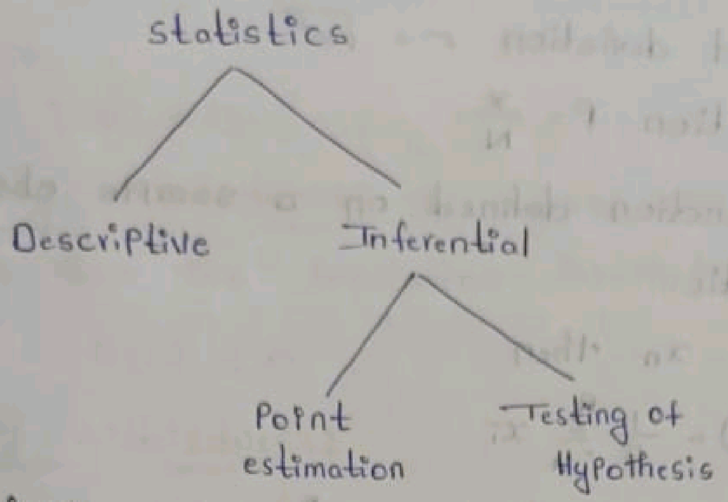
1. Basic concepts



Some definitions:

① **Population:** All items (or) elements (or) observations of interest having similar properties are known as "Population".

Ex: Sensus of Population.

The size of a Population is denoted as 'N'. i.e, $x_1, x_2 \dots x_N$.

Types of Population:

1. **Finite Population:** The finite Population are countable elements which are similar Properties.

Ex: No. of students in a class.

2. **Infinite Population:** Infinite Population means which are uncountable elements that are similar Properties.

Ex: No. of cells in our Body.

② **Sample:** A small group of items drawn from Population is known as sample. That is sample is subset of entire Population. The size of a sample is denoted by 'n'. i.e $x_1, x_2 \dots x_n$.

Ex: A Group of students selected from a college.

③ **Parameter:** Population constants is called Parameters.

Ex: Let $x_1, x_2, \dots x_N$ be a Population units then

① Population mean $\mu = \frac{1}{N}\sum_{i=1}^{n} x_i$

② Population Variance $\sigma^2 = \frac{1}{N}\sum_{i=1}^{N}(x_i - \mu)^2$

③ Population standard deviation $\sigma = \sqrt{\sigma^2}$

④ Population Proportion $P = \frac{X}{N}$

④ **Statistic:** A Function defined on a sample observations is called statistic.

Ex: Let $x_1, x_2, \ldots x_n$ then

① Sample Mean $(\bar{x}) = \frac{1}{n}\sum_{i=1}^{n} x_i$

② Sample Variance $s^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{x})^2$

③ Sample standard deviation $S = \sqrt{s^2}$

④ Sample Proportion $P = \frac{X}{n}$

Characteristic of a Good Estimator:

To find a Good estimator we have to know about Point Estimation.

① **Point estimation:** The Process to estimate a single Value of Population Parameter on the basis of sample observations. By this we have the Properties of good estimators they are:

* Unbaisdness
* Consistency
* Efficiency
* Sufficiency

① **Unbaisdness:** If Expected Value of an estimator is equal to its corresponding Population Parameter then it is unbaisdness.

$$E(\theta) = \theta$$

Where θ = Population Parameter

Baisd

equa

Baisd

Posi-

is

Neg

con

An

to

i.e

th

E

Baisdness: If expected value of an estimator is not equals to its Population Parameter then it is Baisdness.

$$E(\theta) \neq \theta$$

Positivity unbaisdness: It means expected Parameter is less than the Population Parameter.

$$E(\theta) \leq \theta$$

Negativity unbaisdness:

$$E(\theta) \geq \theta$$

consistency:

An Estimator is consistant if the estimator tends to get closer to the Parameter 'θ' it is, estimating as sample size increases

i.e., $x = \dots x_n \rightarrow \mu$ where $n - - - - \rightarrow \infty$

Here, the consistent estimator is expectation of $E(x) = \mu$

$Var(x) = s^2$

Efficiency:

Efficiency means the estimator taking as two i.e., $\hat{\theta}_1$ & $\hat{\theta}_2$ by comparing two Parameters of estimators is known as "efficiency". To find the Population Parameter for efficiency is

$$Var(\hat{\theta}_1) = Var(\hat{\theta}_2)$$

sufficiency:

The last Property that a good estimator should Possess sufficiency an estimator $\hat{\theta}$ is said to be sufficient estimator of Parameter θ if it contains all the information in the sample regarding the Parameter. In other words, a sufficient estimator utilizes all the information that the given sample about the Population.

These are the characteristics of a good estimator.

~~Invalid~~ Property of consistent estimator:

Invariant

Statement: If T_n is consistent estimator of γ and f is a continuous function then $f(T_n)$ is a consistent estimator of $f(\gamma)$.

Proof: T_n is the consistent estimator of γ i.e., $T_n \xrightarrow{P} \gamma$

$\Rightarrow P(|T_n-\gamma|<\varepsilon) \to 1 ; n \to \infty$

Since f is continuous function, so for every $\epsilon>0$, ϵ_1 is the positive integer such that $|T_n-\gamma|<\epsilon \Rightarrow |f(T_n)-f(\gamma)|<\epsilon_1$

We know that for any two events A & B, if $A=B$ then

$A \subseteq B = P(A) \leq P(B)$

$P(|T_n-\gamma|<\epsilon) \leq P(|f(T_n)-f(\gamma)|) < \epsilon_1$

$P(|f(T_n)-f(\gamma)|) < \epsilon_1 \geq P(|T_n-\gamma|<\epsilon$

Thus $f(T_n) \xrightarrow{P} f(\gamma)$

i.e., $f(T_n)$ is a consistent estimator of $f(\gamma)$.

Hence the theorem is proved.

Note:

If A is unbaised estimator of B then $f(A)$ is not unbaised of $f(B)$. (In General)

Example:

1. If x follows B Variance with Probability p then show that $\frac{x}{n}(1-\frac{x}{n})$ is a consistent estimator of $P(1-P)$.

Sol: Given x follows a Binomial distribution

so, $E(x)=np$ and $Var(x)=np(1-P)$

Given that $\frac{x}{n}(1-\frac{x}{n})$ is a consistent estimator of $P(1-P)$

Taking $G=\frac{x}{n}$

To prove that $G(1-G)$ is a consistent estimator of $P(1-P)$

firstly G is a consistent estimator of P.

$E(G) \to P$ and $Var(G) \to 0$, as $n \to \infty$

$E(G) = E\left(\frac{x}{n}\right)$

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

$= \frac{1}{\cancel{n}}(\cancel{n}P)$

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

$Var(G) = Var\left(\frac{x}{n}\right)$

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

$= \frac{P(1-P)}{n} \to 0$

as $n \to \infty$

$\therefore \frac{x}{n}\left(1-\frac{x}{n}\right)$ is a consistent estimator of $P(1-P)$.

Sufficient condition for consistency:

Statement: Let T_n be a sequence of estimator such that for all $\theta \in \Theta$

(i) $E_\theta(T_n) \to \gamma(\theta)$, $n \to \infty$

(ii) $Var_\theta(T_n) \to 0$, $n \to \infty$

then T_n is consistent estimator of $\gamma(\theta)$.

Proof: To Prove that T_n is a consistent estimator of $\gamma(\theta)$, then

$T_n \xrightarrow{P} \gamma(\theta)$, $n \to \infty$

$P(|T_n - \gamma(\theta)| < \varepsilon) > 1 - \eta \;\forall\, n \geq m(\varepsilon, \eta) \to (1)$

Where ε and η are arbitarily small Positive numbers and m is large value of n.

Applying chebyche's Inequality theorem

$P[|T_n - E_\theta(T_n)| \leq \delta] \geq 1 - \frac{Var_\theta(T_n)}{\delta^2} \to (2)$

$P\{|x-\mu| \geq k\sigma\} \leq \frac{1}{k^2}$; $P\{|x-\mu| < k\sigma\} \geq 1 - \frac{1}{k^2}$

By Triangular Inequality

$|T_n - \gamma(\theta)| = |T_n - E(T_n) + E(T_n) - \gamma(\theta)| \leq |T_n - E_\theta(T_n)| + |E_\theta(T_n) -$

$\gamma(\theta)| \to (3)$

Now

$|T_n - E_\theta(T_n)| \le \delta \to |T_n - \gamma(\theta)| \le \delta + |E_\theta(T_n) - \gamma(\theta)| \to (4)$

from two events A & B, if $A \to B$ then $A \subseteq B$ then $P(A) \le P(B)$

(or) $P(B) \ge P(A)$

By above formula

$$P\{|T_n - \gamma(\theta)| \le \delta + E_\theta(T_n) - \gamma(\theta)$$

$$\ge P|T_n - E_\theta(T_n)| \le \delta$$

$$\ge 1 - \frac{Var_\theta(T_n)}{\delta^2} \quad [\text{from } (2)] \to (5)$$

$\therefore E_\theta(T_n) \to \gamma(\theta) \ \forall \ \theta \in \Theta$ as $n \to \infty$

Hence for every $\delta_1 > 0$, $\exists$ a possible integer $n \ge n$.

$|E_\theta(T_n) - \gamma(\theta)| \le \delta_1, \forall n \ge n_0(\delta_1) \to (6)$

Also, $Var_\theta(T_n) \to 0$ as $n \to \infty$

$$\frac{Var_\theta(T_n)}{\delta^2} \le \eta, \ \forall n \ge n_0(\eta) \to (7)$$

substituting (6) & (7) in (5)

$P[|T_n - \gamma(\theta)| \le \delta + \delta_1] \ge 1 - \eta \ ; \ n \ge m(\delta, \eta)$

where $m = \max(n_0, n_1, ...)$ and $c = \delta + \delta_1 > 0$

$P[|T_n - \gamma(\theta)| \le \varepsilon] \ge 1 - \eta, \ n \ge m$

$T_n \xrightarrow{P} \gamma(\theta)$ as $n \to \infty$

T_n is the consistent estimator of $\gamma(\theta)$.

Examples

1. Find the sample Mean is an unbaised estimator of Population mean?

Sol: Sample Mean $= \bar{x}$

Population mean $= \mu$

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

Sample mean, $\overline{x} = \frac{1}{n} \sum_{i=1}^{n} x_i$

Applying expectation on both sides

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

$= \frac{1}{n} E\left[x_1 + x_2 + \cdots + x_n\right]$

Where $x_1, x_2, x_3, \ldots, x_n$ are independent variables.

$E(\overline{x}) = \frac{1}{n}\left[E(x_1) + E(x_2) + \cdots + E(x_n)\right]$

$= \frac{1}{n}\left[\mu + \mu + \cdots + \mu\right]$

$= \frac{1}{\cancel{n}}\left[\cancel{n}\mu\right]$

$\boxed{E(\overline{x}) = \mu}$

2) If $x_1, x_2, \ldots x_n$ is a random sample from a normal population with mean μ and variance 1, then show that $t = \frac{1}{n} \sum x_i^2$ is an unbaised estimator of μ^2+1.

<u>Sol:</u> Given $x_1, x_2, \ldots, x_n$

$x \sim N(\mu, 1)$

Sample Mean $E(x_i) = \mu$

$Var(\overline{x}) = E(x_i) - [E(x_i)]^2$

$1 = E(x_i^2) - \mu^2$

$E(x_i^2) = 1 + \mu^2$

To show that,

$t = \frac{1}{n} \sum x_i^2$

Applying expectation on b.s

$E(t) = E\left[\frac{1}{n} \sum x_i^2\right]$

$E(t) = \frac{1}{n}\left[\sum x_i^2\right]$

$E(t) = \frac{1}{n} \sum_{i=1}^{n} \left[E(x_i^2)\right]$

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

Consistency:

An estimator $T_n = T[x_1, x_2, \ldots, x_n]$ based on a random sample of size n is said to be a consistent estimator of $\gamma(\theta)$ where θ belongs to Θ or $\theta \in \Theta$ (Population Parameter). If T_n converges to $\gamma(\theta)$ (or) i.e $T_n \to \gamma(\theta)$ where $n \in \infty$. There exists a positive integer $n \geq m$ such that (ε, n) Probability of $P[|T_n - \gamma(\theta)| < \eta] \geq n \to \infty$ where m is large value of n $P[|T_n - \gamma(\theta)| < \varepsilon] \geq 1 - n \to \infty$

Ex: $\overline{x_n} = \frac{1}{n} \Sigma x_i = \mu_i$

Sample Mean $\overline{x_n}$ is always a consistent estimator of Population mean μ.

Most efficient estimator:

In a class of consistent estimator for a parameter there exist one sampling variance is less than that of any such estimator it is called most efficient estimator. Such estimator provides a criterion for measurement of efficiency of other estimators. To define this we have to know about efficiency.

Efficiency:

If T_n is the most efficient estimator with variance V_1 and t_2 is any other estimator with variance V_2 then the efficiency of t_2 is defined as $E = \frac{V_1}{V_2}$

obviously E cannot excede unity.

Examples

1. If $t_1, t_2, \ldots, t_n$ are all estimators of $\gamma(\theta)$ and the variance of t (or) $Var(t)$ is minimum then efficiency $E_i(T_i)$ where $i = 1, 2, \ldots, n$ it can be defined as

$$E_i = \frac{Var_1(T_i)}{Var_2(T_i)}$$

$\therefore E_i \leq 1$

Note:

In normal sample mean $\bar{x}$ is most efficient estimator of population of μ.

Factorisation theorem (Neymann):

Statement: $T = t(x)$ is sufficient for θ, if and only if the joint density function L (say) of the sample values can be expressed in the form.

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

where $g_\theta(t(x))$ depends on θ and x only to the value of $t(x)$ and $h(x)$ is independent variable.

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

Fisher, Neymann criterion:

A static $t_1 = t_1(x_1, x_2, \ldots, x_n)$ is sufficient estimator of parameter θ if and only if the likelyhood function can be expressed as

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

$$= g_1(t_1, \theta) \cdot k(x_1, x_2, \ldots x_n)$$

where $g_1(t_1, \theta)$ is the probability density function of statistics t_1 and $k(x_1, x_2, \ldots, x_n)$ is a function of sample observations and independent of θ. This method requires the working out of probability density function of statistic t_1 which is not always easy.

Sufficiency:

An estimator is said to be sufficient for a parameter if it constains all the information in the sample regarding the parameter. If $T = t(x_1, x_2, \ldots, x_n)$ is an estimator of parameter θ based on a sample size n

from the Population with density $f(x), \theta$ such that the conditional distribution $(x_1, x_2, \ldots, x_n)$ given to T and it is independent of θ then T is sufficient estimator of θ.

Rao-BlackWell theorem:

Let X and Y be random variables such that $E(y) = \mu$ and $Var(y) = \sigma_Y^2 > 0$

let $E[Y/X = x] = \phi(x)$ then

i) $E[\phi(x)] = \mu$

ii) $Var[\phi(x)] \leq Var(y)$

Proof:

Let $f_{xy}(x,y)$ be a joint Probability density function of random variables of X and Y. $f_1(\cdot)$ and $f_2(\cdot)$ are the marginal Probability density function of X and Y respectively and $h(y/x)$ be the conditional Probability function of y for given $X = x$ such that $h(y|x) = \frac{f(x,y)}{f_1(x)}$

$$E(y|x) = x) = \int_{-\infty}^{\infty} y \cdot h(y|x)\, dx$$

$$= y \int_{-\infty}^{\infty} \frac{f(x,y)}{f_1(x)}\, dx$$

$$= y \cdot \frac{1}{f_1(x)} \int_{-\infty}^{\infty} f(x,y)\, dx = \phi(x)$$

$$\Rightarrow \int_{-\infty}^{\infty} y f(x,y)\, dy = \phi(x) f_1(x) \rightarrow eq\ 1$$

From eq 1 we observe that the conditional distribution of y given $X = x$ doesn't depend on Parameter μ. Hence x is sufficient statistic of μ.

$E[\phi(x)] = E[E(y|x)] = E[y] = \mu$

we have $Var(y) = E[y - E(y)]^2 = E[y-\mu]^2$

$$= E[y - \phi(x) + \phi(x) - \mu]^2$$

$$= E[y-\phi(x)]^2 + E[\phi(x)-\mu]^2 + 2[y-\phi(x)][\phi(x)-\mu] \rightarrow eq\ 2$$

This Product term gives us,

$$E[y-\phi(x)][\phi(x)-\mu] = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} (y - \phi(x))(\phi(x)-\mu) \cdot f(x,y)\, dx dy$$

$$= \int_{-\infty}^{\infty} [\phi(x)-\mu]\left[\int_{-\infty}^{\infty} [y \cdot \phi(x)] \cdot h(y|x) \cdot h(y|x)\, dy\right] dx$$

$$= \int_{-\infty}^{\infty} (y - \phi(x)) \cdot h(y|x)\, dy = 0$$

$$\therefore E[(y-\phi(x))(\phi(x)-\mu)] = 0$$

By substituting in eq 2, we get

$Var(y) = E[y-\phi(x)]^2 + Var[\phi(x)]$

$var(y) \geq Var[\phi(x)]$

where $E[y-\phi(x)^2] = 0$

$\Rightarrow Var[\phi(x)] \leq Var(y)$

Hence the theorem is Proved

Here we have Proved the theorem for continuous random Variable this result can be Proved for discrete case by replacing integration as summation.

* Rao-Black Well theorem enables to obtain estimators through sufficient statistic if a sufficient estimator exists a Parameter then our search for MVU (minimum Variance unboised) estimator we may Restrict ourselves to functions of the sufficient statistic and the theorem can be stated slightly different as follows:

Let $u = v[x_1, x_2, \ldots x_n]$ be an unbaised estimator of Parameter $\gamma(\theta)$ and let $T = T(x_1, x_2, \ldots, x_n)$ be the sufficient statistic for $\gamma(\theta)$. consider the function ϕ of sufficient statistic defined as

$$\phi(t) = E[u / T = t]$$

which is independent of θ

Then $E[\phi(T)] = \gamma(\theta)$

$$Var(\phi(T)) \leq Var(u)$$

Consistent Problems

1. If $x_1, x_2, \ldots x_n$ on Bernoulli Variate X taking the Value 1 which Probability P and the Value 0 with Probability $(1-P)$ then show that $\frac{\Sigma x_i}{n}\left(1 - \frac{\Sigma x_i}{n}\right)$ is a consistent estimator of $P(1-P)$.

Sol: Given $x_1, x_2, \ldots, x_n$ are Bernoulli Variates With Parameter P

$$T = \sum_{i=1}^{n} x_i \sim B(n \cdot P)$$

$$E(T) = nP, \; Var(T) = nPq$$

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

$$E(\bar{x}) = E\left(\frac{T}{n}\right)$$

$$= \frac{1}{n} E(T)$$

$$= \frac{1}{\cancel{n}}(\cancel{n}P)$$

$$E(\bar{x}) = P$$

$$Var(\bar{x}) = Var\left(\frac{T}{n}\right) = \frac{1}{n} Var(T)$$

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

$$= \frac{Pq}{n}$$

$\therefore E(\bar{x}) = P$ and $Var(x) = 0$ as $n \to \infty$

$\therefore \bar{x}$ is a consistent estimator of P:

By this $\Sigma \frac{x_i}{n}\left(1-\frac{\Sigma x_i}{n}\right) = \bar{x}(1-\bar{x})$ being a Polynomial in $\bar{x}$. Since $\bar{x}$ is consistent estimator of P.

$\therefore$ By invariance Property of consistent estimator $\bar{x}(1-\bar{x})$ is a consistent estimator of $P(1-P)$.