Paper - II   II year

Sampling Distⁿ & Elements of Estimation

Unit I :- Univariate Sampling Distⁿ.

Univariate :- Having one variable only.

Univariate Distⁿ :- The distⁿ which carry one variable only are known as univariate distⁿ.

Population :- The word "popⁿ" or "universe" in statistics is used to collection of individuals or of their attributes which can be numerically specified. A Popⁿ is aggregate of objects animate or inanimate under study. The popⁿ may be finite or infinite.

eg: The popⁿ of weights, heights, prices of wheat, mileages of automobile tyres etc.

Finite popⁿ :- A popⁿ containing a finite no. of individuals or member is called a finite popⁿ.

eg: popⁿ of boys ~~ages of twenty in a~~ class.

Infinite popⁿ :- A popⁿ with infinite no. of members is known as infinite popⁿ.

eg:- popⁿ of pressures at various points in the atmosphere is an example of this type of popⁿ.

Hypothetical popⁿ :- The collection of all possible ways

in which an event can materialize as the hypothetical population.
eg:- popn of heads and tails obtained by tossing a coin an infinite no. of times is a hypothetical popn.

Sample :- A finite subset of statistical individuals in the popn or a part or small section selected from the popn is called a sample and the process of ~~of~~ such selection is called sampling.

Sample size :- The no of individuals in a sample is called the sample size.

Why need sampling :- Sampling is resorted or (required) to when either it is impossible to enumerate the whole popn or when it is too costly to enumerate in terms of time and money or when uncertainity inherent in sampling is more than compensated by the possibilities of errors in complete enumeration.
To serve a useful purpose sampling should be unbiased or representative.
The main aim of theory of sampling is to get as much information as possible ideally about the popn from which sample has been drawn.

In particular given the form of the parent popⁿ. we would like to estimate the parameters of the popⁿ or specify the limits within which the popⁿ parameters are expected to lie with a specified degrees of confidence.

Parameters are unknown statistical constant of the popⁿ.

eg: Popⁿ mean (μ), popⁿ variance (σ^2).

Statistics :- Statistical measures computed from the sample obserⁿ.

eg: mean ($\bar{x}$), variance (s^2)

Unbiased Estimate :-

$$E(\text{statistic}) = \text{Parameter}$$

Then statistic is called unbiased estimate of parameter.

Sampling Distⁿ :- A part or small section selected from the popⁿ is called a sample and the process of such selection is called sampling.

If we draw a sample of size n from a popⁿ of size N then the total no. of possible sample is $\binom{N}{n} = k$ (say). for each of these k we compute some statistic like mean ($\bar{x}$), variance (s^2) etc. which constitute a freqⁿ distⁿ.

Computation of values or set of values so obtained is called sampling distribution of the statistic.

Types of sampling :-

i) Purposive sampling
ii) Random sampling
iii) Simple sampling
(iv) Stratified sampling.

Random sampling :- In random sampling each unit of popⁿ have equal and indept. chance to being included in the sample. The process of drawing sample by such method is known as random sampling.

Sampling Error :- Sampling error arises due to the fact that only a part of popⁿ (i.e) sample has been used to estimate the popⁿ parameter & draw inferences about the popⁿ which is unavoidable & inherent in any and every sampling scheme.

Standard Error :- The standard deviation of a the sampling distⁿ of a statistic is known as its standard error.

$\therefore$ S.E.(x) = σ = $\sqrt{V(x)}$

Page
Date

Standard Error of sample mean:— Let x_i $(i = 1, 2, \ldots n)$ be a r.s. of size n from a popⁿ with variance σ^2, then the sample mean $\bar{x}$ is given by

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

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

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

$$= \frac{1}{n^2} V(x_1 + x_2 + \ldots + x_n)$$

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

$$= \frac{1}{n^2}\left[\sigma^2 + \sigma^2 + \ldots + \sigma^2) + 0 + 0 + \ldots 0\right]$$

(covariance term vanishes since x_i's are indept.)

$$= \frac{1}{n^2} n \cdot \sigma^2 = \frac{\sigma^2}{n}$$

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

Standard Error of residual variance or Standard error of estimate:— The equation of the line of regression of Y on X is

$$(Y - \bar{y}) = r\frac{\sigma_y}{\sigma_x}(X - \bar{x})$$

$$\text{or} \quad \left(\frac{Y - \bar{y}}{\sigma_y}\right) = r\left(\frac{X - \bar{x}}{\sigma_x}\right)$$

The residual variance or S.E. of estimate S_y^2 is expected value of the square of deviations of the observed value of y from the expected value as given by the line of regression of y on x. Thus

$$S_y^2 = E\left[(y-\bar{y}) - r\frac{\sigma_y}{\sigma_x}(x-\bar{x})\right]^2$$

$$= \sigma_y^2 E\left[\frac{(y-\bar{y})}{\sigma_y} - r\left(\frac{x-\bar{x}}{\sigma_x}\right)\right]^2$$

$$= \sigma_y^2 E\left[\frac{y-\bar{y}}{\sigma_y} - r\left(\frac{x-\bar{x}}{\sigma_x}\right)\right]^2$$

$$= \sigma_y^2 E\left[y^* - r x^*\right]^2$$

Where x^* & y^* are standardised variate so that

$$E(x^*) = E(y^*) = 0$$
$$E(x^{*2}) = E(y^{*2}) = 1$$
$$E(x^* y^*) = \cancel{E}\, r$$

$$\therefore S_y^2 = \sigma_y^2 E\left[y^{*2} - 2r x^* y^* + r^2 x^{*2}\right]$$

$$= \sigma_y^2\left[E(y^{*2}) - 2r E(x^* y^*) + r^2 E(x^{*2})\right]$$

$$= \sigma_y^2\left[\cancel{E}\, 1 - 2r.r. + r^2.1\right]$$

$$= \sigma_y^2\left[1 - r^2\right]$$

$$\therefore S_y = \sigma_y\sqrt{(1-r^2)}$$