Course Title: Statistics for Economics
Course Code:ECON4008
Topic:Theory of Estimation
M.A. Economics (2nd Semester)

Dr. Kailash Chandra Pradhan

Mahatma Gandhi Central University,

Department of Economics

Theory of Estimation

- In this topic, we will generalize the results of sample to the population; to find how far these generalization are valid, and also to estimate the population parameters along with degree of confidence.
- To answer to these are provided by the statistical inference and it is classified into the following two.

  (i) Theory of estimation

  (ii) Testing of Hypothesis

Estimation Theory:

- The theory of estimation was founded by Prof. R.A. Fisher in a series of fundamental papers round about 1930 and is divided into two groups.

  (i) Point Estimation

  (ii) Interval Estimation

- In point estimation, a sample statistic is used to provide an estimate of the population parameter whereas in interval estimation, probable range is specified within which the true value of parameter might be expected to lie.

Point Estimation

- A particular value of a statistic which is used to estimate a given parameter is known as point estimate or estimator of parameter.
- A good estimator is one which is as close to the true value of the parameter as possible.
- The following are some of the criteria which should be satisfied by a good estimator.

  (1) Unbiasedness

  (2) Consistency

  (3) Efficiency

  (4) Sufficiency

Point Estimation

1. Unbiasedness

- A statistic $t=t(x_1, x_2, x_3, \ldots, x_n)$, a function of the sample observations $x_1, x_2, x_3, \ldots, x_n$ is said to be an unbiased estimate of the corresponding population parameter θ, if

$$E(t)= \theta$$

- This means the mean value of the sampling distribution of a statistic is equal to the parameter.
- For example, the sample mean $\bar{x}$ is an unbiased estimate of the population mean μ; the sample proportion of p is an unbiased estimate of the population proportion P.

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

$$E(p)=P$$

Point Estimation

- Sample variance

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

is not unbiased estimate of the population variance σ^2. However,

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

provides an unbiased estimate of the population variance σ^2. Thus

$E(s^2) \neq \sigma^2$ but $E(S^2) = \sigma^2$

- If the sample size is large, then n-1 can be approximately by n. Thus, for large sample, the sample variance gives an unbiased estimate of the population variance σ^2.

Point Estimation

2. Consistency

- A statistic $t=t_n=t(x_1, x_2, x_3,\ldots..,x_n)$ based on a sample of size n is said to be a consistent estimator of the parameter θ. If $t_n \rightarrow \theta$ as $n \rightarrow \infty$. Symbolically,

$$\lim_{n\to\infty} p(t_n \to \theta) = 1$$

- For any distribution, sample mean $\bar{x}$ is a consistent estimator of the population mean, sample proportion p is a consistent estimator of the population proportion P and sample variance s^2 is a consistent estimator of the population variance σ^2.
- A consistent estimator need not be unbiased.

Theorem:

- A statistic $t=t_n=t(x_1, x_2, x_3,\ldots..,x_n)$ is a consistent estimator of the parameter θ if

$$\left.\begin{array}{l} E(t_n) \to \theta \\ Var(t_n) \to 0 \end{array}\right\} as..n \to \infty$$

Point Estimation

3. Efficiency

- If we have more than one consistent estimators of a parameter θ, then efficiency is the criterion which enables us to choose between them by considering the variances of the sampling distributions of the estimators. Thus, if t_1 and t_2 are consistent estimators of a parameter θ such that

  $$Var(t_1)<Var(t_2), \text{ for all } n.$$

  then t_1 is said to be more efficient than t_2. In other words, an estimator with lesser variability is said to be more efficient and consequently more reliable than the other.
- If there exist more than two consistent estimators for the parameter θ, then considering the class of all such possible estimators we can choose the one whose sampling variance is minimum.
- Such an estimator is known as the most efficient estimator and provides a measure of the efficiency of the other estimators.
- If t is the most efficient estimator of a parameter θ with variance v and t_1 is any other estimator with variance v_1, then the efficiency E of t_1 is defined as:

  $E=v/v_1$ which less than equal to 1.

Point Estimation

4. Sufficiency

- A statistic $t=t(x_1, x_2, x_3,.....,x_n)$ is said to be a sufficient estimator of parameter θ if it contains all the information in the sample regarding the parameter.
- In other words, a sufficient statistic utilises all the information that a given sample can furnish about the parameter.
- The sample mean $\bar{x}$ is sufficient estimator of population mean μ and sample proportion p is a sufficient estimator of population proportion P.

Properties of Sufficient estimator

1. If a sufficient estimator exists for some parameter then, it is also the most efficient estimator.
2. It is always consistent.
3. It may or may not be unbiased.
4. A minimum variance unbiased estimator for a parameter exists if and only if there exists a sufficient estimator for it.

Central Limit Theorem

- We know that if x_1, x_2, x_3,.....,x_n is a random sample from normal population with mean μ and variance σ^2, then the sample mean $\bar{x}$ is also normally distributed with mean μ and variance σ^2/n, i.e. $\bar{x} \sim N(\mu, \sigma^2/n)$.
- The result is true even if the population from which samples are drawn is not normal, provided the sample size is sufficiently large.
- The larger sample size, better will be the approximation. The approximation is fairly good if n>30.

Central Limit Theorem

- If x_1, x_2, x_3,.....,x_n are independent random variables following any distribution, then under certain very general condition, their sum $\sum x = x_1+x_2+ x_3+\ldots +x_n$ is asymptotically normal distributed, i.e. $\sum x$ follows as $n \rightarrow \infty$.

Interval Estimation

- In point estimation, a single value of statistic used as an estimate of population parameter.
- But even the best possible point estimate may deviate enough from the true parameter value to make the estimate unsatisfactory.
- Thus, having obtained a statistic t from given random sample, the problem arises, "Can we make some reasonable probability statements about the unknown parameter θ in the population from which the sample has been drawn"?
- The answer is provided by the technique of interval estimation, pioneered by Neyman. This consists in the determination of two constants c_1 and c_2.

$$P[c_1 < \theta < c_2, \text{ for the given value of } t] = 1-\alpha$$

- Where, α is the *level of significance*. The interval $[c_1, c_2]$, within which the unknown value of the parameter θ is expected to lie is known as *confidence interval*; the limits c_1 and c_2 so determined are known as *confidence limit* and $1-\alpha$, is *confidence coefficient*.

Interval Estimation cntd…

- For example, α=0.05 (or 0.01) gives the 95% (or 99%) confidence limits.
- If t is the statistic used to estimate the parameter θ, then

  (1-α) *confidence limit for* $\theta = t \pm S.E.(t) \times t_{\alpha/2}$
- Where $t_{\alpha/2}$ is the significance or critical value.
- The computation of confidence limits for a parameter θ, involves the following three steps.
  1. Compute the appropriate sample statistic t
  2. Obtain the standard error of the sampling distribution of the statistic t i.e. S.E. (t)
  3. Choose appropriate confidence coefficient (1-α), depending on thee precision of the estimation.

Interval estimation for large sample

- For the large samples, the underlying distribution of the standardised variate corresponding to the sampling distribution of the statistic t will be asymptotically normal by Central Limit Theorem and $n \to \infty$

$$Z = \frac{t - E(t)}{S.E.(t)} \sim N(0,1)$$

Reference:

Gupta, S. C. (2015), *Fundamentals of Statistics*, Himalaya Publishing House.