28/09/23

(unit-4) Special Continuous distribution

Topic

- ★ Normal distribution. and its MGF.
- Gamma distribution and its MGF.
- Exponential distribution and its MGF.

Normal dis^{tri}

- n = large, $n \to \infty$

for ↳ continuous data.

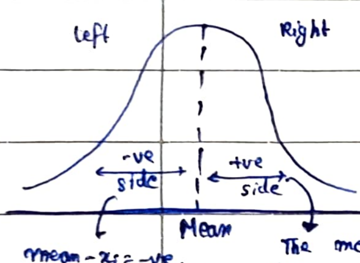
$n \to$ small ↳ B.D

$n \to$ large ↳ P.D ↓ Discrete

mean → avg value or mid value.

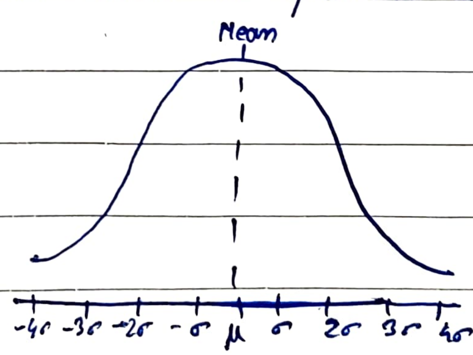
variance → Difference / deviation b/w original data and mean data.

★ it will tell you above the curve



↳ both are +ve but only for representation it is (-ve).

- Normal $dist^{ri}$ is also known as "Gaussian $dist^{ri}$" or "Bell shape curve dist."



- A normal Random $Vari^{ble}$ is said to have a normal $dist^{ri}$ with parameter μ(mean) and σ^2(variance); if its Pdf is given by:

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

__/__/__

we denote the value of the RV 'x' as :-

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

or

$$n(x; \mu, \sigma)$$

i.e, $X \sim N(\mu, \sigma^2)$ or $X \sim n(x; \mu, \sigma)$

★ **Properties of Normal distribution.**

•) The mean, median and mode are all equal or coinciding.

$\boxed{\text{mean} = \text{median} = \text{mode}}$ → equal for normal distri.

↳ arrange in ascending

↳ then the middle term.

↳ even = $\frac{\cancel{n^{th} + n+1^{th}}}{2}$ → ~~one~~ avg of two middle most value

odd = $\frac{n}{2}$ → middle most value.

some sort of thing

29/09/23

•) The curve is symmetric about the origin (μ = mean) ★

•) The total area under the curve and above the ~~Hozi~~ Horizontal axis is equal = 1.

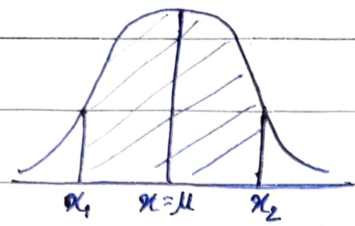
i.e $\int_{-\infty}^{\infty} f(x).dx = 1$

•) Area to the left and to the right of the mean (μ) is 0.5 ★

•) The mode, which is the point of the horizontal axis, where the curve is maximum occurs at $\boxed{x = \mu\,(\text{mean})}$

↳

Area under the curve :-



$$P(x_1 < x < x_2) = \int_{x_1}^{x_2} f(x)\,dx$$

$x_1 < x < x_2$

$$= \int_{x_1}^{x_2} \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}\,dx$$

$$\boxed{= \frac{1}{\sigma\sqrt{2\pi}} \int_{x_1}^{x_2} \exp\left\{\frac{-(x-\mu)^2}{2\sigma^2}\right\}.dx.} \quad \text{—①}$$

It is a very lengthy approach so, we do have another method.

⇒ ★ If $x \sim N(\mu, \sigma^2)$ or $x \sim n(x; \mu, \sigma)$ for normal distri variable X with mean = 0 and var = 1 ; then the new variable Z is given by,

$\mu = 0$
$\sigma^2 = 1$

$$\boxed{Z = \frac{X-\mu}{\sigma}}$$

, where Z is called standard normal variate

whenever, X assume a value x, the corresponding value of Z is given by $\boxed{Z = \frac{x-\mu}{\sigma}}$

∴ If X takes b/w the values, $x = x_1$ and $x = x_2$, then the random variable z, will fall b/w the corresponding values,

$$Z_1 = \frac{x_1-\mu}{\sigma} \quad \& \quad Z_2 = \frac{x_2-\mu}{\sigma}$$

from eq (1)

$$-\frac{(x-\mu)^2}{2\sigma^2} \rightarrow -\frac{z^2}{2} \quad —(2)$$

↳ so from (iii) → $P(x_1<x<x_2) = \int_{z_1}^{z_2} \frac{1}{\sigma\sqrt{2\pi}} \cdot e^{-\frac{z^2}{2}} \cdot dz$

for z → mean = 0
variance = 1

$$= \frac{1}{\underset{\downarrow\, 1}{\sigma}\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} \cdot dz$$

$$\boxed{P(x_1<x<x_2) = \frac{1}{\sqrt{2\pi}} \int_{z_1}^{z_2} e^{-z^2/2} \cdot dz}$$

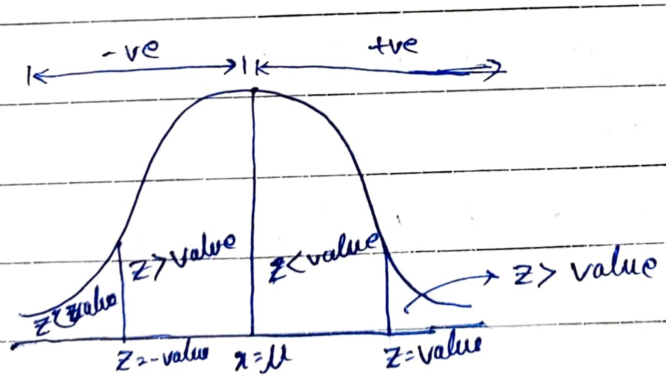
$P(x_1<x<x_2) = P(z_1<x<z_2)$ ← $\boxed{\int_{z_1}^{z_2} n(z;0,1)}$

Q. Given a standard normal distribution, find the area given under the curve, which lies to the right of z,

a) to the right of z = 1.84 and
b) b/w z = -1.97 and z = 0.86

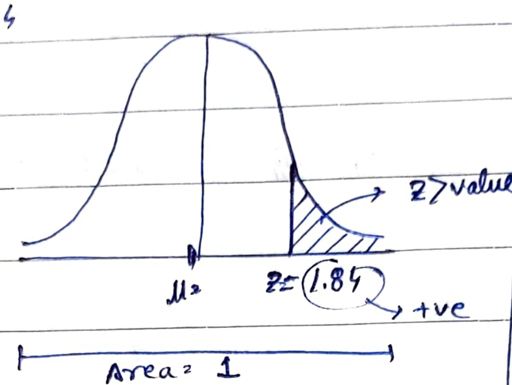
< > → 1st Table
0 → value → third table.

Ex:-



↳

a) to the right of $Z = 1.84$

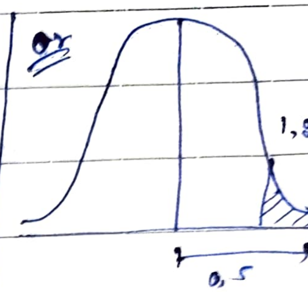


$= 1 - \underbrace{P(Z<1.84)}$

$= 1 - 0.9671$

$= 0.0329$

Or



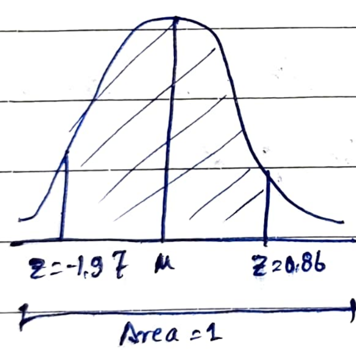
$P(Z>1.84)$

$= 0.5 - P(0<Z<1.84)$

$= 0.5 - 0.4671$

$= 0.0329$

b) b/w $Z = -1.97$ and $Z = 0.86$.



z table

$\cancel{P(Z<0.86)} \rightarrow P(0<z<0.86) \rightarrow 0.3051$

$= \cancel{0.8}$

$P(Z>-1.97) \rightarrow P(-1.97<z<0) \rightarrow 0.4756$

$= 0.7807$

or

1 table

$P(z<0.86) - P(z<-1.97)$

$= 0.8051 - 0.0244$

$= 0.7807$

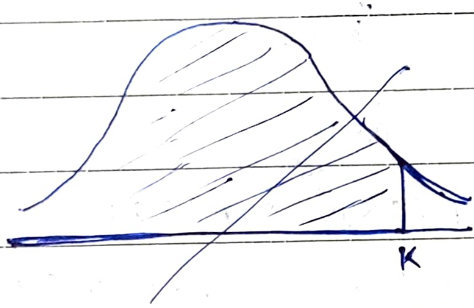
05/10/23

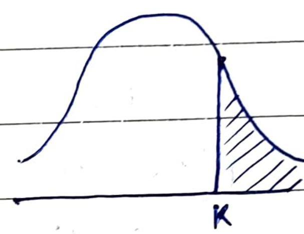
Q. Given a standard normal distribution, find the value of K such that
$\hookrightarrow Z$

a) $P(Z > K) = 0.3015$

b) $P(K < Z < -0.18) = 0.4197$

a)





$Z <$

1st Table

$Z > K = 0.3015.$

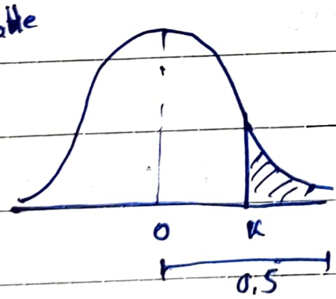
$Z < K = 1 - 0.3015$

$= 0.6985$

≈ 0.6985

$\boxed{K = 0.52}$ ✓

2nd Table



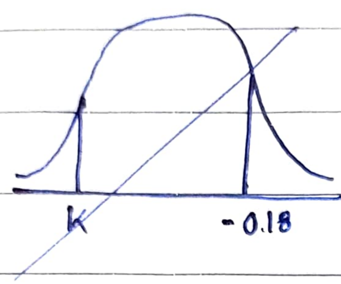
$Z > K = 0.3015$

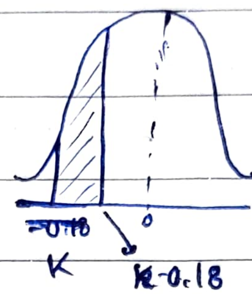
$P(0 < Z < K) = 0.5 - 0.3015$

$= 0.1985$

$\Rightarrow K = \boxed{0.52}$

b) $P(K < Z < -0.18) = 0.4197$





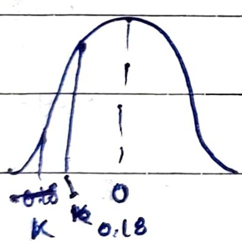
$P(Z < -0.18) = 0.4286$

$P(K < Z < -0.18) - P(Z < \overset{-0.18}{K}) = P(Z \overset{<}{>} K)$

$0.4197 - 0.4286 = -0.0089$

~~K = 0.1867~~ $\boxed{K = 2,37}$

Table 2



$P(K < Z < -0.18) = 0.4197$

~~P(K< 0~~

~~P(K<Z<-0.18)~~ $P(0 < Z < K) - P(0 < Z < \overset{-0.18}{\cancel{K}}) = 0.4197$

0.714

$P(0 < Z < K) = 0.4197 + 0.714$

$= \cancel{0.2943}$ ~~0.3483~~ (.4911)

$K = 2.37$

Q. X is normally distributed, and mean of X $= 12$, $\sigma = 4$.
find the Probability of the following.

(i) $X \geq 20$

(ii) $X \leq 20$

(iii) $0 \leq X \leq 12$

$\mu = 12$; $\sigma = 4$

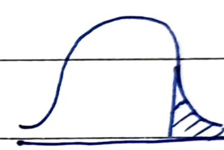
(i) ~~P(X~~ $X \geq 20$

$X = 20 \rightarrow \quad Z = \dfrac{X - \mu}{\sigma} = \dfrac{20-12}{4} = 8/4 = 2$

$P(X \geq 20) \Rightarrow P(Z \geq 2)$

Table 1

$P(Z \geq 2) = 1 - P(Z < 2)$



$Z=2$

$= 1 - 0.9772$

$= 0.0228$

(ii) $P(X \leq 20) = P(Z \leq 2)$

$= 1 - P(Z \geq 2)$

$= 1 - 0.0228$

$= 0.9772$

Table 2

$P(Z \geq 2 = 0.5 -$

$P(0 < Z < 2)$

$\downarrow$

$0.5 - 0.4772$

0.0228

(iii) $X_1 = 0$, $X_2 = 12$

$X_1 = X_1 \Rightarrow Z_1 = \dfrac{X_1 - 12}{4} = -12/4 = -3$

$-3 < Z < 0$

$X_2 = X_2 \Rightarrow Z_2 = \dfrac{X_2 - 12}{4} = 0/4 = 0$

Table 2

$P(-3 < Z < 0) = 0.4987$

MGF of Normal Distribution

$$M_x t = \int_{-\infty}^{\infty} e^{tx} f(x).dx$$

$$= \int_{-\infty}^{\infty} e^{tx}.\frac{1}{\sigma\sqrt{2\pi}}.e^{\frac{-(x-\mu)^2}{2\sigma^2}}.dx$$

$$= \frac{1}{\sigma\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{tx}.\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}.dx$$

$$= \frac{\cancel{\sigma}}{\cancel{\sigma}\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{t\overset{\sigma z+\mu}{x}}.\exp\left\{-\frac{z^2}{2}\right\}.dz$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{t(\mu+\sigma z)}.e^{-z^2/2}.dz$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{\sigma zt - z^2/2}.dz$$

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

$$\sigma z = x-\mu$$

$$\sigma z+\mu = x$$

differentiate

$$\sigma dz + 0 = dx$$

$$\sigma dz = dx$$

$$= \frac{e^{t\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2-2\sigma zt)}.dz$$

performing, Now

06/10/23

$$\Rightarrow \frac{e^{t\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z^2+\sigma^2t^2-2\sigma zt-\sigma^2t^2)}.dz$$

completeing Square

↳ coff of z^2 must be 1

↳ coff of $z = 2\sigma t$

↳ Divide by $\frac{1}{2}$

$$\Rightarrow (2\sigma t)\frac{1}{2}$$

$$= \sigma t$$

$$\Rightarrow \frac{e^{t\mu}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}((z-\sigma t)^2-\sigma^2t^2)}.dz$$

$$\Rightarrow \frac{e^{t\mu}.e^{\frac{1}{2}(\sigma^2t^2)}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(z-\sigma t)^2}.dz$$

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\sqrt{2\pi}}\int_{-\infty}^{\infty} e^{-\frac{1}{2}(b)^2}\,.\,db$$

$z - \sigma t = b$

diff

$dz - 0 = db$

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\sqrt{2\pi}}\,.\,2\int_{0}^{\infty} e^{-\frac{b^2}{2}}\,.\,db$$

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

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\sqrt{2\pi}}\,.\,2\int_{0}^{\infty} e^{-c}\,.\,\frac{1}{\sqrt{2c}}\,.\,dc$$

$\frac{b^2}{2} = c$

$\cancel{\frac{2bb}{2} = dc}$ $\quad \frac{2b.db}{2} = dc$

$db = \frac{dc}{b} \rightarrow \sqrt{2c}$

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\sqrt{2\pi}}\,\frac{\cancel{2}}{\sqrt{2}}\int_{0}^{\infty} e^{-c}\,.\,c^{-\frac{1}{2}}\,.\,dc$$

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\sqrt{\pi}}\int_{0}^{\infty} \cancel{e^{-c}}\,c^{-\frac{1}{2}}\,.\,e^{-c}\,.\,dc$$

$\alpha = -\frac{1}{2}+1$

$= \frac{1}{2}$

$$\Gamma\tfrac{1}{2} \Rightarrow \int_{0}^{\infty} c^{-\frac{1}{2}}\,.\,e^{-\frac{1}{2}}\,.\,dc$$

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

$$\Rightarrow \frac{e^{(tu+\frac{1}{2}\sigma^2t^2)}}{\cancel{\sqrt{\pi}}}\,.\,\Gamma\tfrac{1}{2} \rightarrow \cancel{\sqrt{\pi}}$$

** $$\boxed{M_x(t) = e^{(tu+\frac{1}{2}\sigma^2t^2)}}$$

Gamma function.

$\Gamma\frac{1}{2} = \sqrt{\pi}$

value of Gamma $\frac{1}{2} = \sqrt{\pi}$

** $$\boxed{\Gamma\alpha = \int x^{\alpha-1}\,.\,e^{-\alpha}\,.\,dx}\,,\ x>0$$

ex:- $\Gamma 3 = \int x^2\,.\,e^{-3}\,.\,dx$

Normal Distribution as Limiting case of Binomial Distribution.

↳ Normal Distribution is the another Limiting case of Binomial Distribution just like poison's Distribution.

↳ The conditions are :-

★★ (i) n, the no. of trials is indefinately very large. (ie $n \to \infty$)

(ii) neither p nor q is small. $(p, q \not\to 0)$

If ~~capital~~ "X" is a binomial r.v with mean $(\mu) = np$ and variance $(\sigma^2) = npq$, then the Limiting form of the distribution is

$$\boxed{Z = \frac{X - np}{\sqrt{npq}}}$$

as $n \to \infty$, is the standard Normal Distribution, $n(Z; 0, 1)$

Remark :- The ND with $\mu = np$ and $var = npq$ not only provides a very accurate approximation to the binomial distribution, when n is large, and p is not extremely close to 0 or 1, but also provides a fairly good approximation, even when n is small and p is close to $\frac{1}{2}$.

↳

Q The prob that a patient recovers from a rare blood disease. (0.4) If 100 people are known to have contacted this disease, what is the prob that fewer than 30 patient survive.

→ $n = 100$

$p = 0.4$ $\quad q = 0.6$

possion → (statement) on an averag

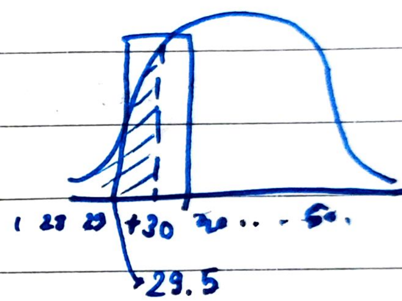
we can convert all Dist to Normal DB

$P(X < 30)$

$np = 100 \times 0.4 = 40$

$\sqrt{npq} = \sqrt{100 \times 0.4 \times 0.6} = 2\sqrt{6}$

$Z = \frac{X - np}{\sqrt{npq}}$ $\quad x = ? = \cancel{30}\ 29.5$



$Z = \frac{29.5 - 40}{2\sqrt{6}} = -2.14$

$P(X \leq 30) = P(Z \leq -2.14)$

$= 0.0162$

Q A MCQ Quiz has 200 Question each with 4 possible answer of which only one is correct. What is the prob. of sheer (guesswork) yeilds from 25→30 correct answer for the 80 of the 200 Q problems about which the player has no. Knowledge.

→ $n = 80$

x : correct Q/A.

$p = 1/4$, $q = 3/4$

$np = 80 \times 1/4 = 20$

$\sqrt{npq} = \sqrt{80 \times 1/4 \times 3/4}$ $\sqrt{15} = 3.872$

$P(25 \leq x \leq 30)$

$= P(1.291 \leq Z \leq 2.582)$

$Z_1 = \frac{25 - 20}{\sqrt{15}} = 5/\sqrt{15} = 1.291$

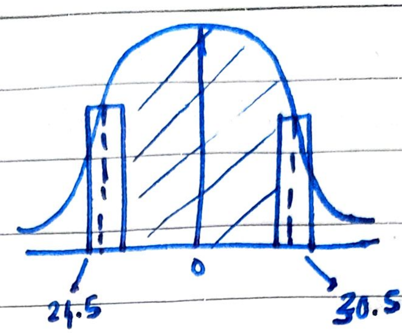
$Z_2 = \frac{30 - 20}{\sqrt{15}} = 10/\sqrt{15} = 2.582$

$\Rightarrow n = 80$

$p = 1/4 \ , \ q = 3/4$

$np = 20$

$\sqrt{npq} = \sqrt{15}$



$x_1 = 24.5$ $\qquad Z_1 = \dfrac{24.5 - 20}{\sqrt{15}} = \dfrac{4.5}{\sqrt{15}} = \dfrac{4.5}{3.872} \approx 1.162$

$x_2 = 30.5$

$$Z_2 = \frac{30.5 - 20}{\sqrt{15}} = \frac{10.5}{\sqrt{15}} = \frac{10.5}{3.872} \approx 2.711$$

$$P(25\cancel{6} \leq x \leq 30\cancel{6}) = P(1.162 \leq Z \leq 2.71)$$

$$\downarrow$$

$$= P(0 \leq Z \leq 2.71) - P(0 \leq Z \leq 1.162)$$

$$\downarrow$$

$$= 0.4966 - 0.3770$$

$$= 0.1196$$

* **Mean and Variance of Normal Distribution using MGF.**

↳ Derivation of Mean and Variance.

$$\mu'_r = E[x^r] = \frac{d^r}{dx^r} M_x t \Big|_{t=0}.$$

(i) Mean

$$\mu = E(x) = \frac{d}{dt} M_x(t) \Big|_{t=0}$$

$Var,\ \sigma^2 = E[x^2] = \frac{d^2}{dx^2} M_x t \Big|_{t=0}$

(i) Mean $\Rightarrow$ $E[x] = \frac{d}{dt} M_x t$

$= \frac{d}{dt}\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]_{t=0}$

$= e^{(\mu t + \frac{1}{2}\sigma^2 t^2)} \cdot \left(\mu + \cancel{\frac{2}{2}}\sigma^2 t\right)_{t=0}$

$= \cancel{e^0 \cdot \mu}\ \cancel{e^{0 \cdot \mu}}\ e^0 \cdot \mu$

$\boxed{= \mu}$ μ

(ii) Variance

$\Rightarrow \sigma^2 = E[x^2] = \frac{d^2}{dx^2} M_x t \Big|_{t=0}$

$\Rightarrow \frac{d^2}{dt^2}\left[e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right]_{t=0} = \frac{d}{dt}\left(e^{(\mu t + \frac{1}{2}\sigma^2 t^2)} \cdot \left(\mu + \cancel{\frac{2}{2}}\sigma^2 t\right)\right)$

$\cancel{\Rightarrow \left(\mu + \frac{1}{2}\sigma^2 t\right)\left\{\frac{d}{dt}\left(e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right)\right\}}$

$\Rightarrow \left(e^{\mu t + \frac{1}{2}\sigma^2 t^2}\right) \cdot \left(\cancel{\frac{2}{2}}\sigma^2\right) + \left(\mu + \cancel{\frac{2}{2}}\sigma^2 t^2\right) \cdot \left\{e^{(\mu t + \frac{1}{2}\sigma^2 t^2)} \cdot \mu + \cancel{\frac{2}{2}}\sigma^2 t^2\right\}_{t=0}$

$= \left(e^0 \cdot \sigma^2 + \cancel{e^0 \mu^2}\right) = \boxed{\sigma^2 + \cancel{\mu^2}.}$

$-\mu^2$

12/10/23

Gamma and Exponential Distribution

Discrete Data				Continuous Data		
B.D	NGD	Geo	Posi	N.D	Gamma	Exp
n = trial → small, p → given → small	check for ~~kth trial~~ k^{th} success in n^{th} trial	Check for 1st success	"on an avg term" given question	$n \to$ large, $p \to$ Large, "Normally distributed term"	k^{th} event occurs in what time?!	after k^{th} event the time period for the $k+1^{th}$ event

timer after kth event goes to 0

k^{th} event occurs in what particular time = ??

* Gamma Distribution is used to find time until K events occur.

Ex:- The time until K customers arrived in a particular restaurant.

Ex:- The time you have been invited to cake party.

Ex:- The time when 50% student will pass probability and statistics.

Gamma function is defined.

** $$\boxed{\Gamma\alpha = \int_0^\infty x^{\alpha-1} \cdot e^{-x} \cdot dx}, \quad x>0$$

properties.

i) $\Gamma n = (n-1).(n-2)\ldots\ldots 1.$ for +ve int n.

ii) $\Gamma n = (n-1)!$

iii) $\Gamma n = \Gamma 1 = 1$ 0!

iv) $\Gamma_{1/2} = \sqrt{\pi}$

Defination :- The continuous RV 'X' has gamma distribution with parameters α and β, if its PDF is given by :-

**

where $\alpha > 0$, $\beta > 0$

$$f(x;\alpha,\beta) = \begin{cases} \dfrac{1}{\beta^\alpha \Gamma\alpha} x^{\alpha-1} \cdot e^{-x/\beta}, & x>0 \\ 0, & \text{elsewhere} \end{cases}$$

Possion distribution.

Remark :- The relation b/w λ and β

$\Rightarrow \boxed{\lambda = 1/\beta}$ where λ = Average no. of events per unit time. rate

α = No. of Events. like n

β = Avg time b/w no. of events.

Exponential Distribution :- The continous RV 'X' has an exp. distribution with parameter "β", if its pdf is given by.

It occurs when $\boxed{\alpha = 1}$

$$\boxed{f(x;\beta) = \begin{cases} \frac{1}{\beta} e^{-x/\beta}, & x > 0 \\ 0, & \text{elsewhere} \end{cases}}$$ ★★

	Mean	Variance.
Gamma distribution	$\alpha\beta$	$\alpha\beta^2$
Exponential distribution	β	β^2.

MGF of Gamma distribution.

$$\boxed{M_x t = E[e^{tx}] = \int_0^\infty e^{tx} \cdot f(x).dx}$$

$$M_x t = E[e^{tx}] = \int_0^\infty e^{tx} \cdot f(x).dx$$

$$= \int_0^\infty e^{tx} \cdot \left(\frac{1}{\beta^\alpha \Gamma\alpha} x^{\alpha-1} \cdot e^{-x/\beta} \right).dx$$

$$= \frac{1}{\beta^\alpha \Gamma\alpha} \int_0^\infty e^{-(1/\beta - t)x} \cdot x^{\alpha-1} \cdot dx.$$

Put $\left(\frac{1}{\beta} - t\right) x = u$

$$x = \frac{u}{1/\beta - t}$$

$$dx = \frac{du}{1/\beta - t}$$

$$= \frac{1}{\beta^{\alpha}\Gamma\alpha}\int_0^{\infty} e^{-u}\left(\frac{u}{1/\beta - t}\right)^{\alpha-1} \cdot \frac{du}{1/\beta - t}$$

$$= \frac{1}{\beta^{\alpha}\Gamma\alpha} \cdot \frac{1}{(1/\beta - t)^{\alpha-1}} \cdot \frac{1}{(1/\beta - t)} \cdot \int_0^{\infty} e^{-u} \cdot u^{\alpha-1} \cdot du$$

$$= \frac{1}{\beta^{\alpha}\Gamma\alpha} \cdot \frac{\beta^{\alpha-1}}{1-}$$

$$\boxed{M_x t = \frac{1}{(1-\beta t)^{\alpha}}}$$

13/11/23

Derivation of Mean and Variance of Gamma distribution by using MGF.

$$\mu_r' = E[X^r] = \frac{d^r}{dt^r} M_x(t)\Big|_{t=0}$$

$$\mu = E[X] = \frac{d}{dt} M_x(t)\Big|_{t=0}$$

$$\Rightarrow \frac{d}{dt}\left[\frac{1}{(1-\beta t)^{\alpha}}\right]\Bigg|_{t=0}$$

$x' = 1$

$1/x = x^{-1}$ or $= -1 x^{-2}$

$$\Rightarrow \cancel{\left(\frac{\frac{1}{(1-\beta t)^{\alpha-1}}}{\alpha-1}\right)}$$

$$\Rightarrow \frac{d}{dt}\left[(1-\beta t)^{-\alpha}\right]\Big|_{t=0}$$

$$\Rightarrow (-\alpha) \cdot (1-\beta t)^{-\alpha-1} \cdot (0-\beta)\Big|_{t=0}$$

$\Rightarrow (-\alpha).(1-0).(-\beta)$

$$\boxed{\mu = \alpha\beta}^{**}$$

Variance by MGF

$$\sigma^2 = E[x^2] - (E[x])^2$$

$$= \frac{d^2}{dt^2}\left[\frac{1}{(1-\beta t)^\alpha}\right]\Bigg|_{t=0} \Rightarrow \frac{d^2}{dt^2}\left[(1-\beta t)^{-\alpha}\right]\Bigg|_{t=0}$$

$$\Rightarrow \frac{d}{dt}\left[\alpha\beta(1-\beta t)^{-\alpha-1}\right]\Bigg|_{t=0}.$$

$$\Rightarrow \alpha\beta.\left((-\alpha-1)(1-\beta t)^{-\alpha-2}.(0-\beta)\right)$$

$$\Rightarrow \alpha\beta.\left((-\alpha-1)(1-0).(-\beta)\right)$$

$$\Rightarrow \alpha\beta[\alpha\beta+\beta]$$

$$\Rightarrow \alpha^2\beta^2 + \alpha\beta^2$$

Now, $\sigma^2 = \cancel{\alpha^2\beta^2} + \alpha\beta^2 - \cancel{\alpha^2\beta^2}$

$$\boxed{\sigma^2 = \alpha\beta^2}^{**}$$

MGF of Exponential Distribution

$$\Rightarrow M_x(t) = E[e^{tx}] = \int_0^\infty e^{tx}.f(x).dx$$

$$\Rightarrow \int_0^\infty e^{tx}.\left(\frac{1}{\beta}.e^{-x/\beta}\right).dx \Rightarrow \frac{1}{\beta}\int_0^\infty e^{tx}.e^{-x/\beta}.dx \Rightarrow \frac{1}{\beta}\int_0^\infty e^{-(1/\beta - t)x}.dx$$

put $\Rightarrow (1/\beta - t)x = u$

$$x = \frac{u}{1/\beta - t}$$

$$dx = \frac{du}{1/\beta - t}$$

$$\Rightarrow \frac{1}{\beta}\int_0^\infty e^{-u}.\frac{du}{1/\beta - t}$$

$$\Rightarrow \frac{1}{\beta}.\frac{1}{1/\beta - t}\int_0^\infty u^{\alpha-1}e^{-u}.du$$

(Since in Expo ($\alpha = 1$) we insert $u^{\alpha-1} \Rightarrow u^{1-1} = u^0$)

$$\Rightarrow \frac{1}{\beta} \cdot \frac{\beta}{1-\beta t} \cdot \underbrace{\int_0^{\infty} e^{-u} \cdot u^{\alpha-1} \, dx}_{\rightarrow \sqrt{1}}$$

$$\Rightarrow \boxed{M_x(t) = \frac{1}{1-\beta t}} \; **$$

Mean of Exponential by using MGF

$$\mu_1' = E[x] = \frac{d}{dt} M_x t \Big|_{t=0}$$

$$= \frac{d}{dt}\left(\frac{1}{1-\beta t}\right)\Big|_{t=0} = \frac{d}{dt}\left((1-\beta t)^{-1}\right)\Big|_{t=0}$$

$$= \left((-1).(1-\beta t)^{-1-1}(-\beta)\right)\Big|_{t=0}$$

$$\Rightarrow (-1)(-\beta)$$

$$\boxed{\mu = \beta} \; **$$

Variance of Exponential by using MGF

$$\sigma^2 = E[x^2] - (E(x))^2$$

$$= \frac{d^2}{dt^2} M_x(t)\Big|_{t=0} = \frac{d^2}{dt^2}\left[\frac{1}{1-\beta t}\right]\Big|_{t=0}$$

$$= \frac{d}{dt}\left((-1)(1-\beta t)^{-2}(-\beta)\right)\Big|_{t=0} = \frac{d}{dt}\left(\beta.(1-\beta t)^{-2}\right)\Big|_{t=0}$$

$$= -2\beta.(1-\beta t)^{-3}(-\beta)\Big|_{t=0}$$

$$= -2\beta.(1).(-\beta)$$

$$E[x^2] = \boxed{2\beta^2}$$

$$\boxed{\sigma^2 = 2\beta^2 - \beta^2} \; **$$

$$\boxed{\sigma^2 = \beta^2}$$

16|10|23

Q Suppose that telephone calls arriving at a call centre with an avg of 5 calls coming per minute. What is the prob that up to a minute will ellapse by the time two calls have come in to the telephone calls.

→ $\lambda = 5$

$\alpha = 2$

$\beta = 1/5$

X: Time will elapse before 2 calls.

$P(X \leq 1) = ?$

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

$\Gamma 2 = 1!$
$= 1$

$$\Rightarrow \int_0^{1} \frac{1}{(1/5)^2\,\Gamma 2\,(1)} \cdot x^{1} \cdot e^{-5x} \cdot dx$$

$$\Rightarrow \frac{1}{\left(\frac{1}{25}\right)} \int_0^{1} x^{\alpha-1 \to 1} \cdot e^{-5x} \cdot dx$$

function by part.
ILATE

$\int I.II = I\int II - \int\left[\frac{d}{dx} I \int II\right] dx$

$$\cancel{\Rightarrow \frac{1}{25} \int_0^1 \left[x^1 \cdot \frac{e^{-5x}}{-5} \Big|_0^1 - \int_0^1 1. \right]}$$

$$\Rightarrow 25\left[x^1 \cdot \frac{e^{-5x}}{-5} \Big|_0^1 - \int_0^1 1 . e^{-5x} \cdot dx \right]$$

$I\int II - \int\left[\frac{d}{dx} I \int II\right] . dx$

$$\Rightarrow 25\left[-5\left[e^{-5}\right] - \frac{1}{25} \cdot e^{-5x}\Big|_0^1 \right\}$$

$\boxed{e^0 = 1}$

Ans = 0.96

Q. In a bio-medical study with rats are those responds investigation is used to determine the effect of the dose of the toxicants on their survival. The toxicant is one that is frequently discharge into the atmosphere from the jet fuel. For a certain dose of the toxicants the study determine that the survival time in week as a gamma distribution $\alpha = 5$, $\beta = 10$.

What is the prob that a rat survives no longer that 60 weeks.

$\Rightarrow$ $\alpha = 5$, $\beta = 10$.

$$P(X \leq 60\text{week}) = ?$$

$$\int_0^{\infty \to 60} \frac{1}{\beta^{\alpha} \Gamma\alpha} \cdot x^{\alpha-1} \cdot e^{-x/\beta} dx$$

$$P(X \leq 60) = \int_0^{60} \frac{1}{(10)^5 (24)} \cdot x^4 \cdot e^{-x/10} \cdot dx$$

$$= \frac{1}{(10^5).(24)} \int_0^{60} \underbrace{x^4}_{\text{diff}} \cdot \underbrace{e^{-x/10}}_{\text{integrate}} dx$$

technique when e is a variable

$$= \frac{1}{(10^5).24} \left[\frac{+x^4 . e^{-x/10}}{-1/10} - \frac{4x^3 . e^{-x/10}}{-1/10^2} + \frac{12x^2 . e^{-x/10}}{-1/10^3} - \frac{24x . e^{-x/10}}{-1/10^4} + \frac{24 . e^{-x/10}}{-1/10^5} \right]_0^{60}$$

$$\Rightarrow \frac{1}{10^5(24)} \left[-10x^4 . e^{-x/10} - 400 . x^3 . e^{-x/10} + 12000 . x^2 . e^{-x/10} - 240000 . x . e^{-x/10} + 2400000 . e^{-x/10} \right]_0^{60}$$

$\Rightarrow$

Ans $= 0.72$

16/10/23

{Unit - 5}

{CLT and Estimator.}

↳ "Central Limit Theorem"

Data Centre

Sample → Statistic

A small part of large population

Population — a large group

↳ categories
- ↳ Schools
- ↳ Years
- ↳ Sessions

Parameter
↳ Mean, Variance, Standard Deviation, Median, Mode

\# population → it is a term which we use to represent large data.
↳ Sample is just a small part of the population.

\# Parameter → A numerical value summarizing all the data of an entire population.
{{ Ex:- pop mean(μ), pop var(σ^2). }}

~~Static~~

\# Statistic → A numerical value summarizing the sample data.

Remark

	Mean	Variance	S.D	size
pop →	μ	σ^2	σ	N
Sample →	$\bar{x}$	s^2	s	n

↳

Central Limit Theorem.

LPU → Avg no of placement 2021 — population

Schools → CSE → S_1 → $x_1 + x_2 + x_3 + \ldots x = \bar{x}_1$

Samples — ECE → S_2 → $\bar{x}_2$

Civil → S_3 → $\bar{x}_3$

Managment → S_4 → $\bar{x}_4$

$$\mu = \bar{x}_1 + \bar{x}_2 + \bar{x}_3 + \bar{x}_4$$

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

$x \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$ SD

CLT [Liapounoff's form]

If $x_1, x_2, \ldots x_n$ be a sequence of independent RV with $E(x_i) = \mu_i$ and $Var(x_i) = \sigma_i^2$, $i = 1, 2, \ldots$;

then as $n \to \infty$, the distribution of the sum of these 'n' RV's, namely,

$S_n = x_1 + x_2 + \cdots + x_n$ tends to the normal distribution with mean 'μ' and variance(σ^2)

such that

$$\boxed{\mu = \sum_{i=1}^{n} \mu_i}$$

and

$$\boxed{\sigma^2 = \sum_{i=1}^{n} \sigma_i^2}$$

mean and variance of each sample is distinct

CLT (Lindberg - lavy's Theorem)

If $x_1, x_2, \ldots, x_n$ be a seq of independent identically distributed r.v's with $E(x_i) = \mu_i$ and $Var(x_i) = \sigma_i^2$; $i = 1, 2 \ldots$

and if $S_n = x_1 + x_2 + \ldots + x_n$;

then under certain general conditions, S_n follows a normal distribution with mean '$n\mu$' and var '$n\sigma^2$' as $n \to \infty$

Corr:- If $\bar{x} = \frac{1}{n}(x_1 + x_2 + \ldots + x_n)$

$E(\bar{x}) = \mu$. and (/mu * n)/n = /mu

$$Var(\bar{x}) = \frac{1}{n^2}(n\sigma^2)$$

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

mean of variance of each sample is same

* The lifetime of a certain brand of a electric bulb may be considered a random variable with mean 1200 hrs and SD. 250 hrs find the prob. of using CLT of Levy's form that the avg lifetime of 60 bulbs exceeds 1250 hrs.

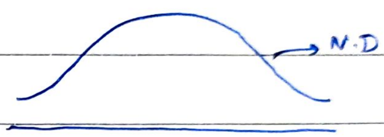
μ_i

X_i = The lifetime of the bulb.

$\bar{X}_i \rightarrow E[X_i] = 1200$ hrs.

$\sigma_i \rightarrow Var(X_i) = (250)^2$ hrs.

$$x \sim N\left(\mu, \frac{\sigma}{\sqrt{n}}\right)$$



here will we Normal Distribution

$$P(\bar{X} > 1250) = ? = P(Z > 1250)$$

$$= P\left(z > \frac{\bar{X} - \mu_{\bar{X}}}{\left(\frac{\sigma}{\sqrt{n}}\right)}\right)$$

1250, 1200, $\frac{250}{\sqrt{60}}$

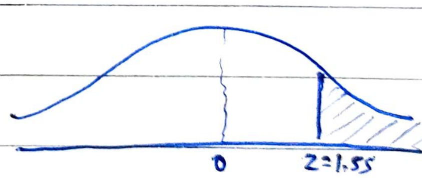
$Z = \frac{X - \mu_X}{\sigma_X}$ → RV

$Z = \frac{\bar{X} - \mu_{\bar{X}}}{\sigma_{\bar{X}}}$ → Mean

$$= P\left(Z > \frac{50}{250/\sqrt{60}}\right)$$

$$P(Z > 1.55)$$

Given $[P(Z < 1.55 = 0.9394)$
or $P(0 < Z < 1.55) = 0.4394)$



1st $P(Z > 1.55) = 1 - P(Z < 1.55)$
$= 0.0606$

IInd

$P(Z > 1.55) = 0.5 - P(0 < Z < 1.55)$

$= 0.5 - 0.4394$

0.0606

Estimator :-

↳ we use to find the info. of population with the help of the sample.

mean
var
S.D

⇒ statistic → used to find → parameter

statistic ↓ numeric value to summarize sample

parameter ↓ numeric value to summarize population.

any function of the random sample $x_1, x_2, x_3, \dots x_n$, say $T_n(x_1, x_2, \dots$ is called statistic.

A statistic is a R.V, if it is used to Estimate an unknown parameter θ (theta) of the distribution. It is called an estimator.

Type of Estimator :-
↳ 4 types

characteristic of Estimator
1) Unbiasedness
2) Consistency
3) Efficiency
4) Sufficiency

If an Estimator have there it is the best

Unbiased Estimator.

An Estimator $T_n(x_1, x_2, \dots x_n)$ is said to be an unbiased estimator if

↳ $\boxed{E[T_n] = f(\theta)}$ ★ → unbaised

E(estimator)= parameter of population

↳ $E[\bar{x}] = \mu$

$\bar{x}$ → mean of sample

μ → population mean

$E[\bar{x}] \neq \mu$ → baised

normal. __/__/__

Q If $X_1, X_2, \ldots X_n$ is a random sample for a normal population $N(\mu, 1)$, show that $t = \underbrace{\frac{1}{n}\sum_{i=1}^{n} x_i^2}_{\text{statistic}}$, is an unbaised estimator of $\underbrace{\mu^2+1}_{\text{parmeter}}$.

$\Rightarrow$ pop mean $\rightarrow \mu$

pop var $\rightarrow 1$

to prove $\rightarrow$ $E[t] = \mu^2+1$

$$E\left[\frac{1}{n}\sum_{i=1}^{n} x_i^2\right] = \mu^2+1$$

$Var(x_i) = E[x_i^2] - (E[x_i])^2$

$1 = E[x_i^2] - \mu^2$

$E[x_i^2] = \mu^2+1$

$$= \frac{1}{n}\sum_{i=1}^{n} E[x_i^2] = \frac{1}{n}\sum_{i=1}^{n}(\mu^2+1)$$

$= 1$

$$= \frac{1}{n} \times n(\mu^2+1)$$

$$\boxed{E[t] = \mu^2+1}$$

Proved

Q If T is an unbaised estimator for θ show that T^2 is a biased estimator for θ^2.

Given $\rightarrow$ $E[T] = \theta$

Show $\rightarrow$ $E[T^2] \neq \theta^2$

$(E(T))^2 = \theta^2$

Variance $= E(T^2) - (E(T))^2$

$E(T^2) - \theta^2$

$E(T^2)$ cannot be θ^2 becoz if we put $E[T^2] = \theta^2$ then variance becomes 0

and we all known variance is always greater than 0.

26/10/23

+ve and −ve biased.

if $E[T_n] > \theta$, T_n is said to be +vely biased.

$E[T_n] < \theta$, T_n is said to be −vely biased.

statistic ← $E[T_n]$, θ → parameter.

Q Show that $\dfrac{\sum x_i(\sum x_i - 1)}{n(n-1)}$ is a unbaised estimator of θ^2 for the sample $x_1, x_2, \ldots x_n$ drawn on X which takes the values 1 or 0 with respective prob. θ & $1-\theta$.

To show → $E\left[\dfrac{\sum x_i(\sum x_i - 1)}{n(n-1)}\right] = \theta^2$

Value ⇒ 1 , 0

Prob ⇒ θ , $1-\theta$

Solve ⇒

Let $x_1, x_2, \ldots x_n$ be a random sample from binomial distribution with $E[x_i] = n\theta$

$var(x_i) = n\theta(1-\theta)$

$Var(T) = n\theta(1-\theta)$

let $T = \sum_{i=1}^{n} x_i$

then → $E\left[\dfrac{T(T-1)}{n(n-1)}\right] = \dfrac{1}{n(n-1)}\left[E(T^2) - E(T)\right]$

$= \dfrac{1}{n(n-1)}\left(Var(\sum x) + [E[\sum x]]^2 - E[\sum x]\right)$

$(n(\theta)(1-\theta) + n^2\theta^2 - n\theta)$

$n\theta(1-\theta) + n\theta(n\theta - 1)$

$n\theta(1-\theta + n\theta - 1)$

$n\theta(n\theta - \theta)$

$n\theta(\theta(n-1))$

$\dfrac{1}{\cancel{n(n-1)}}\ \cancel{n}\theta^2\cancel{(n-1)}$

$= \theta^2$

Q. Let t_1 and t_2 be unbiased estimator of θ with variance σ_1^2 and σ_2^2. Consider the estimator $\hat{\theta} = \alpha t_1 + (1-\alpha) t_2$, check whether the $\hat{\theta}$ is an unbiased estimator of θ or not

find $E[\hat{\theta}] = \theta$.

$\Rightarrow$ Given $\rightarrow$ $E[t_1] = E[t_2] = \theta$

$var(t_1) = \sigma_1^2$, $var(t_2) = \sigma_2^2$ to find $E[\hat{\theta}] = ?$

$\hat{\theta} = \alpha t_1 + (1-\alpha) t_2 \Rightarrow E[\hat{\theta}] = E[\alpha t_1 + (1-\alpha) t_2]$

$$= \alpha E[t_1] + (1-\alpha) E[t_2]$$
$$= \alpha\theta + (1-\alpha)\theta$$
$$= \theta(\alpha + 1 - \alpha)$$
$$= \theta$$

//

(II) Consistency.

↳ An estimator $T_n = T(x_1, x_2, x_3, \dots x_n)$ based on random sample of size 'n' is said to be consistent estimator of $f(\theta)$, if T_n converges to $f(\theta)$ in prob, i.e, $T_n \xrightarrow{p} f(\theta)$.

$\Rightarrow$ Sufficient condition for Consistent Estimator.

→ Let T_n be a sequence of estimator such that :-

(i) $E[T_n] \rightarrow f(\theta)$ as $n \rightarrow \infty$

(ii) $Var(T_n) \rightarrow 0$ as $n \rightarrow \infty$

that means

any statistic; any parameter

$$\left.\begin{array}{l} \text{if } E[\bar{x}] \rightarrow \mu \quad \text{as } n \rightarrow \infty \\ Var(\bar{x}) \rightarrow 0 \quad \text{as } n \rightarrow \infty \end{array}\right\} \text{consistent Estimator}$$

Q. Prove that in sampling from $N(\mu, \sigma^2)$ population the sample mean $(\bar{x})$ is a consistent estimator of μ.

To prove $\rightarrow E[\bar{x}] = \mu$

$\rightarrow$ Let $x_1, x_2, \ldots x_n$ be a random sample of size n from $N(\mu, \sigma^2)$

$E[x_i] = \mu$

$Var[x_i] = \sigma^2$

$$\bar{x} = \frac{x_1 + x_2 + \ldots + x_n}{n} \qquad E[\bar{x}] = E\left[\frac{x_1 + x_2 + \ldots + x_n}{n}\right]$$

$$= \frac{1}{\cancel{n}} \times \cancel{n}(\mu)$$

mean of $X_1 \rightarrow \mu_1$

mean of $X_2 \rightarrow \mu_2$

$$\boxed{E[\bar{x}] = \mu \quad as \quad n \rightarrow \infty}$$

Variance: $(\bar{x}) = V\left[\frac{1}{n}(x_1 + x_2 + \ldots + x_n)\right]$

$Var(ax+b)$

a^2x

$$= \frac{1}{n^2}\left[V(x_1) + V(x_2) + \ldots V(n)\right]$$

$$= \frac{1}{n^{\cancel{2}}}(\cancel{n}\sigma^2)$$

$$= \frac{\sigma^2}{\cancel{n}\infty} \quad as \quad n \rightarrow \infty \quad will \quad 0$$

$$\frac{\sigma^2}{n} = 0$$

$$\boxed{V[\bar{x}] = 0 \quad as \quad n \rightarrow \infty}$$

So it is a consistent estimator

27/10/23

Q If $X_1, X_2, \ldots X_n$ are random observation on a bernoulli-variate X, taking the value 1 (P) with prob P and the value 0 with prob $(1-P)$. Show that :-

$$\frac{\sum X_i}{n}\left(1-\frac{\sum x_i}{n}\right) \text{ is an consistent estimator of } p(1-p)$$

$\bar{x}$ $\bar{x}$

$\Rightarrow$ let $T = \sum X_i$ $\qquad X \sim B(n, p)$

$$\bar{x} = \frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum x_i}{n} = T/n$$

$$E[\bar{x}] = E\left[T/n\right] = \frac{1}{n} E[T] = \frac{1}{\cancel{n}} \times \cancel{n}p = p$$

$E[\bar{x}] \to p$ as $n \to \infty$

$$V[T] = np(1-p)$$

$$V[\bar{x}] = V\left[T/n\right] = \frac{1}{\cancel{n}^2} \times \cancel{n}p(1-p)$$

$$= \frac{p(1-p)}{n}$$

$V[\bar{x}] \to 0$ as $n \to \infty$

$\Rightarrow$ $\bar{x}$ is C.E of p.

Now to check;

$\bar{x}(1-\bar{x})$ is a C.E of $p(1-p)$.

$$E[\bar{x}(1-\bar{x})]$$
$$\Rightarrow E[\bar{x} - \bar{x}^2]$$
$$\Rightarrow E[\bar{x}] - E[\bar{x}^2]$$
$$\downarrow$$
$$p$$

$\Rightarrow$ $\bar{x}(1-\bar{x}) = \bar{x} - (\bar{x})^2$ is a poly function.

T $\bar{x}$ is C.E of $p \cap \theta$

{ $\bar{x}(1-\bar{x})$ is a conti function}

$\Rightarrow$ $\bar{x}(1-\bar{x})$ is a C.E of $p(1-p)$ by using in-variance property

Remark : InVariance property of Consistent Estimator.

If T_n is a consistent estimator of $f(\theta)$ & $\psi\{f(\theta)\}$ is a continuous function of $f(\theta)$.

then $\psi\{T_n\}$ is a consistent Estimator of $\psi\{f(\theta)\}$.

Eg:- T is C.E of θ

f is a conti function of $f(\theta)$

then $f(T)$ is a C.E of $f(\theta)$.

(III) Efficient Estimator.

→ Most Efficient Estimator.

If in a class of consistent Estimator for a parameter there exist one, whose sampling variance is less than that of any such estimator, it is called most-Efficient Estimator.

T_1, T_2, T_3

$V(T_1), V(T_2), V(T_3)$ } → which is less is the most efficient estimator.

~~Condition.~~

Unbiased ← i) E[statistic] = population parameter.

Consistency ← ii) $\left.\begin{array}{l} E[\text{stat}] \to \text{pop} \\ V[\text{stat}] \to 0 \end{array}\right\}$ as $n \to \infty$.

efficiency ← iii) var[stat] → less

Q. A Random sample $X_1, X_2, X_3, \dots X_5$ is drawn from a normal population with unknown mean (μ).
Consider the following estimator to estimate μ.

(i) $t_1 = \frac{X_1 + X_2 + X_3 + \dots X_5}{5}$

(ii) $t_2 = \frac{X_1 + X_2}{2} + X_3$

(iii) $t_3 = \frac{2X_1 + X_2 + \lambda X_3}{3}$ where λ is such that t_3 is an unbaised Estimator of μ.

Find (λ), are t_1 and t_2 unbaised, state giving reasons the estimator which is best among t_1, t_2, t_3.

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

t_3 = unbaised $\rightarrow E[t_3] = \mu$ $\quad\quad E[x_i] = \mu,\ V[x_i] = \sigma^2$

$i \Rightarrow 1 \rightarrow 5$

(1) $E[t_3] = \mu$

$$E\left[\frac{2x_1 + x_2 + \lambda x_3}{3}\right] = \mu$$

$$\frac{1}{3}\left(2E[x_1] + E[x_2] + \lambda[x_3]\right)$$

$$\frac{1}{3}(2\mu + \mu + \lambda\mu)$$

$$\frac{\mu(3+\lambda)}{3} = \mu$$

$$3 + \lambda = 3$$

$$\boxed{\lambda = 0}$$

/ /

(II) $E[t_1] = \mu$

$E\left[\frac{X_1+X_2+X_3+X_4+X_5}{5}\right] = \mu$

$\frac{1}{5} E[X_1+X_2+X_3+X_4+X_5] = \mu$

$\frac{1}{\cancel{5}} \times \cancel{5}(\mu) = \mu$ $\Rightarrow$ t_1 is an unbaised estimator.

$E[t_2] = \mu$

$E\left[\frac{X_1+X_2}{2} + X_3\right] = \mu$

$\frac{1}{2} \times E[X_1+X_2] + E[X_3] = \mu$

$\frac{1}{\cancel{2}} \times \cancel{2}(\mu) + \mu = \mu$

$2\mu \neq \mu$ $\Rightarrow$ t_2 is a baised estimator.

(III) $V(t_1)$, $V(t_2)$, $V(t_3)$

$V\left(\left(\frac{1}{5}\right)(X_1+X_2 \dots + X_5)\right)$ σ^2

$\frac{1}{\cancel{25}_5}[\cancel{5}(\sigma^2)]$

$\frac{\sigma^2}{5}$

(Less) $\Rightarrow$ Most - Efficient Estimator

$V\left(\frac{X_1+X_2}{2} + X_3\right)$

$\frac{1}{\cancel{4}2}(\cancel{2}(\sigma^2)) + \sigma^2$

$\frac{\sigma^2}{2} + \sigma^2$

$\frac{3\sigma^2}{2}$

$V\left(\frac{2X_1+X_2+X_3}{3}\right)$

$\frac{1}{9}[4\sigma^2 + \sigma^2]$

$\frac{5\sigma^2}{9}$

$a\bar{x}+b$

$V(a\bar{x}+b)$

a^2

___/___/_____

30/10/23

Q If X_1, X_2, X_3 is a random sample of size 3 from a population with mean 'μ' and variance 'σ^2'. Normal distribution.

Consider the following estimators :-

$T_1 = X_1 + X_2 - X_3$

$T_2 = 2X_1 + 3X_3 - 4X_2$

$T_3 = \frac{1}{3}(\lambda X_1 + X_2 + X_3)$

where λ is such that T_3 is unbaised Estimator of μ.

find λ, T_1 and T_2 are unbaised ?, which is best estimator ?

⇒ (i) $E(T_3) = \mu$

$\frac{1}{3}\left(E(\lambda X_1) + E(X_2) + E(X_3)\right) = \mu$

$\frac{1}{3}\left[\lambda\mu + 2\mu\right] = \mu$

$= \left(\frac{\lambda}{3} + 2\right)\mu = \mu$

$\Rightarrow \frac{\lambda}{3} + 2 = 1 \rightarrow \frac{\lambda}{3} = -1$ $\boxed{\lambda = -3}$

(ii) $E(T_1) \rightarrow E\left[x_1 + x_2 - x_3\right] = 3\mu$

T_1 is not unbaised Estimator of μ.

↳ T_1 is baised.

$E(T_2) \rightarrow E(2x_1 + 3x_3 - 4x_2) = 2\mu + 3\mu - 4\mu = \mu$

T_2 is unbiased estimator of μ

(iii) $V(T_1) = 3\sigma^2$, $V(T_2) = 29\sigma^2$, $V(T_3) = \frac{11}{3}\sigma^2$

So, T_1 is most efficient estimator.

(IV) Sufficiency

↳ An estimator is said to be sufficient for a parameter if it contains all the info. in the sample regarding the parameter.

Definition → If $T_n = T(x_1, x_2, \ldots, x_n)$ is an estimator of parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size 'n' from the population, with density function → $f(x,\theta)$ such that the conditional distribution, $\underbrace{\text{density function}}_{\text{Probability}}$ of $x_1, x_2, \ldots, x_n$ given T_n is independent of θ, then T is sufficient estimator of θ.

$$\text{i.e, } P(x_1, x_2, \ldots, x_n / T_n) = \frac{P((x_1, \ldots, x_n) \cap T_n)}{P(T_n)}$$

factorisation Theorem →

If $T = t(x)$ is sufficient for θ iff (if and only if) the joint density function (L) of the sample values can be used in the form.

$$\boxed{L = g_\theta[t(x)]\, h(x)} \quad \text{or} \quad \boxed{L = \prod_{i=1}^{n} f(x,\theta)}$$

$\prod$ → for multiplication

where $g_\theta[t(x)]$ depends on θ and 'x' only through the value of $t(x)$ and $h(x)$ is independent of θ.

↳

Q. Let $X_1, X_2, \ldots, X_n$ be a random sample from $N(\mu, \sigma^2)$. Find sufficient estimator for σ^2 and μ.

$\rightarrow$ Let $\theta = \mu, \sigma^2$, $-\infty < \mu < \infty$, $0 < \sigma^2 < \infty$.

$$L = \prod_{i=1}^{n} f(x, \theta)$$

$$L = \prod_{i=1}^{n} \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_i-\mu)^2}{2\sigma^2}}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left\{\frac{-1}{2\sigma^2}\sum_{i=1}^{n}(x_i-\mu)^2\right\}$$

$$= \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left\{\frac{-1}{2\sigma^2}\left(\sum_{i=1}^{n} x_i^2 + \mu^2\sum_{i=1} 1 - 2\mu\sum_{i=1} x_i\right)\right\}$$

$$L = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left\{\frac{-1}{2\sigma^2}\left\{\sum_{i=1}^{n} x^2 + n\mu^2 - 2\mu \underbrace{\frac{\sum x_i}{n}}_{\bar{x}} \times n\right\}\right\}$$

$$L = \{g_\theta\, t(x)\}\, h(x) \quad \text{compare}$$

cannot contain parameter

$$[g_\theta\, t(x)] = \left(\frac{1}{\sigma\sqrt{2\pi}}\right)^n . \exp\left\{\frac{-1}{2\sigma^2}\left\{\sum_{i=1}^{n} x^2 + n\mu^2 - 2\mu n \bar{x}\right\}\right\}$$

$h(x) = 1.$

$$t(x) = \{t_1(x), t_2(x)\}$$

$$= \left\{\sum x_i^2, \sum x_i\right\}$$

$\therefore$ μ & σ^2 are sufficient estimator of $N(\mu, \sigma^2)$

Q. Let $x_1, x_2, \ldots x_n$ be a random sample from a population with pdf $f(x,\theta) = \theta x^{\theta-1}, 0 < x < 1, \theta > 0,$

Show that

$$t_1 = \prod_{i=1}^{n} x_i \text{ is sufficient estimator for } \theta.$$

$\rightarrow$ $L = \prod_{i=1}^{n} f(x,\theta)$

$$= \prod_{i=1}^{n} \theta x^{\theta-1} = \theta^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{\theta} \cdot \frac{1}{\left(\prod_{i=1}^{n} x_i\right)^1}$$

$$L = [g_\theta t(x)]\, h(x)$$

$F(x,y) = x^3 + y^2 + 1$

$\left.\frac{dF}{dx}\right|_{y=const}$; $\left.\frac{dF}{dy}\right|_{x=const}$

partial derivative.

where $\rightarrow [g_\theta t(x)] = \theta^n \cdot \left(\prod_{i=1}^{n} x_i\right)^{\theta}$

$$h(x) = \frac{1}{\left(\prod_{i=1}^{n} x_i\right)^1}$$

$\therefore \prod_{i=1}^{n} x_i$ is a S.E of θ.

complete

__/__/____

Maximum Likelihood Estimator (MLE).

↳ Let $x_1, x_2, \ldots, x_n$ be a random sample from a population with density function, $f(x_i/\theta)$ or $f(x_i;\theta)$ whose likelihood function is given by :-

$$L = f(x_1, x_2, \ldots, x_n | \theta) = \prod_{i=1}^{n} f(x_i | \theta) \quad \text{—— ①.}$$

For finding the maximum likelihood estimator, we have to maximize the likelihood function given in eq ①.

condition :-

(i) $\frac{dL}{d\theta} = 0$

(ii) $\frac{d^2L}{d\theta^2} < 0$

$N(\mu, \sigma^2)$ → in normal distribution we have two ~~vatie~~ variables partial derivate

$$\left.\frac{dL}{d\mu}\right|_{(\sigma^2 = \text{const})} \qquad \left.\frac{dL}{d\sigma^2}\right|_{\mu = \text{const}}$$

** . In a random sampling from a normal population, $N(\mu, \sigma^2)$, find the max LH estimators for :-

(i) μ when σ^2 is known.

(ii) σ^2 when μ is known.

(iii) Simultaneous estimation of μ & σ^2.

↳

3/11/23

(unit-6) {Hypothesis Testing}

all game of assumption.

first Assumption is called NULL Hypothesis.

Step → (i) $H_0 \rightarrow \mu = 17.2$

$H_0 \rightarrow \mu \neq 17.2$

(ii) Test statistics → Z-test [$n \geq 30$]
→ t-test [$n < 30$]
→ χ^2-test [Goodness of fit]
→ F-test → (var)

chi

(iii) Conclusion → Tabulated, Calculated

Tabulated → Reject H_0, Accept H_0

Hypothesis – It is a proposed explanation for a phenomenon.
The processdure of testing hypothsis involves the following steps :-

(i) To lay down the hypothesis.

(ii) To choose the level of significance, → 1% 5% 10% → 5% → α ↳ for risk factor.

(iii) choice of test statistic

(iv) To find the critical value.

[conclusion]

NULL Hypo → (i) It says that their is no significant statistical difference b/w the population value and sample value.

(ii) It is denoted by (H_0).

(iii) H_0 will be that sample static which doesn't differ significantly from the Hypothetized parameter value.

Alternative Hypothesis. :-

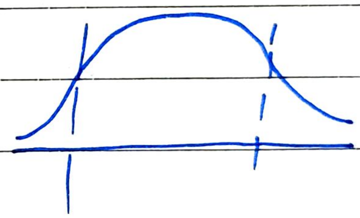
→ Any Hypothesis which is complimentary to the NULL Hypo. is called Alternative Hypothesis, denoted by (H_1).

H_0 $\mu = \mu_0$

H_1 $\mu \neq \mu_0$ [Two- tailed test]

$\mu > \mu_0$ [Left - tailed test]

$\mu < \mu_0$ [Right - tailed test]



Errors in Sampling / Test Statistics.

level of Significance

Decision

		H_0 is True	H_0 is False
True state	Reject H_0	Type - I Error [Prob = α]	Correct decision
	Accept H_0	Correct decision	Type - II Error [Prob = β]

level of Confidence

- Type I → Reject H_0, when H_0 is True.
  Type II → Accept H_0, when H_0 is False

- $P[\text{Type I}] = \alpha$
  $P[\text{Type II}] = \beta$

9/11/23

Degree of freedom

↳ The max number of independent values which have the freedom to vary in sample data, i.e, $\boxed{(n-1) \text{ is degree,}}$ where n = no. of independent obs. in the sample.

* mean → single mean.
  mean → Difference mean.
  ↳ in mean do only Z-test or T-test

Z-test: $\boxed{n \geq 30}$ T-test: $\boxed{n < 30}$

Z-Test

Step-I → To Take Hypo ie H_0 and H_1

Step-II → Choose level of significance.

Step-III → Test statics -

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

[σ^2 is Known] used when.

or

$$\boxed{Z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}}$$

sample variance

[σ^2 is not known]

T-Test

→ To Take hypothesis i.e H_0 & H_1.

→ Choose level of significance

~~→ let x_i be a random~~

↳ Next page

$$\boxed{\text{where } s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Z-Test

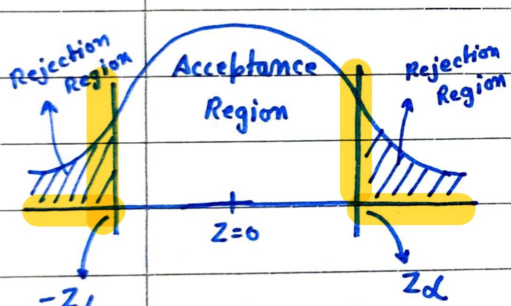
Step-IV → Conclusion

→ if calculated $|Z| > Z_\alpha$ → Reject H_0

→ " " $|Z| < Z_\alpha$ → Accept H_0

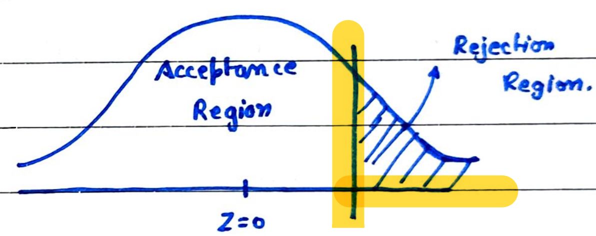
level of significance

Critical Region	1%	5%	10%
Two-Tailed Test	$\lvert Z_\alpha \rvert = 2.58$	$\lvert Z_\alpha \rvert = 1.96$	$\lvert Z_\alpha \rvert = 1.645$
Right-Tailed Test	$Z_\alpha = 2.33$	$Z_\alpha = 1.645$	$Z_\alpha = 1.28$
left-Tailed Test	$Z_\alpha = -2.33$	$Z_\alpha = -1.645$	$Z_\alpha = -1.28$



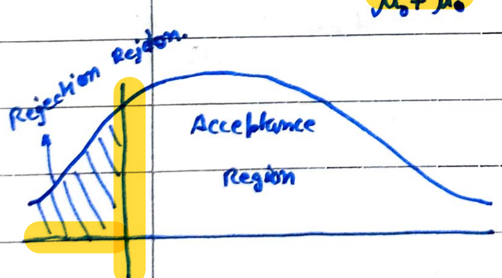
Two-Tailed Test.

$\mu_0 \neq \mu_0$



Right-Tailed Test.

$\mu_0 > \mu$



Left-Tailed Test $\mu_0 < \mu$.

09/11/23

T-Test

Step-III → let x_i $(i = 1,2,3,\ldots n)$ be a random sample of size n, from $N(\mu, \sigma^2)$ then the student's T-test is given by :-

$$\boxed{t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}}} \sim t(n-1)\, df$$

df → degree of freedom.

where S^2 is the unbaised estimator of σ^2

i.e $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$.

Step-IV → Conclusion.

- calculated $|t|$ < tabulated of t, then reject H_0.
- calculated $|t|$ > tabulated of t, then accept H_0.

Remark :- (i) Z-Test.

let $s^2 = \frac{1}{n} \sum_{i=1}^{n} (x_i - \bar{x})^2$ and $S^2 = \frac{1}{n-1} \sum_{i=1}^{n} (x_i - \bar{x})^2$

$\Rightarrow ns^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$ — ① and $(n-1)S^2 = \sum_{i=1}^{n} (x_i - \bar{x})^2$ — ⓘⓘ

① and ②

$$ns^2 = (n-1)S^2$$

$$s^2 = \left(\frac{n-1}{n}\right) S^2$$

$$s^2 = \left(1 - \frac{1}{n}\right) S^2$$

As $n \to \infty$

$s^2 = S^2$

Hence, for large sample $s^2 \approx S^2$

$\therefore \sigma^2 \approx S^2$

/ /

(ii) T-Test

from eq^n ① & ②

$$ns^2 = (n-1)S^2$$

$$\frac{s^2}{n-1} = \frac{S^2}{n} \quad \text{or} \quad \frac{S^2}{n} = \frac{s^2}{n-1}$$

$$t = \frac{\bar{x} - \mu}{\frac{S}{\sqrt{n}}} \Rightarrow \boxed{\frac{\bar{x}-\mu}{\sqrt{\frac{S^2}{n}}} \Rightarrow \frac{\bar{x}-\mu}{\sqrt{\frac{s^2}{n-1}}}}$$ //.

Q. (i) A sample of 900 members has a mean 3.4 cm and sd = 2.61 cm. In the sample from a large population of 3.25 cm and std = 2.61 cm

μ σ

(ii) If the population is normal and its mean is unknown then find the 95% judicial limit ~~and~~/ confidence limit of the mean.

risk = 5%

→ $n = 900$, $\bar{x} = 3.4$ cm, $s = 2.61$ cm;

$\mu = 3.25$ cm, $\sigma = 261$ cm;

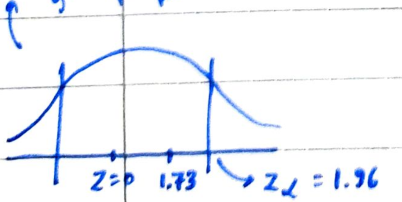
Normal pop

Z-Test $[n \geq 30]$

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{3.4 - 3.25}{\frac{2.61}{\sqrt{900}}} = \frac{0.15}{\frac{2.61}{30}} = \cancel{15} \times \frac{1}{\cancel{2.61}} \times \frac{1}{\cancel{30}_2} \times \frac{1}{100}$$

$= 1.73$

if not given greater or lesser than two tailed.



calculated $z < Z_\alpha$ → Accept H_0.

Two-Tailed

___/___/______

(ii) Fudical Limit = $\boxed{\bar{x} \pm \frac{\sigma}{\sqrt{n}} Z_{\alpha}}$ → Tabulated value.

$$3.4 \pm \frac{2.61}{\sqrt{900}} (1.96)$$

$$= \{3.2295, 3.5705\}.$$

Q If a sample of 100 recorded disk in the US during last year represent an average life span of 71.8 years. Assuming a population s.d of 8.9 years. Does this seem to indicate the average lifespan today is greater than 70 year? use 0.05 level of significance? → 5% → $Z_{\alpha} = 1.96$

→ $n = 100$, $\bar{x} = 71.8$ yrs, $\sigma^2 = 8.9$ yrs, $\mu = 70$.

↳ $H_0 \Rightarrow \mu = 70$ yrs.

$H_1 \Rightarrow \mu > 70$ yrs.

[Right tailed]

Z-test.

$$Z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}} = \frac{71.8 - 70}{\frac{8.9}{10}} = \frac{1.8}{\frac{8.9}{10}} = 2.02.$$

$Z > 1.96645 \; Z_{\alpha}$.

↳ Reject H_0.

10/11/23

Q. A mechanist is making Engine parts with axle diameter of ~~0.10~~ 0.700 inch. A random sample of 10 parts shows a mean diameter of 0.742 inch. with a sd of 0.04 inch. Compute the statistic you would use whether the work is meeting the specifications at 5% level of Significance with 9 degree of freedom?
Given t_9 at 5% 2.262.

$H_0 = \mu = 0.700$ inch

t → test $\quad n = 10, \quad \bar{x} = 0.742 \quad s = 0.04$

$H_1 = \mu \neq 0.700$ inch

[Two-Tailed Test]

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n}}} = \frac{\bar{x} - \mu}{\sqrt{\frac{s^2}{n-1}}}$$

$$\frac{0.742 - 0.700}{\sqrt{\frac{(0.04)^2}{9}}} = \frac{0.042}{\frac{0.04}{3}} = \boxed{3.15}$$

calculated value is $3.15 > 2.262$

→ ⇒ Reject H_0

Q. The mean weekly sales of soap bars in department store is $146.\bar{3}$ bars per store. After advertising campaign the mean weekly sales for 22 stores increase to 153.7 with ~~std~~ sd of 17.2. was the adverting successful?

→ [21 df $t_{0.05} = 1.72$]

$n = 22 \quad \bar{x} = 153.7 \quad \mu = 146.3 \quad$ (s) sd $= 17.2$

$H_0 = \mu = 146.3$

$H_1 = \mu > 146.3$

$$t = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n-1}}} = \frac{153.7 - 146.3}{\frac{17.2}{\sqrt{21}}} = \frac{7.4}{\frac{17.2}{\sqrt{21}}} = \frac{7.4}{3.75} = 1.97$$

[Right Tailed Test]

calculated value is $1.97 > 1.72$

→ Reject H_0

* Difference of Mean.

↳ z - test

i) Let $\bar{x}_1$ be the mean of a sample size n_1 from a pop μ_1 and $\bar{x}_2$ be the mean of sample size n_2 from pop μ_2 with var σ_1^2 and σ_2^2 then under the $H_0 \rightarrow \mu_1 = \mu_2$.

The test statistic (for large Sample) becomes

$$\boxed{Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \sim N(0,1)}$$

or

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

↳ t - Test

i) Under the H_0 the sample has be drawn from the normal pop with mean μ_x and μ_y and under assumption that the pop variance are equal ie, $\sigma_x^2 = \sigma_y^2 = \sigma^2$ (say).

then the test statistic (for small sample) is

$$\boxed{t = \frac{(\bar{x} - \bar{y}) - (\mu_x - \mu_y)}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}}$$

$$s^2 = \frac{1}{n_1 + n_2 - 2}\left[\sum_{i=1}^{n} (x_i - \bar{x})^2 + \sum_{i=1}^{n} (y_i - \bar{y})^2\right]$$

with $\boxed{(n_1 + n_2 - 2) \text{ df}}$

↳

14.29

Book

SC Gupta and VK Kapur.

Q The means of 2 ~~signal~~ single large sample 1000 and 2000 members are 67.5 inch and 68 inch resp. can the samples be regarded as drawn from ~~sample~~ the same pop. of s.d = 2.5 inch, Test at 5% level of significance.

$\sigma_1^2 = \sigma_2^2$

$\sigma_1 = \sigma_2$

→ $n_1 = 1000$ $\quad \bar{x}_1 = 67.5$ $\quad$ s.d = 2.5. $\quad$ 5% l.o.s.

$n_2 = 2000$ $\quad \bar{x}_2 = 68$

$H_0 = \mu_1 = \mu_2$

$H_1 = \mu_1 \neq \mu_2$

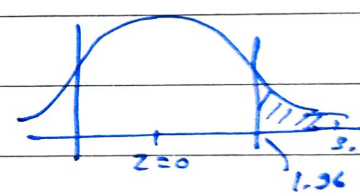
[Two-Tailed Test)

$$Z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{(2.5)^2}{1000} + \frac{(2.5)^2}{2000}}} = -0.5$$

$$= \frac{-0.5 \cancel{\times 1000}}{2.16 \times}$$

$$\frac{6.25}{1000} + \frac{6.25}{2000}$$

$$\frac{12.5 + 6.25}{2000} \Rightarrow \sqrt{\frac{18.75}{2000}} = -5.2$$



Calculated $|Z| = 5.1 > 1.96$.

Reject H_0

Paired t-test :- Pair t-test for difference of means.

consider the case when:-

(i) the sample size are equal $n_1 = n_2$.

(ii) the two sample are not ~~ind~~ independent but the sample observation are paired together.

Under the null H_0, the test statistic is :-

$$\boxed{\bar{d} = \frac{1}{n}\sum_{i=1}^{n} d_i} \qquad \boxed{t = \frac{\bar{d}}{\frac{S}{\sqrt{n}}}} \qquad \boxed{\text{where } S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(d_i - \bar{d})^2} \text{ with}$$

$(n-1)$ df

__/__/__

Q. In a certain exp. to compare two types of animals foods A and B, the following results if increase in weights were obtained in animals.

Animal number	1	2	3	4	5	6	7	8	Total
increased weight: food A →	49	53	57	52	47	50	52	53	407
food B →	52	55	52	53	50	54	54	53	423

a) assuming that the two samples of animals are independent. Can we conclude food B is better than food A.

b) Also, examine the case where the same set of animals were used in both foods.

{Given $t = 0.05$ with 14 deg of $f = 1.76$;
$t = 0.05$ with 7 deg of $f = 1.90$, }

→ X: food A, Y: food B

$H_0 = \mu_x = \mu_y$

$H_1 = \mu_x < \mu_y$ [Left tailed test]

t-test will done as $n < 30$

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

Food A

X	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
49	-1.8	+3.26
53	2.2	4.64
51	0.2	0.4
52	1.2	1.44
47	-3.8	14.44
50	-0.8	6.4
52	1.2	1.44
53	2.2	4.64

$\bar{x} = \frac{407}{8} = 50.8$

Food B

Y	$(y_i - \bar{y})$	$(y_i - \bar{y})^2$
52	-0.8	
55	2.2	
52	-0.8	
53	~~-2.8~~ 0.2	
50	-2.8	
54	1.2	
54	1.2	
53	0.2	

$\bar{y} = 52.8$

$t = -2.17 \Rightarrow |-2.17| = 2.17$

a) for left tailed test, $t_{0.05}$ with 14 f is -1.76

$\Rightarrow$ calc > tabulated

$2.17 > -1.76$

Reject H_0

b) Paired t-test, $n_1 = n_2 = n$

$t = \frac{d/s}{\sqrt{n}}$ with $(n-1)$ d.o.f where $s^2 = \frac{1}{n-1}\Sigma(d - d_i)^2$

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

= 2

★ χ^2- test :-   Mean ⇒ t-test , z-test

variance ⇒ f-test - χ^2 test

Degree of freedom = $(n-1)$

↳ Two tests if the hypothetical values of population Variance is :-

let → (i) $\sigma^2 = \sigma_0^2$ (say)

(ii) To test goodness of fit.

★ Goodness of fit :- It is used for testing the significance of the discrepency b/w theory and experiment for which we use χ^2-test of goodness of fit.

$$\boxed{\chi^2 = \sum_{i=1}^{n} \left(\frac{[f_i - e_i]^2}{e_i}\right)}$$

observed frequency   expected frequency.

Q. The demand for a particular spare part in a factory found to vary from day to day. In a Sample study, the following information was obtained.

Days :	Mon	Tue	Wed	Thu	Fri	Sat	~~Sun~~
Demand of parts :	1124	1125	1110	1120	1126	1115	

Test the hypothesis that the no. of spare parts demanded does not depend on the day of week.

{Given value → chi-sq significance at 5, 6, 7 df are resp. 11.07, 12.59, 14.07 at 5% level. }

H_0 ⇒ ~~H_1~~ H_0 ⇒ $\sigma^2 = 11.07$

$\sigma^2 \neq 11.07$

→ H_0 : no. of spare parts does not depends on days of week.

Days	frequency f_i	e_i	$f_i - e_i$	$(f_i - e_i)^2$	$(f_i - e_i)^2/e_i$
Mon	1124	1120	4	16	
Tue	1125	1120	5	25	
Wed	1110	1120	-10	100	
Thu	1120	1120	0	0	
Fri	1126	1120	6	36	
Sat	1115	1120	-5	25	

$\frac{\Sigma f_i}{n} = 1120$ $\qquad \Sigma(f_i - e_i)^2 = 202$

$$\chi^2 = \frac{\Sigma (f_i - e_i)^2}{e_i} = \frac{202}{1120} = 0.18 \approx (0.179)$$

The degree of freedom is 5 whose value is 11.07

Calc $\left(\chi^2_{0.05}\right) < \left(\text{Tabulated } \chi^2_{0.05}\right)$

$0.179 < 11.07$

Accept H_0.

Q. The following figure shows the distribution of digits in numbers choosen at random from a telephone directory.

Digits :	0	1	2	3	4	5	6
freq :	1026	1107	997	966	1075	933	1107

7	8	9
972	964	853

{Given : $\chi_{0.05}$ for 9 df $= 16.919$}

Test whether the digits may be taken to occur equally frequent in directory.

→

$(f_i - e_1)$	26	107	-3	-34	75	-67	107	-28	-36	-147
$(f_i - e_1)^2$	676	11449	9	1156	5625	4489	11449	784	1296	21609

$e_1 = 1000$

$\therefore \quad \chi^2 = \dfrac{\Sigma (f_i - e_1)^2}{e_1} = \dfrac{58542}{1000} = 58.542$

$\text{Calc}\left(\chi^2_{0.05}\right) = 58.54 > \left(\text{Tab of } \chi^2_{0.05}\right) = 16.919$

$= 58.54 > 16.919$

$=$ Reject H_0.

★ **f-test** :-

↳ we use f-test for equality of two population Variance.

↳ under the null-Hypothesis (H_0), the population Variance are equal $\Rightarrow \quad H_0 : \sigma_x^2 = \sigma_y^2 = \sigma^2$ (say);

↳ Two independent estimators of the population Variance are homogeneous, then the test statistics of F is given by :-

$$\boxed{F = \frac{S_x^2}{S_y^2}}$$

→ unbaised

→ always greater

→ smaller than S_x

$$S_x^2 = \frac{1}{n_1 - 1} \sum_{i=1}^{n} (x_i - \bar{x})^2$$

and $$S_y^2 = \frac{1}{n_2 - 1} \sum_{i=1}^{n} (y_i - \bar{y})^2$$

↳

__/__/__

Q. In one sample of 8 obs, the sum of the squares of deviation of the sample values from the sample mean was 84.4 and in the other sample of 10 obs, it was 102.6.

Test whether these difference is significant at 5% Level. Given that Dof for $n_1 = 7$ and $n_2 = 9$ is 3.29.

$H_0 = \cancel{x_1} = \cancel{y}$ $\sigma_x^2 = \sigma_y^2$

$H_1 = \cancel{x} \neq \cancel{y}$

→ $\Sigma(x_i - \bar{x})^2 = 84.4, \; n_1 = 8$

$\Sigma(y_i - \bar{y})^2 = 102.6, \; n_2 = 10$

$S_x^2 = \frac{84.4}{7} = 12.05 \qquad S_y^2 = \frac{102.6}{9} = 11.4$

$F = \frac{S_x^2}{S_y^2} = \frac{12.05}{11.4} = 1.057$

$\text{Calc}\left(F_{0.05}\right) = 1.057 < \text{Tab}\left(F_{0.05}\right) = 3.29$

⇒ Accept H_0.

Q. Two random samples, where

Sample	Size	Sample Mean	Sum of sq. of deviation	
1	10 (n_1)	15 ($\bar{x}_1$)	90	$\Sigma(x_i - \bar{x})^2$
2	12 (n_2)	14 ($\bar{x}_2$)	108	$\Sigma(y_i - \bar{y})^2$

Test whether the both samples comes from the same normal population at 5% Level of Significance.

$H_0 \Rightarrow \mu_1 = \mu_2$

$H_1 \Rightarrow \mu_1 \neq \mu_2$

sample size < 30

⇒ t-test

↳ Paired test → when $n_1 = n_2$

✓↳ Difference of mean $\mu_1 = \mu_2$

$$t = \frac{\bar{x}-\bar{y}}{s\sqrt{\frac{1}{n_1}+\frac{1}{n_2}}}$$

$$s^2 = \frac{1}{n_1+n_2-2}\left[\Sigma(x_i-\bar{x})^2 + \Sigma(y_i-\bar{y})^2\right]$$

$$= \frac{1}{10+12-2}\left[90+108\right]$$

$$s = \frac{1}{20}\left[198\right]$$

$$t = \frac{15-14}{\frac{1}{20}\times 198\sqrt{\frac{1}{10}+\frac{1}{12}}} \rightarrow \frac{22}{120}$$

$$= \frac{20}{198\times}\frac{1}{13.74} = 0.742$$

$$\Rightarrow \boxed{t = 0.742}$$

So calc at $df=20$ $(t_{0.05} = 0.742)$ < tab of $df=20$ $(t_{0.05} = 2.086)$

$\Rightarrow H_0$ is accepted

Given → $F_{0.05}(9,11) = 2.90$;
$F_{0.05}(11,9) = 5.10$;
$t_{0.05}(20) = 2.086$;
$t_{0.05}(22) = 2.07$;

df → degree of freedom

$$\left\{\frac{1}{n_1+n_2-2}\right\} = \frac{1}{20} \Rightarrow 20$$

(ii) for Variance,

$$S_1^2 = \frac{1}{n_1-1}\times\sum_{i=1}^{n}(x_i-\bar{x})^2 = \frac{1}{9}\times 90 = 10$$

$$S_2^2 = \frac{1}{n_2-1}\times\sum_{i=1}^{n}(y_i-\bar{y})^2 = \frac{1}{11}\times 108 = 9.81$$

$$F = \frac{S_1^2}{S_2^2} = \frac{10}{9.81} = 1.019$$

Calc value of (9,11) df < Tab value of (9,11)

$= 1.019 < 2.90$

$\Rightarrow H_0$ is accepted

Q. Below are the given "gains" in weights (in Kg) of pigs fed on two diets A and B.

⇒ <u>Gain in Weights.</u>

Diet A → 25, 32, 30, 34, 24, 40, 32, 24, 30, 31, 35, 25.

Diet B → 44, 34, 22, 10, 47, 31, 40, 30, 32, 35, 18, 51, 21, 35, 29, 22

→ Test if the two diet significantly in regard their effect on increase in weight.

Given → { tab value = 2.06 }

⇒ $H_0 \rightarrow \mu_A = \mu_B$ $\qquad$ $n_1 = 12$, $n_2 = 15$ } since $n < 30$ ↳ we will do t-test ↓ and $\mu_A = \mu_B$ ↓ so (difference of mean)

$H_1 \rightarrow \mu_A \neq \mu_B$

[Two Tailed Test] ↓ Here value can't be -ve.

$\bar{x}_1 = 336/12 = 28.$

$\bar{X}_2 = 450/15 = 30.$

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

$$S^2 = \underbrace{\frac{1}{n_1 + n_2 - 2}}_{25} \left[\Sigma(x_i - \bar{x})^2 + \Sigma(y_i - \bar{y})^2\right]$$

3	12, 15
2	4, 5
2	2, 5
5	1, 5
	1, 1

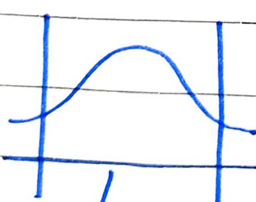
$$\Rightarrow t = \frac{28 - 30}{8.46\sqrt{\frac{1}{12} + \frac{1}{15}}}$$

$\sqrt{\frac{9}{60}}$

$S^2 = 71.6$

$\boxed{S = 8.46}$

Two tailed →



$$= \frac{-2}{8.46 \times 0.38} = -0.622 \approx (-0.609)$$

↳ t can't be -ve

$\Rightarrow t = |-0.609| = \boxed{0.609}$

Calc value < tab value

0.609 < 2.06 $\qquad$ ⇒ H_0 is accepted

———×——— ———×———