By Sangita Acharya

Random variable

It is a function which maps random process to the actual number.

X, Y, Z

x, y, z

E.g.

$$X = \begin{cases} 20 & \text{if it rains tommorrow.} \\ 30 & \text{if it doesn't rain tomorrow.} \end{cases}$$

$$X = \begin{cases} 1 & \text{if it turns up head} \\ 0 & \text{" " " tail.} \end{cases}$$

1. Discrete random variable

→ Things that you can count is discrete.

→ It maps to either a finite set or countably infinite set.

$S = \{\text{rain, no rain}\}$ $\quad$ $S = \{\text{head, tail}\}$

$X = \{20, 30\}$ $\quad$ $X = \{1, 0\}$

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

$X = \{0, 1, 2 \ldots\}$

2. Continuous RV

→ Things that you measured but not by counting.

→ It maps to uncountably infinite set.

$X \rightarrow$ amount of rainfall.

1"

1.001", 1.00001", 1.0098"

Continuous Probablity Distribution

P.d.f. → probablity density function

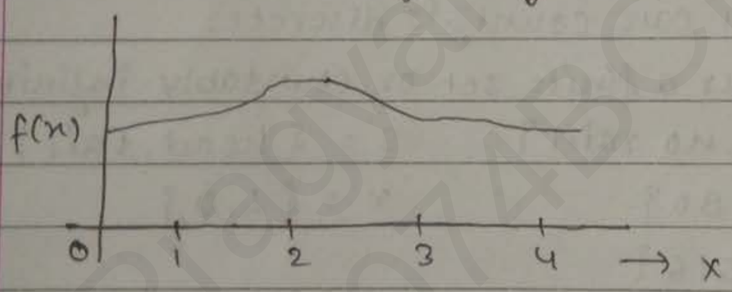
It is a function $f(x)$ which is associated with a continuous random variable x such that for ~~every~~ any two numbers a and b with $a \le b$.

$$\text{Prob.}(a \le X \le b) = \int_a^b f(x)\,dx$$

if it satisfies following properties.

1. $f(x) \ge 0$ for all x
2. $\int_{-\infty}^{\infty} f(x)\,dx = 1.$

$x \rightarrow$ amount of rainfall



$$\text{prob}(x=2) = \int_2^2 f(x)\,dx = 0$$

$$\text{prob}(2 \le x \le 3) = \int_2^3 f(x)\,dx$$

$$\text{Prob}(x>4) = \int_4^{\infty} f(x)\,dx = 1 - \int_0^4 f(x)\,dx$$

By Sangita Acharya

Date:
Page:

Cumulative distribution function

It also refers to the probablity that the value of a random variable falls within some specified value.

$$F(x) = \text{Prob.}(X \le x) = \int_{-\infty}^{x} f(x)\,dx$$

Properties of F(x)

① $0 \le F(x) \le 1$

② $F(-\infty) = \int_{-\infty}^{-\infty} f(x)\,dx = 0$

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

④ $\text{Prob.}(a \le X \le b) = \int_{a}^{b} f(x)\,dx$

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

$$= F(b) - F(a)$$

⑤ $X, f(x), F(x)$

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

$f(x) = \frac{dF(x)}{dx}$

By Sangita Acharya

Date:
Page:

Expected value and variance of a continuous RV
(Mean)

Mean denoted by $E(X)$.
variance denoted by $V(X)$.

The average value of a random variable is called mathematical expectation or mean of random variable.

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

$$\text{variance} = E[X - E(X)]^2$$
$$= E(X^2) - [E(X)]^2$$

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

\# If the probablity density of a random variable is given by

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

Find (a) the value of k (b) $P(1/4 \le X \le 3/4)$ (c) $P(X > 2/3)$ (d) Mean and variance.

ⓐ $k = 4$
ⓑ 0.3125
ⓒ 0.8025
ⓓ $0.8, 0.0267$

⇒ Solution,

If $f(x)$ is p.d.f. then

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

By Sangita Acharya

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

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

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

or $k \cdot \frac{x^4}{4}\Big|_0^1 = 1$

$\therefore k = 4.$

$$P(1/4 \leq X \leq 3/4) = \int_{1/4}^{3/4} 4x^3\,dx = 0.3125$$

$$P(X \geq 2/3) = \int_{2/3}^{\infty} f(x)\,dx = \int_{2/3}^{1} 4x^3\,dx + \int_{1}^{\infty} 0\,dx$$

$= 0.8025$

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

$= \int_{0}^{1} x \cdot 4x^3\,dx$

$= 0.8$

$$E(x^2) = \int_{0}^{1} x^2 \cdot 4x^3\,dx$$

$= 0.67$

By Sangita Acharya

$$\text{variance} = E(x^2) - [E(x)]^2$$
$$= 0.67 - (0.8)^2$$
$$= 0.0267$$

The length of time (in minutes) that a certain lady speaks on the telephone is found to be random phenomenon, with a Probablity function specified by the Prob. density function

$$f(x) = \begin{cases} A e^{-x/5} & \text{for } x \geq 0 \\ 0 & \text{elsewise} \end{cases}$$

(a) Find the value of A (b) what is the Prob. that the number of minutes that she will take over the phone is
i) more than 10 minutes ii) less than 5 minutes
iii) between 5 and 10 minutes

(a) $A = 1/5$
(b) 0.1354
0.6321
0.2325

⇒ Solution,

$$\int_{-\infty}^{0} 0 + \int_{0}^{\infty} A e^{-x/5}\, dx = 1$$

Date:

Page:

a. $A \cdot \dfrac{-e^{-x/5}}{1/5} \Big|_0^\infty = 1.$

$\dfrac{A}{1/5} = 1$

$\therefore A = \dfrac{1}{5}$

$P(x > 10) = \int_{10}^{\infty} \frac{1}{5} e^{-x/5}\, dx$

$= \frac{1}{5} \dfrac{-e^{-x/5}}{1/5} \Big|_{10}^{\infty} = \dfrac{1}{e^2} = 0.1354$

$P(x < 5) = \int_{-\infty}^{0} f(x)\, dx + \int_0^5 f(x)\, dx$

$= \int_0^5 \frac{1}{5} e^{-x/5}\, dx = 0.6321$

$P(5 < x < 10) = \int_5^{10} \frac{1}{5} e^{-x/5}\, dx$

$= 0.2325$

Date:
Page:

\# If p.d.f. of x is

$$f(x) = \begin{cases} ax & \text{for } 0 \le x < 1 \\ a & \text{for } 1 \le x < 2 \\ -ax + 3a & \text{for } 2 \le x \le 3 \\ 0 & \text{otherwise} \end{cases}$$

find a and $P(X \le 1.5)$

⇒ Solution,

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} f(x)\,dx + \int_{1}^{2} f(x)\,dx + \int_{2}^{3} f(x)\,dx + \int_{3}^{\infty} f(x)\,dx = 1$$

or, $0 + \int_{0}^{1} ax\,dx + \int_{1}^{2} a\,dx + \int_{2}^{3} (-ax+3a)\,dx + 0 = 1$

or, $\frac{a}{2} + a + 3a - \frac{5a}{2} = 1$

$\therefore a = \frac{1}{2}$.

$$P(X \le 1.5) = \int_{-\infty}^{0} f(x)\,dx + \int_{0}^{1} ax\,dx + \int_{1}^{1.5} a\,dx$$

$$= \frac{1}{2}\left[\frac{x^2}{2}\Big|_0^1 + x\Big|_1^{1.5} \right]$$

$$= \frac{1}{2}\left[\frac{1}{2} + 0.5 \right]$$

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

By Sangita Acharya

Date:
Page:

If x has a probablity density function

$$f(x) = \begin{cases} ke^{-3x} & \text{for } x>0 \\ 0 & \text{otherwise} \end{cases}$$

Find (i) k (ii) $P(0.5 \le x \le 1)$ (iii) The distribution of random variable x.

⇒ Solution,

$$\int_{-\infty}^{0} f(x)\,dx + \int_{0}^{\infty} ke^{-3x}\,dx = 1$$

or, $k \frac{-e^{3x}}{3} \Big|_0^{\infty} = 1$

or, $\frac{k}{3} = 1$

$\therefore k = 3$

$$P(0.5 \le x \le 1) = \int_{0.5}^{1} ke^{-3x}\,dx$$

$$= \frac{ke^{-3x}}{-3}\Big|_{0.5}^{1} = 3 \times 0.0578$$

$$= 0.1733$$

Normal Distribution

Definition :- A continuous random variable x is said to follow normal distribution with parameter mean μ and variance σ^2 if its probability density function is given by

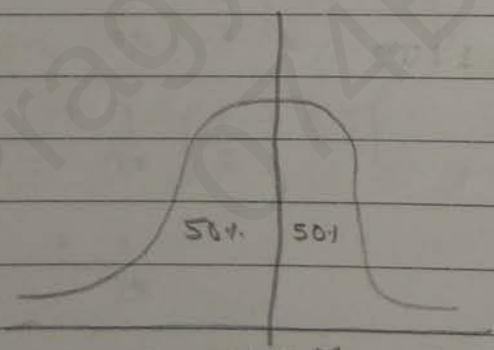
$$f(x, \mu, \sigma^2) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}$$

$$X \sim N(\mu, \sigma^2)$$

Properties

1. The curve of normal distribution is bell shaped and symmetrical about vertical line $x = \mu$

2. The mean, median and mode of normal distribution lie at same point at center.



$X = \mu$

$M_d = M_0$

3. The x-axis is asymptotes to the normal curve.

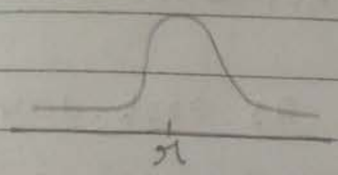
4. Area gives probability

5. Total area of normal curve is equal to one.

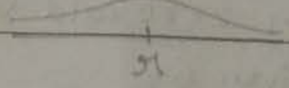
Date:
Page:

6. The width of normal curve is controlled by S.D. σ.

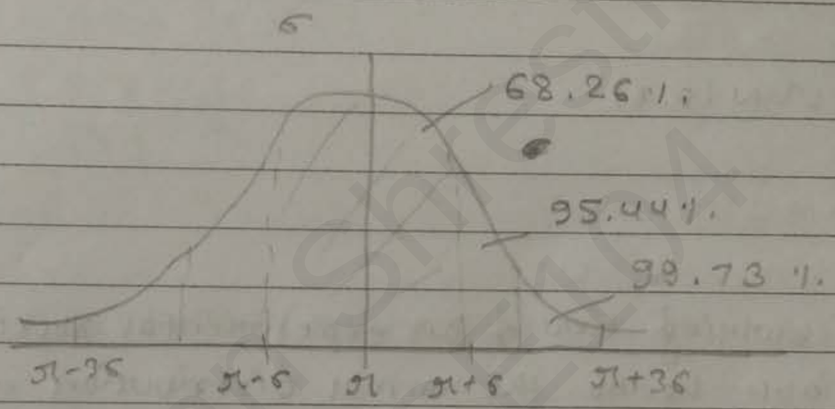
σ = min



σ = max.



7. Area property.



8. Quartile deviation = $\frac{2}{3}\sigma$

Mean deviation =

- σ is scale parameter.

By Sangita Acharya

Standard normal Distribution

If a continuous random variable x flows normal distribution with parameters mean μ and variance σ^2 then $z = \frac{x-\mu}{\sigma}$ follows standard normal distribution with mean zero and variance one if its prob. density function is given by,

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

$$z \sim N(0,1)$$

Q.1. The burning time of an experimental rocket is a random variable having the normal distribution with $\mu = 4.76$ sec and $\sigma = 0.04$ sec. What is the probability that this kind of rocket will burn?

a) less than 4.66 second ?

b) more than 4.80 sec ?

c) anywhere from 4.70 to 4.80 sec?

⇒ solution,

Let x be a continuous random variable which represents burning time of an experimental rocket.

$$x \sim N(\mu, \sigma^2)$$

$\mu = 4.76$ sec

$\sigma = 0.04$ sec.

a) Prob. $(x < 4.66)$

By Sangita Acharya

$= prob\left(\frac{x-\mu}{\sigma} < \frac{4.66-4.75}{0.04}\right)$

$= prob(Z < -2.5)$

$= 0.0062$

b) $prob(x > 4.80)$

$= prob.\left(\frac{x-\mu}{\sigma} > \frac{4.80-4.76}{0.04}\right)$

$= prob.(Z > 1)$

$= 1 - prob(Z \le 1)$

$= 1 - 0.8413$

$= 0.1587.$

c) $prob.(4.70 < x < 4.80)$

$prob(x < 4.80) = 0.8413$

~~prob (x > 4.70)~~

$= prob.\left(\frac{4.70-4.76}{0.04} < \frac{x-\mu}{\sigma} < \frac{4.80-4.76}{0.04}\right)$

$= prob(-1.5 < Z < 1)$

$= 0.8413 - 0.0668$

$= 0.7745$

By Sangita Acharya

Page:

Q.8. In an intelligence test administrated to 1000 student the average score was 42 and standard deviation 24. Find a) the no. of student exceeding a score 50 b) the no. of student having between 30 and 54. c) the score exceeded by the top 100 students.

⇒ solution

$\mu = 42$

$\sigma = 24.$

a) $x > 50.$

$prob(x > 50)$

$= prob.\left(\frac{x-\mu}{\sigma} > \frac{50-42}{24}\right)$

$= prob\left(z > 0.33\right)$

$= 1 - Prob.(z \le 0.33)$

$= 1 - 0.6293$

$= 0.3707.$

b) $Prob(30 < x < 54)$

$= prob.\left(\frac{30-42}{24} < \frac{x-\mu}{\sigma} < \frac{54-42}{24}\right)$

$= prob.(-0.5 < z < 0.5)$

$= 0.6915 - 0.3085$

$= 0.383$

Date:
Page:

c) No. of student exceeding a.

$$\text{prob.}(x > a) = 0.1$$

$$\text{prob.}\left(\frac{x-\mu}{\sigma} > \frac{a-42}{24}\right) = 0.1$$

$$\text{prob}\left(z > \frac{a-42}{24}\right) = 0.1$$

$$1 - \text{prob.}\left(z \le \frac{a-42}{24}\right) = 0.1$$

$$\text{prob}\left(z \le \frac{a-42}{24}\right) = 0.9$$

$$\therefore \frac{a-42}{24} = 1.28$$

$$\Rightarrow a = 72.72$$

Q.2.

⇒ solution

$$\mu = 52.4,\ \sigma = ?$$

$$\text{prob}(x > 79.2) = 0.2$$

a, $$\text{prob.}\left(\frac{x-52.4}{\sigma} > \frac{79.2-52.4}{\sigma}\right) = 0.2$$

$$\text{prob}\left(z > \frac{26.8}{\sigma}\right) = 0.2$$

$$1 - \text{prob}\left(z \le \frac{26.8}{\sigma}\right) = 0.2$$

$$\text{prob}\left(z \le \frac{16.8}{\sigma}\right) = 0.8$$

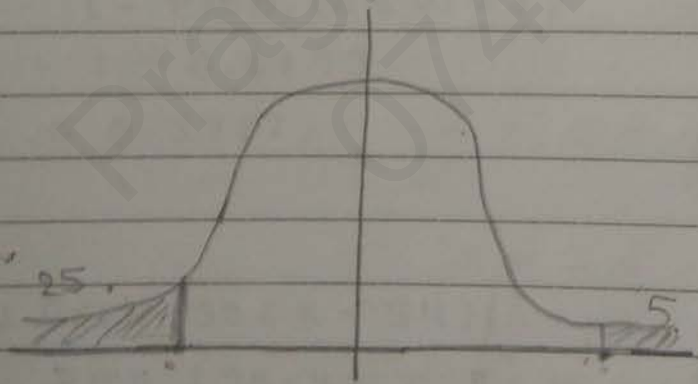
$$\frac{16.8}{\sigma} = 0.84$$

$$\sigma = 20.$$

Q.11 A set of examination marks ~~fa~~ is approximately normally distributed with mean of 75 and standard deviation of 5. If the top 5% of student get grade A and bottom 25% get grade F. what marks is the lowest A and what marks is the highest F?

⇒ Solution

$\mu = 75 \quad \sigma = 5$



$$\text{prob}(x < a) = 0.95$$

$$\text{prob}\left(\frac{x-\mu}{\sigma} < \frac{a-75}{5}\right) = 0.95$$

$$\text{prob}\left(z < \frac{a-75}{5}\right) = 0.95$$

Date:
Page:

$$\frac{a-75}{5} = 1.64$$

$$\therefore a = 83.2$$

$$\text{prob}(x < b) = 0.25$$

$$\text{prob}\left(\frac{x-\mu}{\sigma} < \frac{b-75}{5}\right) = 0.25$$

$$\text{prob}\left(z < \frac{b-75}{5}\right) = 0.25$$

$$\frac{b-75}{5} = -0.67$$

$$b = 71.65$$

Q. 4. The distribution is exactly normal, 7% of the items are under 35 and 63% are under 63. what are the

→ soln,

$$\mu = ?, \sigma = ?$$

$$\text{prob}(x < 35) = 0.07.$$

$$\text{prob.}\left(z < \frac{35-\mu}{\sigma}\right) = 0.07$$

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

Date:

Page:

$prob(x < 63) = 0.63$

$prob.\left(z < \frac{63-\mu}{\sigma}\right) = 0.63$

$\frac{63-\mu}{\sigma} = 0.33.$

on solving,

$\mu = 15.47$

$\sigma = 57.89.$

Gamma Distribution

A continuous random variable x is said to follow Gamma distribution with parameters α and β if its probability density function is given by

$$f(x:\alpha,\beta) = \frac{1}{\beta^{\alpha}\,\Gamma\alpha}\, x^{\alpha-1} e^{-x/\beta}, \quad x>0, \ \alpha>0, \ \beta>0.$$

$\alpha \rightarrow$ shape parameters

$\beta \rightarrow$ scale parameter

Mean $E(x) = \alpha\beta$

Variance $V(x) = \alpha\beta^2$

Uses

1. ~~Quen~~ Queuing theory :- waiting time distribution
2. Reliability test :- life span, failure time
3. To study distribution of water supply, electricity, petrol supply in a city. (volume).

Exponential distribution is special case of Gamma distribution when $\alpha = 1$.

$\alpha = 1$, the p.d.f. of exponential distribution is

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

Mean $E(x) = \beta$ $\quad V(x) = \beta^2$

By Sangita Acharya

Date:

Page:

Q. Suppose that the time in hours taken to repair a heat pump is a random variable having a gamma distribution $\alpha = 2,\ \beta = 1/2$. What is the probability that the next service call will require

a) at most one hour to repair the pump.

b) At least 2 hours to ~~remain~~ repair.

⇒ solution,

a) $$\text{prob.}(X \le 1) = \int_0^1 f(x)\,dx$$

$$= \int_0^1 \frac{1}{\beta^\alpha \Gamma\alpha}\, x^{\alpha-1}\, e^{-x/\beta}\,dx$$

$$= \int_0^1 \frac{1}{(1/2)^2\,\Gamma 2}\, x^{2-1}\, e^{-x/(1/2)}\,dx$$

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

$$= \cancel{4\left[e^{-2x} \cdot \frac{x^2}{2}\right.}$$

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

$=$

By Sangita Acharya

b) $\text{prob}(x > 2) = 1 - \int_0^2 f(x)\,dx$

$= 1 - \int_0^4 4 \cdot x e^{-2x}\,dx$

$=$

Q. In a certain city the dialy consumption of water (in millions of gallons) follows approximately a gamma distribution with $\alpha = 2$ and $\beta = 3$. If the daily capacity of this city is 9 millions of gallons of water. show that the probability that on any given day, the water supply is in adequate is $4e^{-3}$.

=) solution,

$\text{prob.}(X \geq 9) = 1 - \int_0^9 \frac{1}{3^2 \Gamma 2} x^1 e^{-x/3}\,dx$

$= 1 - \frac{1}{9} \int_0^9 x e^{-x/3}\,dx$

$= 1 - \frac{1}{9}\left[x \cdot \frac{e^{-x/3}}{-1/3} - \frac{e^{-x/3}}{(-1/3)^2}\right]_0^9$

$= 1 - 0.8$

$= 0.2$

$= 4e^{-3}.$

Date:

Page:

Q. Define exponential distribution. Suppose that the service life of a semiconductor is expotentially distributed with an average of 60 hrs. Find the probability that the semiconductor will

a) still working after 90 hrs

b) fail within 120 hrs.

→ Solⁿ.

$\beta = 60$

a). $P(x > 90) = 1 - \int_0^{90} \frac{1}{60} e^{-x/60} \, dx$

$= 1 - \frac{1}{60} \left[- e^{-x/60} \times 60 \right]_0^{90}$

$= \cancel{1 - 0.2864}$

$= \cancel{0.8}$

By Sangita Acharya

Date:

Page:

Chi-square Distribution

It is special case of Gamma distribution when $\alpha = \frac{\nu}{2}$ and $\beta = 2$. A continuous random variable

x is said to follow chi-square distribution with parameter ν degrees of freedom if its probability density function is given by.

$$f(x, \nu) = \frac{1}{2^{\nu/2} \Gamma \nu/2} x^{\nu/2 - 1} e^{-x/2}$$

Mean $E(X) = \nu$

$V(X) = 2\nu$

Uses

- To study sampling distribution of sample variance
- To test goodness of fit
- To test independence of attributes.
-

Chapter - 4.

Date:
Page:

Population, sample, parameter, statistic.

population unit
population data

(1) Finite population
(2) Infinite "

subset of population is sample.

parameter is function of population data. parameter is constant.

statistic is function of sample data. variable.

parameter (θ)	($\hat{\theta}$) statistic (Estimator)
population mean μ	$\bar{x}$
" S.D σ	s
" proportion P	p

sampling distribution of mean (sample mean)

The frequency distribution or probability distribution of all possible values of sample means computed from different samples drawn from same population is called sampling distribution of sample mean.

By Sangita Acharya

population $N = 4$ $\{3, 7, 11, 15\}$
sample $n = 2$ (WOR)

$^{N}C_{n} = {}^{4}C_{2} = 6$ ways.

Random sample

		$\bar{x}$
S_1	= 3, 7	= 5
S_2	= 3, 11	= 7
S_3	= 3, 15	= 9
S_4	= 7, 11	= 9
S_5	= 7, 15	= 11
S_6	= 11, 15	= 13

sample mean $\bar{x}$	frequency f	prob. $(\bar{x})$
5	1	1/6
7	1	1/6
9	2	2/6
11	1	1/6
13	1	1/6
	6	1

population mean $\mu = \dfrac{\Sigma x}{N} = 9$

population variance $\sigma^2 = \dfrac{\Sigma (x-\mu)^2}{N} = 20$

Mean of sample mean $E(\bar{x}) = \dfrac{\Sigma f \bar{x}}{\Sigma f} = 9$

$$\Sigma(\bar{x}) = \mu$$

variance of sample mean $V(\bar{x})$

$$\sigma_{\bar{x}}^{2} = \frac{\Sigma f(\bar{x}-\mu)^2}{\Sigma f} = \frac{\sigma^2}{n}\left(\frac{N-n}{N-1}\right)$$

finite population correction factor.

By Sangita Acharya

The standard deviation of sampling distribution of mean is called standard ~~me~~ error of mean.

$$S.E.(\bar{x}) = \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$$

WR.

N^n ways $= 4^2$ ways $= 16$ ways.

(3, 3) → 3	(7, 3) → 5	(11, 3) → 7	(15, 3) → 9
(3, 7) → 5	(7, 7) → 7	(11, 7) → 9	(15, 7) → 11
(3, 11) → 7	(7, 11) → 9	(11, 11) → 11	(15, 11) → 13
(3, 15) → 9	(7, 15) → 11	(11, 15) → 13	(15, 15) → 15

Sample Mean $\bar{x}$	frequency f	prob $(\bar{x})$
3	1	1/16
5	2	2/16
7	3	3/16
9	4	4/16
11	3	3/16
13	2	2/16
15	1	1/16
	16	1

Mean of sample mean $E(\bar{x}) = \frac{\Sigma f \bar{x}}{\Sigma f}$

$= 9$

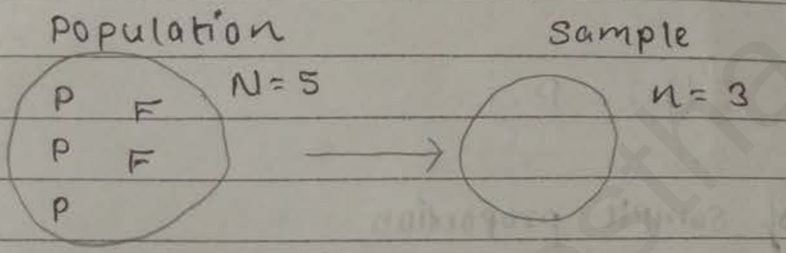
By Sangita Acharya

Date:
Page:

variance of sample mean $v(\bar{x}) = \frac{\sum f(\bar{x}-\mu)^2}{\sum f}$

$= 10.$

Sampling distribution of proportion



WOR

${}^{N}C_n$ ways $= {}^{5}C_3$ ways $= 10$ ways.

		p
S_1	= PFP	= 2/3
S_2	= PFF	= 1/3
S_3	= PFP	= 2/3
S_4	= PPF	= 2/3
S_5	= PPP	= 3/3
S_6	= PFP	= 2/3
S_7	= FPF	= 1/3
S_8	= FPP	= 2/3
S_9	= FFP	= 1/3
S_{10}	= PFP	= 2/3

sample proportion (p)	frequency (f)	prob. (P)
2/3	6	6/10
1/3	3	3/10
3/3	1	1/10
	10	1

Page:

Population proportion for pass $P = \frac{3}{5} = 0.6$

population porportion for fail $Q = 0.4$

Mean of sample proportion $E(p) = \frac{\Sigma f p}{\Sigma f}$

$= 0.6$

$\boxed{E(p) = P}$

variance of sample proportion

$V(p) = \sigma_p^2 = \frac{\Sigma f (p - P)^2}{\Sigma f} = \frac{PQ}{n}\left(\frac{N-n}{N-1}\right)$

$= 0.04.$

standard error of sample proportion

$S.E(p) = \sigma p = 0.2$

WR

N^n ways $= 5^3 = 125$ ways.

Population proportion for pass $= 0.6 = P$

" " " fail $Q = 0.4$

Mean of sample proportion $E(p) = \frac{\Sigma f p}{\Sigma f}$

$= 0.6$

$E(p) = P$

By Sangita Acharya

Date:
Page:

variance of sample proportion $v(p) =$

$$\sigma_p^2 = \frac{PQ}{n}$$

standard error of sample proportion

$$SE(p) = \sigma_p = \sqrt{\frac{PQ}{n}}$$

Q. A sample of size 64 is drawn from a population consisting of 256 units. If the population standard deviation is 16, find the standard error of sample mean when the sample drawn is without replacement and with replacement.

⇒ solution.

$n = 64$

$N = 256$

$\sigma = 16$

without replacement.

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

$$= 1.74$$

with replacement.

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

$$= 2$$

By Sangita Acharya

Date:
Page:

Q. A random sample of 196 apples was taken from a large consinment and 49 of them were found to be bad. find the error of the proportion of bad apple.

=) solution

$P = 0.75$

$Q = 0.25$.

$n = 196$.

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

$$= 0.031$$

Central Limit Theorem (CLT)

Let $x_1, x_2, x_3, \ldots x_n$ be a random sample of size n taken from any population with mean μ and variance σ^2, if sample size is sufficiently large, then sampling distribution of mean follows normal distribution with mean μ and variance $\frac{\sigma^2}{n}$.

Large sample $(n \geq 30)$

\# $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$

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

\# The distribution of sum S_n also follows normal distribution with mean $n\mu$ and variance $n\sigma^2$

$$S_n \sim N(n\mu, n\sigma^2)$$

$$\Rightarrow Z = \frac{S_n - n\mu}{\sigma\sqrt{n}} \sim N(0,1)$$

$p \sim N(P, PQ/n)$

$$Z = \frac{p - P}{\sqrt{PQ/n}} \sim N(0,1)$$

$(\bar{x} - \bar{y}) \sim N(\mu_1 - \mu_2, \sigma_1^2/n_1 + \frac{\sigma_2^2}{n_2})$

$$E(\bar{x} - \bar{y}) = E(\bar{x}) - E(\bar{y}) = \mu_1 - \mu_2$$
$$V(\bar{x} - \bar{y}) = V(\bar{x}) + V(\bar{y}) = \sigma_1^2/n_1 + \sigma_2^2/n_2$$

$$Z = \frac{(\bar{x} - \bar{y}) - (\mu_1 - \mu_2)}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)$$

$\cancel{(p_1 - p_2) \sim N(E(p_1) - E(p_2))}$,

$(p_1 - p_2) \sim N(P_1 - P_2, \frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2})$

$$Z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\left(\frac{P_1Q_1}{n_1} + \frac{P_2Q_2}{n_2}\right)}} \sim N(0,1)$$

Q. The time at the counter for a costomer to be served at a post office can be model as a random variable having mean 176 sec and variance 256. The sample mean $\bar{x}$ will be obtained from a random sample of 100 costomer. what is the probability that the $\bar{x}$ in between 175 and 178 sec.

⇒ Solution.

$\bar{x} = 176$ sec $\qquad n = 100$

$\sigma^2 = 256$

Date:
Page:

$$P(175 < x < 178) = \text{Prob.}\left(\frac{175-176}{16/10} < \frac{x-\mu}{\sigma/\sqrt{n}} < \frac{178-176}{16/10}\right)$$

$= \cancel{\text{prob.}(-0.0625 < z < 0.125)}$

$= \cancel{-1.15-}$

$= \text{Prob.}(-0.625 < z < 1.25)$

$= \text{Prob}(z \le 1.25) - \text{Prob}(z \le -0.625)$

$= 0.8944 - 0.2676$

$= 0.628$

Q. If the distribution of the weights of all men travelling by air between kathmandu and Nepalgunj has a mean of 163 pounds and σ of 18 pounds. What is the probability that the combined gross wt. of 36 men travelling on the plane between these two cities is more than 6000 pounds.

⇒ solution

$\bar{x} = 163 \quad \sigma = 18 \quad n = 36$

~~Prob. (x > [illegible])~~

$\text{Prob.}(S_n > 6000) = 1 - \text{Prob.}(S_n < 6000)$

$= 1 - \text{P}\left(\frac{6000 - 36 \times 163}{18\sqrt{36}}\right)$

$= 1 - \text{Prob.}(z < 1.222)$

$= 1 - 0.888$

$= 0.112$

By Sangita Acharya

Large sample	Small sample.
① $n \geq 30$	① $n < 30$
② The value of population S.D(σ) is generally known.	② The value of σ is always unknown.
③ If σ is unknown, we use sample S.D. S $S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n}}$	③ we use sample S.D. $S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2}{n-1}}$
④ we use normal distribution.	④ t-distribution
⑤ z-table	⑤ t-table.

If $n < 30$ and σ is known, we use normal distribution.

Estimation

The process of estimating population parameters using sample data is called estimation. Two types of estimation.

1. Point estimation
2. Interval estimation

If a single value computed from sample data is used to estimate population parameter θ, then the process is point estimation.

If an interval or range of values computed from sample data is used to estimate population

Date:
Page:

parameter θ then the process is called interval estimation.

The confidence interval can be expressed in the probability form as

$$\text{prob.}(C_1 < \theta < C_2) = 1-\alpha$$

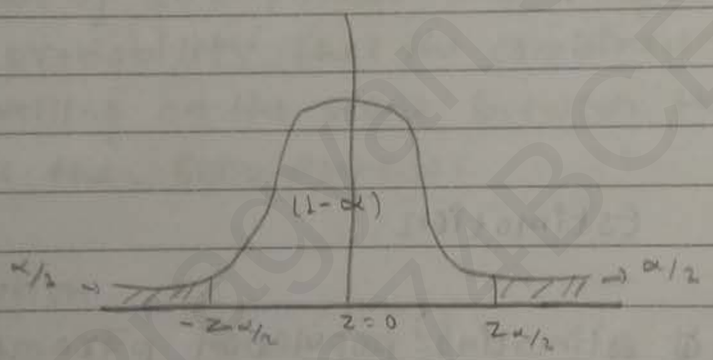
$1-\alpha \rightarrow$ confidence level
$\alpha \rightarrow$ significance level

$(1-\alpha)\,100\% = 95\%,\ 99\%.$

confidence interval for population mean μ for large sample $(n \geq 30)$.

By CLT, $\bar{x} \sim N(\mu, \sigma^2/n)$

$$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} \sim N(0,1)$$



$$\text{Prob}(-z_{\alpha/2} < z < z_{\alpha/2}) = 1-\alpha$$

$$\text{or prob.}\left(-z_{\alpha/2} < \frac{\bar{x}-\mu}{\sigma/\sqrt{n}} < z_{\alpha/2}\right) = 1-\alpha$$

$$\text{or prob.}\left(-z_{\alpha/2}\cdot\frac{\sigma}{\sqrt{n}} < \bar{x}-\mu < z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

$$\text{or prob}\left(\bar{x}-z_{\alpha/2}\frac{\sigma}{\sqrt{n}} < \mu < \bar{x}+z_{\alpha/2}\frac{\sigma}{\sqrt{n}}\right) = 1-\alpha$$

Date:
Page:

Then the confidence interval for population mean

$$\mu = \left(\bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \; , \; \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \right)$$

$$= \bar{x} \pm z_{\alpha/2} \cdot S.E.(\bar{x})\}$$

$|\bar{x} - \mu| =$ Max error (E).

$$E = z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

$n = ?$

$$\text{width of interval} = 2 z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$

confidence interval

Large same	small interval
$\mu \to \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$ for infinite population	$\mu \to \bar{x} \pm t_{\alpha/2, n-1}$
$\to \bar{x} \pm z_{\alpha/2} \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}}$ for finite population	$\mu_1 - \mu_2 = (\bar{x} - \bar{y}) \pm t_{\alpha/2, n_1+n_2-2}$
$P \to p \pm z_{\alpha/2} \cdot \sqrt{\frac{pq}{n}}$ - infinite	$s = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}}$
$\to p \pm z_{\alpha/2} \sqrt{\frac{pq}{n} \left(\frac{N-n}{N-1} \right)}$ finite.	
$\mu_1 - \mu_2 \to (\bar{x} - \bar{y}) \pm z_{\alpha/2} \cdot \sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}$	

By Sangita Acharya

Page:

$$P_1 - P_2 = (p_1 - p_2) \pm z_{\alpha/2} \cdot \sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}$$

Q. 1. A sample of 450 items is taken from a population whose standard deviation is 20. The mean of the sample is 30. Calculate 95% confidence limits for the mean age of all mean in that population.

=) solution

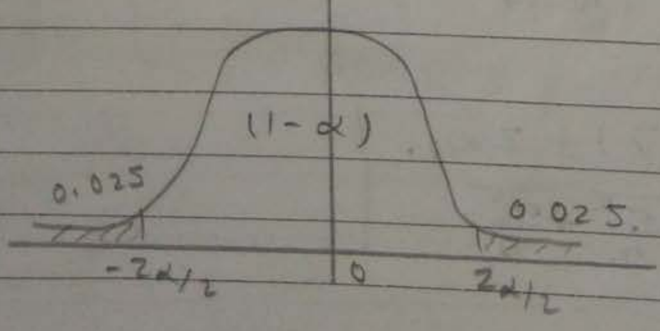
$(1-\alpha) = 95\%$

95% Lower confidence limit for population mean

$$\mu = \bar{x} - z_{\alpha/2} \frac{\sigma}{\sqrt{n}}$$

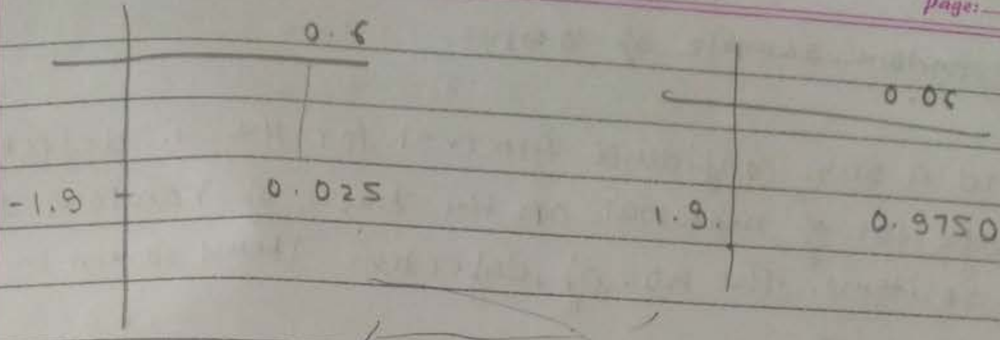
95% upper confidence limit for population mean

$$\mu = \bar{x} + z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}}$$



Date:

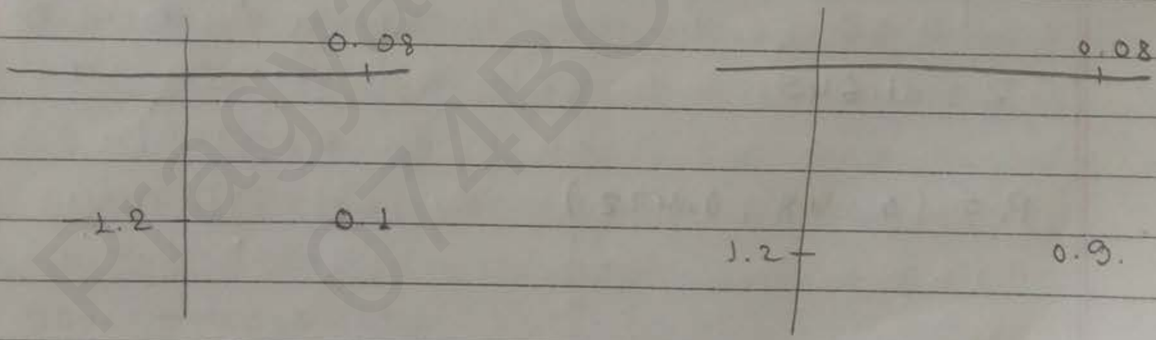
Page:



$r = 31.85$ $\quad$ $r = 28.152$.

2. In a random sample of 100 of 500 experimental batteries produced by a firm the lifetimes have a mean of 148.2 hours with a standard deviation of 24.9 hr. Find a 80% confidence interval for the mean life of the 500 batteries. (Using finite population N = 500)

⇒ solution,



$r = \bar{x} - z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} \sqrt{\frac{N-n}{N-1}} = \cancel{xxx.873}\ 145.346$

$r = \cancel{155.85}.\ 151.053$

3. ~~A random sample of 10 boys~~

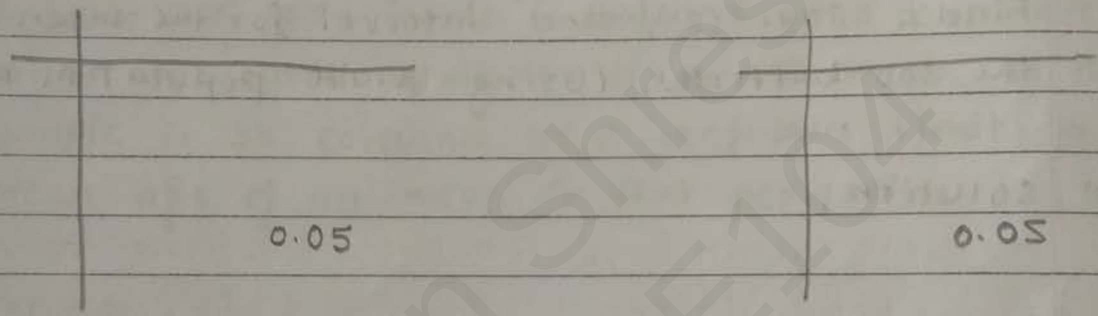
5. Find a 90% confidence interval for the % defective in a large lot of material on the basis of random sample of 50 items. The no. of defective items found in the sample is 18.

⇒ Soln.

$n = 50.$

$p = \frac{18}{50} = 0.36.$

$q = 0.64,$



$Z = 1.645$

$P = (0.248, 0.472)$

6. The dean of a college wants to use the mean of a random sample to estimate the average amount of time students takes to get from one class to the next, and she wants to be able to assert with 99% confidence that the error is at most 0.25 min. If it can presumed from experience that $\sigma = 1.4$ minutes, how large a sample will she have to take ?

Date:
Page:

⇒ solution

$$\text{Max. error} = 0.25 \text{ min}$$

$$z_{\alpha/2} \cdot \frac{\sigma}{\sqrt{n}} = 0.25.$$

99% confidence.

$$z_{\alpha/2} = z_{0.005} = 2.575.$$

Now.

$$2.575 \times \frac{1.4}{\sqrt{n}} = 0.25$$

$$n = 207.9$$

7. What is the maximum error one can expect to make with probability 0.9 when using the mean of a random sample of size $n = 64$ to estimate the mean of population with $\sigma^2 = 2.56$?

⇒ soln.

$$n = 64$$

$$\sigma = 1.6.$$

90% probability

$$z_{\alpha/2} = z_{0.05} = 1.645.$$

Now,

$$1.645 \times \frac{1.6}{\sqrt{64}}$$

$$= 0.329.$$

9. Random samples drawn from two countries given the following data relating to the heights of a adults male.

Males.	Country A	Country B
Mean ht. in inches	67.42	67.25.
S.D. (in inches)	2.58	2.50
No. of samples	1000	1200

find 98% confidence interval for $(\mu_1 - \mu_2)$.

⇒ soln.

$$Z_{\alpha/2} = Z_{0.01} = 2.33.$$

$$\mu_1 - \mu_2 \rightarrow (67.42 - 67.25) \pm 2.33 \sqrt{\frac{2.58^2}{1000} + \frac{2.5^2}{1200}}$$

$$= (-0.084, 0.4237)$$

10. A research worker wants to determine the average time it takes a mechanic to rotate the tires of a car, and she wants to be able to assert with 95% confidence that mean of her sample is off by at most 0.5 minute. If she can presume from the past experiment that $\sigma = 1.6$ min. how large a sample will she have to take?

⇒ soln.

$E. = 0.5.$

$\sigma = 1.6$ min.

$n = ?$

$Z_{0.025} = 1.96.$

So,

$$0.5 = \frac{1.96 \times 1.6}{\sqrt{n}} \quad \Rightarrow n = 39.$$

Date:

Page:

12. The campus chief wants to estimate the proportion of smokers among his students. What size of sample should be selected so as to have the proportion of smokers not to exceed by 10% with atmost certainty? It is believed from previous records that the proportion of smoker was 0.30.

2) Sol^n.

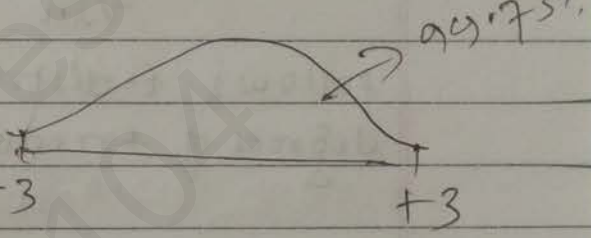
$(1-\alpha) = I$

$p = 0.30$ $\quad q = 0.70$

$n = ?$

$E = 0.1.$

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



$n = 189.$

Date:
Page:

t- distribution,

Q. 3.

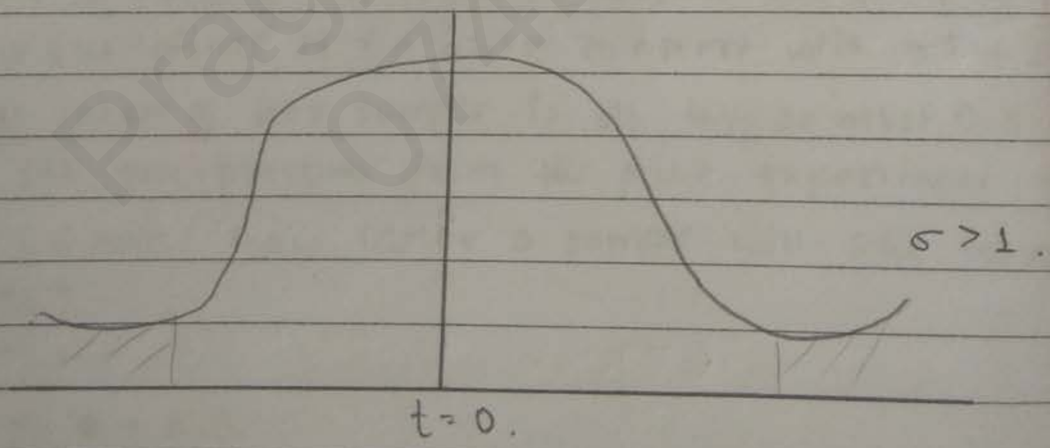
Assumptions.

1. Small sample $(n < 30)$
2. σ is always unknown
3. parent population should follow normal distribution
4. samples are drawn randomly and independent.

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \quad , \quad s = \sqrt{\frac{\Sigma(\ )^2}{n-1}}$$

follows t-distribution with parameter $\nu = n-1$ degrees of freedom.

The total number of values that can be choosen freely is called degree of freedom.



By Sangita Acharya

Date:
Page:

Q. 3. A random sample of 10 boys had the following IQ's: 70, 120, 110, 101, 83, 88, 95, 98, 107, 100. Find the reasonable range in which most of the mean IQ values of samples of 10 boys lie. ($\alpha = 5\%$)

→ Soln.

$n = 10$ (small sample)

$\bar{x} = \frac{\Sigma x}{n} = 97.2$.

Sample SD. $S = \sqrt{\frac{\Sigma (x_i - \bar{x})^2}{n-1}}$

$= \sqrt{\frac{\Sigma x_i^2 - n\bar{x}^2}{n-1}}$

$= \sqrt{\frac{96312 - 10 \times 97.2^2}{9}}$

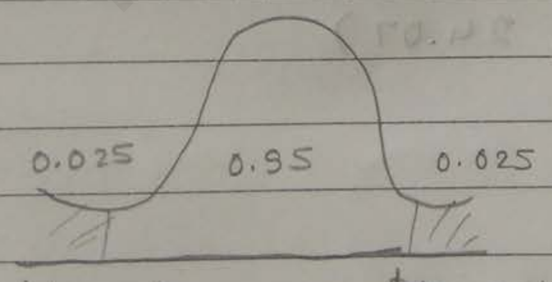
$= 14.27$

95% lower confidence limit for population mean

$r = \bar{x} - t_{\alpha/2, n-1} S/\sqrt{n}$

95% upper " " " " "

$r = \bar{x} + t_{\alpha/2, n-1} \frac{S}{\sqrt{n}}$



$-t_{\alpha/2, n-1}$ $t_{\alpha/2, n-1}$

-2.262 2.262

$r = (86.99, 107.41)$

8) Soln,

$n_1 = 6 \quad n_2 = 6.$

$\bar{x} = 127.33$

$\bar{y} = 129$

$$S = \sqrt{\frac{\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(\Sigma x_i^2 - n_1\bar{x}^2) + (\Sigma y_i^2 - n_2\bar{y}^2)}{n_1+n_2-2}}$$

$$= \sqrt{\frac{(100272 - 6\times127.33^2) + (100856 - 6\times129^2)}{6+6-2}}$$

$= 20.011$

95% lower confidence limit for different of population mean $(\mu_1-\mu_2) = (\bar{x}-\bar{y}) - t_{\alpha/2,\, n_1+n_2-2}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

95% upper " " " " :-

$(\mu_1-\mu_2) = (\bar{x}-\bar{y}) + t_{\alpha/2,\, n_1+n_2-2}\, S\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}$

$t_{\alpha/2} = 2.228.$

$(\mu_1-\mu_2) = (-27.41,\ 24.07)$

soln,

unbiased $\rightarrow E(\bar{x}) = \mu.$

Date:
Page:

Hypothesis.

It is a statement or assumption about population parameters developed for purpose of testing.

Example

$\mu = 1000$ hrs.

$\bar{x} = 995$ hrs.

The difference is insignificant.

$\bar{x} = 500$ hrs.

The difference is significant.

Testing of hypothesis

It is a procedure based on sample data and probability theory to determine whether a given statement is reasonable or not.

Procedure of testing of hypothesis

Step 1:- State Null hypothesis H_0 and Alternative hypothesis H_1

Null hypothesis → It is statement about population parameter θ.

Example:-

$H_0 : \mu = \mu_0 = 1000$ hrs

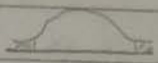
The average life of bulb is not different from 1000hr.

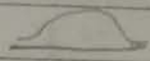
By Sangita Acharya

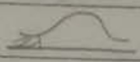
Date:
Page:

Alternative hypothesis → It is also statement about population parameter θ. which is accepted if sample data provides sufficient evidence to reject null hypothesis.

Example:-

$H_1 : \mu \neq 1000$ hrs  Two tailed test

$H_1 : \mu > 1000$ hrs  Right tailed test

$H_1 : \mu < 1000$ hrs.  left tailed test.

Step 2 :- Select significance level α. It is a max. value of probability of rejecting null hypothesis when it is true. Two types of errors in testing of hypothesis.

Type I error → Rejecting Null hypothesis when it is true.

α = Prob(Type I error) = Prob. (Rejecting null hyp. when it is true)

= Prob(Reject Ho | Ho)

Type II error ⇒ Accepting null hypothesis when it is false

β = Prob. (Type II error) = Prob. (Accepting null hyp. when Ho is false)

= Prob. (Accepting Ho | H_1)

Decision taken from sample.

Actual state	Accept Ho	Reject Ho
Ho true	Correct	Type I error
Ho false.	Type II error	Correct

By Sangita Acharya

Date:
Page:

Example :-

$H_0 : P = 6\%$

$H_1 : P > 6\%$

$50 \rightarrow 4 = 8\% \rightarrow \alpha$

$50 \rightarrow 2 \rightarrow 4\% \rightarrow \beta.$

Step 3 :- Select appropriate test statistic value.

It is a value computed from sample data which is used to make decision about accepting or rejecting null hypothesis.

$\mu \quad n \geq 30 \qquad z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

$n < 30 \qquad t = \dfrac{\bar{x} - \mu_0}{s/\sqrt{n}}$

$P \quad n \geq 30 \qquad z = \dfrac{p - P_0}{\sqrt{P_0 Q_0/n}}$

Test statistic value depends on

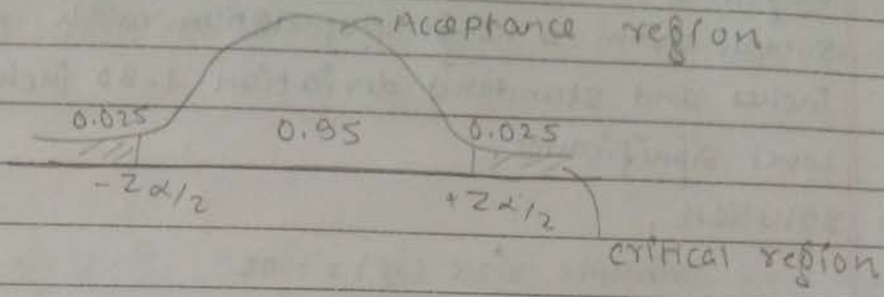
1. sample size
2. Type of test
3. probability distribution

Step 4 :- formulate decision rule based on step 1, 2 and 3.

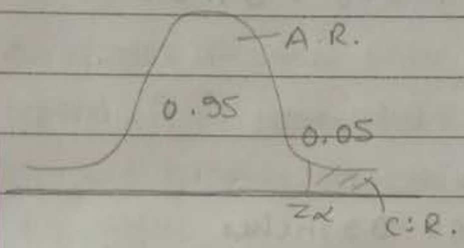
By Sangita Acharya

Two tailed test.

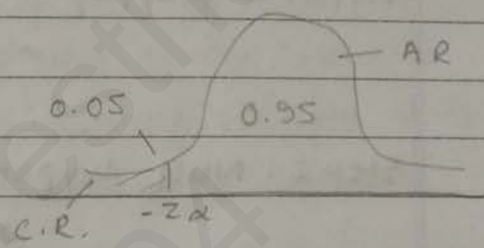
μ,
$n \geq 30$
$\alpha = 5\%$



Right tail test



Left tail test.



Critical value depends on
a) significance level (α)
b) Type of alternative hypothesis
c) Sample size
d) Probability distribution.

Step 5 :- Make a decision about Null hypothesis using test statistic value and critical value.

Test statistic $z = \dfrac{\bar{x} - \mu_0}{\sigma/\sqrt{n}}$

e.g. $z = 3.28$
critical value $z_\alpha = 1.645$

→ If $|z_{cal}| \leq z_\alpha$, we accept null hypothesis (H_0)
→ If $|z_{cal}| > z_\alpha$, we reject H_0 & accept H_1.

By Sangita Acharya

Date:

Page:

Q. 1. A sample of 400 male students is found to have a mean height 67.47 inches. Can it be reasonly regarded as a sample from a large population with mean height 67.39 inches and standard deviation 1.30 inches. Test at 5% level significance.

→ Solution,

Sample size (n) = 400

$\bar{x} = 67.47$

$\sigma = 1.30$

Hypothesized population mean $\mu_0 = 67.39$

$\alpha = 5\%$.

Step 1 : Null hypothesis : H_0

$\mu = \mu_0 = 67.39$ inches

Alt. hyp. $H_1 : \mu_1 \neq 6.39$

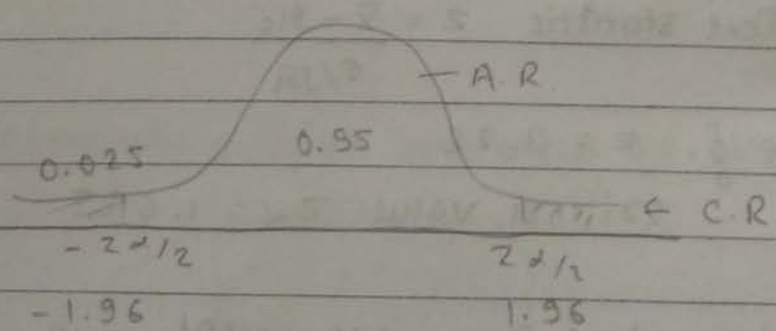
H_0 → There is no significant difference between population mean and sample mean

H_1 → There is significant difference between population mean and sample mean.

Step 2 : sig. level $\alpha = 5\%$.

Step 3 : Test statistic value

$$z = \frac{\bar{x} - \mu_0}{\sigma / \sqrt{n}} = 1.2307$$



By Sangita Acharya

Step 5 : The test statistic value 1.23 lies in acceptance region so we accept Null hypothesis H_0. So there is no significant difference between population mean and sample mean.

Q. 2. The mean breaking strength of cable supplied by a ~~manf.~~ manufacture is 1800 with a standard deviation of 100. By a new technique in the manufacturing process, it is claim that the breaking strength of the cable have increased. In order to test this claim a sample of 50 cables is tested, it is found that the mean breaking strength is 1850. can we support the claim at 0.01 level of significance.

=) Solution,

$n = 50$

$\bar{x} = 1850$

$\sigma = 100$

$\mu_0 = 1800$

$\alpha = 0.01$

Step 1:

Null hypothesis : $H_0 : \mu_0 = 1800$

Alt. hypothesis : $H_1 \neq 1800$

Step 2 :

sig. level $\alpha = 0.01$

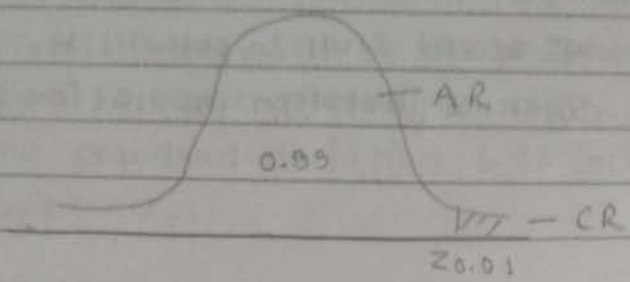
Step 3 :

Test statistic value.

$$Z = \frac{\bar{x} - \mu_0}{\sigma/\sqrt{n}} = 3.54$$

Date:
Page:

Step 4:



Right tail test = 2.33

Step 5:

The test statistic value is 3.54 lies in the critical region. So, we reject null hypothesis H_0 and accept alt. hyp. H_1. So, the mean breaking strength of cables have increased by new technique. in manufacturing process.

Q.4. The specimen of copper wires drawn from a large lot have the following breaking strength in (kg weight) 578, 572, 570, 568, 572, 573, 570, 572, 596, 544. Test whether the mean breaking strength of the lot may be taken to be 578 kg weight. Test at 5% level of significance.

=) Solution,

$\bar{x} = 571.5$ kg

$n = 10$ (small sample)

S.D., $S = 12.554$

$$S.D. = \sqrt{\frac{\sum x_i^2 - n\bar{x}^2}{n-1}}$$

$$= \sqrt{\frac{3267541 - 10 \times 571.5^2}{9}}$$

$$= 12.554$$

$\mu_0 = 578$ kg.

Date:
Page:

t-test.

Step 1: Null hypothesis : $H_0: \mu_0 = 578$ kg.

Alternate hypothesis : $H_1: \mu_1 \neq 578$ kg.

Step 2: sig. level = 5%.

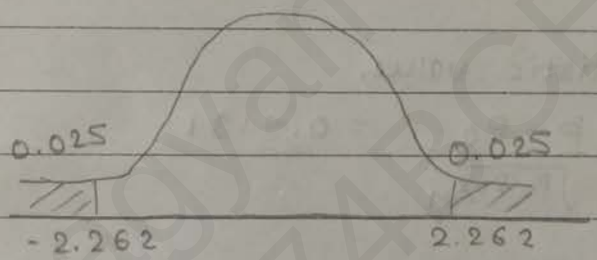
Step 3: Test statistic value

$$t = \frac{\bar{x} - \mu_0}{s/\sqrt{n}}$$

$$= \frac{571.5 - 578}{12.554/\sqrt{10}}$$

$$= -1.637$$

Step 4:



$Z = 2.262$

Step 5: The test static value lies in acceptance region. So, we accept Null Hypothesis. So, there is no significance difference between ~~poput~~ population mean and sample mean.

Date:

Page:

Q.8. A manufacture of transistors claims that the defective pieces can not be 10% in any lot. A sample of 60 transistors was drawn randomly. On testing ~~of~~ it was found that 7 transistors were out of orders. Test whether the manufacturer's claim is correct at $\alpha = 0.01$.

⇒ Solution,

$P_0 = 0.1$

$n = 60$

$Q_0 = 0.9$

$p = 7/60 = 0.116$

Step 1: Null hypothesis : $H_0 : P_0 = 0.1$

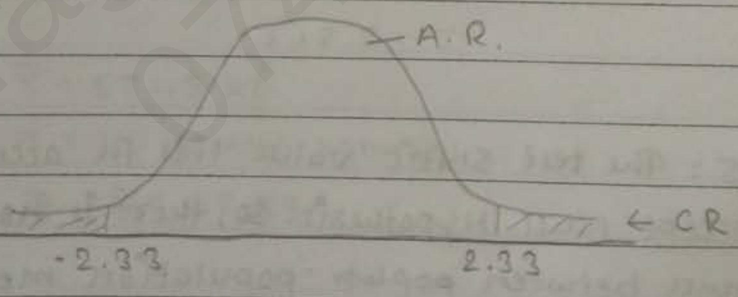
Alt. hypothesis : $H_1 : P_1 \neq 0.1$.

Step 2: sig. level $= 0.01$

Step 3: Test statistic value.

$$Z = \frac{p - P_0}{\sqrt{P_0 Q_0 / n}} = 0.4131$$

Step 4:

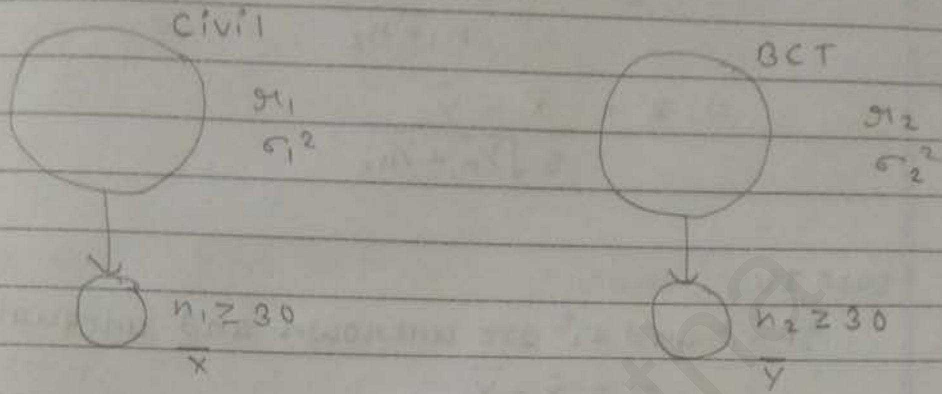


Step 5: lies in acceptance region. Hence, we accept Null hypothesis.

By Sangita Acharya

Date:
Page:

Significance test of difference of population means $(\mu_1 - \mu_2)$ for large sample $(n \geq 30)$



Step 1: Null Hyp. : $H_0 : \mu_1 = \mu_2$

Alt. Hyp. : $\left.\begin{array}{l} H_1 : \mu_1 \neq \mu_2 \\ H_1 : \mu_1 > \mu_2 \\ H_1 : \mu_1 < \mu_2 \end{array}\right\}$

Step 2: Sig. level $\alpha = 5\%, 1\%$.

Step 3: Test statistic

$$z = \frac{(\bar{x} - \bar{y}) - \overset{0}{(\mu_1 - \mu_2)}}{\sqrt{\sigma_1^2/n + \sigma_2^2/n_2}}$$

(for Null $\rightarrow \mu_1 = \mu_2$)

Case I

If σ_1^2 and σ_2^2 are known and equal.

$$z = \frac{\bar{x} - \bar{y}}{\sigma\sqrt{1/n_1 + 1/n_2}}$$

Case II

If σ_1^2 and σ_2^2 are known and unequal

$$z = \frac{\bar{x} - \bar{y}}{\sqrt{\sigma_1^2/n_1 + \sigma_2^2/n_2}}$$

Page:

Case III

If σ_1^2 and σ_2^2 are unknown and equal.

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2}$$

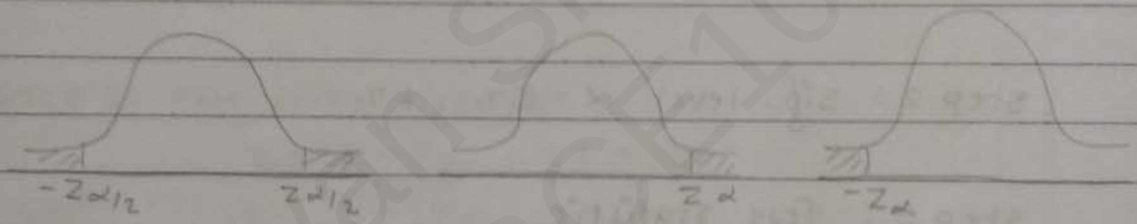
$$\Rightarrow Z = \frac{\bar{x} - \bar{y}}{S\sqrt{1/n_1 + 1/n_2}}$$

Case IV

If σ_1^2 and σ_2^2 are unknown and unequal.

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

Step 4:



Step 5: Decision

For small sample

use t-distribution

$$t = \frac{(\bar{x} - \bar{y})}{S\sqrt{1/n_1 + 1/n_2}}$$

$$S = \sqrt{\frac{\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2}{n_1 + n_2 - 2}}$$

By Sangita Acharya

Date: ________
Page: ________

Q. 11. A simple random sample height of 6400 English men has mean of 67.85 inches and S.D. 2.56. Australians has mean of 68.55 inches and S.D. of 2.52 inches. Do the data indicates that Australians are, on the average, taller than English men? Use $1 - \alpha = 99\%$.

=) Solution,

$\alpha = 1\%$.

English

$\bar{x} = 67.85$

$\sigma_1 = 2.56$

Australian

$\bar{y} = 68.55$

$\sigma_2 = 2.52$

$n_1 = n_2 = 6400$

Step 1: Null hypothesis : $H_0 : \mu_1 = \mu_2$

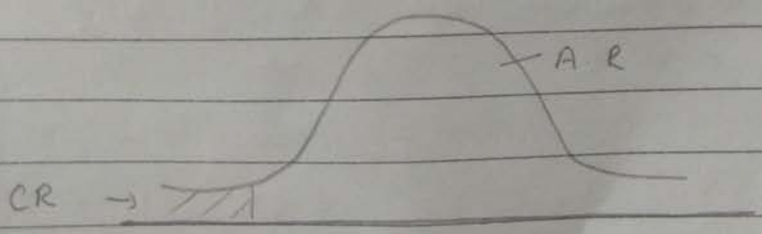
Alt. hypothesis : $H_1 : \mu_1 < \mu_2$

Step 2: sig. level $\alpha = 1\%$.

Step 3: Test statistic value

$$z = \frac{(\bar{x} - \bar{y})}{\sqrt{\sigma_1^2/n + \sigma_2^2/n}} = -15.59$$

Step 4: Left tail test.



$-z_\alpha = -2.33$

$|z| \geq z_\alpha$

Step 5: So, we reject Null hypothesis and hence we accept alternate hypothesis. Hence, Australians are taller than English men.

By Sangita Acharya

Significance test of difference of population proportion $(P_1 - P_2)$ for large sample $(n \geq 30)$

Case I:

Null hypothesis $H_0 : P_1 = P_2$

Step I:-

Alt. hypothesis $H_1 : P_1 \neq P_2$

$H_1 : P_1 > P_2$

$H_1 : P_1 < P_2$

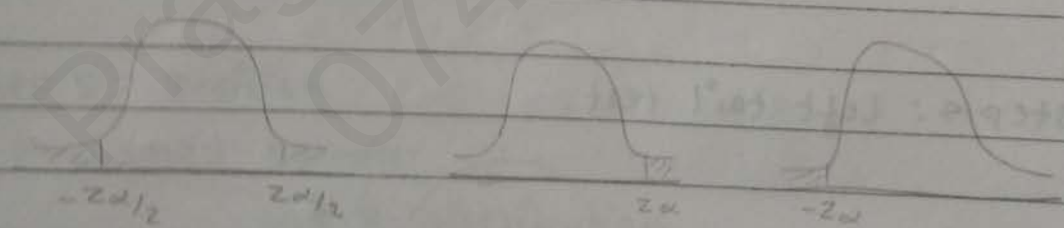
Step 2:

sig. level $\alpha = 5\%, 1\%$

Step 3:

Test statistic $z = \dfrac{(p_1 - p_2) - (P_1 - P_2)}{\sqrt{\hat{P}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$

Step 4:-



Case II

$H_0 : P_1 - P_2 = d_0$

$H_1 : P_1 - P_2 \neq d_0$

$H_1 : P_1 - P_2 > d_0$

$H_1 : P_1 - P_2 < d_0$

$$z = \frac{(p_1 - p_2) - d_0}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

Q.6. In a simple random sample of 600 high school students from a city, 400 are found to use dot pens. In one of 900 from a neighbouring city, 450 are found to use dot pens. Do the data indicates that the cities are significantly different with respect to the habit of using dot pens among the students.

⇒ Solution,

$n_1 = 600$ $\qquad$ $n_2 = 900$

$p_1 = \frac{400}{600} = \frac{2}{3}$ $\qquad$ $p_2 = \frac{450}{900} = 0.5$

$= 0.667$ $\qquad$ $q = 0.5$.

$q = 0.33$

$$\hat{p} = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = 0.567$$

$$\hat{Q} = 0.433.$$

Step 1:

Null hypothesis : $H_0 = P_1 = P_2$

Alt. hyp. : $H_1 = P_1 \neq 2$.

Step 2:

sig. level $\alpha = 5\%$.

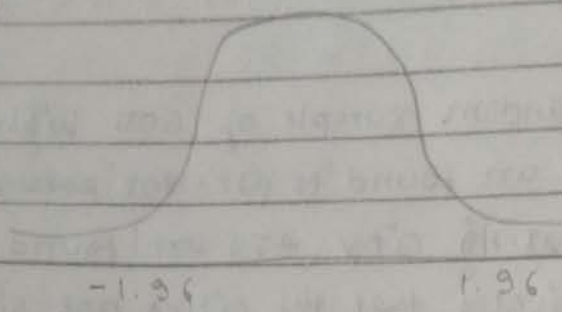
Step 3:

$$z = \frac{p_1 - p_2}{\sqrt{\hat{p}\hat{Q}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = 6.394$$

By Sangita Acharya

$z_{0.025} = 1.96.$

Step 4.



Step 5

Decision: The test statistic value 6.39 lies in the critical region so we reject null hyp. H_0 and accept alt. hyp H_1 so the cities are significantly different w.r.t. the ~~b~~ habit of using dot pens.

Q. 7. A cigarette-manufacturing firm claims that its brand A of the cigarettes out sells its brand B by 8%. If it is found that 42 out of sample of 200 smokers prefer brand A and 18 out of another random sample of 100 smokers prefer brand B, test whether the 8% difference is a valid claim. ($\alpha = 5\%$.)

⇒ solution,

$n_1 = 200$ $\qquad$ $n_2 = 100$

$p_1 = \dfrac{42}{200} = 0.21$ $\qquad$ $p_2 = \dfrac{18}{100} = 0.18$

$q_1 = 0.79$ $\qquad$ $q_2 = 0.82$

Step 1: Null hyp: $H_0: P_1 = P_2$

Alt. hyp: $H_1: P_1 \neq P_2$.

Step 2:

$\alpha = 5\%$.

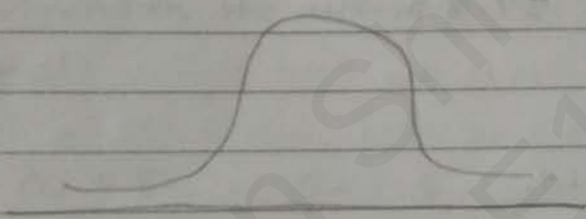
Step 3.

$$Z = \frac{p_1 - p_2 - d_0}{\sqrt{\frac{p_1 q_1}{n_1} + \frac{p_2 q_2}{n_2}}}$$

$$= -1.041$$

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

Step 4:



-1.96 1.96

Step 5:

$|Z| = 1.041$ lies in the acceptance region.

Hence we accept null hypothesis.

Q14. Sample of types of electric light bulbs were tested for length of life and following data were obtained.

	Type I	Type II
Sample size	$n_1 = 8$	$n_2 = 7$
Sample mean	$\bar{x}_1 = 1234$	$\bar{x}_2 = 1036$
Sample S.D	$s_1 = 36$	$s_2 = 40$

By Sangita Acharya

If the difference in the means sufficient to warrant that type I is superior to type II regarding length life? (use $\alpha = 1\%$)

⇒ solution.

Step 1:

Null hyp. $H_0 : \mu_1 = \mu_2$

Alt. hyp. $H_1 : \mu_1 > \mu_2$

$$S = \sqrt{\frac{\Sigma (x_i - \bar{x})^2 + \Sigma (y_j - \bar{y})^2}{n_1 + n_2 - 2}}$$

$$= \sqrt{\frac{(n_1 - 1) s_1^2 + (n_2 - 1) s_2^2}{n_1 + n_2 - 2}}$$

$$= 37.89.$$

Step: 2

$\alpha = 1\%.$

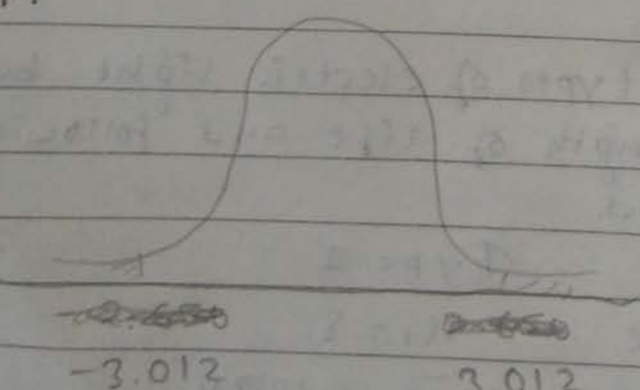
Step 3

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - (\mu_1 - \mu_2)}{S \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= 10.036$$

$\frac{\alpha}{2}, n_1 + n_2 - 2$

Step 4.



Step 5

Decision: t lies in C.R. hence we reject Null hyp. and accept alt. hyp. H_1. So, the type I are superior than type II regarding length of life.

Q15. The following random samples are measurements of the heat producing capacity (in millions of calories per ton) of specimens of coal from two mines.

Mine 1: 8260 8130 8350 8070 8340

Mine 2: 7250 7890 7900 8140 7920 7840

Use the 0.01 level of significance to test whether the difference between the means of these two samples is significant.

⇒ Solution,

$$n_1 = 5 \qquad n_2 = 6$$

$$\bar{x}_1 = 8230. \qquad \bar{x}_2 = 7823.33$$

$$s_1^2 = \frac{\Sigma (x_i - \bar{x})^2}{n_1 - 1} \Rightarrow s_1 = 125.499$$

$$s_2 = 299.644$$

$$S = \sqrt{\frac{(n_1-1)s_1^2 + (n_2-1)s_2^2}{n_1+n_2-2}} = 238.495.$$

Step 1:

Null hyp. $H_0: \mu_1 = \mu_2$

Alt. hyp $H_1: \mu_1 \neq \mu_2$

Step 2:

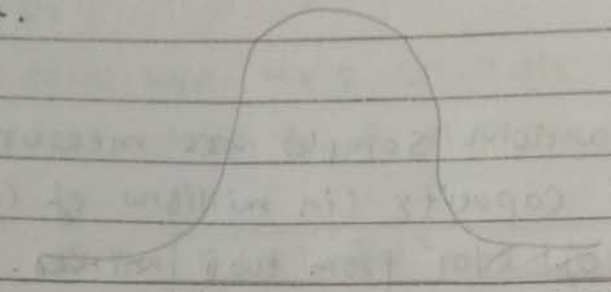
$$\alpha = 0.01.$$

Step 3:

$$t = \frac{(\bar{x}_1 - \bar{x}_2) - 0}{S\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$t = 2.8142$

Step 4.



-3.250 $\quad$ $+3.250$

Step 5

Value of t lies in the acceptance region. Hence we accept null hyp. H_0 and reject alt hyp H_0. Hence there is no significant difference between the means of these two samples.

paired t-test.

Assumptions

1. Small sample $(n < 30)$
2. Equal sample size for pair of data
3. Every pair of data (x_i, y_i) is dependent.

Before	After	Difference = Before - after
x_i	y_i	$d_i = x_i - y_i$

By Sangita Acharya

$$\bar{d} = \frac{\Sigma d_i}{n}$$

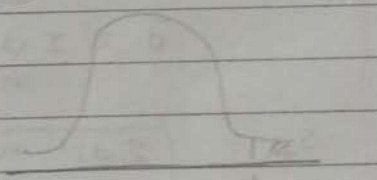
$$S_d = \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n-1}}$$

Step 1: Null hyp. H_0 $\bar{D} = 0$
The program is not effective.
Alt hyp. $H_1: \bar{D} \neq 0$
$H_1: \bar{D} > 0$
$H_1: \bar{D} < 0$ The program is effective.

Step 2:
sig. level $\alpha = 5\%, 1\%$

Step 3
Test statistic $t = \dfrac{\bar{d}}{S_d/\sqrt{n}}$

Step 4:



Step 5:
Decision.

By Sangita Acharya

Q. Marks of 8 students before and after tuition is given below.

Before tution	50	54	52	53	48	51	53	54
After tution	54	57	54	55	52	56	56	55

Can you conclude that the tution has benifited the students? ($\alpha = 5\%$)

⇒ Solution,

Before x_i	After y_i	Difference $d_i = x_i - y_i$
50	54	-4
54	57	-3
52	54	-2
53	55	-2
48	52	-4
51	56	-5
53	56	-3
54	55	-1
		$\Sigma d_i = -24.$

$$\bar{d} = \frac{\Sigma d_i}{n} = -3$$

$$s_d = \sqrt{\frac{\Sigma d_i^2 - n\bar{d}^2}{n-1}} = 1.309$$

Step 1:

Null hyp. $H_0 : \bar{D} = 0$

Alt hyp. $H_1 : \bar{D} < 0$

Step 2:

$\alpha = 5\%$

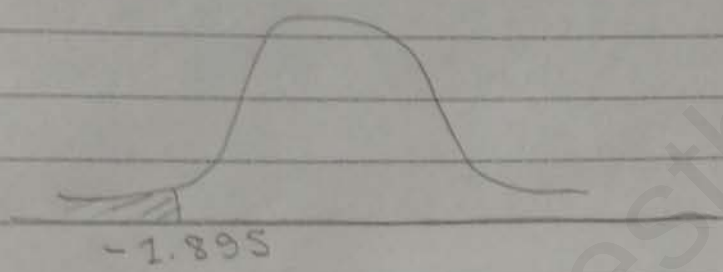
By Sangita Acharya

Step 3:

$$t = \frac{\bar{d}}{s_d/\sqrt{n}} = -6.482$$

Step 4:

left tail test.



By Sangita Acharya

Probability & statistics for engineers & scientists.
- J.L. Devore

Experiment

- Operation that can yield at least an ~~op~~ output.

a) Deterministics
b) Random

Deterministics mean such experiments whose outcomes are fixed or can already be predicted.
i) Immersion of blue litmus paper into an acidic sol^n.
ii) Determination of value of 'g' at pulchowk campus lab.

Random mean such experiments whose outcomes are unknown out of known set of outcomes.
i) Drawing a ball from a box containing 5 white, 3 blue and 4 red balls.
ii) Drawing (or selecting) an item from a lot containing 7 good items and 3 defective items.

Difference
$P(\text{Deterministics}) = 1$
$0 \le P(\text{random}) \le 1.$

By Sangita Acharya

Sample space

The set of all possible outcomes of a random experiment.

Expt : Rolling a dice

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

finite | infinite

countably infinite

Event

Any subsets of a sample space.

Define

A = getting an even number in rolling a dice.

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

$A \subset S$.

$\therefore$ A is an event.

$B = \{8, 9, 10\}$

$B \not\subset S$

B cannot be an event in S.

C = getting a number more than 6 in rolling a dice

$= \emptyset$

$C \subset S \quad (\because \emptyset \subset S)$

$\therefore$ C is an event in S.

Page:

Trial

In an experiment each successive operation is known as a trial.

$P(E) = \frac{m}{n} = \frac{\text{No. of fav. cases}}{\text{Total no. of cases}}$

↓

PRIORI
or
classical definition

Q. what are the shortcoming of priori defination of probability?

→ can not address natural phenomenon or the cases whose sample space is not determined in advance

If E be an event, then probability of event E is :

$P(E) = \frac{\text{No. of cases favourable to E in S.}}{\text{Total no. of cases}}$

where S = Sample space

This is PRIORI or CLASSICAL definition of probability.

Exhaustive cases

S = sample space

A, B, C, D are events in S.

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

$A = \{1, 3, 5\}$

$B = \{2, 4, 6\}$

$C = \{1\}$

$D = \{1, 3, 5, 6\}$

we find $A \cup D = S$.

$\therefore$ A and D are exhaustive cases.

But

$A \cup C = \{1, 3, 5\} \neq S$

So A and C are not exhaustive cases.

Let S be the sample space of a random experiment. ~~Every~~ atleast any two or more events are said to be exhaustive if their union is sample space

odds in favour of an event E

$$= \frac{\text{No. of cases fav. to } E}{\text{No. of cases not fav. to } E}$$

odds against an event E

$$= \frac{\text{No. of cases not fav. to } E}{\text{No. of cases fav. to } E}$$

Mutually exclusive and independent events:

Q. Can two events be mutually exclusive and independent simultaneously? Give support to your answer.

$\rightarrow$ Let S be sample space of a random experiment. Let A and B be any two events in S. These events are mutually exclusive if $A \cap B = \phi$. In other words two events in a trial are mutually exclusive if occurance of one completely

~~banned~~ bans the occurance of other. Eg :-

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

A = {1,

Independent events

A set of events in a multiple trail or atleast in two trial are said to be independent if occurance of one does not affect the occurance of other or is not affected by the occurance of other.

Ex :-

No, two events can not be mutually exclusive and independent simultaneously. As we know to measure the mutually exclusiveness of the event we required events from the same sample space in the single trial. In the other hand to measure the independency of the event we required more than one trial and they may be from some or different sample space.

Draw

- Successive (one by one or one at a time)
  - with replacement
  - without replacement
- Random (more than one at a time)
  - use of combination

1. A bag contains 3 red, 2 white & 4 black-balls. A ball is drawn. what is the probability that it is red?

$P(R) = \frac{3}{9}$

2. A bag contains 3 red, 2 white & 4 black-balls. 2 balls are drawn at random. what is the prob that both are red?

→ Total no. of cases = $C(9,2)$

favourable cases for red = $C(3,2)$

$\therefore P(\text{both red}) = \frac{C(3,2)}{C(9,2)}$

Addition theory of probability

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

proof:-

let S be the sample space of a random experiment. Now, probability of any event E in S is,

$$P(E) = \frac{\text{no. of cases fav. to E}}{\text{Total no. of cases}} = \frac{n(E)}{n(S)}$$

From set theory. we have.

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

Dividing throughout by $n(S)$. we have.

$$\frac{n(A \cup B)}{n(S)} = \frac{n(A)}{n(S)} + \frac{n(B)}{n(S)} - \frac{n(A \cap B)}{n(S)}$$

$$P(A \cup B) = P(A) + P(B) - P(A \cap B) \quad (\text{from } 1).$$

NOTE:- Here $A \cap B$ is the simultaneous occurrence of A and B.

Q. A card is drawn from a deck of 52 cards. what is the probability that either it is red or Jack.

=) let,

A and B be the events of getting a red and a Jack respectively.

P(either red or a Jack) = $P(A \cup B)$.

The event of simultaneous occurrance = $A \cap B$ = getting a red jack.

Now,

$$P(A) = \frac{26}{52} = \frac{n(A)}{n(S)} = \frac{1}{2}.$$

$P(B) = \frac{n(B)}{n(S)} = \frac{4}{52} = \frac{1}{13}$

$P(A \cap B) = \frac{n(A \cap B)}{n(S)} = \frac{2}{52} = \frac{1}{26}.$

By addition theorem of probability.

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

$= \frac{1}{2} + \frac{1}{13} - \frac{1}{26}$

Q. 4 defective items are suddenly mixed with 7 good items. 3 items are drawn at random. what is the probability that the 2 are defective and 1 is good in the drawn sample?

⇒ Total items = 4+7 = 11 items

Total no. of cases = No. of ways in which 3 items are selected from 11 items
= C(11, 3).

favourable cases for 2 defective and one good item = No. of ways of selecting 2 items out of 4 & no. of ways of selecting 1 items from 7. simultaneously.

$= C(4, 2) \times C(7, 1)$

∴ ~~P(2 defective)~~

$P(2 \text{ defective } \& \text{ } 1 \text{ good}) = \frac{C(4, 2) \times C(7, 1)}{C(11, 3)}$

Page:

Conditional Probability

Two events A and B are ~~not~~ said to be conditional if occurance of one of them is followed by the occurance of other.

Example :-

if B occurs only when the event A has already occured then we say it is the conditional event with probability, $P(B/A)$.

Multiplicative law of probability.

$P(A \cap B) = P(A) \cdot P(B/A)$. (conditional dependent) — ①

Q. A box contains 5 white 3 black and four red balls. 2 balls are drawn in succession. what is the probability of getting red ball in first drawn & white ball in 2nd drawn if the second drawn is performed without replacement of the ball in First drawn ?

$\Rightarrow$ $P(R \text{ and } W) = P(R) \cdot P(W/R)$.

By Sangita Acharya

NOTE.

when A and B are independent then

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

- $P(A)$ and $P(B)$ are the probabilities of independent events A & B.

1. probability of happening both simultaneous

$= P(A) \cdot P(B)$.

2. probability of happening at least one of them (either A or B).

$= 1 - P(\bar{A}) \cdot P(\bar{B})$

3. Exactly one of them

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

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

$= P(A) \cdot P(\bar{B}) \text{ or } P(\bar{A}) \, P(B)$

Q. A speaks the truth 75% of a time and B 80% a time. what is the probability that they will contradic in narrating the same massage.

By Sangita Acharya

Q. Distinguish between conditional and marginal probabilities with example of your own.

⇒ Suppose a group of 100 persons from mountain, hilly and terai region were asked whether they smoke. The result of inquiry has been tabulated as below. (Suppose).

	Smoker	Non-smoker	Total
Mountain	15	28	43
Hilly	12	22	34
Terai	13	10	23
Total	40	60	100

Suppose that one person is selected from these 100 persons and we wish to find the probability that the person is smoker.

The probability is $\frac{40}{100} = \frac{2}{5}$.

Since this is the probability obtained from margin of the table, it is called marginal probability. Again if it is known that the person is smoker and we wish to find the probability of him/her from terai region (say), it becomes the case of conditional probability. The conditional in this sense that the one of the event has already occured and within it we are finding the probability of the event associated with it. This probability is denoted by $P(\text{Terai}/\text{smoker}) = \frac{13}{40}$

Q. A couple has two children. find the probability that both are boys if it is known that
i) one of the children is a boy
ii) the older child is a boy.

⇒ solution,

Denoting B_1, B_2 & G_1, G_2 the boys and girls in their birth order, then sample space is

$$S = \{B_1B_2, BG_2, G_1B_2, G_1G_2\}$$

Let A = event of getting both boys = $\{B_1B_2\}$

B = event of getting one of two children is boy
$= \{B_1B_2, B_1G_2, G_1B_2\}$

$A \cap B = \{B_1B_2\}$

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

$$= \frac{\frac{n(A\cap B)}{n(S)}}{\frac{n(B)}{n(S)}}$$

$$= \frac{n(A\cap B)}{n(B)}$$

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

Q. prove that $P(A|B) = \frac{P(A\cap B)}{P(B)}$.

⇒ Soln,

for events A and B in S,

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

BG GG
BB GG

By Sangita Acharya

$$P(A \cap B) = \frac{n(A \cap B)}{n(B)} \cdot \frac{n(B)}{n(S)}$$

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

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

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

proved

Q. A natural number is choosen from 500 initial numbers what is the probability that no. so choosen is dividible by 3 or 5.

→ solution,

$S = \{1, 2, 3, \ldots 500\}$

A = event of getting a no. divisible by 3

B = " " " " " 5

$A \cap B$ = " " " divisible by both 3 & 5

i.e. L.C.M of 3, 5 i.e. 15

$n(A) = 166$

$n(B) = 100$

$n(A \cap B) = 33$

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

Odd against A = $\frac{\text{No. of cases not fav. to A}}{\text{No. of cases fav. to A}}$.

Date:
Page:

Q. The probabilities of occurance of two events E and F are 0.25 and 0.50 respectively. The probability of their simultaneous occurance 0.14. find the probability that neither E nor F occur.

Q. Let E and F be the events such that $P(E)=0.3$, $P(E \cup F)=0.4$ and $P(F)=x$. find the value of x such that
a) E & F are mutually exclusive
b) E & F are mutually independent.

=) Soln.

b) $P(E \cap F) = P(E).P(F)$

Q. Probability that A hits a target is 1/3 and that B hits it is 2/5. If both of them hit the target what is the probability that
a) target will be hit
b) both will hit the target
c) ~~exact~~ exactly one of them will hit.

=) Soln.

a) $1 - P(\bar{A})\,P(\bar{B})$
b) $P(A).P(B)$
c) $P(A)\,P(\bar{B}) + P(\bar{A}).P(B)$.

Q. The odds against a husband who is 45 years old, living till he is 70 are 7:5. And the odds against wife who is now 36 living till 61 are 5/3. find the probability that exhaustive → सबैको union ले ~~sp~~ sample space दिन्छ।

Date:
Page:

a) the couple will be alive 25 years hence
b) at least one of them will be alive 25 years hence

→) Soln.

a) $P(A) \cdot P(B)$

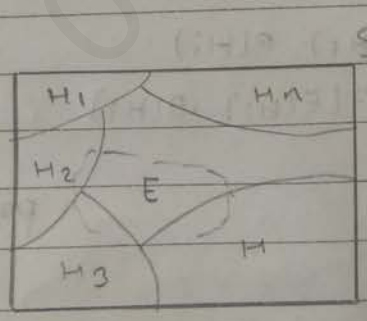
b) $1 - P(\bar{A}) P(\bar{B})$

Baye's Theorem
(Reversal law of probability)

Statement : Let $H_1, H_2, H_3, \ldots, H_n$ be mutually exclusive and exhaustive events of a random experiment with $P(H_i) \neq 0$; $i = 1, 2, 3, \ldots n$ and E be any event in the same sample space; with $P(E) \neq 0$. Then,

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

proof :-



Let S be the sample space of a random experiment. Given that $H_1, H_2, H_3, \ldots, H_n$ are mutually exclusive and exhaustive events in S and E is any event in S.

Then,

$$H_i \cap H_j = \emptyset \text{ for } i \neq j \text{ \&}$$

$$H_1 \cup H_2 \cup H_3 \cup \dots H_n = S$$

Now,

$$E = E \cap S \quad [\because E \subset S]$$

$$E = E \cap [H_1 \cup H_2 \cup H_3 \cup \dots \cup H_n]$$

$$\text{a } E = (E \cap H_1) \cup (E \cap H_2) \cup (E \cap H_3) \cup \dots \cup (E \cap H_n)$$

$$\therefore P(E) = P[(E \cap H_1) \cup (E \cap H_2) \cup (E \cap H_3) \cup \dots \cup (E \cap H_n)]$$

$\because$ $H_i \cap H_j \neq 0$; we find $(E \cap H_i)$ and $(E \cap H_j)$ are mutual exclusive.

$$\therefore P(E) = P(E \cap H_1) + P(E \cap H_2) + P(E \cap H_3) \dots + P(E \cap H_n)$$

$$\boxed{P(E) = \sum_{i=1}^{n} P(E \cap H_i)} \quad \text{--- (1)}$$

Next.

$$P(H_i/E) = \frac{P(H_i \cap E)}{P(E)}$$

$$P(H_i/E) = \frac{P(E/H_i) \cdot P(H_i)}{\sum_{i=1}^{n} P(E \cap H_i)}$$

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

proved

NOTE

The probability of E as indicated in eq. (1) is called total probability.

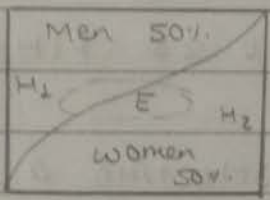
Q. In a group of equal no. of men & women, 20% of men and 30% of women are unemployed. If a person is selected at random, what is the probability that the selected person is employed?

⇒ Solution,

let H_1 = event of getting a men

H_2 = " " " " women

& E = " " " an employed person.



$$P(E) = \sum_{i=1}^{2} P(E \cap H_i)$$

$$= \sum_{i=1}^{2} P(E/H_i) \cdot P(H_i)$$

$$= P\left(\frac{E}{H_1}\right) \cdot P(H_1) + P\left(\frac{E}{H_2}\right) \cdot P(H_2)$$

$$= 0.8 \times 0.5 + 0.7 \times 0.5$$

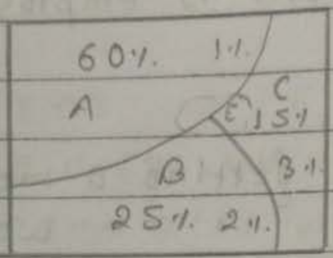
$=$

Q. In a ~~ball~~ bulb factory 3 machines A, B & C manufactured 60%, 25% and 15% of total production. Of the total output 1%, 2% & 3% are respectively defective from the machine. The ~~ba~~ bulb is drawn at random from the product & it is found to be defective. find the probability that it was

Page:

manufactured by machine C.

⇒ Solution.



Let H_1, H_2, H_3 be events of selecting the machines A, B and C respectively.

$\therefore P(H_1) = 0.6 \quad P(H_2) = 0.25 \quad P(H_3) = 0.15$

Let E = probability of selecting a defective bulb.

$P(E/H_1) = 0.01$

$P(E/H_2) = 0.02$

$P(E/H_3) = 0.03$

Now, to find

$P(H_3/E) = ?$

We know,

$$P(H_3/E) = \frac{P(E/H_3) \cdot P(H_3)}{\sum_{i=1}^{3} P(E/H_i) \cdot P(H_i)}$$

$$= \frac{P(E/H_3) \cdot P(H_3)}{P(E/H_1) \cdot P(H_1) + P(E/H_2) \cdot P(H_2) + P(E/H_3) \cdot P(H_3)}$$

$$= \frac{0.03 \times 0.15}{0.01 \times 0.6 + 0.02 \times 0.25 + 0.03 \times 0.15}$$

$=$

Page.

Random variables

Q. Define random variable with example.

→ Let,

$S = \{\omega_1, \omega_2, \omega_3, --- \omega_n\}$ be the sample space of a random experiment.

Let, x be the rule ~~with~~ which assigns each element of S a real number. In other words, let

$x(\omega_i) = x_i$ where $\omega_i \in S$ &

$x_i \in R$

Range of x is called the random variable.

Example.

1) let us toss a coin twice. Let H & T denotes the events of getting a head & tail in single toss. Then the sample space

$S = \{HH, HT, TH, TT\}$

suppose that x is the rule or represents the no. of head in the elements of S. Then clearly.

$x(HH) = 2$, $x(HT) = 1$, $x(TH) = 1$, $x(TT) = 0$

Now, x is a random with values 0, 1, 2.

2) Let us measure the distance between two consecutives bus stops along Arniko highway. Let x represents the distance between two consequetive bus stops in kilometers then

x can assume the values like 0.66 km, 2 km, 2.06, 3.2 km, 4 km etc in the range say 0.5 km to 5 km (0.5 km, 5 km). Then x is a random variable assuming all values lying in the interval (0.5, 5)

Distinguish between discrete and continuous random variables with examples.

A random variable x is said to be discrete random variable (DRV). If it takes only integral value from the

On the other hand random variable is said ~~b~~ to be continuous if it can assume all the decimal and integral value in the given range or interval.

Example :-

1. Let us record the attendence of BCE EF 2074 per day for a month in the first period. Let ~~x~~ denotes ~~d~~ or measure ~~o~~ the attedence per day. Then x can take the value 0, 15, 29, 48 etc ... Now since x takes only integral value, x is a DRV.

2. Let us measure the weight of individual student of a class EF. Let x measure the wt. of the individual student in kilogram. Then x can take the value like 45, 52.25, 57.0, 60 etc. in a given range or interval. (43, 72) say. Then x is a continuous random variable since it can take all real no.

By Sangita Acharya

value in given range.

Q. Identify the following are discrete or continuous random variable

1. Inerrse in length of life by cancer patient as a result of surgry.

2. Quantity of ball pen sold by stationary per day during last month.

3. Blood pressure measure (lower) of individual 20F students

4. Applications registered for commersical loan during the fist 15 days in the ~~motha~~ month of 2019 at Nabil Bank.

=) 1) let x denotes increase in length of life in year due to the stated reason. Then x can take the values like 2 years, 4 years, 3.5 years etc in the range (1,5) years say. Since x can take every real number in the given range. So, x is a continuous random variable.

Q8. Define probability mass function of a discrete random variable. Three defective balls are mixed with seven good one. Find the probability distribution of the no. of defective balls if three balls are drawn at random. And find the expected no. of defective balls drawn out.

=) Let x be a ~~de~~ discrete random variable. Define on the elements of sample space of random variable. Let x assumes the values $x_1, x_2, x_3, \ldots x_n$ and let $p_1, p_2, p_3, \ldots p_n$ be corresponding probabilities when x assumes these values. It can be expressed as a function in the following way.

$$P(x = x_i) = p_i \quad ; \quad i = 1, 2, 3, \ldots n \qquad \text{—} (1)$$

The function $P(x)$ as defined in (1) is called probability mass function (pmf) if following conditions are satisfied.

i) $P(x_i) = p_i \geq 0$

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

Moreover $P(x = x_i) = p_i$ can be represented in tabular form which shows the values of random variable x together with probability. In this connection, the function,

$$P(x = x_i) = p_i$$

is called probability distribution. It looks like

x	x_1	x_2	x_3	x_n
p(x)	p_1	p_2	p_3	p_n

By Sangita Acharya

The mean of the distribution is defined as the

$$\bar{x} = \sum_{i=1}^{n} p_i x_i$$

The mean is also called mathematical expectation. further the variance of the distribution is defined as

$$\sigma^2 = \left(\sum_{i=1}^{n} x_i^2 p_i \right) - \bar{x}^2$$

where σ is S.D. of the distribution.

Numerical part.

Let x denotes the discrete RV representating the no. of defective items in the drawn sample. Clearly x will take the value 0, 1, 2, 3 according as non or few or all taken items are defective

Now,

$P(X=0)$ = probability of getting non-defective items

$$= \frac{C(7,3) \times C(3,0)}{C(10,3)}$$

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

$P(x=1)$ = probability of getting 1 defective and 2 good items

$$= \frac{C(7,2) \times C(3,1)}{C(10,3)}$$

$$= \frac{21}{40}$$

$$P(x=2) = \frac{C(7,1) \times C(3,2)}{C(10,3)} = \frac{7}{40}$$

$P(X=3)$ = probability of getting all defective items

$$= \frac{C(3,3) \times C(7,0)}{C(10,3)}$$

$$= \frac{1}{120}$$

∴ Probability distribution is.

X	0	1	2	3
$P(X)$	$7/24$	$21/40$	$7/40$	$1/120$

Now,

expected no. of defective items = mean of the distribution

$$\bar{x} = \sum_{i=1}^{4} x_i \, p_i$$

$$= x_1 . p_1 + x_2 p_2 + x_3 p_3 + x_4 p_4$$

$$= 0 \times \frac{7}{24} + 1 \times \frac{21}{40} + 2 \times \frac{7}{40} + 3 \times \frac{1}{120}$$

$$= 0.9$$

Q. A random variable X has the following p.m.f.

X	0	1	2	3	4	5	6	7
$P(X)$	0	k	$2k$	$2k$	$3k$	k^2	$2k^2$	$7k^2+k$

Find,

i) $P(X \geq 6)$ ii) $P(2 < x < 6)$

iii) $P(X \leq 5)$ iv) P(at least one)

⇒ Solution,

Since given is probability distribution

$$\Sigma P(X) = 1$$

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

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

$\Rightarrow k = \frac{1}{10}, -10$

since each P(X) must be positive. we choose

$k = \frac{1}{10}$

i) $P(X \geq 6) = P(X = 6) + P(X = 7)$

$= 2k^2 + 7k^2$

$=$

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

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

$= 1 - [P(X = 6) + P(X = 7)]$

$=$

iv) P(at least one) $= P(X \geq 1)$

$= 1 - P(X = 0)$

$= 1 - 0$

$= 1$

Q. which of the following is legitimate for pmf and explain your answer.

a) $f(x) = \frac{x}{5}; \quad 0 \leq x \leq 5$

what happens if $x = 0, 1, 4$?

By Sangita Acharya

b) $f(x) = \frac{5-x^2}{6}$; $x = 0, 1, 2, 3$

c) $f(x) = \frac{x}{6}$; $x = 0, 1, 2, 3$

obtain mean and variance of x for which f(x) is defined to be p.m.f.

⇒ solution,

The distribution for each f(x) is as given below.

	0	1	2	3	4	5	$\Sigma f(x)$
$f(x) = x/5$	0	1/5	2/5	3/5	4/5	1	>1
$f(x) = x/5$	0	1/5	–	–	4/5	–	1
$f(x) = (5-x^2)/6$	5/6	4/6	1/6	-4/6	–	–	1
$f(x) = x/6$	0	1/6	2/6	3/6	–	–	1

Analysis:

for $f(x) = x/5$; $0 \le x \le 5$

Although each entry is positive, we find $\Sigma f(x) > 1$.

∴ f(x) is not a p.m.f. under this condition

for $f(x) = x/5$, $x = 0, 1, 4$.

Since each entry is positive. and $\Sigma f(x) = 1$, $f(x) = x/5$ under this condition is legitimate for p.m.f.

Here, $f(x) = x/5$ for $x = 0, 1, 4$ is a p.m.f

So, mean,

$$r = \Sigma x.f(x)$$

$$= 0.0 + 1.\frac{1}{5} + 4.\frac{4}{5}$$

$$= \frac{17}{5}$$

variance,

$$\sigma^2 = [\Sigma x^2 f(x)] - \mu^2$$

$$= (0^2 \times 0 + 1^2 \times 1/5 + 4^2 \times 4/5) - (17/5)^2$$

$$= \ldots$$

By

Discrete Distribution

1. Binomial Distribution
2. Poisson's "
3. Negative Binomial "
4. Hypergeometric "

Binomial Distribution

Let there be n independent trials in a random experiment and let a discrete random variable x denotes the no. of successes in these trial. Let p and q denote the probabilities of sucess and failure respectively in a single trial. so that $p+q=1$. Then, the probability that x takes the value r or the probability of getting exactly r sucesses is given by,

$$P(X=r) = C(n,r)\, p^r q^{n-r} \quad \text{--- (1)}$$

The probability distribution having p.m.f. as in (1) is called binomial distribution or Bernoulli's distribution.

The mean μ and the variance σ^2 of the distribution are given by,

$$\mu = np \quad \text{and} \quad \sigma^2 = npq$$

Conditions ~~.......~~ underlying Binomial distribution

1. The experiment is performed for fixed and finite no. of times say n. In other words n is assume or predicted in

By

advance.

2. The trials are identical and each trial surely result either in a sucess or failure. In other words the outcomes of the identical trials are dichotomy in nature.

3. Trial should be independent so that outcome of any particular trial does not influence or is not influence by outcome of any other trial.

4. The probability of sucess (and hence that of failure) should remain constant throughout the experiment.

Example:-

(i) Detecting "bad samples" from a lot of finite items containing bad and good items

(ii) Hitting a target with a significant probability for a finite number of times.

problems.

Q. A box contains 100 transistor, 20 of which are defective 10 are selected for inspection. Indicate the probability that (a) all 10 are defective (b) at least one is defective (c) atmost 3 are defective.

Q. Probability of a bomb hitting a target is 2/3. 3 bombs are enough to destroy the target. If 7 bombs are aimed at the target, find the probability that the pillar is destroyed.

Q. Determine the binomial distribution whose mean is 3 and variance is 3/4. ~~It is observed that 4~~

Q. It is observed that the 80% of the television viewers watch entertainment channels. What is the probability that at least 80% of the viewers in a random sample of 5 watch entertainment channel.

Q. 1) Ans

let x represents the no. of defective items (sucess) in the drawn sample. Since 10 items have been selected, the DR ~~V~~ x will take the value 0, 1, 2, 3, --- upto 10.

$$n = 10$$

probability of sucess is,

$$P(\text{getting a defective item}) = p = \frac{20}{100} = \frac{1}{5}$$

$$\text{Probability of failure} = q = 1 - \frac{1}{5} = \frac{4}{5}$$

Since total no. of sample and no. of drawn items are fixed and finite, and probability of sucess (and hence that of failure) is moderate; we can apply binomial distribution

Acc. to it. probability of getting exactly r sucess or probability that x will take the value r is given by,

$$P(X=r) = C(n,r)\, p^r q^{n-r} \quad \text{— (1)}$$

i) $r=10$

from (1)

$$P(X=10) = C(10,10) \left(\frac{1}{5}\right)^{10} \cdot \left(\frac{4}{5}\right)^{10-10}$$

$$= \left(\frac{1}{5}\right)^{10}$$

ii) P (at least one is defective)

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

$$= 1 - C(10,0) \left(\frac{1}{5}\right)^{0} \left(\frac{4}{5}\right)^{10-0}$$

$$= 1 - \left(\frac{4}{5}\right)^{10}$$

$$= \frac{5^{10} - 4^{10}}{5^{10}}$$

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

Q. 3 Solution

Let x denote the random variable representing no. of sucess and n denotes no. of trials, p and q denotes prob. of sucess & failure in a single trial then, we have to determine the binomial distribution for probability of getting exactly r sucess. which is given by

$$p(x=r) = C(n,r)\, p^r q^{n-r} \quad -(1)$$

Given

mean $\mu = 9$

variance $\sigma^2 = 9/4$

i.e. $np = 9 \quad -(2)$

$npq = 9/4 \quad -(3)$

from (2) and (3)

$9q = 9/4$

$\Rightarrow q = \frac{1}{4}$

$\because p+q=1 \Rightarrow p = 3/4$

from (2)

$n \times \frac{3}{4} = 9$

$\Rightarrow n = 12$

from (1) the required distribution is,

$$P(X=r) = C(12,r) \left(\frac{3}{4}\right)^r \left(\frac{1}{4}\right)^{12-r}.$$

By

Page:

Poisson's Distribution

We know from binomial distribution that the probability of getting exactly r sucess in the finite no. of trials (n) is given by

$$P(X=r) = C(n,r)\, p^r q^{n-r} \quad \text{—} (1)$$

where,

p and q are prob. of sucess and failure respectively in a single trial of a random experiment,

So that $p+q=1$

However when p is very small and n is either very large or unknown, the probability mass function (1) for binomial distribution doesn't work good. In order to overcome these difficulties the following conditions are considered in binomial distribution.

i) p is very small ; say $p \to 0$

ii) n is very large or n is unknown ; say $n \to \infty$

iii) however np is finite ; say $np = m$

The limiting form of probability mass function (1) is reduced to

$$P(X=r) = \frac{e^{-m}\, m^r}{r!} \quad \text{—} (2)$$

By Sangita Acharya

Date:
Page:

The probability distribution having pmf ② is called poisson's distribution.

The conditions underlying poisson's distribution

① The process of sucess (p) should be very small, the no. of trials (n) should be sufficiently large but the product np should always remain finite.

② The occurance of the event should be random, rare and they should be independent. The outcome of each trial should either sucess or a failure.

③ The entire experiment (successive trial) is performed under identical condition. Mainly two types of phenomenon called temporal and spatial cases follow poisson's distribution.

④ The probability of sucess, however small, should remain constant throughout the experiment.