3577	0.3599	0.3621
1.1	0.3643	0.3665	0.3686	0.3708	0.3729	0.3749	0.3770	0.3790	0.3810	0.3830
1.2	0.3849	0.3869	0.3888	0.3907	0.3925	0.3944	0.3962	0.3980	0.3997	0.4015
1.3	0.4032	0.4049	0.4066	0.4082	0.4099	0.4115	0.4131	0.4147	0.4162	0.4177
1.4	0.4192	0.4207	0.4222	0.4236	0.4251	0.4265	0.4279	0.4292	0.4306	0.4319
1.5	0.4332	0.4345	0.4357	0.4370	0.4382	0.4394	0.4406	0.4418	0.4429	0.4441
1.6	0.4452	0.4463	0.4474	0.4484	0.4495	0.4505	0.4515	0.4525	0.4535	0.4545
1.7	0.4554	0.4564	0.4573	0.4582	0.4591	0.4599	0.4608	0.4616	0.4625	0.4633
1.8	0.4641	0.4649	0.4656	0.4664	0.4671	0.4678	0.4686	0.4693	0.4699	0.4706
1.9	0.4713	0.4719	0.4726	0.4732	0.4738	0.4744	0.4750	0.4756	0.4761	0.4767
2.0	0.4772	0.4778	0.4783	0.4788	0.4793	0.4798	0.4803	0.4808	0.4812	0.4817
2.1	0.4821	0.4826	0.4830	0.4834	0.4838	0.4842	0.4846	0.4850	0.4854	0.4857
2.2	0.4861	0.4864	0.4868	0.4871	0.4875	0.4878	0.4881	0.4884	0.4887	0.4890
2.3	0.4893	0.4896	0.4898	0.4901	0.4904	0.4906	0.4909	0.4911	0.4913	0.4916
2.4	0.4918	0.4920	0.4922	0.4925	0.4927	0.4929	0.4931	0.4932	0.4934	0.4936
2.5	0.4938	0.4940	0.4941	0.4943	0.4945	0.4946	0.4948	0.4949	0.4951	0.4952
2.6	0.4953	0.4955	0.4956	0.4957	0.4959	0.4960	0.4961	0.4962	0.4963	0.4964
2.7	0.4965	0.4966	0.4967	0.4968	0.4969	0.4970	0.4971	0.4972	0.4973	0.4974
2.8	0.4974	0.4975	0.4976	0.4977	0.4977	0.4978	0.4979	0.4979	0.4980	0.4981
2.9	0.4981	0.4982	0.4982	0.4983	0.4984	0.4984	0.4985	0.4985	0.4986	0.4986
3.0	0.4987	0.4987	0.4987	0.4988	0.4988	0.4989	0.4989	0.4989	0.4990	0.4990
3.1	0.4990	0.4991	0.4991	0.4991	0.4992	0.4992	0.4992	0.4992	0.4993	0.4993

Dr. P. GodhandaRaman

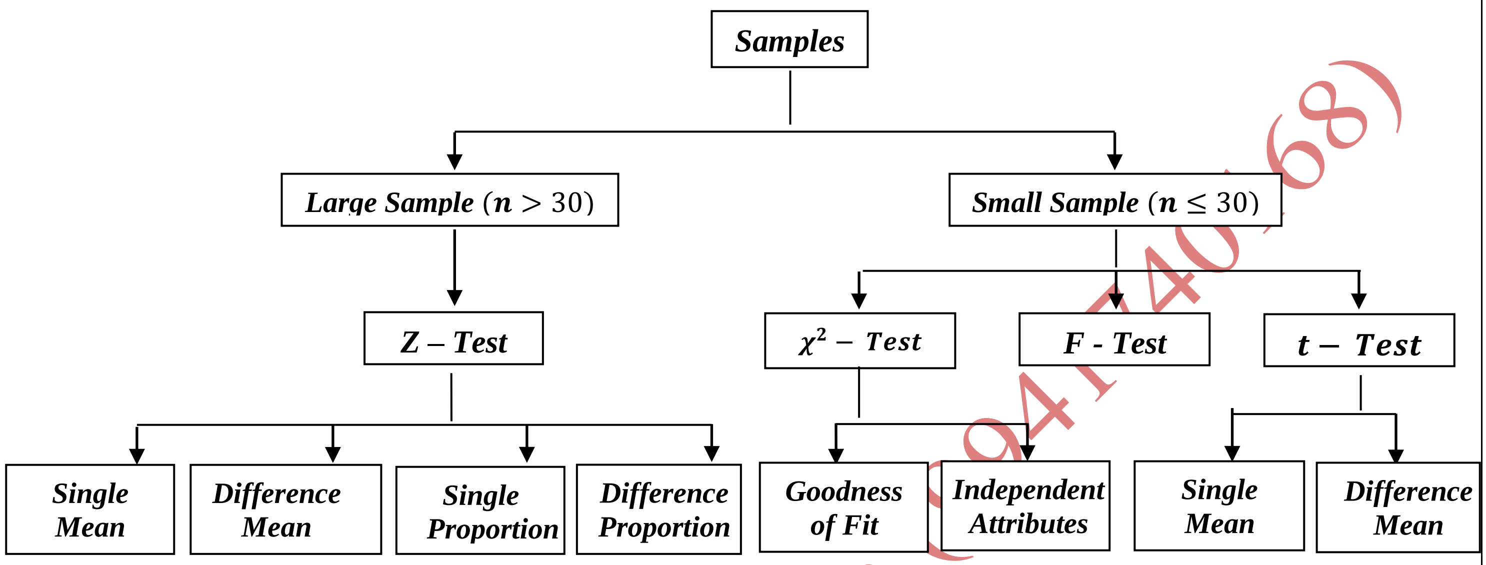
Assistant Professor (Sr.G.), Department of Mathematics, SRM IST, Kattankulathur

Email : godhanda@gmail.com, Mobile : 9941740168

UNIT – 3 : TESTING OF HYPOTHESIS

Syllabus:

- Large Samples - Z – Test
- Small Samples - t – Test



STATISTICS BASIC DEFINITIONS

Population: The word population or Universe in statistics is used to refer to any collection of individuals. The population may be finite or infinite. For example, we may have population of height, weight, ages, etc.

Sample : A finite subset of a population is called a sample.

Sample size: The number of elements in a sample is called the sample size.

Sampling : The process of selection of such samples is called sampling. For example, a housewife normally tests the cooked products to find if they are properly cooked & contain the proper quantity of salt.

Parameters : The statistical constants of the population such as the mean, the variance etc,

Statistics : The statistical concepts of the sample computed from the members or observation of the sample to estimate the parameters of the population from which the sample has been drawn are known as statistics. Population mean and variance are normally referred by μ and σ^2 while the sample mean and variance are referred by $\bar{x}$ and s^2.

Sampling Distribution : If we draw a sample of size n from a given finite population of size N then the total number of possible sample is NC_n. $NC_n = \frac{N!}{n!(N-n)!} = K$.

Standard Error (S.E.) : The standard deviation of sampling distribution of a statistic is known as its standard error.

Test of Significance : A very important aspect of the sampling theory is the study of tests of significance which enable us to decide on the basis of the sample results if

(i) The deviation between the observed sample statistic & the hypothetical parameter value is significant.

(ii) The deviation between two sample statistics is significant.

Null Hypothesis$(\boldsymbol{H_0})$ **:** There is no significant difference between the sample statistic and the corresponding population parameter or between two sample statistics is called Null Hypothesis.

Alternate Hypothesis$(\boldsymbol{H_1})$ **:** A hypothesis that is different from the null hypothesis.

Two Tailed : $H_0: \mu = \mu_0$, $H_1: \mu \neq \mu_0$,

One Tailed : $H_0 : \mu = \mu_0$, $H_1: \mu > \mu_0$ ***(Right Tailed),*** $H_1: \mu < \mu_0$ ***(Left Tailed)***

Error in Sampling: The main aim of the sampling theory is to draw a valid conclusion or a valid inference about the population parameters on the basis of the sample results. ***For example***, the mother at home tests the cooked products by taking and testing a small amount of cooked product. If this small amount of cooked product is good, we accept the lot to be good.

Type I Error : Reject H_0 when it is true. ***Type II Error :*** Accept H_0 when it is wrong.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Critical Region : A region, corresponding to a statics t, in the sample space S which amounts to rejection of the null hypothesis H_0 is called as critical region or region of rejection. The region of the sample space S which amounts to the acceptance H_0 is called acceptance region.

Level of Significance : The probability α that a random value of the statistic t belongs to the critical region is known as the level of significance. In other words, level of significance is the size of the Type I error. The levels of significance usually employed in testing of hypothesis are 5% and 1%.

LARGE SAMPLES

If the size of the sample $n \geq 30$, then that sample is called large sample. If n is large, the distributions, such as Binomial, Poisson, chi – square etc., are closely approximated by normal distributions. There are 4 important test to test the significance of large samples. 1.Single Proportion 2. Difference of proportions 3.Single mean 4. Difference of means.

PROCEDURE FOR TESTING OF HYPOTHESIS

1. Set up the null hypothesis H_0.
2. Set up the alternative hypothesis H_1.
   This will enable us to decide whether we have to use a single tailed (right or left) test or two tailed test.
3. Choose the appropriate level of significance(either 5% or 1% level). This is to be decided before sample is drawn.
4. Compute the test statistics $z = \frac{t-E(t)}{sE(t)}$ under the null hypothesis.
5. We compare the computed value of z in step (4) with the tabulated value z_α at given level of significance α.
   If the calculated value of Z is less than tabulated value z_α then H_0 accepted.
   If the calculated value of Z is greater than tabulated value z_α then H_0 rejected.

Critical Value (z_α) of z :

Critical Value (z_α)	**Level of Significance (α)**	
	1% (0.01)	**5% (0.05)**
Two Tailed Test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$
Right Tailed Test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$
One Tailed Test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$

SINGLE PROPORTION

Suppose a large sample of size n is taken from a normal population. To test the significant difference between the sample proportion p and the population proportion P, we use the statistic $\boldsymbol{z} = \frac{p-P}{\sqrt{\frac{PQ}{n}}}$

1. The probable limits for the observed proportion of successes p are given by $P \pm z_\alpha\sqrt{\frac{PQ}{n}}$.
2. If P is not known, the limits for the population proportion P are given by $p \pm z_\alpha\sqrt{\frac{pq}{n}}$ where $q = 1 - p$.
3. If α is not given, we can take safely 3σ limits. Hence, confidence limits for observed proportion p are $P \pm 3\sqrt{\frac{PQ}{n}}$ and confidence limits for the population proportion P are $p \pm 3\sqrt{\frac{pq}{n}}$ where $q = 1 - p$.
4. 95% confidence limits for population proportion P are given by $p \pm 1.96\sqrt{\frac{pq}{n}}$ where $q = 1 - p$
5. 99% confidence limits for population proportion P are given by $p \pm 2.58\sqrt{\frac{pq}{n}}$ where $q = 1 - p$

PROBLEMS IN SINGLE PROPORTION

Two Tailed Test

1. ***A coin is tossed 256 times and 132 heads are obtained. Would you conclude that the coin is a biased one?***

Solution: $n = 256$, $X =$ No. of success $= 132$, $p =$ proportion of successes in the sample $= \frac{X}{n} = \frac{132}{256} = 0.5156$,

$P =$ populatin proportion $= \frac{1}{2}$, $Q = 1 - P = 0.5$; *Null Hypothesis* H_0 : The coin is unbiased.

Alternative Hypothesis H_1 : The coin is biased. $(P \neq 0.5)$ (two tailed test)

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5156 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{256}}} = 0.4992$, $|z| = 0.4992 < 1.96$; H_0 is accepted. Hence the coin is unbiased.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

2. ***A random sample of 400 mangoes was taken from a large consignment and 40 were found to be bad. In this a sample from a consignment with proportion of bad mangoes 7.5%?***

Solution: $n = 400$, $p =$ sample proportion of bad mangoes $= \frac{40}{400} = 0.1$

$P =$ populatin proportion of bad mangoes $= 7.5\% = \frac{7.5}{100} = 0.075$, $Q = 1 - P = 0.925$

$H_0 : P = 0.075$, $H_1 : P \neq 0.075$ (two tailed test); z_α *at* 1% *LoS* $= 2.58$

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.1 - 0.075}{\sqrt{\frac{(0.075)(0.925)}{400}}} = 1.89$; $z < z_\alpha$; H_0 accepted at 1%.

3. ***In a city, a sample of 1000 people were taken & out of them 540 are vegetarians & the rest are non vegetarians. Can we say that both habits of eating are equally popular in the city at 1% & 5% level of significance?***

Solution: $n = 1000$, $p =$ sample proportion of vegetarians $= \frac{540}{1000} = 0.54$

$P =$ populatin proportion of vegetarians $= \frac{1}{2} = 0.5$, $Q = 1 - P = 0.5$

$H_0 : P = 0.5$ (Both habits are equally popular in the city). $H_1 : P \neq 0.5$ (two tailed test)

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.54 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{1000}}} = 2.5298$

(i) $|z| = 2.5298 < 2.58$, H_0 accepted at 1%. (ii) $|z| = 2.5298 > 1.96$, H_0 rejected at 5% level of significance. Both types of eaters are popular at 1% level and not so at 5% level of significance.

Right Tailed Test

4. ***In a big city 325 men out of 600 men were found to be smokers. Does this information support the conclusion that the majority of men in this city are smokers?***

Solution: $n = 600$, $p = \frac{325}{600} = 0.5417$, $P =$ Population proportion of smokers in the city $= 0.5$, $Q = 0.15$

Null Hypothesis $H_0 : P = 0.5$, *Alternative Hypothesis* $H_1 : P > 0.5$ (right tailed test)

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.5417 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{600}}} = 2.04$, $z_\alpha = 1.645$ at 5% level of significance

$|z| > z_\alpha$, H_0 is rejected. $z_\alpha = 2.33$ at 1% level of significance, $|z| < z_\alpha$, H_0 is accepted

Left Tail Test

5. ***A manufacturer claimed that at least 95% of its products supplied confirms to the specifications. Out of a sample of 200 numbers, 18 are defective. Test the claim at 5% level of significance.***

Solution: $n = 200$, $p =$ proportion of products confirming to specifications $= \frac{200-18}{200} = \frac{182}{200} = 0.91$

$P =$ population proportion $= \frac{95}{100} = 0.95$, $Q = 1 - P = 0.05$

Null Hypothesis $H_0 : P = 0.95$, *Alternative Hypothesis* $H_1 : P < 0.95$. (Left tailed test)

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.91 - 0.95}{\sqrt{\frac{(0.95)(0.05)}{200}}} = -2.596$, $|z| = 2.596 > 1.645$, H_0 is rejected at 5% level of significance.

6. ***The fatality rate of typhoid patients is believed to be 17.26 per cent. In a certain year 640 patients suffering from typhoid were treated in a metropolitan hospital and only 63 patients died. Can you consider the hospital efficient?***

Solution: $n = 640$, $p = \frac{63}{640} = 0.0984$, $P = 0.1726$, $Q = 1 - P = 0.8274$

Null Hypothesis $H_0 : p = P$, *i.e.* The hospital is not efficient.

Alternative Hypothesis $H_1 : p < P$. (Left tailed test)

Test statistics $z = \frac{p-P}{\sqrt{\frac{PQ}{n}}} = \frac{0.0984 - 0.1726}{\sqrt{\frac{(0.1726)(0.8274)}{640}}} = -4.96$, $|z| = 4.96 > 2.33$, H_0 is rejected at 1% level of significance.

The hospital is efficient in bringing down the fatality rate of typhoid patients.

Confidence Limits for the Population Proportion

7. ***A random sample of 500 toys was taken from a large consignment and 65 were found to be defective. Find the percentage of defective toys in the consignment.***

Solution: $n = 500$, $p = \frac{65}{500} = 0.13$, $q = 1 - p = 0.87$,

Confidence Limits for the Population Proportion P: $P = p \pm 3\sqrt{\frac{pq}{n}} = 0.13 \pm 3\sqrt{\frac{0.13 \times 0.87}{500}} = 0.175$ *and* 0.085

The percentage of defective toys in the consignment lies between 17.5 and 8.5.

8. ***A biased coin was thrown 400 times and 240 heads turned up. Find the probability of throwing heads in a single trial almost certainly lies between 0.53 and 0.67.***

Solution: $n = 400,\ p = \frac{240}{400} = 0.6,\ q = 1 - p = 0.4$

Confidence Limits for the Population Proportion P : $P = p \pm 3\sqrt{\frac{pq}{n}} = 0.6 \pm 3\sqrt{\frac{0.6\times0.4}{400}} = 0.6735\ and\ 0.5265$

The percentage of defective toys in the consignment lies between 52.65 and 67.35.

DIFFERENCE OF PROPORTIONS

Suppose 2 large samples of sizes n_1 & n_2 are taken respectively from 2 different populations. To test the significant difference between the sample proportions p_1 and p_2 find $\boldsymbol{z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}}$ where $\boldsymbol{P = \frac{n_1p_1 + n_2p_2}{n_1+n_2}}$ and $\boldsymbol{Q = 1 - P}$.

Two Tailed Test

1. ***Random samples of 400 men and 600 women were asked whether they would like to have a school near their residence. 200 men and 325 women were in favour of the proposal. Test the hypothesis that the proportions of men and women in favour of the proposal are same, at 5% level of significance.***

Solution: $n_1 = 400,\ n_2 = 600,\ p_1 =$ proportion of men $= \frac{200}{400} = 0.5,\ p_2 =$ proportion of women $= \frac{325}{600} = 0.54$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 \neq p_2$ (two tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.525,\ Q = 1 - P = 0.475$

$z = \frac{0.5-0.54}{\sqrt{(0.525\times0.475)\left(\frac{1}{400}+\frac{1}{600}\right)}} = -1.29,\ |z| = 1.29,\quad z_\alpha = 1.96$ at 5% level of significance, $|z| < z_\alpha$,

H_0 is accepted. i.e. men and women do not differ significantly in their attitude as regards the proposal.

2. ***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 between the proportions significant?***

Solution: $n_1 = 900,\ n_2 = 1600,\ p_1 = 20\% = 0.2,\ p_2 = 18.5\% = 0.185$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 \neq p_2$ (two tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2}$, $Q = 1 - P$

$P = \frac{(900\times0.2)+(1600\times0.185)}{900+1600} = \frac{180+296}{2500} = 0.1904,\ Q = 1 - P = 0.8096$

$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.2-0.185}{\sqrt{(0.1904\times0.8096)\left(\frac{1}{900}+\frac{1}{1600}\right)}} = 0.92$

$z_\alpha = 1.96$ at 5% level of significance, $|z| < z_\alpha,\ H_0$ is accepted.

Therefore, the difference between p_1 and p_2 is not significant at 5% level.

3. ***In a sample of 400, proportion of tea drinkers is 0.0125 and in another sample of 1200, proportion of tea drinkers is 0.0083. Test whether the samples are taken from a population in which proportion of tea drinkers is 0.01.***

Solution: $n_1 = 400,\ n_2 = 1200,\ p_1 = 0.0125,\ p_2 = 0.0083$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 \neq p_2$ (two tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.01,\ Q = 1 - P = 0.99$

$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.0125-0.0083}{\sqrt{(0.01\times0.99)\left(\frac{1}{400}+\frac{1}{1200}\right)}} = 0.73,\ z_\alpha = 1.96$ at 5%, $|z| < z_\alpha,\ H_0$ is accepted.

4. ***A company has the head office at Kolkata and a branch at Mumbai. The personnel director wanted to know if the workers at the two places would like the introduction of a new plan of work and a survey was conducted for this purpose. Out of a sample of 500 workers at Kolkata, 62% favoured the new plan. At Mumbai out of a sample 400 workers, 41% were against the new plan. Is there any significant difference between the two groups in their attitude towards the new plan at 5% level?***

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Solution: $n_1 = 500,\ n_2 = 400,\ p_1 = 62\% = 0.62,\ p_2 = 1 - 0.41 = 0.59$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 \neq p_2$ (two tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.607,\ Q = 1 - P = 0.393$

$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.62-0.59}{\sqrt{(0.607\times 0.393)\left(\frac{1}{500}+\frac{1}{400}\right)}} = 0.917,\ z_\alpha = 1.96$ at 5% level of significance, $|z| < z_\alpha,\ H_0$ is accepted. Therefore, the difference between p_1 and p_2 is not significant at 5% level.

5. ***In a random sample of 1000 people from city A, 400 are found to be consumers of wheat. In a sample 800 from city B, 400 are found to be consumers of wheat. Do these data reveal a significant difference between the two cities, so far as the proportion of wheat consumers is concerned?***

Solution: $n_1 = 1000,\ n_2 = 800,\ p_1 = \frac{400}{1000} = 0.4,\ p_2 = \frac{400}{800} = 0.5$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 \neq p_2$ (two tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = \frac{4}{9},\ Q = 1 - P = \frac{5}{9}$

$z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{0.4-0.5}{\sqrt{\left(\frac{4}{9}\times\frac{5}{9}\right)\left(\frac{1}{1000}+\frac{1}{800}\right)}} = 4.24,\ z_\alpha = 1.96$ at 5% LoS, $|z| > z_\alpha,\ H_0$ is rejected.

Right Tailed Test

6. ***Out of a sample of 1000 persons were found to be coffee drinkers. Subsequently, the excise duty on coffee was increased. After the increase in excise duty of coffee seeds, 800 people were found to take coffee out of a sample 1200. Test whether there is any significant decrease in the consumption of coffee after the increase in excise duty***

Solution: $n_1 = 1000,\ n_2 = 1200,\ p_1 = \frac{800}{1000} = 0.8,\ p_2 = \frac{800}{1200} = 0.67$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 > p_2$ (right tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.7273,\ Q = 1 - P = 0.2727$

$z = \frac{0.8-0.67}{\sqrt{(0.7273\times 0.2727)\left(\frac{1}{1000}+\frac{1}{1200}\right)}} = 6.82,\ z_\alpha = 2.33$ at 1% level of significance, $|z| > z_\alpha,\ H_0$ is rejected.

That is, there is significant decrease in the consumption of tea after the increase in duty.

7. ***A machine produced 20 defective units in a sample of 400. After overhauling the machine, it produced 10 defective units in a batch of 300. Has the machine improved in production due to overhauling. Test at 5% level of significance.***

Solution: $n_1 = 400,\ n_2 = 300,\ p_1 = \frac{20}{400} = 0.05,\ p_2 = \frac{10}{300} = 0.033$

Null Hypothesis $H_0 : p_1 = p_2$, *Alternative Hypothesis* $H_1 :\ p_1 > p_2$ (right tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.0427,\ Q = 1 - P = 0.9573$ $z = \frac{0.05-0.033}{\sqrt{(0.0427\times 0.9573)\left(\frac{1}{400}+\frac{1}{300}\right)}} = 1.1,\ z_\alpha = 1.645$ at 5% level of significance, $|z| < z_\alpha,\ H_0$ is accepted.

Left Tailed Test

8. ***15.5% of a random sample of 1600 UG students were smokers, whereas 20% of a random sample of 900 PG students were smokers in a state. Can we conclude that less number of UG students are smokers than PG?***

Solution: $n_1 = 1600,\ n_2 = 900,\ p_1 = 15.5\% = 0.155,\ p_2 = 20\% = 0.2$

Null Hypothesis $H_0 : p_1 = p_2$

Alternative Hypothesis $H_1 :\ p_1 < p_2$ (left tailed test)

Test statistics $z = \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $P = \frac{n_1p_1+n_2p_2}{n_1+n_2} = 0.1712,\ Q = 1 - P = 0.8288$

$z = \frac{0.155-0.2}{\sqrt{(0.1712\times 0.8288)\left(\frac{1}{1600}+\frac{1}{900}\right)}} = -2.87$, $z_\alpha = -1.645$ at 5% level of significance, $|z| > |z_\alpha|$, H_0 is rejected.

SINGLE MEAN

Suppose we want to test whether the given sample of size n has been drawn from a population with mean μ. We set up null hypothesis that there is no difference between $\bar{x}$ and μ, where $\bar{x}$ is the sample mean.

The statistic is $\boldsymbol{z} = \frac{\bar{\boldsymbol{x}}-\boldsymbol{\mu}}{\frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}}$ where σ is the standard deviation of the population. If the population S.D. is not known, then use the statistic $\boldsymbol{z} = \frac{\bar{\boldsymbol{x}}-\boldsymbol{\mu}}{\frac{\boldsymbol{s}}{\sqrt{\boldsymbol{n}}}}$ where s is the standard deviation of the population.

1. The limits of population mean μ are given by $\bar{\boldsymbol{x}} - \boldsymbol{z}_{\boldsymbol{\alpha}} \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}} \le \boldsymbol{\mu} \le \bar{\boldsymbol{x}} + \boldsymbol{z}_{\boldsymbol{\alpha}} \frac{\boldsymbol{\sigma}}{\sqrt{\boldsymbol{n}}}$.
2. At 5% level of significance, 95% confidence limits are $\bar{x} - 1.96 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 1.96 \frac{\sigma}{\sqrt{n}}$.
3. At 1% level of significance, 99% confidence limits are $\bar{x} - 2.58 \frac{\sigma}{\sqrt{n}} \le \mu \le \bar{x} + 2.58 \frac{\sigma}{\sqrt{n}}$.

Two Tailed Test

1. *The heights of college students in a city are normally distributed with S.D. 6 cms. A sample of 100 students has mean height 158 cms. Test the hypothesis that the mean height of college students in the city is 160 cms.*

Solution: $n = 100,\ \mu = 160,\ \bar{x} = 158,\ \sigma = 6,$

Null Hypothesis $H_0 : \mu = 160$, *Alternative Hypothesis* $H_1 : \mu \ne 160$ (two tailed test)

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{158-160}{6/\sqrt{100}} = 3.333,\ \ z_\alpha = 2.58$ at 1%, $z_\alpha = 1.96$ at 5%. $|z| > z_\alpha,\ H_0$ is rejected at both 1% & 5%.

2. *A sample of 900 members is found to have a mean 3.5 cm. Can it be reasonable regarded as a simple sample from a large population whose mean is 3.38 cm and a standard deviation 2.4 cm?*

Solution: $n = 900,\ \mu = 3.38,\ \bar{x} = 3.5,\ \sigma = 2.4,$

Null Hypothesis $H_0 : \mu = 3.38$, *Alternative Hypothesis* $H_1 : \mu \ne 3.38$ (two tailed test)

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.5-3.38}{2.4/\sqrt{900}} = 1.5,\quad z_\alpha = 2.58$ at 1%, $|z| < z_\alpha,\ H_0$ is accepted.

3. *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 165cm, and the S.D. is 10 cm?*

Solution: $n = 100,\ \mu = 165,\ \bar{x} = 160,\ \sigma = 10,$

Null Hypothesis $H_0 : \bar{x} = \mu$, *Alternative Hypothesis* $H_1 : \bar{x} \ne \mu$ (two tailed test)

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{160-165}{10/\sqrt{100}} = -5,\ |z| = 5,\ \ z_\alpha = 2.58$ at 1%, $|z| > z_\alpha,\ H_0$ is rejected.

4. *A random sample of 100 steel rods was drawn from a population of rods whose length are normally distributed with mean 4 feet and standard deviation 0.6 feet. If the sample mean is 4.2 feet, can the sample be regarded as a random sample drawn from the above population?*

Solution: $n = 100,\ \mu = 4,\ \bar{x} = 4.2,\ \sigma = 0.6,$

Null Hypothesis $H_0 : \mu = 4$, *Alternative Hypothesis* $H_1 : \mu \ne 4$ (two tailed test)

$z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{4.2-4}{0.6/\sqrt{100}} = 3.33,\quad z_\alpha = 1.96$ at 5%, $|z| > z_\alpha,\ H_0$ is rejected.

Right Tailed Test

5. *The mean breaking strength of the cables supplied by a manufacturer is 1800 with a S.D. of 100. By a new technique in the manufacturing process, it is claimed that the braking strength of the cable has increased. In order to test this 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?*

Solution: $n = 50,\ \mu = 1800,\ \bar{x} = 1850,\ \sigma = 100$

Null Hypothesis $H_0 : \bar{x} = \mu$, *Alternative Hypothesis* $H_1 : \bar{x} > \mu$ (right tailed test)

Test statistics $z = \frac{\bar{x}-\mu}{\frac{\sigma}{\sqrt{n}}} = \frac{1850-1800}{100/\sqrt{50}} = 3.54$, $z_\alpha = 2.33$ at 1% level of significance

$|z| > z_\alpha,\ H_0$ is rejected. i.e., based on the sample data, we may support the claim of increase in breaking strength.

Confidence Limits

6. *The mean value of a random sample of 60 items was found to be 145 with a S.D. of 40. Find the 95% confidence limits for the population mean.*

Solution: $n = 60,\ \bar{x} = 145,\ s = 40$, 95% confidence limits for μ : $\frac{|\mu-\bar{x}|}{\sigma/\sqrt{n}} \le 1.96$

If σ is not given, we can approximate it by the sample S.D. therefore 95% confidence limits for μ: $\frac{|\mu-\bar{x}|}{s/\sqrt{n}} \le 1.96$

$\bar{x} - 1.96\frac{s}{\sqrt{n}} \le \mu \le \bar{x} + 1.96\frac{s}{\sqrt{n}} \Rightarrow 145 - 1.96\frac{40}{\sqrt{60}} \le \mu \le 145 + 1.96\frac{40}{\sqrt{60}} \Rightarrow 134.9 \le \mu \le 155.1$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

DIFFERENCE OF MEAN

Let $\overline{x_1}$ be the mean of a sample of size n_1 from a population with mean μ_1 and variance σ_1^2. Let $\overline{x_2}$ be the mean of a sample of size n_2 from a population with mean μ_2 and variance σ_2^2.

To test whether there is any significant difference between $\overline{x_1}$ and $\overline{x_2}$ we have to use the statistic $\boldsymbol{z} = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$

Note: If the samples have been drawn from the same population then $\sigma_1^2 = \sigma_2^2 = \sigma^2$, $z = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\frac{\sigma^2}{n_1}+\frac{\sigma^2}{n_2}}}$

If σ is not known we can use a estimate of σ^2 given by $\sigma^2 = \frac{n_1 s_1^2+n_2 s_2^2}{n_1+n_2}$

Two Tailed Test

1. ***In a random sample of size 500, the mean is found to be 20. In another independent sample of size 400, the mean is 15. Could the samples have been drawn from the same population with S.D. 4?***

Solution: $n_1 = 500,\ n_2 = 400, \overline{x_1} = 20,\ \overline{x_2} = 15,\ \sigma = 4$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

Test statistics $z = \frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{20-15}{4\sqrt{\frac{1}{500}+\frac{1}{400}}} = 18.6$; $z_\alpha = 2.58$ at 1%, $|z| > z_\alpha$, H_0 is rejected.

That is, the samples could not have been drawn from the same population.

2. ***The mean of 2 large samples 1000 and 2000 members are 67.5 inches and 68.0 inches respectively. Can the samples the regarded as drawn from the same population of S.D. 2.5 inches.***

Solution: $n_1 = 1000,\ n_2 = 2000, \overline{x_1} = 67.5,\ \overline{x_2} = 68,\ \sigma = 2.5$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

Test statistics $z = \frac{\bar{x}_1-\bar{x}_2}{\sigma\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}} = \frac{67.5-68}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}} = -5.1,\ |z| = 5.1$

$z_\alpha = 1.96$ at 1% level of significance, $|z| > z_\alpha$, H_0 is rejected and H_1 is accepted.

3. ***In a survey of buying habits, 400 women shoppers are chosen at random in super market A located in a certain section of the city. Their average weekly food expenditure is Rs. 250 with a S.D. of Rs. 40. For 400 women shoppers chosen at random in super market B in another section of the city, the average weekly food expenditure is Rs. 220 with a S.D. of Rs. 55. Test at 1% level of significance whether the average weekly food expenditure of the two populations of shopper are equal.***

Solution: $n_1 = 400,\ n_2 = 400, \overline{x_1} = 250,\ \overline{x_2} = 220,\ s_1 = 40, s_2 = 55$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

Test statistics$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$, Since σ_1 & σ_2 the population S.D. are not known, we can take $\sigma_1^2 = s_1^2$ *and* $\sigma_2^2 = s_2^2$,

Test statistics $z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{250-220}{\sqrt{\frac{40^2}{400}+\frac{55^2}{400}}} = 8.82$; $z_\alpha = 2.58$ at 1% level of significance, $|z| > z_\alpha$, H_0 is rejected.

4. ***Test the significance of difference between the means of the samples, drawn from two 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*

Solution: $n_1 = 100,\ n_2 = 200, \overline{x_1} = 61,\ \overline{x_2} = 63,\ s_1 = 4, s_2 = 6$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 \neq \bar{x}_2$ (two tailed test)

Test statistics$z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1}+\frac{\sigma_2^2}{n_2}}}$, Since σ_1 & σ_2 the population S.D. are not known, we can take $\sigma_1^2 = s_1^2$ *and* $\sigma_2^2 = s_2^2$,

Test statistics $z = \frac{\bar{x}_1-\bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{61-63}{\sqrt{\frac{4^2}{100}+\frac{6^2}{200}}} = -3.02, |z| = 3.02$; $z_\alpha = 1.96$ at 5%, $|z| > z_\alpha$, H_0 is rejected.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Right Tailed Test

5. ***The average marks scored by 32 boys is 72 with a S.D. of 8, while that for 36 girls is 70 with a S.D. of 6. Test at 1% level of significance whether the boys perform better than girls.***

Solution: $n_1 = 32,\ n_2 = 36, \overline{x_1} = 72,\ \overline{x_2} = 70,\ s_1 = 8,\ s_2 = 6$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2\ (or\ \mu_1 = \mu_2)$,

Alternative Hypothesis $H_1 : \bar{x}_1 > \bar{x}_2$ (right tailed test)

Test statistics $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{72-70}{\sqrt{\frac{64}{32} + \frac{36}{36}}} = 1.15$, $z_\alpha = 2.33$ at 1% level of significance

$|z| < z_\alpha,\ H_0$ is accepted. That is, we cannot conclude that boys perform better than girls.

6. ***A random sample of 100 bulb from a company A showed a mean life 1300 hours and standard deviation 82 hours. Another random sample of 100 bulbs from company B showed a mean life 1248 hours and standard deviation of 93 hours. Are the bulbs of company A superior to bulbs of company B at 5% level of significance.***

Solution: $n_1 = 100,\ n_2 = 100, \overline{x_1} = 1300,\ \overline{x_2} = 1248,\ s_1 = 82,\ s_2 = 93$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2\ (or\ \mu_1 = \mu_2)$,

Alternative Hypothesis $H_1 : \bar{x}_1 > \bar{x}_2$ (right tailed test)

Test statistics $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{1300-1248}{\sqrt{\frac{82^2}{100} + \frac{93^2}{100}}} = 4.19$, $z_\alpha = 1.645$ at 1%, $|z| > z_\alpha,\ H_0$ is rejected.

Left Tailed Test

7. ***The average hourly wage of a sample of 150 workers in plant A was Rs. 2.56 with a S.D. of Rs. 1.08. The average wage of a sample of 200 workers in plant B was Rs. 2.87 with a S.D. of Rs. 1.28. Can an applicant safely assume that the hourly wage paid by plant B are higher than those paid by plant A?***

Solution: $n_1 = 150,\ n_2 = 200, \overline{x_1} = 2.56,\ \overline{x_2} = 2.87,\ s_1 = 1.08, s_2 = 1.28$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 < \bar{x}_2$ (left tailed test)

Test statistics$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, Since σ_1 & σ_2 the population S.D. are not known, we can take $\sigma_1^2 = s_1^2\ and\ \sigma_2^2 = s_2^2$, Test statistics $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{2.56-2.87}{\sqrt{\frac{(1.08)^2}{150} + \frac{(1.28)^2}{200}}} = -2.46$; $|z| = 2.46$; $z_\alpha = 1.645$ at 5% LoS, $|z| > z_\alpha,\ H_0$ is rejected.

Conclusion: The hourly wage paid by plant B are higher than those paid by plant A.

8. ***A simple sample of heights of 6400 English men has a mean of 170 cm and a S.D. of 6.4 cm, while a simple sample of heights of 1600 Americans has a mean of 172 cm and a S.D. of 6.3 cm. Do the data indicate that Americans are on the average, taller than the Englishmen?***

Solution: $n_1 = 6400,\ n_2 = 1600, \overline{x_1} = 170,\ \overline{x_2} = 172,\ s_1 = 6.4, s_2 = 6.3$

Null Hypothesis $H_0 : \bar{x}_1 = \bar{x}_2$, *Alternative Hypothesis* $H_1 : \bar{x}_1 < \bar{x}_2$ (left tailed test)

Test statistics$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}$, Since σ_1 & σ_2 the population S.D. are not known, we can take $\sigma_1^2 = s_1^2\ and\ \sigma_2^2 = s_2^2$, Test statistics $z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}} = \frac{170-172}{\sqrt{\frac{(6.4)^2}{6400} + \frac{(6.3)^2}{1600}}} = -11.32$; $|z| = 11.32$; $z_\alpha = 2.33$ at 1% LoS, $|z| > z_\alpha,\ H_0$ is rejected.

SMALL SAMPLES ($n \leq 30$)

Students 't' – Test - *SINGLE MEAN*

Standard deviation given directly: $t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n-1}}\right)}$, where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$,

$\bar{x}$ - sample mean, μ – population mean, n – sample size, s^2 – sample variance

Standard deviation not given directly: $t = \frac{\bar{x}-\mu}{\left(\frac{S}{\sqrt{n}}\right)}$, where $S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1}$

$\bar{x}$ - sample mean, μ – population mean, n – sample size, S^2 – population variance σ^2, ***Degree of freedom : $n-1$***

DIFFERENCE MEAN

$t = \frac{\bar{x}-\bar{y}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}}$, where $\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n}$, $\bar{y} = \frac{\sum_{j=1}^{n} y_j}{n}$, ***Degree of freedom : $n_1 + n_2 - 2$***

Standard deviation given directly: $s^2 = \frac{[n_1 s_1^2 + n_2 s_2^2]}{n_1+n_2-2}$

Standard deviation not given directly: $s^2 = \frac{\sum(x-\bar{x})^2+\sum(y-\bar{y})^2}{n_1+n_2-2}$,

Confidence or Fiducial Limits for μ:

95% Confidence limits for μ : $\bar{x} - t_{0.05}\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + t_{0.05}\left(\frac{S}{\sqrt{n}}\right)$

99% Confidence limits for μ : $\bar{x} - t_{0.01}\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + t_{0.01}\left(\frac{S}{\sqrt{n}}\right)$

Application of t – distribution:

The t – distribution has a wide number of applications in statistics, some of which are enumerated below

(i) To test if the sample mean $\bar{x}$ differs significantly from the hypothetical value μ of the population mean.
(ii) To test the significance of the difference between two sample means.
(iii) To test the significance of an observed sample correlation coefficient and sample regression coefficient
(iv) To test the significance of observed partial correlation coefficient.

Assumption for students's t – test:

(i) The parent population from which the sample is drawn is normal.
(ii) The sample observations are independent, that is, the sample is random.
(iii) The population standard deviation σ is unknown.

PROBLEMS

Students 't' – Test : Single Mean and Standard Deviation Given Directly

1. ***The mean weekly sales of soap bars in departmental stores was 146.3 bars per store. After an advertising campaign the mean weekly sales in 22 stores for a typical week increased to 153.7 and showed a S.D. of 17.2. Was the advertising campaign successful?***

Solution: $n = 22$, $\bar{x} = 153.7$, $\mu = 146.3$, $s = 17.2$

Null Hypothesis H_0 : $\mu = 146.3$, *i.e.* The advertising campaign is not successful.

Alternate Hypothesis H_1 : $\mu > 146.3$ *(Right tailed)*

$t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n-1}}\right)} = \frac{153.3-146.3}{\left(\frac{17.2}{\sqrt{22-1}}\right)} = 1.96$**,** degree of freedom: $\mathrm{n} - 1 = 21$ at 5% level of significance = 1.72.

Calculate value t > Tabulated t. **H_0 is rejected.**

Conclusion: The advertising campaign was definitely successful in promoting sales.

2. ***A sample of 20 items has mean 42 units and S.D. 5 units. Test the hypothesis that it is a random sample from a normal population with mean 45 units.***

Solution: $n = 20$, $\bar{x} = 42$, $\mu = 45$, $s = 5$

Null Hypothesis H_0 : $\mu = 45$, ***Alternate Hypothesis H_1 :*** $\mu \neq 45$ *(Two tailed)*

$t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n-1}}\right)} = \frac{42-45}{\left(\frac{5}{\sqrt{20-1}}\right)} = 2.615$**,** degree of freedom: $\mathrm{n} - 1 = 19$ at 5% level of significance = 2.09.

Calculate value t > Tabulated t. **H_0 is rejected.** The sample could not have come from this population.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

3. ***A spare part manufacturer is making spare parts with axle diameter of 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch with a S.D. of 0.040 inch. Verify whether the work satisfies the specifications.***

Solution: $n = 10, \ \bar{x} = 0.742, \ \mu = 0.700, \ s = 0.040$

Null Hypothesis $\boldsymbol{H_0}$: $\mu = 0.700$, ***Alternate Hypothesis*** $\boldsymbol{H_1}$ **:** $\mu \neq 0.700$ *(Two tailed)*

$t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n-1}}\right)} = \frac{0.742-0.700}{\left(\frac{0.040}{\sqrt{10-1}}\right)} = 3.15$, degree of freedom: $\mathrm{n} - 1 = 9$ at 5% level of significance = 2.26.

Calculate value t > Tabulated t. $\mathbf{H_0}$ **is rejected.** ***Conclusion:*** The product is not meeting the specifications.

4. ***The mean life time of a sample of 25 bulbs is found as 1550 hours with a S.D. of 120 hours. The company manufacturing the bulbs claims that the average life of their bulbs is 1600 hours. Is the claim acceptable at 5% level of significance?***

Solution: $n = 25, \ \bar{x} = 1550, \ \mu = 1600, \ s = 120$

Null Hypothesis $\boldsymbol{H_0}$: $\bar{x} = \mu$, ***Alternate Hypothesis*** $\boldsymbol{H_1}$ **:** $\bar{x} < \mu$ *(Left Tailed)*

$t = \frac{\bar{x}-\mu}{\left(\frac{s}{\sqrt{n-1}}\right)} = \frac{1550-1600}{\left(\frac{120}{\sqrt{25-1}}\right)} = -2.04$, degree of freedom: $\mathrm{n} - 1 = 24$ at 5% level of significance = 1.71.

Calculate value t > Tabulated t. $\mathbf{H_0}$ **is rejected.** The claim of the company cannot be accepted at 5% LOS.

Students 't' – Test : Single Mean and Standard Deviation Not Given Directly

5. ***A random sample of 10 boys had the following I.Q.'s 70, 120, 110, 101, 88, 83, 95, 98, 107, 100. Do these date support the assumption of a population mean I.Q.'s of 100? Find a reasonable range in which most of the mean I.Q. values of samples of 10 boys lie.***

Solution: $n = 10, \ \mu = 100, \ \boldsymbol{H_0} : \mu = 100, \ \boldsymbol{H_1} : \ \mu \neq 100$ *(Two tail)*

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{972}{10} = 97.2, \quad S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{1833.6}{10-1} = 203.73$

$\boldsymbol{x}$	$\boldsymbol{x - \bar{x}}$	$\boldsymbol{(x - \bar{x})^2}$
70	-27.2	739.84
120	22.8	519.84
110	12.8	163.84
101	3.8	14.44
88	-9.2	84.64
83	-14.2	201.64
95	-2.2	4.84
98	0.8	0.64
107	9.8	96.04
100	2.8	7.84
$\boldsymbol{\sum x = 972}$		$\boldsymbol{\sum (x - \bar{x})^2 = 1833.60}$

$t = \frac{\bar{x}-\mu}{\left(\frac{S}{\sqrt{n}}\right)} = \frac{97.2-100}{\left(\sqrt{\frac{203.73}{10}}\right)} = \frac{-2.8}{\left(\sqrt{20.373}\right)} = -0.62 \Rightarrow |\boldsymbol{t}| = \mathbf{0.62}$; d.f.= $\mathrm{n} - 1 = 10 - 1 = 9$ at 5% LOS = 2.262.

Calculate value t < Tabulated t. $\mathbf{H_0}$ **is accepted.**

Conclusion: The data are consistent with the assumption for a mean I.Q. of 100 in population.

95% Confidence Limits for $\boldsymbol{\mu}$***:*** $\bar{x} - t_{0.05}\left(\frac{S}{\sqrt{n}}\right) \leq \mu \leq \bar{x} + t_{0.05}\left(\frac{S}{\sqrt{n}}\right)$

$97.2 - 2.262\left(\sqrt{\frac{203.73}{10}}\right) \leq \mu \leq 97.2 + 2.262\left(\sqrt{\frac{203.73}{10}}\right)$

$86.99 \leq \mu \leq 107.41$, Hence the required 95% confidence interval is [86.99, 107.41].

6. ***The wages of 10 workers taken at random from a factory are given as Wages: 578, 572, 570, 568, 572, 578, 570, 572, 596, 584. Is it possible that the mean wage of all workers of this factory could be Rs. 580***

Solution: $n = 10, \ \mu = 580, \ \boldsymbol{H_0} : \mu = 580, \boldsymbol{H_1} : \ \mu \neq 580$ *(Two tailed)*

$\bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{5760}{10} = 576, \quad S^2 = \frac{\sum_{i=1}^{n}(x_i-\bar{x})^2}{n-1} = \frac{656}{10-1} = 72.89$

d.f.= $\mathrm{n} - 1 = 10 - 1 = 9$ at 5% LoS = 2.262.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

x	$x-\bar{x}$	$(x-\bar{x})^2$
578	2	4
572	-4	16
570	-6	36
568	-8	64
572	-4	16
578	2	4
570	-6	36
572	-4	16
596	20	400
584	8	64
$\sum x = 5760$		$\sum(x-\bar{x})^2 = 656$

$t = \frac{\bar{x}-\mu}{\left(\frac{S}{\sqrt{n}}\right)} = \frac{576-580}{\left(\sqrt{\frac{72.89}{10}}\right)} = \frac{-4}{(\sqrt{7.289})} = -1.48 \Rightarrow |t| = \mathbf{1.48}$; Calculate value t < Tabulated t. $\mathbf{H_0}$ **is accepted.**

Conclusion: Is it possible that the mean wage of all workers of this factory could be Rs. 580.

Students 't' – Test : Difference Mean and Standard Deviation Given Directly

7. ***Sample of two types of electric light bulbs were tested for length of life and following data were obtained***

	Type I	*Type II*
Sample Number	$n_1 = 8$	$n_2 = 7$
Sample Means	$\overline{x_1} = 1234\ hrs$	$\overline{x_2} = 1036\ hrs$
Sample S.D.	$s_1 = 36\ hrs$	$s_2 = 40\ hrs$

Is the difference in the means sufficient to warrant that type I is superior to type II regarding length of life

Solution: Null Hypothesis : $H_0: \overline{x_1} = \overline{x_2}$, **i.e.** The two types I and II of electric bulbs are identical.

Alternate Hypothesis : $H_1: \overline{x_1} > \overline{x_2}$ (Right tail)

Standard deviation given directly: $s^2 = \frac{[n_1 s_1^2 + n_2 s_2^2]}{n_1+n_2-2} = \frac{[8(36)^2+7(40)^2]}{8+7-2} = 1659.08$

$t = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{S^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{1234-1036}{\sqrt{1659.08\left(\frac{1}{8}+\frac{1}{7}\right)}} = 9.39$, degree of freedom = $n_1 + n_2 - 2 = 3$ at 5% = 1.77.

Calculate value t > Tabulated t. $\mathbf{H_0}$ **is rejected.** The type I is definitely superior to type II regarding length of life.

Students 't' – Test : Difference Mean and Standard Deviation Not Given Directly

8. ***Below are given the gain in weights (in kgs) of pigs fed on two diets A and B.***

Diet A	*25*	*32*	*30*	*34*	*24*	*14*	*32*	*24*	*30*	*31*	*35*	*25*	-	-	-
Diet B	*44*	*34*	*22*	*10*	*47*	*31*	*40*	*30*	*32*	*35*	*18*	*21*	*35*	*29*	*22*

Test if the two diets differ significantly as regards their effect on increase in weight.

Solution: Null Hypothesis $H_0: \overline{x_1} = \overline{x_2}$, **i.e.** There is no significant difference between the mean increase in weight due to diets A & B, ***Alternate Hypothesis*** : $H_1: \overline{x_1} \neq \overline{x_2}$ (Two tailed)

x	$x-\bar{x}$	$(x-\bar{x})^2$	y	$y-\bar{y}$	$(y-\bar{y})^2$
25	-3	9	44	14	196
32	4	16	34	4	16
30	2	4	22	-8	64
34	6	36	10	-20	400
24	-4	16	47	17	289
14	-14	196	31	1	1
32	4	16	40	10	100
24	-4	16	30	0	0
30	2	4	32	2	4
31	3	9	35	5	25
35	7	49	18	-12	144
25	-3	9	21	-9	81
			35	5	25
$\sum x = 336$		$\sum(x-\bar{x})^2 = 380$	29	-1	1
			22	-8	64
			$\sum y = 450$		$\sum(y-\bar{y})^2 = 1410$

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

$n_1 = 12,\ n_2 = 15,\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{336}{12} = 28,\ \bar{y} = \frac{\sum_{j=1}^{n} y_j}{n} = \frac{450}{15} = 30$

Standard deviation not given directly: $s^2 = \frac{\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2}{n_1+n_2-2} = 71.6$

$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{28-30}{\sqrt{71.6\left(\frac{1}{12}+\frac{1}{15}\right)}} = -0.609 \Rightarrow |\boldsymbol{t}| = \mathbf{0.609},$ d.f.= $n_1 + n_2 - 2 = 25$ at 5% LOS = 2.06.

Calculate value t < Tabulated t. **$\mathbf{H_0}$ is accepted.**

Conclusion: The two diets do not differ significantly as regards their effect on increase in weight.

9. ***A group of five patients treated with medicine A weigh 42, 39, 48, 60 and 41 kg.; a second group of 7 patients from the same hospital treated with medicine B weigh 38, 42, 56, 64, 68, 69 and 62 kg. Do you agree with the claim that medicine B increases the weight significantly?***

Solution: Null Hypothesis : $H_0: \boldsymbol{\mu_1} = \boldsymbol{\mu_2}$ **,** ***Alternate Hypothesis*** **:** $H_1: \boldsymbol{\mu_1} < \boldsymbol{\mu_2}$

$\boldsymbol{x}$	$\boldsymbol{x-\bar{x}}$	$\boldsymbol{(x-\bar{x})^2}$	$\boldsymbol{y}$	$\boldsymbol{y-\bar{y}}$	$\boldsymbol{(y-\bar{y})^2}$
42	-4	16	38	-19	361
39	-7	49	42	-15	225
48	2	4	56	-1	1
60	14	196	64	7	49
41	-5	25	68	11	121
$\sum \boldsymbol{x} = \mathbf{230}$		$\sum(\boldsymbol{x-\bar{x}})^2 = \mathbf{290}$	69	12	144
			62	5	25
			$\sum \boldsymbol{y} = \mathbf{399}$		$\sum(\boldsymbol{y-\bar{y}})^2 = \mathbf{926}$

$n_1 = 5,\ n_2 = 7,\ \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} = \frac{230}{5} = 46,\ \bar{y} = \frac{\sum_{j=1}^{n} y_j}{n} = \frac{399}{7} = 57$

Standard deviation not given directly: $s^2 = \frac{\sum(x-\bar{x})^2 + \sum(y-\bar{y})^2}{n_1+n_2-2} = \frac{290+926}{5+7-2} = \frac{1216}{10} = 121.6$

$t = \frac{\bar{x}-\bar{y}}{\sqrt{s^2\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{46-57}{\sqrt{121.6\left(\frac{1}{5}+\frac{1}{7}\right)}} = \frac{46-57}{\sqrt{41.69}} = \frac{-11}{6.46} = -1.703 \Rightarrow |\boldsymbol{t}| = \mathbf{1.703},$ d.f.= $n_1 + n_2 - 2 = 10$ at 5% = 1.81.

Calculate value t < Tabulated t. **$\mathbf{H_0}$ is accepted.** Medicines A and B do not differ significantly.

10. ***The table below represent the values of protein content from cow's milk and buffalo's milk at a certain level. Examine if these differences are significant.***

Cow's milk	***1.82***	***2.02***	***1.88***	***1.61***	***1.81***	***1.54***
Buffalo's milk	***2***	***1.83***	***1.86***	***2.03***	***2.19***	***1.88***

Solution: Null Hypothesis $H_0: \overline{\boldsymbol{x_1}} = \overline{\boldsymbol{x_2}}$ **,** ***Alternate Hypothesis*** **:** $H_1: \overline{\boldsymbol{x_1}} \neq \overline{\boldsymbol{x_2}}$ (Two tailed)

$\boldsymbol{x_1}$	$\boldsymbol{x_1^2}$	$\boldsymbol{x_2}$	$\boldsymbol{x_2^2}$
1.82	3.3124	2	4.0040
2.02	4.0804	1.83	3.3489
1.88	3.5344	1.86	3.4596
1.61	2.5921	2.03	4.1209
1.81	3.2761	2.19	4.761
1.54	2.3716	1.88	3.5344
$\sum \boldsymbol{x_1} = \mathbf{10.68}$	$\sum \boldsymbol{x_1^2} = \mathbf{19.1670}$	$\sum \boldsymbol{x_2} = \mathbf{11.79}$	$\sum \boldsymbol{x_2^2} = \mathbf{23.2599}$

$n_1 = 6,\ n_2 = 6,\ \overline{x_1} = \frac{\sum x_1}{n_1} = \frac{10.68}{6} = 1.78;\ \ \overline{x_2} = \frac{\sum x_2}{n_2} = \frac{11.79}{6} = 1.965$

Sample S.D.: $s_1^2 = \frac{\sum x_1^2}{n_1} - (\overline{x_1})^2 = \frac{19.167}{6} - (1.78)^2 = 0.0261;\ \ s_2^2 = \frac{\sum x_2^2}{n_2} - (\overline{x_2})^2 = \frac{25.2599}{6} - (11.79)^2 = 0.0154$

$t = \frac{\overline{x_1}-\overline{x_2}}{\sqrt{\left(\frac{n_1 s_1^2 + n_2 s_2^2}{n_1+n_2-2}\right)\left(\frac{1}{n_1}+\frac{1}{n_2}\right)}} = \frac{1.78-1.965}{\sqrt{\left(\frac{6\times0.0261+6\times0.0154}{6+6-2}\right)\left(\frac{1}{6}+\frac{1}{6}\right)}} = -2.03 \Rightarrow |\boldsymbol{t}| = \mathbf{2.03},$ d.f.= $n_1 + n_2 - 2 = 10$ at 5% = 2.26.

Calculate value t < Tabulated t. **$\mathbf{H_0}$ is accepted.**

Conclusion: The difference between the mean protein values of the two varieties of milk is not significant at 5%.

Dr. P. Godhandaraman, Assistant Professor (Sr.G.), Dept of Maths, SRM IST, Kattankulathur, Mobile: 9941740168, Email: godhanda@gmail.com

Paired Student 't' – Test

$t = \frac{\bar{d}}{\left(\frac{s}{\sqrt{n-1}}\right)}$, where $s^2 = \frac{\sum d^2}{n} - \left(\bar{d}\right)^2$; $d = x_1 - x_2$; $\bar{d} = \frac{\sum d}{n}$; $d.f = n - 1$; ***Degree of freedom : n − 1***

11. ***The following data relate to the marks obtained by 11 students in two tests, one held at the beginning of a year and the other at the end of the year after intensive coaching. Do the data indicate that the students have benefited by coaching?***

Test I	19	23	16	24	17	18	20	18	21	19	20
Test II	17	24	20	24	20	22	20	20	18	22	19

Solution: Null Hypothesis $H_0: \overline{x_1} = \overline{x_2}$ **;** ***Alternate Hypothesis :*** $H_1: \overline{x_1} < \overline{x_2}$ (left tailed)

x_1	x_2	$d = x_1 - x_2$	d^2
19	17	2	4
23	24	-1	1
16	20	-4	16
24	24	0	0
17	20	-3	9
18	22	-4	16
20	20	0	0
18	20	-2	4
21	18	3	9
19	22	-3	9
20	19	1	1
		$\sum d = -11$	$\sum d^2 = 69$

$n_1 = 11,\ n_2 = 11,\ \bar{d} = \frac{\sum d}{n} = \frac{-11}{11} = -1;\ \ s^2 = \frac{\sum d^2}{n} - \left(\bar{d}\right)^2 = \frac{69}{11} - (-1)^2 = 5.27;\ \ s = 2.296$

$t = \frac{\bar{d}}{\left(\frac{s}{\sqrt{n-1}}\right)} = -\frac{1}{\left(\frac{2.296}{\sqrt{11-1}}\right)} = -1.38$; $\Rightarrow |t| = 1.38$, d.f.= $n - 1 = 10$ at 5% = 1.81.

Calculate value t $<$ Tabulated t. $\mathbf{H_0}$ **is accepted.** ***Conclusion:*** The students have not benefitted by coaching.

All the Best

Dr. P. Godhanda Raman, *M.Sc., M.A., M.B.A., Ph.D.*

Assistant Professor (Senior Grade)

Department of Mathematics

SRM Institute of Science and Technology

Kattankulathur – 603 203, Chengalpattu District

Email : godhanda@gmail.com, Mobile : 9941740168

Values of |t| with probability P and degrees of freedom ν

P / ν	0.50	0.10	0.05	0.02	0.01
1	1.000	6.34	12.71	31.82	63.66
2	0.816	2.92	4.30	6.96	9.92
3	0.765	2.35	3.18	4.54	5.84
4	0.741	2.13	2.78	3.75	4.60
5	0.727	2.02	2.57	3.36	4.03
6	0.718	1.94	2.45	3.14	3.71
7	0.711	1.90	2.36	3.00	3.50
8	0.706	1.86	2.31	2.90	3.36
9	0.703	1.83	2.26	2.82	3.25
10	0.700	1.81	2.23	2.76	3.17
11	0.697	1.80	2.20	2.72	3.11
12	0.695	1.78	2.18	2.68	3.06
13	0.694	1.77	2.16	2.65	3.01
14	0.692	1.76	2.14	2.62	2.98
15	0.691	1.75	2.13	2.60	2.95
16	0.690	1.75	2.12	2.58	2.92
17	0.689	1.74	2.11	2.57	2.90
18	0.688	1.73	2.10	2.55	2.88
19	0.688	1.73	2.09	2.54	2.86
20	0.687	1.72	2.09	2.53	2.84
21	0.686	1.72	2.08	2.52	2.83
22	0.686	1.72	2.07	2.51	2.82
23	0.685	1.71	2.07	2.50	2.81
24	0.685	1.71	2.06	2.49	2.80
25	0.684	1.71	2.06	2.48	2.79
26	0.684	1.71	2.06	2.48	2.78
27	0.684	1.70	2.05	2.47	2.77
28	0.683	1.70	2.05	2.47	2.76
29	0.683	1.70	2.04	2.46	2.76
30	0.683	1.70	2.04	2.46	2.75

Regards!
Dr. P. Godhanda Raman, M.Sc., M.A., M.B.A., Ph.D.
Assistant Professor(Senior Grade), Department of Mathematics, SRM IST, Kattankulathur
Email : godhanda@gmail.com, Mobile : 9941740168