Statistics: Date: 11.12.2023

বই: Gupta- Kapoor
Fundamentals of Math. Stat.

* Variable
* Random variable
  ↳ Probability add হয়।
* Normal distribution
  ↳ probality distribution

* $\boxed{X \sim N(\mu, \sigma^2)}$

variable, follows, Normal, Mean, variance^2

* (μ, σ^2) → parameter গুলো (value exact বের হয় না)

estimate value

$(\bar{X}, S)$ → sample wise model এর একটা কাছাকাছি মান পাবো। estimate value.

* $\bar{X} - \mu$ → মান নেবার সময় error, estimated
  নইলে sample এ সমস্যা

* $S^2 - \sigma^2$ → error

Inference

(i) Estimation:

$\hat{\mu}$ ← Estimated value

$(\hat{\mu} = \bar{x}) > 0$
< 0

(ii) Test: Estimated value
প্রত্যাশিত মানের সমান আছে কিনা Test করবো।

* Inference = Estimation + Test

$H_0: \mu = 1800$
$H_1: \mu < 1800$
$H_2: \mu > 1800$

$\bar{X} = 1750$

* Sampling distribution:
  - χ^2 - (chi-square)
  - t - distribution
  - f - distribution
  - z - distribution → standard Normal distribution

* Question:

Section A: 3টা

- Estimation < Point estimation, properties
- sampling distribution
  - χ^2 (Imp.)
  - t-dis.
  - F-dis (2020 এসেছে)
  - z-dis.

Section-B: 3টা

- Contigency table.
- (parametic) Test of signi-ficance
  - χ^2
  - t
  - f
  - z
- non-parametic test.

-Aminul sir : 01715 361973
aminul @ nu.ac
mdaminul@gmail.com.

Scanned with OKEN Scanner

Date: 11.12.23

* Z-distribution (disⁿ):

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

$Z \sim N(0,1) \rightarrow$ standard Normal disⁿ

$$\boxed{Z = \frac{X-\mu}{\sqrt{V(x)}} = \frac{X-\mu}{\sigma}}$$

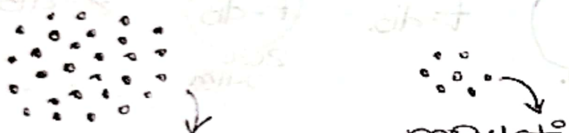
$\Rightarrow Y \sim N(10, 16)$

$$Z = \frac{y-10}{\sqrt{16}} = \frac{y-10}{4} \sim N(0,1)$$

* Population:

* Random variable

→ যদি sample আসার probability থাকে।



population

population representing sample

· $\hat{\mu} = \bar{x}$ · sample $\bar{x}$?

→ parameters:

parameters is a constant that specifies the population distribution.

Example:

· $X \sim N(\mu, \sigma^2)$

μ, σ are two parameters of Normal disⁿ

· $P(X=x) = P(x) = \dfrac{e^{-\lambda}\lambda^x}{x!}$

λ is the parameter of poisson disⁿ

$\bar{X}$: $\bar{X} = \dfrac{\sum_{i=1}^{n} X_i}{n} = \dfrac{x_1+x_2+\dots+x_n}{n}$

This is the function of sample observation.

* Statistic:

Statistic is the function of sample observation.

Example: $\bar{X}, S^2, r$ etc.

* parameter vs statistic:

দিন এর pdf.

* population probabity disⁿ:

* Sampling disⁿ: Imp

$\boxed{\chi^2, t, f}$

xm এর জন্য

defination, derivation, application
variance, Mean, β_1, β_2,
Moment generating function (MGF), properties (mgf)

range:

$Z \sim N(0,1) \longrightarrow -\infty < Z < \infty$

$Z = \dfrac{X-\mu}{\sigma}$ → X এর range: $-\infty < x < \infty$.

$Z^2 \longrightarrow \chi^2$ disⁿ

0 হলে meaningless

$0 \le \chi^2 \le \infty$

$Z^2 \longrightarrow \chi^2$

Here, degree of freedom (d.f) = 1.

~~$Z_1^2 + Z_2^2 + Z_3^2 + \dots + Z_n^2 \rightarrow \chi^2$~~

$\therefore Z_1^2 + Z_2^2 \longrightarrow \chi^2$

d.f = 1+1.

$\therefore \sum_{i=1}^{n} z_i^2 \sim \chi^2 \to df : n.$

Date: 12.12.2023

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

Probability density function. (pdf)

$$f(x) = \frac{1}{\sqrt{2\pi\sigma v}} e^{-1/2\left\{\frac{x-\mu}{\sigma}\right\}^2} \quad ; \; -\infty < x < \infty, \; -\infty < \mu < \infty, \; \sigma^2 > 0$$

$= 0$; otherwise

$0 \le x \le 1.$

- $\int_x f(x)\,dx = 1.$

- $\int_{-\infty}^{\infty} f(x)\,dx = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi\sigma v}} e^{-1/2\left\{\frac{x-\mu}{\sigma}\right\}^2} \cdot dx$

$= 1.$ (Integation করলে 1)
সরাসরি

Mean, $\bar{X} = \frac{\sum_{i=1}^{n} x_i}{n} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$

- Chi, $Z = \frac{x-\mu}{\sigma} \sim N(0,1)$
Standard Normal

- Chi-square,

$$Z^2 = \left(\frac{X-\mu}{\sigma}\right)^2 \sim \chi^2 \text{ (chi-sqr)}$$

with (df). → degrees of freedom
(স্বাধীনতার মাত্রা)

$$* \sum_{i=1}^{n} z_i^2 = \sum_{i=1}^{n} \left(\frac{x_i - \mu}{\sigma}\right)^2 \sim \chi^2 \text{ with } n\,df.$$

χ^2 : * defination

- $f(\chi^2) = \frac{1}{2^{n/2}\,\Gamma_{n/2}} \cdot e^{-\chi^2/2} \cdot (\chi^2)^{n/2-1} \; ; \; \chi^2 \geq 0$

$= 0$; otherwise

- $\Gamma \to$ gmmma

- If $x \sim \chi^2$,

then pdf of x is,

$$f(x) = \frac{1}{2^{n/2}\,\Gamma_{n/2}} \cdot e^{-x/2} \cdot x^{n/2-1} \; ; \; x \geq 0$$

$= 0$; otherwise

- Mean $= E(x) = \int_0^\infty x f(x)\,dx.$
- variance $= V(x) = E(x^2) - \{E(x)\}^2$
- Moment, $\mu_r' = E(x^r) = \int_0^\infty x^r f(x)\,dx$
- $E(x^2) = \int_0^\infty x^2 f(x)\,dx.$

*Q: Find the moment generating (mgf) function of χ^2-dis^n. Hence or otherwise, find mean, variance, β_1 and β_2.
xm এ আসবে না

Solve: By defination of mgf,

$$M_x(t) = E(e^{tx}) = \int_x e^{tx} f(x)\,dx$$

Expected value of e^{tx}

$$= \int_0^\infty e^{tx} \frac{1}{2^{n/2}\,\Gamma_{n/2}} \cdot e^{-x/2} \cdot x^{n/2-1} \cdot dx$$

efunction থাকার জন্য সরাসরি Integration করেও 1 আসবে না।

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty e^{tx} \cdot e^{-x/2} \cdot x^{n/2-1} \cdot dx$$

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \int_0^\infty \exp\left[-\left(\frac{1-2t}{2}\right)x\right] x^{n/2-1} \cdot dx$$

$$= \frac{1}{2^{n/2}\Gamma(n/2)} \cdot \frac{\Gamma(n/2)}{[(1-2t)/2]^{n/2}}$$

$$= (1-2t)^{-n/2}$$ [Using gamma Integral]

; $|2t|<1$.

$\therefore M_x(t) = (1-2t)^{-n/2}$.

Date: 18.12.2023

* Moment generating function:
function that generates moment.

- $X \sim \chi^2_n$
- $\boxed{M_X(t) = (1-2t)^{-n/2}}$; $2t<1$
  $\therefore t < \frac{1}{2}$.

$\mu_r' = \frac{\partial^r}{\partial t^r} M_x(t) \; |t=0$.

$\mu_1' = \frac{\partial}{\partial t} M_x(t) \; | t=0$.

$= \frac{\partial}{\partial t} \{(1-2t)^{-n/2}\} \; / t=0$.

$= -\frac{n}{2} (1-2t)^{-\frac{n}{2}-1} \cdot (-2) / t=0$

$= n(1-2t)^{-n/2-1} \quad | t=0$.

$= n(1-2\cdot 0)^{-n/2-1}$

$= n(1)$

$= n$.

যাতে কোনো
ঝামেলা
ছাড়া
$t=0$
বসানো

$\mu_1' =$ Mean

$\mu_2' = \mu_2' - (\mu_1')^2$ = $n(n+2) - n^2$
variance
$= n^2+2n-n^2$
$= 2n$.

$\frac{\partial^2}{\partial t^2} = n(-\frac{n}{2}-1)(1-2t)^{-n/2-2}(-2) \; |t=0$

$= 2n(-\frac{n+2}{2})(1)^{-n/2-2}$

$= n(n+2)$

$\therefore \mu_2' =$ variance

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

$= n(n+2) - n^2$

$= n^2 + 2n + n^2$

$= 2n$.

Cumulant generating function:

$K_x(t) = \log M_x(t)$

$= \log (1-2t)^{-n/2}$

$= -\frac{n}{2} \log(1-2t)$

$= \frac{n}{2}\left[2t + \frac{(2t)^2}{2} + \frac{(2t)^3}{3} + \cdots\right]$

$\left[\because \log(1-x) = -(x + \frac{x^2}{2} + \frac{x^3}{3} + \cdots)\right]$

K_r = Coefficient of $\frac{t^r}{r!}$ in $K_x(t)$
Kappa.

$\therefore K_1$ = coeff. of $\frac{t}{1!}$ in $K_x(t)$.

$\therefore K_2$ = co.eff. of $\frac{t^2}{2!}$ in $K_x(t)$

? $= \frac{n}{2} \cdot 4\left[\frac{(2t)^2}{4} + \frac{(2t)^2}{4\times 2}t + \frac{(2t)^3}{4\times 3}t + \cdots\right]$

$= \frac{n}{2} \cdot 4 \left[\frac{2t^2}{2} + \frac{2t^3}{2} + \frac{2t^4}{3} + \cdots\right]$

$= 2n$

$\therefore K_2 = 2n$ = variance

$\therefore K_3 = \text{coeff. of } \frac{t^3}{3!} \text{ in } K_x(t)$

$= \frac{n}{2}\left[\frac{(2t)^3}{3\times 2}\times 6 + \frac{(2t)^4}{3} + \frac{(2t)^5}{1} + \cdots\right]$

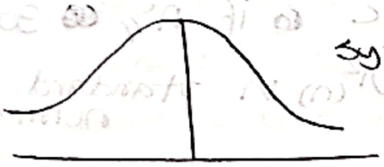
$= \frac{n}{2} \times 16 = 8n$

$\mu_3 = K_3 \qquad K_4 = 48n$

$\mu_4 = K_4 + 3K_2^2 = 12n(n+4)$

* Coefficient of skewness.

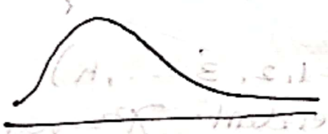
$\beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{64n^2}{8n^3} = \frac{8}{n}$



symmetric

$\beta_1 = 0$

negative skew



positive skew

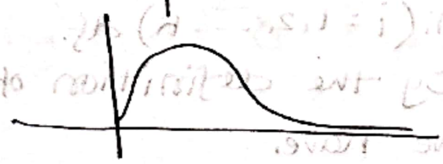
$\mu_3 = (+)\text{ve} \quad (+)\text{ve skew}$

$= 8n \rightarrow (+)\text{ve skew}$

$\because \mu_3 = 8n \quad n \rightarrow \text{positive}$

and $\beta_1 = \frac{8}{n}$

$\therefore$ chi-square skew positive skew



* Coefficient of Kurtosis:

$\beta_2 = \frac{\mu_4}{\mu_2^2} = \frac{12n(n+4)}{4n^2}$

$= 3 + \frac{12}{n} \rightarrow$ (pletokartic)

$\beta_2 = 3$ (Mesokartic)

$\beta_2 < 3$ (Leptokartic)

$\beta_2 > 3$ (pletikartic)

* value 30 এর বেশিহলে χ^2 এর মান normal dis^n হবে

* Limiting case of χ^2

if $n \rightarrow \infty$ $(n > 30)$

then $\chi^2_{(n\rightarrow\infty)} \rightarrow$ Normal dis^n

$X_1 \rightarrow \chi^2_{n_1}$

$X_2 \rightarrow \chi^2_{n_2}$

$X_1 + X_2 \sim \chi^2_{n_1+n_2}$

$X_1 - X_2 \sim \chi^2_{n_1-n_2}$

$\frac{X_1}{X_2} \rightarrow$ F-dis^n (n_1, n_2)

* দুটি chi-square ratio F dis^n

Date: 16.1.2024

* Moment generating function থেকে distribution কোনটা বুঝা যায়।

$\lim_{n\to\infty}$ ($n \geq 30$) $\chi^2 \to$ Normal (MGF $n\to\infty$)

If $X \sim \chi^2(n)$

$$M_X(t) = (1-2t)^{-n/2}$$

The standard mgf of χ^2-variate z is

$$M_{\frac{x-\mu}{\sigma}}(t) = M_{(\frac{x}{\sigma}+(\frac{-\mu}{\sigma}))}(t)$$

$$= M_{\frac{x}{\sigma}}(t) \cdot M_{\frac{\mu}{\sigma}}(t).$$

$[M_{(x_1 \pm x_2)}(t) = M_{x_1}(t) M_{x_2}(t)$

$$M_x(t) = E(e^{tx}) = e^{\frac{\mu t}{\sigma}}$$

$$= e^{-\frac{\mu t}{\sigma}} M_{\frac{x}{\sigma}}(t)$$

$$= e^{-\frac{\mu t}{\sigma}} M_x\left(\frac{t}{\sigma}\right)$$

$$= e^{-\frac{\mu t}{\sigma}} \cdot \left(1-\frac{2t}{\sigma}\right)^{-n/2}$$

[Mean $= \mu = n$.

variance $= \sigma^2 = 2n$.

$\therefore \sigma = \sqrt{2n}$]

$$= e^{-\frac{nt}{\sqrt{2n}}} \left(1-\frac{2t}{\sqrt{2n}}\right)^{-n/2}.$$

$$M_z(t) = e^{-t\sqrt{n/2}} \left(1 - t\sqrt{\frac{2}{n}}\right)^{-n/2}.$$

$$K_z(t) = \log M_z(t)$$

$$= -t\sqrt{n/2} - n/2 \log\left(1-t\sqrt{2/n}\right)$$

$$= -\cancel{t\sqrt{n/2}} + \cancel{\frac{n}{2}}\left(\cancel{\sqrt{2/n}} + \frac{t^2}{2}\cdot\frac{2}{n} + \frac{t^3}{3}\left(\frac{2}{n}\right)^{3/2} + \cdots\right)$$

$$= \frac{t^2}{2} + \frac{t^3}{3}\left(\frac{2}{n}\right)^{3/2} + \cdots$$

$$= \frac{t^2}{2} + w\left(\frac{2}{n}\right)$$

$$\lim_{n\to\infty} K_z(t) = \frac{t^2}{2} + \lim_{n\to\infty} w\left(\frac{2}{n}\right)$$

$$= \frac{t^2}{2} + 0.$$

$$\lim_{n\to\infty} M_x(t) = e^{t^2/2} \text{ on } n\to\infty$$

which is the mgf of standard distn.

In practice if $n \geq 30$ then $\chi^2_{(n)} \sim$ standard normal.

Additive property of χ^2-distn.

Theorem: The sum of independent χ^2-variate is also χ^2-variate.

Suppose, x_i $(i = 1, 2, 3 \ldots, K)$ are K-independent χ^2-variate with $n_i (i = 1, 2, \ldots, K)$ degrees of freedom respectively, then.

$\sum_{i=1}^{K} x_i$ follows χ^2-variate with

$$n = \sum_{i=1}^{K} n_i \text{ d.f.}$$

Proof: Since $x_i \sim \chi^2$-variate with $n_i (i = 1, 2, \ldots K)$ d.f. then by the definition of mgf we have,

$M_{x_i}(t) = (1-2t)^{-\frac{n_i}{2}} \quad (i = 1, 2, \cdots K)$

$$M_{\sum_{i=1}^{K} x_i}(t) = M_{(x_1+x_2+\cdots x_K)}(t)$$
$$= M_{x_1}(t)\, M_{x_2}(t) \ldots M_{x_K}(t)$$
$$= (1-2t)^{-\frac{n_1}{2}} (1-2t)^{-\frac{n_2}{2}} \ldots (1-2t)^{-\frac{n_K}{2}}$$
$$= (1-2t)^{-1/2\,(n_1+n_2+\cdots+n_K)}$$
$$= (1-2t)^{-n/2} \qquad [n = n_1+n_2+n_3+\cdots n_K = \sum n]$$

which is the mgf of χ^2-variate with $n = \sum_{i=1}^{K} n_i$ d.f.

Hence by the uniqueness theorem of mgf, it is proved that.

The sum of independent χ^2-variate is also χ^2-variate.

Mode of χ^2-variate:

গুরুত্ব

The mode of χ^2-variate is the maximum of χ^2-function.

We know the pdf

$$f(x) = \frac{1}{2^{n/2}\,\Gamma n/2}\, e^{-x/2}\, x^{n/2-1} \; ; \; x \geq 0$$

$\cancel{f'(x) \neq 0}$.

~~log~~ [লগারিদম এর value এর 1st derivate আর normally value এর derivative same, exp থাকলে লগারিদম করে নিতে হবে]

$$\therefore \log f(x) = \log(c) - \frac{x}{2} + \left(\frac{n}{2}-1\right)\log x$$

$$\frac{f'(x)}{f(x)} = \log f'(x) = 0 - \frac{1}{2} + \left(\frac{n}{2}-1\right)\cdot\frac{1}{x}.$$

since, $f(x) \neq 0$ then $f'(x) = 0$.

$$\therefore -\frac{1}{2} + \left(\frac{n}{2}-1\right)\cdot\frac{1}{x} = 0.$$

$$\Rightarrow x = n-2.$$

$$f''(x) = 0 - \left(\frac{n}{2}-1\right)\cdot\frac{1}{x^2}$$
$$= \frac{1}{x^2}\left(1-\frac{n}{2}\right)$$
$$= \frac{1}{x^2}\left(\frac{2-n}{2}\right)$$
$$= -\frac{1}{x^2}\left(\frac{n-2}{2}\right)$$
$$= -\frac{1}{(n-2)^2}\cdot\frac{(n-2)}{2}$$
$$= -\frac{1}{2(n-2)} < 0 \text{ for } n.$$

Date: 22.01.2024

Sampling t-distribution:

Sampling t-distribution
- student's t-dis. $\lambda = n-1$ df.
- Fisher's t-dis. n d.f.

Q. * Define ~~sampling~~ t-distribution with it's major applications.

Answer: Student's t-distribution is a sampling distribution of continuous types range's from $-\infty$ to ∞.

Let $x_1, x_2, x_3, \ldots, x_n$ be a random sample of size n from a normal population with mean μ and variance σ^2. Then student's t-distribution is defined as,

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

where $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$ is the sample mean and

$s^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{x})^2$ is an unbiased estimate of the population variance of σ^2.

and it follows the student's t-dist^n. with $\lambda = (n-1)$ d.f having the pdf

$$f(t) = \frac{1}{\sqrt{\lambda}\,\beta\left(\frac{1}{2}, \frac{\lambda}{2}\right)\left(1+\frac{t^2}{\lambda}\right)^{\frac{\lambda+1}{2}}} \; ; \quad -\infty < t < \infty$$

$= 0$; otherwise

If $\lambda = (n-1) = 1$

$$f(t) = \frac{1}{\sqrt{\pi}\,\beta\left(\frac{1}{2}, \frac{1}{2}\right)} \cdot \frac{1}{(1+t^2)} \; ; -\infty < t < \infty.$$

~~where~~ which is standard chauchy distribution.

Fisher's t-dist^n.

$$t = \cancel{\frac{z}{x/\sqrt{n}}} = \frac{z}{\sqrt{x/\sqrt{n}}}$$

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

and $x \sim \chi^2$ - with n d.f.

$$f(t) = \frac{1}{\sqrt{n}\,\beta\left(\frac{1}{2}, \frac{n}{2}\right)} \cdot \frac{1}{\left(1+\frac{t^2}{n}\right)^{\frac{n+1}{2}}}$$

$= 0$; other wise.

- n এর মান 30-৩৫ এর বেশি হলে normal dis.

Application of t-dist^n:

1. It is used to test the single sample mean $\bar{x}$ with population mean μ.
2. To ~~test~~ compare the population means we can use t-dis.
3. It is used to test the correlation and regression coefficients.
4. To test for proportion tests.
5. To test for multiple correlation coefficients.

✓ joint density function.

- $f(x,y) = f(x)\cdot f(y)$

অথবা, x, y independent হলে

উপর ভালোমত লিখতে হবে

$f(x,y) = x^2 y$

$f(x) = \int f(x,y)\, dy.$

$f(y) = \int f(x,y)\, dx.$

Date: 23.1.2024

t-distn:

Moments/constants of t-distn:

As t-distn. symmetrical at about origin, therefore all odd ordered moments are zero.

Then $\mu_{2r+1} = \mu'_{2r+1} = 0$;
$r = 0, 1, 2, \ldots\ldots$

$\therefore \mu'_1 = \text{Mean} = 0.$

$\therefore \mu_3 = \mu_5 = \ldots = \mu_{2r+1} = 0.$

Now, the even ordered moments

$$\mu_{2r} = \mu'_{2r} = E(t^{2r})$$

$$= \int_t t^{2r} f(t)\, dt. \qquad (r = 1, 2, \ldots)$$

$$= \int_{-\infty}^{\infty} t^{2r} \frac{1}{\sqrt{n}\, \beta(\frac{1}{2}, \frac{n}{2})} \cdot \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} dt$$

$$= \frac{1}{\sqrt{n}\, \beta(\frac{1}{2}, \frac{n}{2})} \int_{-\infty}^{\infty} t^{2r} \frac{1}{(1+\frac{t^2}{n})^{\frac{n+1}{2}}} dt$$

$$= \frac{2}{\sqrt{n}\, \beta(\frac{1}{2}, \frac{n}{2})} \int_0^{\infty} t^{2r} \frac{1}{\{1+t^2/n\}^{\frac{n+1}{2}}} dt \quad \ldots\ldots (1)$$

put $1 + \frac{t^2}{n} = x^{-1}$

$\Rightarrow t^2 = (x^{-1} - 1) n$

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

$2t \cdot dt = -\frac{n}{x^2}$

$\therefore dt = -\frac{n}{2tx^2} \cdot dx.$

If $t = 0$ then $x = 1$.
$t = \infty$ then, $x = 0$.

$$= \frac{2}{\sqrt{n}\, \beta(\frac{1}{2}, \frac{n}{2})} \int_1^0 \frac{t^{2r}}{(\frac{1}{x})^{\frac{n+1}{2}}} \cdot \frac{-n}{2tx^2} \cdot dx.$$

$$= \frac{\sqrt{n}}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^1 \frac{t^{2r}}{(\frac{1}{x})^{\frac{n+1}{2}}} \cdot \frac{1}{tx^2} \cdot dx.$$

$$= \frac{\sqrt{n}}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^1 (t^2)^{(2r-1)/2} \cdot x^{\{(n+1)/2\}-2} dx.$$

$$= \frac{\sqrt{n}}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^1 \left[\frac{n(1-x)}{x}\right]^{r-\frac{1}{2}} \cdot x^{\frac{n+1}{2}-1} \cdot dx.$$

$$= \frac{n^r}{\beta(\frac{1}{2}, \frac{n}{2})} \int_0^1 x^{(\frac{n}{2}-r)-1} \cdot (1-x)^{(r+\frac{1}{2})-1} dx.$$

- $\left[\int_0^1 y^{m-1} (1-y)^{n-1} dx = \beta(m,n)\right]$

$\left[\frac{1}{\beta(m,n)} \int_0^1 y^{m-1} (1-y)^{n-1} dy = \beta(m,n)\right]$

$$= \cancel{\frac{\sqrt{n}}{\beta(m,n)}}\ \frac{n^r}{\beta(\frac{1}{2}, \frac{n}{2})} \beta\left(\frac{n}{2} - r, r + \frac{1}{2}\right)$$

$$\left[\{n\}^{r-1/2}\cdot(n)^{\frac{2r-1}{2}} = n^{r-1/2+1/2}\right]$$

$$= \frac{n^r}{\beta(1/2, n/2)}\,\beta\left(\frac{n}{2}-r,\ r+1/2\right)$$

$$= n^r\,\frac{\Gamma\left(\frac{n}{2}-r\right)\Gamma\left(r+1/2\right)}{\Gamma\left(\frac{n}{2}+\frac{1}{2}\right)}\times\frac{\Gamma\left(\frac{1}{2}+\frac{n}{2}\right)}{\Gamma 1/2\cdot\Gamma n/2}$$

$$= n^r\,\frac{\Gamma\left(\frac{n}{2}-r\right)\Gamma\left(r+1/2\right)}{\Gamma(1/2)\,\Gamma(n/2)}$$

$$= \frac{n^r\left(r-\frac{1}{2}\right)\left(r-3/2\right)\cdots\frac{3}{2}\cdot\frac{1}{2}\cdot\cancel{\Gamma 1/2}\cdot\cancel{\Gamma\left(\frac{n}{2}-r\right)}}{\cancel{\Gamma\frac{1}{2}}\left(\frac{n}{2}-1\right)\left(\frac{n}{2}-2\right)\cdots\left(\frac{n}{2}-r\right)\cancel{\Gamma\left(\frac{n}{2}-r\right)}}$$

$$\mu_{2r} = \frac{n^r(2r-1)(2r-3)\cdots 3\cdot 1}{(n-2)(n-4)\cdots(n-2r)},\quad \frac{n}{2}>r,\quad r=1,2,\cdots$$

In particular,

$$\mu_2 = n\cdot\frac{1}{n-2} = \frac{n}{n-2}.$$

$$\mu_4 = n^2\cdot\frac{3\cdot 1}{(n-2)(n-4)} = \frac{3n^2}{(n-2)(n-4)};\ n>4$$

$$\therefore \beta_1 = \frac{\mu_3^2}{\mu_2^3} = \frac{0}{\mu_2^3} = 0.$$

$$\therefore \beta_2 = \frac{\mu_4}{\mu_2^2} = 3\left(\frac{n-2}{n-4}\right)$$

$$\lim_{n\to\infty}\beta_2 = 3\lim_{n\to\infty}\left\{\frac{1-2/n}{1-4/n}\right\} = 3.$$

for large $n\,(n\to\infty)$ t-distn becomes normal

* For large $n\,(n\to\infty)$ distn tends to standard normal distn

proof: For n d.f the pdf of t distn

$$f(t) = \frac{1}{\sqrt{n}\,\beta(1/2, n/2)}\left\{1+\frac{t^2}{n}\right\}^{-n/2-1/2}$$

$$\lim_{n\to\infty} f(t) = \lim_{n\to\infty}\frac{1}{\sqrt{n}}\,\frac{\Gamma\frac{n+1}{2}}{\Gamma n/2\cdot\Gamma 1/2}$$

$$\left[\beta(m,n) = \frac{\Gamma m\,\Gamma n}{\Gamma m+n}\right]$$

$$\lim_{n\to\infty}\left[\left\{1+\frac{t^2}{n}\right\}^{n}\right]^{-1/2}\times\lim_{n\to\infty}\left\{1+\frac{t^2}{n}\right\}^{-1/2}$$

$$\left[\lim_{n\to\infty}\frac{1}{\sqrt{n}}\cdot\frac{1}{\sqrt{\pi}}\left(\frac{n}{2}\right)^{1/2} = \frac{1}{\sqrt{2\pi}}\right]$$

$$= \frac{1}{\sqrt{2\pi}}\times 1\cdot e^{-1/2}\quad \lim_{n\to\infty} f(t)$$

$$= \frac{1}{\sqrt{2\pi}}\,e^{-t^2/2};\ -\infty<t<\infty.$$

which is the pdf of standard Normal distn.

F-statistic:

• Sampling distn:
① χ^2 ② t ③ F
④ S-N

$$F = \frac{X/n_1}{Y/n_2} \sim F\text{-dist}^n \text{ with } (n_1, n_2) \text{ d.f.}$$

$X \sim \chi^2$-distn with n_1 d.f.
and $y \sim \chi^2$ distn with n_2 df.
If x and y are statistically independent

which has the following pdf, as,

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}} \cdot (F)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \; ; F \geq 0.$$

Application:

1. For testing for equality of two popular variances.
2. Testing for significance of an observed multiple correlation coefficient.
3. Testing for significance of observed sample correlation.
4. Testing for the linearity of Regression.
5. Testing for equality of several means
6. Testing for ANOVA (Analysis of variances)

$$f(F) = \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \cdot \frac{F^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} \; ; F \geq 0$$

$$\left[\int_0^\infty \frac{x^{m-1}}{(1+x)^{m+n}}\, dx = \beta(m,n) \right.$$

$$\left. \beta(m,n) = \frac{\Gamma m \, \Gamma n}{\Gamma(m+n)} \right]$$

Mean: $\mu_1' = E(F)$

Variance: $= \mu_2 = V(F)$

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

$$= E(F^2) - \{E(F)\}^2$$

$$E(X) = \int_x x f(x)\, dx$$

$$E(F) = \int_F F f(F)\, dx.$$

Mean =

$$E(F) = \frac{(n_1/n_2)^{n_1/2}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty F \cdot \frac{(F)^{\frac{n_1}{2}-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}}\, dF$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{(F)^{\left(\frac{n_1}{2}+1\right)-1}}{\left\{1+\frac{n_1}{n_2}\cdot F\right\}^{\frac{n_1+n_2}{2}}}$$

Put $\frac{n_1}{n_2} F = x$.

$$\Rightarrow F = \frac{n_2}{n_1} x.$$

$$dF = \frac{n_2}{n_1} dx.$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{n_1/2}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{\left(\frac{n_2}{n_1}x\right)^{\left(\frac{n_1}{2}+1\right)-1}}{\{1+x\}^{\frac{n_1+n_2}{2}}} \cdot \frac{n_2}{n_1}\, dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+1-1+1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\frac{n_1}{2}+\frac{n_2}{2}}}\, dx$$

$$= \frac{\left(\frac{n_2}{n_1}\right)^{-\frac{n_1}{2}} \left(\frac{n_2}{n_1}\right)^{\frac{n_1}{2}+1-1+1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\left(\frac{n_1}{2}+1\right)\left(\frac{n_2}{2}-1\right)}}$$

$$= \frac{\frac{n_2}{n_1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \beta\left(\frac{n_1}{2}+1, \frac{n_2}{2}-1\right)$$

$$= \left(\frac{n_2}{n_1}\right) \frac{\Gamma\frac{n_1+n_2}{2}}{\Gamma\frac{n_1}{2}\Gamma\frac{n_2}{2}} \cdot \frac{\Gamma\left(\frac{n_1}{2}+1\right)\Gamma\left(\frac{n_2}{2}-1\right)}{\Gamma\left(\frac{n_1}{2}+\frac{n_2}{2}\right)}$$

$$= \frac{n_2}{n_1} \cdot \frac{\frac{n_1}{2}\Gamma\frac{n_1}{2} \cdot \Gamma\frac{n_2}{2}-1}{\Gamma\frac{n_1}{2} \cdot \left(\frac{n_2}{2}-1\right)\Gamma\frac{n_2}{2}-1}$$

$$= \frac{n_2}{n_1} \cdot \frac{\frac{n_1}{2}}{\frac{n_2-2}{2}}$$

$$= \frac{n_2}{2} \times \frac{2}{n_2-2}$$

$$= \frac{n_2}{n_2-2}, \quad n_2 > 2.$$

Now,

$$\mu_2' = E(F^2)$$

$$= \int_0^\infty F^2 f(F)\, dF$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty F^2 \frac{(F)^{\frac{n_1}{2}-1}}{\left\{\left(1+\frac{n_1}{n_2}\right)F\right\}^{\frac{n_1+n_2}{2}}} dF$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{F^2 F^{\left(\frac{n_1}{2}+1\right)-1}}{\left(1+\frac{n_1}{n_2}F\right)^{\frac{n_1+n_2}{2}}} dF.$$

Put

$$\frac{n_1}{n_2} F = x.$$

$$\Rightarrow F = \frac{n_2}{n_1} \cdot x.$$

$$\Rightarrow dF = \frac{n_2}{n_1} dx.$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \left(\frac{n_2}{n_1}x\right)^2 \frac{\left(\frac{n_2}{n_1}x\right)^{\left(\frac{n_1}{2}+1\right)-1}}{(1+x)^{\frac{n_1+n_2}{2}}} dx$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{\frac{n_1}{2}-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{\left(\frac{n_2}{n_1}\right)^2 \cdot \left(\frac{n_2}{n_1}\right)^{\left(\frac{n_1}{2}+1\right)-1} x^{\frac{n_1}{2}+1-1+2}}{(1+x)^{\frac{n_1+n_2}{2}}} dx$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^{-\frac{n_1}{2}+1+2+\frac{n_1}{2}+1-1}}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+2\right)}}{(1+x)^{\frac{n_1+n_2}{2}}}$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+3-1\right)}}{(1+x)^{\frac{n_1+n_2}{2}}}$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\left(\frac{n_1}{2}+3-1\right)}}{(1+x)^{\frac{n_1}{2}+\frac{n_2}{2}}}$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)} \int_0^\infty \frac{x^{\frac{n_1}{2}+3-1}}{(1+x)^{\left(\frac{n_1}{2}+2\right)+\left(\frac{n_2}{2}-2\right)}}$$

$$= \frac{\left(\frac{n_1}{n_2}\right)^2}{\beta\left(\frac{n_1}{2}, \frac{n_2}{2}\right)}$$

Date: 9.1.2024

Estimation:

X

$x_1 - P_1$
$x_2 - P_2$
$\vdots$
$x_n - P_n$

$P \rightarrow$ Probability
$x \rightarrow$ variable

* variable with probability is called Random variable.

$$P(x) = \frac{e^{-\lambda}\lambda^{x}}{x!} \; ; \; x = 0, 1, 2, 3 \ldots \infty$$

$E(x) = \mu = \lambda \leftarrow$ সরাসরি accurate মান বের করা যায় না

$\hat{\lambda}$

$\bar{X} = \frac{\sum x_i}{n} \rightarrow$ statistic

Sample mean

$(\bar{X} - \mu) \neq 0$

↳ পার্থক্য minimize করা আমাদের target.

$\hat{\mu} = \bar{x}$
$\mu \neq \bar{x}$

* Inference
↓
Estimated মান ঠিক কি না, তখন নেয়া যাবে কি না।

* What is estimator?

⇒ যে parameter দিয়ে estimation করা হয়।

$t_n = (x_1, x_2, \ldots x_n)$
↳ estimator.

Imp

* A good estimator has four criteria:–

1. Consistency
2. Unbiasedness
3. Efficiency
4. Sufficiency.

1. Consistency:

$n \rightarrow \infty, \quad \bar{X} \rightarrow \mu$

2. Unbiasedness:

$E(\bar{X}) = \mu$
$\bar{X}$ is an unbiased estimator.

3. Efficiency:

যে estimation এর variance least.

4. Sufficiency:

যার মধ্যে estimation এর আমাদের।

Point Estimation:

↳ $\hat{\mu} = 3.5$

↳ MLE
(Maximum Estimation)

Interval Estimation:

↳ $\hat{\mu} = (2, 5)$
2 to 5

* **Consistancy:**

An estimator $t_n = (x_1, x_2 \ldots, x_n)$ based on a random sample of size n is said to be consistent estimator of the parameter.

$g(\theta), \theta \in \Omega$ (Ω parameter space)

if t_n converges to $g(\theta)$ in tends to (दिखाना) probability i.e.

if $t_n \xrightarrow{P} g(\theta)$

In other words,

t_n is an consistant of $g(\theta)$ if ever $\epsilon > 0$ & $\eta > 0$ there exists a positive integer $n \geq m$ [where m is large such

$$P[\{t_n - g(\theta)\} < \epsilon] > 1 - \eta \;\forall\; n \geq m$$

Example:

If $x_1, x_2 \ldots x_n$ be a r.s. of size n from a population with finite mean $E(X_i) = \mu$ then the weak law of large number, we have

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \xrightarrow{P} E(x_i) = \mu$$

as $n \to \infty$

Statistic
parameter
Random variable
Estimator
Estimation.

Unbiasedness:

An estimator of a given parameter (θ) is said to be an unbiased if its expected value is equal to the true value of the parameter.

In other words, an estimator is unbiased if it produces parameter estimates that are on average correct.

If t_n be a statistic calculated from a sample $(x_1, x_2 \ldots x_n)$ of size n from a density $f(x, \theta)$ ie. $E(t_n) = \theta$. then t_n is an unbiased estimator of θ.

Ex:

If $X \sim N(\mu, \sigma^2)$ variate then sample mean $\bar{x}$ is an unbiased estimator of μ.

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

Solve:

We have to prove

$$E(\bar{x}) = \mu \quad \ldots (i)$$

we know, $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \ldots (ii)$

Scanned with OKEN Scanner

Date: 06.03.2029

$$E(\bar{x}) = E\left\{\frac{1}{n}\sum_{i=1}^{n} x_i\right\}$$

$$= \frac{1}{n}\left\{E\left(\sum_{i=1}^{n} x_i\right)\right\}$$

$$= \frac{1}{n}\left\{E(x_1+x_2+\dots+x_n)\right\}$$

$$= \frac{1}{n}\left\{E(x_1)+E(x_2)+\dots+E(x_n)\right\}$$

$$= \frac{1}{n}\left\{\mu+\mu+\dots+\mu\right\}$$

$$= \frac{1}{n}\times n\cdot\mu$$

$$= \mu$$

$$\therefore E(\bar{x}) = \mu$$

Hence, $\bar{x}$ is an unbiased estimator of μ.

Ex. If t is an unbiased estimator of θ then t^2 is biased estimator for θ^2.

Solution: As t is an unbiased estimator of θ then we get,

$$E(t) = \theta \quad \cdots (1)$$

We know,

$$V(t) = E(t^2) - \{E(t)\}^2$$

$$= E(t^2) - \{\theta\}^2$$

$$\therefore V(t) = E(t^2) - \theta^2$$

$$E(t^2) = \theta^2 + V(t),$$

As $V(t) > 0$

$$\therefore E(t^2) \neq \theta^2$$

$\therefore$ t^2 is not unbiased of θ^2.

Thus t^2 is biased estimator for θ.

Ex-2: A random sample (x_1, x_2, x_3, x_4 and x_5) of size 5 is from a normal population with unknown mean μ. Consider the following estimators to estimate μ.

(i) $t_1 = \dfrac{x_1+x_2+\dots+x_5}{5}$

(ii) $t_2 = \dfrac{x_1+x_2}{2} + x_3$

Are t_1 and t_2 unbiased?

(iii) Find λ for unbiasedness of

$$t_3 = \frac{2x_1 + x_2 + \lambda x_3}{3}$$

(i) We know.

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i$$

$$E(\bar{x}) = E\left\{\frac{1}{5}\sum_{i=1}^{n} x_i\right\}$$

$$= \frac{1}{5} E\left(\sum_{i=1}^{5} x_i\right)$$

$$E(t_1) = E\left\{\frac{1}{5}\sum_{i=1}^{5} x_i\right\}$$

$$= \frac{1}{5}\left(E\sum_{i=1}^{5} x_i\right)$$

$$= \frac{1}{5}\left\{E(x_1 + \dots + x_5)\right\}$$

$$= \frac{1}{5}\left\{E(x_1) + \dots + E(x_5)\right\}$$

$$= \frac{1}{5}(\mu_1 + \mu_2 + \mu_3 + \mu_4 + \mu_5)$$

$$= \frac{1}{5} \times 5\mu$$

$$= \mu$$

$$\therefore E(t_1) = \mu.$$

Hence t_1 is unbiased.

(ii) We know.

$$\bar{x} = \frac{1}{2}\sum_{i=1}^{2} x_i$$

$\cdot E(t_1) \neq \mu$
$\hookrightarrow$ biased

$$E(\bar{x}) = \frac{1}{2}\left\{\frac{1}{2}\sum_{i=1}^{2} x_i\right\} + \sum_{i=1}^{1} x_i$$

$$= \frac{1}{2}\left(E(x_1 + x_2)\right) + E(x_3)$$

$$= \frac{1}{2}(\mu_1 + \mu_2) + \mu_3$$

$$= \frac{1}{2} \times 2\mu + \mu_3$$

$$= \mu + \mu = 2\mu \rightarrow$$ not unbiased = biased.

(iii) $$t_3 = \frac{2x_1 + x_2 + \lambda x_3}{3}$$

$$E(t_3) = \frac{1}{3}\left\{2E(x_1) + E(x_2) + \lambda E(x_3)\right\}$$

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

As t_3 is an unbiased.

Therefore,

$$E(t_3) = \mu.$$

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

$$\Rightarrow 3\mu + \lambda\mu = 3\mu$$

$$\Rightarrow \lambda\mu = 3\mu - 3\mu = 0$$

$$\lambda = 0$$

Efficiency:

$$t_1 \xrightarrow{\text{unbiased}} \theta \longrightarrow v(t_1)$$

$$t_2 \xrightarrow{\text{unbiased}} \theta \longrightarrow V(t_2)$$

say, $v(t_1) < v(t_2)$

then t_1 is more efficient than t_2.

t_i $(i = 1, 2, 3 \dots t_k)$ all are unbiased estimator for a given parameter $g(\theta)$

And their respective variances $v(t_i)$.

Now,
$$E = \frac{V(t_i)}{V(t_j)} \quad \{j \neq i\}$$

if $E < 1$ then t_i is the most efficient estimation of $g(\theta)$.

Minimum variance Bound (MVB)

$$E = \frac{MVB}{V(t_i)} \; ; \; i = 1, 2, \ldots, k.$$

সর্বোচ্চ মান ১ হয়, normally ১ এর কম হবে কারণ, $MVB < V(t_i)$

E ১ হলে, বেশি হওয়া যাবেনা। most efficient.

1. Defination of point estimations.
2. Mothod " "
3. Principles of MLE
4. Applications of MLE.

Date: 07.02.2024

* A good estimator has 4 criteria:

1. Unbiasedness
2. Consistency
3. Efficiency
4. Sufficiency

Efficiency:

$$V(t_1) < V(t_2)$$

where t_1 and t_2 are two unbiased estimators of θ.

then t_1 is more efficient.

MVB → Minimum variance Bound.

$$E_i = \frac{MVB}{V(t_i)} \; ; \; i = 1, 2, \ldots, k.$$

* Cramer-Raw Inequality: (C-R)

Statement:

If $T = t(x, x_1, x_2, \ldots x_n)$ be an unbiased estimator for $g(\theta)$ from pdf $f(x, \theta)$. then Cramer-Raw (C-R) Inequality

$$V(t) \geq \frac{\left\{\frac{d}{d\theta} g\theta\right\}^2}{E\left\{\frac{d}{d\theta} \log L\right\}^2} = \frac{\{g'(\theta)\}^2}{I(\theta)}$$

$f(x, \theta) = \theta . x^{\theta}$; $0 < x < 1$.

$f(x_1, \theta) = \theta x_1^{\theta}$

$f(x_2, \theta) = \theta x_2^{\theta}$

$L(x, \theta) = f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$

* L → Liklihood function of the parameter.

parameter

* Joint density function → ~~observation~~ the function of sample obs., x's

observation

Sufficiency:

An estimator is said to be suffient for a parameter if it contains all the information in the sample regarding the parameter.

If $T = t(x_1, x_2, x_3, \ldots x_n)$ is an estimator of a parameter θ, based on a sample $x_1, x_2, \ldots, x_n$ of size n from the population with p.d.f, $F(x, \theta)$ such that the conditional distribution of $x_1, x_2, \ldots x_n$ given T independent of θ, then T is a sufficient estimator of θ.

Factorization / Neyman Theorem:

Statement:

If $T = t(x_1, \dots x_n)$ is sufficient estimator for $g(\theta)$ if and only if the joint density function L (say) of the sample values can be ~~est~~ expressed in the form-

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

where $g_\theta[t(x)]$ depends on θ and x only but $h(x)$ is independent of θ.

Ex: Let $x_1, x_2, \dots, x_n$ be a random sample of size n from $N(\mu, \sigma^2)$.

Find sufficient estimators for μ and σ^2.

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

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

* $X \sim N(0, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2 \frac{x^2}{\sigma^2}}$

$= \frac{1}{\sigma^2\sqrt{2\pi}} e^{-\frac{1}{2\sigma^2}x^2}$

* $X \sim N(0,1) = \frac{1}{2\pi} e^{-\frac{x^2}{2}}$

Standard normal.

Solve:

Given,

$$f(x; \mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-1/2\left\{\frac{x-\mu}{\sigma_1}\right\}^2}$$

$$\therefore L = f(x_1; \mu, \sigma^2) \dots f(x_n; \mu, \sigma^2)$$

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

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

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

where

$$g_\theta[t(x)] = (2\pi\sigma^2)^{-n/2} \cdot \exp\left[-\frac{1}{2\sigma^2} \cdot \left\{t_2(x) - 2\mu t_1(x) + n\mu^2\right\}\right]$$

$$t(x) = \{t_1(x), t_2(x)\}$$
$$= \left\{\sum_{i=1}^{n} x_i, \sum_{i=1}^{n} x_i^2\right\}$$

and $h(x) = 1$.

by the factorization theorem, ~~Thus~~ Thus $t_1(x) = \sum_{i=1} x_i$ is the sufficient statistic for μ.

and $t_2(x) = \sum x_i^2$ is the sufficient statistic for μ.?

~~So that~~

Ex. $f(x,\theta) = \theta x^{\theta-1}$

Show that $t_1 = \prod_{i=1}^{n} x_i$ is the sufficient estimator for θ.

Soln:

$$L = f(x_1,\theta) f(x_2,\theta) \cdots f(x_n,\theta)$$

$$= \theta x_1^{\theta-1} \, \theta x_2^{\theta-1} \cdots \theta x_n^{\theta-1}$$

$$= \theta^n (x_1 x_2 \cdots x_n)^{\theta-1}$$

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

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

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

Here $g_\theta[t(x)]$ is function of θ and x but $h(x)$ is independent of θ. Therefore by factorization theorem,

$f(x) = \prod_{i=1}^{n} x_i$ is the sufficient Statistic (estimator).

Date: 20.7.2024

Outline:

Inference:
- Estimation
  - Point : $\hat{\mu} = 3.5$
  - Interval : $-3 \le \hat{\mu} \le 3$
- Test

Method: MLE
- Principles of MLE
- Properties of MLE
- Math/example

Questions:

Q1: Define:
(i) Estimation
(ii) Point estimation
(iii)

Q2: Write down the methods of Point estimation.

Q3: Discuss or explain the Estimation principles of Maximum Likehood Method (MLE).

Q4: Properties of MLE.

Q5: Example.

Answers:

1. Point Estimation:

In statistics, a point estimation is a process of finding an approximate value of some parameters such as mean (μ) of a population from the random samples of the population.

Methods of point Estimations:

1. Method of Maximum Likelihood Estimation (MLE)
2. Method of Minimum variance
3. Method of Moments.
4. Method of least squares.
5. Method of minimum chi-square (χ^2).
6. Method of Inverse probability.

Principles of MLE:

Let $x_1, x_2 \ldots x_n$ be a random sample of size n drawn from a given population with pdf $f(x;\theta)$ (probability density function) where θ is the population parameter (s).

Now, the likelihood function denoted by $L(x/\theta) = L(x;\theta)$

$$= f(x_1, \theta) f(x_2, \theta) \ldots f(x_n, \theta)$$

$$= \prod_{i=1}^{n} f(x_i, \theta) \quad \ldots \text{①}$$

To maximize ①, we have to get

$$\frac{\partial L(x;\theta)}{\partial \theta} = \frac{\partial}{\partial \theta}\left\{\prod_{i=1}^{n} f(x_i;\theta)\right\} = 0 \quad \ldots \text{②}$$

Again,

$$\frac{\partial^2 L(x;\theta)}{\partial \theta^2} = \frac{\partial^2}{\partial \theta^2}\left\{\prod_{i=1}^{n} f(x_i;\theta)\right\}$$

If $\frac{\partial^2 L(x;\theta)}{\partial \theta^2} < 0$; then the value of θ obtained from ② is the maximum value of θ. which is our required MLE for θ.

For simplicity, due to exponential form of likelihood function we may take log of $L(x;\theta)$

Therefore,

$$\frac{\partial \log L(x;\theta)}{\partial \theta} = 0 \quad \ldots \text{①}$$

$$\frac{\partial^2 \log L(x;\theta)}{\partial \theta^2} < 0 \quad \ldots$$

5. Example:

1. $f(x;\lambda) = \lambda e^{-\lambda x}; \; x \geq 0$
2. $f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}; \; x \geq 0.$
3. different $f(x;m) = \frac{e^{-m} m^x}{x!}; \; x = 0, 1, 2 \ldots$ distrete
4. $f(x;\mu,\sigma^2)$

$$= \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{1}{2}\left\{\frac{x-\mu}{\sigma}\right\}^2}; \; -\infty < x < \infty; \; -\infty < \mu < \infty; \; \sigma^2 \geq 0.$$

* $0 \leq$ function $\leq 1 \rightarrow$ pdf
integration of

Solution:

3. Given,

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \quad ; \quad x = 0, 1, 2 \dots$$

The likelihood function

$$L(x;\lambda) = f(x_1,\lambda)\dots f(x_n,\lambda)$$

$$= \frac{e^{-\lambda}\lambda^{x_1}}{x_1!}\dots\frac{e^{-\lambda}\lambda^{x_n}}{x_n!}$$

$$= e^{-n\lambda}\left\{\frac{\lambda^{x_1+\dots+x_n}}{\prod_{i=1}^{n}(x_i)!}\right\}$$

$\cancel{\log L = -n\lambda + \sum x_i \log\lambda - \sum_{i=1}^{n} \log x_i!}$

Let $\hat{\lambda}$ be the estimated value of λ.

$$= \log L = -n\lambda + \sum x_i \log\lambda - \sum_{i=1}^{n} \log x_i!$$

$$\frac{\partial \log L}{\partial \lambda} = 0$$

$$\Rightarrow -n + \frac{\sum x_i}{\lambda} - 0 = 0$$

$$\Rightarrow -n + \frac{\sum x_i}{\lambda} = 0.$$

$$\hat{\lambda} = \bar{x}$$

$$\Rightarrow \frac{\sum x_i}{\lambda} = n.$$

$$\Rightarrow \frac{\sum x_i}{n} = \hat{\lambda}$$

$$\Rightarrow \hat{\lambda} = \bar{x}_i$$

unbiased প্রমাণ করতে হবে

extra β.

$$\frac{\partial \log L}{\partial \lambda} = n\left(\frac{\bar{x}}{\lambda} - 1\right)$$

$$= g(x,\lambda)\cdot h(x)$$

Here $g(x,\lambda)$ is depended on x and λ but $h(x)$ is independed of λ.

Therefore,

By factorization theorem, $\bar{x}$ is a sufficient estimator for λ.

$$E(\bar{x}) = E\left(\frac{1}{n}\left(\sum_{i=1}^{n} x_i\right)\right) = \frac{1}{n}\left(E\sum_{i=1}^{n} x_i\right)$$

$$= \frac{1}{n}\left(E(x_1 + x_2 + \dots + x_n)\right)$$

$$= \frac{1}{n}\{E(x_1) + E(x_2) + \dots E(x_n)\}$$

$$= \frac{1}{n}(\lambda_1 + \lambda_2 + \dots \lambda_n)$$

$$= \frac{1}{n}\cdot n\lambda$$

$$= \lambda.$$

2

$$\therefore E(\bar{x}) = \lambda$$

$$f(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!} \; : \; x = 0, 1, 2 \dots \infty$$

Poison এর distribution এর

↳ Mean $= \lambda = E(\lambda) = \lambda$

We have to show that

$$E(\bar{x}) = \lambda \dots (1)$$

Now, $$E(\bar{x}) = E\left(\frac{\sum_{i=1}^{n} x_i}{n}\right)$$

$$= \frac{1}{n}\sum_{i=1}^{n} E(x_i) \quad \dots (2)$$

$$E(x_i) = \sum x \cdot p(x)$$

$$= \sum_{i=1}^{n} x_i \frac{e^{-\lambda}\lambda^{x}}{x_i!}$$

না হলে 1

$$= \sum_{i=1}^{n} x_i \frac{e^{-\lambda}\lambda^{x_i - 1}\cdot\lambda}{x_i(x_i - 1)!}$$

$$= \lambda \sum_{x_i-1=0}^{n} \frac{e^{-\lambda}\lambda^{x_i-1}}{(x_i-1)!}$$

$$= \lambda \sum_{y=0} \frac{e^{-\lambda}\lambda^{y}}{y!}$$

$$= \lambda \sum_{y=0} e \cdot \lambda \cdot 1$$

$$= \lambda$$

$$\therefore E(x_i) = \lambda.$$

1. $f(x;\lambda) = \lambda e^{-\lambda x}$; $x \geq 0$

The likelihood function,

$$L(x;\lambda) = f(x_1,\lambda) \ldots\ldots f(x_n,\lambda)$$
$$= \lambda e^{-\lambda x_1} \ldots\ldots \lambda e^{-\lambda x_n}$$
$$= \lambda^n e^{-\lambda(x_1+\ldots\ldots+x_n)}$$
$$= \lambda^n \cdot e^{-\lambda \sum_{i=1}^{n} x_i}$$

Let $\hat{\lambda}$ be the estimated value of λ.

$$\therefore \log L = \log(\hat{\lambda}^n \cdot e^{-\hat{\lambda}\sum_{i=1}^{n} x_i})$$
$$= n \log \hat{\lambda} - \hat{\lambda} \sum_{i=1}^{n} x_i$$

Now, $\frac{\partial \log L}{\partial \hat{\lambda}} = 0$

$$\Rightarrow \frac{\partial}{\partial \hat{\lambda}}(n \log \hat{\lambda}) - \frac{\partial}{\partial \hat{\lambda}}(\hat{\lambda}\sum_{i=1}^{n} x_i) = 0$$
$$\Rightarrow \frac{n}{\hat{\lambda}} = \sum_{i=1}^{n} x_i$$
$$\Rightarrow \hat{\lambda} = \frac{n}{\sum_{i=1}^{n} x_i} = \frac{1}{\bar{x}}$$

2. $f(x;\theta) = \frac{1}{\theta} e^{-\frac{x}{\theta}}$; $x \geq 0$

The likelihood function,

$$L(x;\theta) = f(x_1,\theta) \ldots\ldots f(x_n,\theta)$$
$$= \frac{1}{\theta} e^{-\frac{x_1}{\theta}} \ldots\ldots \frac{1}{\theta} e^{-\frac{x_n}{\theta}}$$
$$= \left(\frac{1}{\theta}\right)^n \cdot e^{-\frac{1}{\theta}(\sum_{i=1}^{n} x_i)}$$

Let $\hat{\theta}$ be the estimated value of θ

$$\therefore \log L = \log\left(\frac{1}{\hat{\theta}}\right)^n \cdot e^{-\frac{1}{\hat{\theta}}(\sum_{i=1}^{n} x_i)}$$
$$= n \log\left(\frac{1}{\hat{\theta}}\right) - \frac{1}{\hat{\theta}} \sum_{i=1}^{n} x_i$$

Now, $\frac{\partial \log L}{\partial \hat{\theta}} = 0$

$$\Rightarrow \frac{\partial}{\partial \hat{\theta}}(n \log \tfrac{1}{\hat{\theta}}) - \frac{\partial}{\partial \hat{\theta}}(\tfrac{1}{\hat{\theta}}\sum_{i=1}^{n} x_i) = 0$$
$$\Rightarrow -\frac{n}{\hat{\theta}} - \left(-\frac{1}{\hat{\theta}^2}\sum_{i=1}^{n} x_i\right) = 0$$
$$\Rightarrow -\frac{n}{\hat{\theta}} + \frac{1}{\hat{\theta}^2}\sum_{i=1}^{n} x_i = 0$$
$$\Rightarrow -\frac{n}{\hat{\theta}} = -\frac{1}{\hat{\theta}^2}\sum_{i=1}^{n} x_i$$
$$\Rightarrow \hat{\theta} = \frac{1}{n}\sum_{i=1}^{n} x_i = \bar{x}.$$

27.02.2024

Section A: (Estimation)

5.

1. Sampling Distribution (χ^2, t, F) — 1/2 টা
2. Estimation criteria — 1/0টা
3. point estimation — 1/0টা

Section B: (Hypothesis testing)

1. Contingency table — 1
2. Test of Hypothesis — 1
3. Non-parametric test — 1.

* Inference = Estimation + Hypothesis testing.
* Alternative hypothesis (H_1) → H_0, এর against এ
* Null hypothesis (H_0) → সাধারণত rejected
  data দিয়ে যে test

Contengency table: (2x2)

		Parents (χ^2)	
		Yes	No
Children	Yes	30	20
	No	20	30

100 of sample

H_0: There is no relation-ship between column and row/ Both parents and children height is independent/

Vs H_1:

Manifold classification:

২ এর অধিক row বা column হলে যে contengency table.

Level of significance: → যা বলা থাকলে 5% ধরে নিবো

যেকোনো percentage আগেই বলে দিয়ে কাজ শুরু করব।

↳ predetermined error

χ^2 → calculated value of χ^2

$Cal(\chi^2) = 8.25$ (say) — sample calculate করে আসবে

$Tab(\chi^2_{(0.05,1)}) = 7.92$ (say)

Table থেকে আসবে

স্বাধীনতার জন্য উল্লেখ করে দিবো।

Degree of freedom $= (2-1)(2-1) = 1 \cdot 1 = 1$

Level of significance $= 5\% = \frac{5}{100}$

$= 0.05$ (normally)

$Tab(\chi^2_{(0.05,1)}) = 7.92$

(α, n)

If $cal(\chi^2) > Tab(\chi^2_{(\alpha,n)})$

then we may reject H_0.

Otherwise, fail to reject H_0.

Date: 11.13.2024

<u>Contingency</u>: row column $r \times c$

- $(A_1B_1) \to$ সংখ্যা
- $A_1B_1 \to$ গুণ

B	A_1	A_2	...	A_i	...	A_c	Total
B_1	(A_1B_1)	(A_2B_1)	...	(A_iB_1)	...	(A_cB_1)	(B_1)
B_2	(A_1B_2)	(A_2B_2)	...	(A_iB_2)	...	(A_cB_2)	(B_2)
⋮							⋮
B_j	(A_1B_j)	(A_2B_j)	...	(A_iB_j)	...	(A_cB_j)	(B_j)
⋮							⋮
B_r	(A_1B_r)	(A_2B_r)	...	(A_iB_r)	...	(A_cB_r)	(B_r)
Total	(A_1)	(A_2)	...	(A_i)	...	(A_c)	N

H_0: The attributes are independent.

H_1: H_0 is not true.

Now, $P(A_i) = \frac{(A_i)}{N}$

$P(B_j) = \frac{(B_j)}{N}$

$$P(A_iB_j) = P(A_i)\,P(B_j)$$

$$= \frac{(A_i)}{N} \cdot \frac{B_j}{N} \; ; \quad i = 1,2\ldots c, \; j = 1,2\ldots r$$

The expected frequency,

$$(A_iB_j)_0 = N \cdot P(A_iB_j)$$

Both independent

$$= N\, P(A_i)\, P(B_j)$$

$$= N \cdot \frac{(A_i)}{N} \cdot \frac{(B_j)}{N}$$

$$= \frac{(A_i)\cdot(B_j)}{N}$$

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

The test statistic,

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \left[\frac{\{(A_iB_j) - (A_iB_j)_0\}^2}{(A_iB_j)_0} \right]$$

$\sim \chi^2$ distribution with $(r-1)(c-1)$ d.f.

If $\chi^2_{cal} \geq$ tab $\chi^2_{(\alpha, (r-1), (c-1))}$

then we reject H_0, otherwise fail to reject H_0 at $\alpha \cdot 100\%$ level of significance with $(r-1)(c-1)$ d.f.

* $\alpha \to \alpha \cdot 100\%$ যে কোনো level of significance ব্যবহারে use হচ্ছে নাহলে সাধারণত 5% বুঝাবে।

* For 2×2 contigency table,

a	b
c	d

then show that

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+a)(b+d)(c+d)}$$

$N = a+b+c+d$.

<u>Solve:</u> ⇒

$(A_iB_j)_0$ = Expected Frequency

$$= E(A_iB_j)$$

$$= \frac{(A_i)(B_j)}{N}$$

$$\chi^2 = \sum_{i=1}^{c} \sum_{j=1}^{r} \left[\frac{\{(A_iB_j) - E(A_iB_j)\}^2}{E(A_iB_j)} \right]$$

$$\Rightarrow \chi^2 = \frac{[a - E(a)]^2}{E(a)} + \frac{[b - E(b)]^2}{E(b)} + \frac{[c - E(c)]^2}{E(c)} + \frac{[d - E(d)]^2}{E(d)} \quad \ldots (1)$$

			Total
	a	b	a+b
	c	d	c+d
Total	a+c	b+d	N

Now, $E(a) = \frac{(a+b)(a+c)}{N}$

$E(b) = \frac{(a+b)(b+d)}{N}$

$E(c) = \frac{(a+c)(c+d)}{N}$

$E(d) = \frac{(c+d)(b+d)}{N}$

$\therefore a - E(a) = a - \frac{(a+b)(a+c)}{a+b+c+d}$

$= \frac{a^2+ab+ac+ad-a^2-ac-ab-bc}{a+b+c+d}$

$= \frac{ad-bc}{N}$

$b - E(b) = b - \frac{(a+b)(b+d)}{N}$

$= \frac{ab+b^2+bc+bd-ab-ad-b^2-bd}{a+b+c+d}$

$= \frac{bc-ad}{N} = -\frac{ad-bc}{N}$

$c - E(c) = -\frac{ad-bc}{N}$

$d - E(d) = \frac{ad-bc}{N}$

$\therefore$ From (i) we get,

$= \left\{\frac{ad-bc}{N}\right\}^2 \cdot \left\{\frac{1}{E(a)} + \frac{1}{E(b)} + \frac{1}{E(c)} + \frac{1}{E(d)}\right\}$

$= \left\{\frac{ad-bc}{N}\right\}^2 \left\{\frac{N}{(a+b)(a+c)} + \frac{N}{(a+b)(b+d)} + \frac{N}{(a+c)(c+d)} + \frac{N}{(c+d)(b+d)}\right\}$

$= \frac{(ad-bc)^2}{N}\left[\left\{\frac{b+d+a+c}{(a+b)(a+c)(b+d)}\right\} + \left\{\frac{b+d+a+c}{(a+c)(c+d)(b+d)}\right\}\right]$

$= \frac{(ad-bc)^2}{N} \cdot N \left\{\frac{1}{(a+b)(c+d)(b+d)} + \frac{1}{(a+c)(c+d)(b+d)}\right\}$

$= (ad-bc)^2 \left\{\frac{c+d+a+b}{(a+b)(c+a)(b+d)(c+d)}\right\}$

কাটা এবং বাদ, যে যে থাকবে তাদ

$= \frac{N(ad-bc)^2}{(a+b)(c+a)(b+d)(c+d)}$

*** Yeat's correction:**

10	4
16	14

$r \times c$ এ less than 5 হলে, degree of freedom 1 loss করলে, যেটা হবে না।

$r \times c$

$(r-1)(c-1) \rightarrow$ d.f.

$(2-1)(2-1) = 1 \times 1 = 1$

To avoid this situation,

$10-1/2$	$4+1/2$
$16+\frac{1}{2}$	$14-1/2$

- $\pm$ a and d
- $\mp$ b and c

marginal total ঠিক থাকে

$\therefore \chi^2 = \frac{N\left[|ad-bc| - \frac{N}{2}\right]^2}{(a+b)(a+c)(a+d)(c+d)}$

*** Types of Error:**

Date: 12.03.2024

	Male	Female
yes	30	13
No	21	19

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(b+d)(c+d)(b+d)}$$

$N = a+b+c+d$

$\sim \chi^2$-distn with

$(2-1)(2-1) = 1 \cdot 1 = 1$ d.f.

Here, $a=30,\ b=13,\ c=21,\ d=19$

$$\therefore \chi^2 = \frac{(30+13+21+19)(30\times19 - 13\times21)^2}{(30+13)(13+21)(21+19)(13+19)}$$

$= 7.197$

	Male	Female	Trans.
yes	30	13	43
no	21	29	50
Total	51	42	93

$$\sum \frac{\{a-E(a)\}^2}{E(a)}$$

$$E(30) = \frac{43\times51}{93}$$

$\text{Tab}\ \chi^2_{(0.05,1)} = 3.84$.

2 types of Error:

① Type-I error: Rejecting H_0 when H_0 is true.

- $P\{x : x \in W \mid H_0\} = \alpha$. $W \to$ ~~[illegible]~~ critical region

⑪ Type-II error: failed to reject H_0 when H_0 is false.

- $P\{x : x \in \bar{W} \mid H_1\} = \beta$.

$\alpha \to$ Probability of type-I error

$\beta \to$ " " " II "

- $P\{x : x \in W \mid H_1\} = 1-\beta$

ভুল বের হবার Probability.

যাচাইকরণ তত্ত্ব ভালো

Power of the test

Date: 27.04.2024

*Statistical Test of Hypothesis:

Defination:

1. Mean Test
2. Variance Test } 3টি
3. Correlation Test – 4টি.
4.

1. Mean Test

a. Single M.T.

b. Equality of M.T

c. Several M.T.

2. Variance Test

- Single V.T.
- Equality of V.T
- Several V.T

3. Correlation Test

a. Zero correlation

b. correlation test

c. Equality of correlation.

d. Several correlation

a. $\rho = 0$

b. $\rho = 0.81$

c. $\rho_1 = \rho_2$

d. $\rho_1 = \rho_2 = \cdots = \rho_K$

1: $\mu = \mu_0$

2. $\mu_1 = \mu_2$

3. $\mu_1 = \mu_2 = \cdots = \mu_K$

1. $\sigma^2 = \sigma_0$

2. $\sigma_1^2 = \sigma_2^2$

3. $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_K^2$

one tailed / right tailed test

$\int_{-\infty}^{\infty} f(z)dz = \alpha = 0.05 = 0.01$

$\int_{-\infty}^{\infty} f(t)dt = 1.$

Two tailed test

Test statistic:

Math: → sample data → sample size = 5

0.027, 0.031, 0.020, 0.025, 0.041

Q. Discuss the necessary steps to conduct the test of hypothesis. Do this data support the claim distance between two holes is 0.25% ?

$H_0 : \mu = 0.025$

$H_1 : \mu \neq 0.025$

sample size → $n \geq 30$ or population variance (σ^2) known

z =

নাহলে $n < 30$ and pop^n variance = 1

t =

Here, $n = 5$ (< 30) and population variance is unknown.

Then the test statistic,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t\text{-dist}^n \text{ with } n-1 \text{ d.f.}$$

calculated value of t

we know,

$$\bar{x} = \frac{\sum x_i}{n} = \frac{0.150}{5} = 0.0288$$

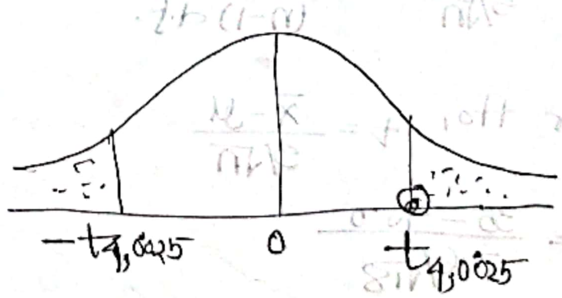
Under Ho,

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

$$= \frac{0.0288 - 0.025}{0.0017/\sqrt{5}}$$

$$= 0.566$$

Df. $= n-1 = 5-1 = 4$.



$t_{4,0.025} = \pm 2.13$

Since the calculated value 0.566 is much less than tabulated value at 5% level of significance with 4 d.f. then it is not significant and we fail to reject our Ho.

Ex. Suppose you are collected a sample of size 18 from a couple to investigate the age of them and found the correlation co-efficient between their age is 0.44. avg. age of husband is 35yrs with SD is 3.6yrs and the average of wives is 32.5yrs. It was known that the SD of wife's age is 4.5yrs. Test the following Ho Hypotheses.

1. $H_0: \mu_H = 40.5$, vs $H_1: \mu_H \neq 40.5$
2. $H_0: \mu_W = 30.2$ vs $H_1: \mu_W > 30.2$
3. $H_0: \rho = 0$ vs $H_1: \rho \neq 0$
4. $H_0: \sigma_W^2 = 2.1$

Solve:

1. The test statistic,

$$t = \frac{\bar{x} - \mu}{s/\sqrt{n}} \sim t\text{-dist}^n \text{ with } (n-1)\text{ d.f.}$$

Under Ho,

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

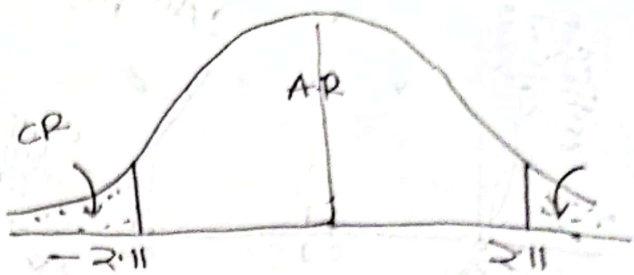
$$= \frac{35 - 40.5}{3.6/\sqrt{18}}$$

$$= -6.48$$

$\therefore |t| = |-6.48| = 6.48$

$\therefore d.f = n - 1 = 18 - 1 = 17$

The tabulated value of t at 5% level of signification with 17 df. is 2.011.



Here, calculated $|t|$ (+6.48) is greater than tabulated $t_{17, .05}$

or Calc $|t|$ (+6.48) > tab $t_{17, .05}$

Therefore it is highly significance and we may reject Ho at 5% level of significance with 17 d.f.

2. Here σ_W^2 is known

$$\therefore z = \frac{\bar{x} - \mu}{\sqrt{\sigma^2/n}} = \frac{\bar{x} - \mu}{\sigma_W/\sqrt{n}}$$

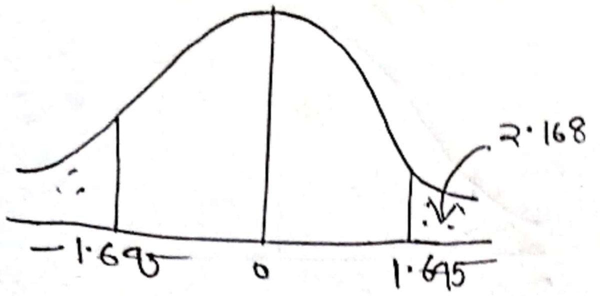
under Ho, $z = \dfrac{\bar{x} - \mu}{\sigma_W/\sqrt{n}}$

$$= \frac{32.5 - 30.2}{4.5/\sqrt{18}}$$

$$= 2.168$$

z-test:

	1%	5%	10%
Two tailed	2.58	1.96	1.645
Right tailed	2.33	1.695	1.28
left tailed	-2.33	-1.695	-1.28



$Cal(z) \geq Tab\ z_{(0.05)}$

It is also significant we reject H_0.

(3) $H_0 : \rho = 0$
$H_1 : \rho \neq 0$

The test statistic,

$$t = \frac{r\sqrt{n-2}}{\sqrt{1-r^2}} \quad \text{n t-dist}^n \text{ with } (n-2) \text{ d.f.}$$

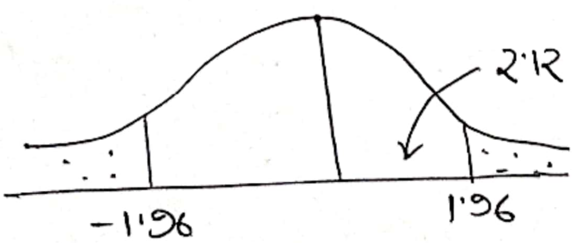
Here, $r = 0.44$
$n = 18$

$$\therefore t = \frac{0.44\sqrt{18-2}}{\sqrt{1-(0.44)^2}}$$

$$= \frac{0.44\sqrt{16}}{\sqrt{1-(0.44)^2}}$$

$$= 1.96$$

Tab $t_{(16; 0.05)} = 2.12$.



Cal $t(1.96) <$ tab $t_{(16; 0.05)}$

Therefore, it is significant we failed to reject H_0.

(4) $H_0 : \sigma_w^2 = 2.1$
$H_1 : \sigma_w^2 \neq 2.1$

The test statistic

$$\chi^2 = \frac{ns}{\sigma} = \frac{18 \times 1}{2.1}$$

তারিখ—২৮/৪/২০২৪

Non parametric Test: (NP)

Test of Hypothesis:
- Parametric → • $H_0 : \mu = 0$ • $H_0 : \sigma^2 = 2.5$ • $H_0 : \rho = 0$
- Non-parametric

এখানে population এর sample নি যা তা Test করে। data আসানো থাকে না random তা নেয়

Exm Q:
1. Definition of Non parametric T.
2. Distinguish between parametric and non parametric.
3. Advantages of NP.
4. Methods (1 OR 2)

CT (5-6 May)

Non-Parametric Methods:

- Rank Test
- Randomness Test
- Run Test
- Median Test
- Sign Test:
- Mann-whitney U-test.

* Another NP test:

- Kolmogorov Smirnov one sample Test:
- " " two " " :

→ * Randomness Test:

$x_1, x_2 \cdots x_n$

Median

$U \sim N(0,1)$ z-test.

H_0: Objective is random observation.

প্রশ্ন আসবে

* Median test:
* Mann – Whitney:

Median

$x_1\ x_2 \cdots x_{n_1}$

$y_1\ y_2 \cdots y_{n_2}$

Total observation: $n_1 + n_2$

$f_1(x)$ $f_2(y)$ $f(x) \rightarrow$ population

$f_1(x) = f_2(y)$

$z_1 \cdots z_{n_1+n_2}$

$m_1 \rightarrow$ কতগুলো x এর Median এর চেয়ে বড়।

Mann – Whitney – Wilcoxon test

x: 5 3 9 10 8

y: 4 7 13 16 12

16 13 12 10 9 8 7 5 ~~5~~ 3