Nepal Journal of Mathematical Sciences (NJMS)
ISSN: **2738-9928** (*online*), **2738-9812** (*print*)
Vol. 3, No. 1, 2022 (February): 125-127
DOI: https://doi.org/10.3126/njmathsci.v3i1.44131

Expository Article
Received Date: December 15, 2021
Accepted Date: February 24, 2022
Published Date: February 28, 2022

Estimating Population Values in the Survey Research

Hari Bhakta Shahi

Tribhuvan University, Public Administration Campus, Kathmandu, Nepal

Email: *haribhaktashahi7@gmail.com*

Abstract: *The significance of sample statistics in the research study is increasing day by day. In the various study as medical science, physical sciences and social sciences total population study is guessed from sample statistics. There are two methods for finding population parameter viz point estimate and interval estimate. Point estimate is error free method and interval estimate always count error as well as confidence level. In research to find out population values there are fixed standard to draw the sample. The sample must be unbiased, consistent, efficient and sufficient because sample is basic foundation of the overall process.*

Key Words: *Statistic, Point- estimate, Interval-estimate, Confidence level*.

1. Introduction

The word Estimation is guess of any things, any matters and any entity. But in quantitative analysis estimation is focus on guess of population parameter from sample statistic. So mathematically, estimation is concerned with the methods by which population parameter are guessed from the sample statistic (Brian and Philip [1], Sharma and Chaudhary, [9]). In the 2010s, estimation methods were increasingly adopted in neuroscience (Hentschke, Harald, Maik and Stüttgen, [4]). Estimation is the process of approximation, which is a value that is usable for some purpose. The value is usable because it is derived from the best information available. Typically, estimation involves using the value of a statistic (sample) to estimate the value of a corresponding population (Enloe, [2]).

Parameter is the characteristic of population such as; population standard deviation (σ), population variance(σ^2), population proportion success (P), population proportion failure (Q) and population mean (μ) etc. statistic is the characteristic of sample such as sample standard deviation (s), sample variance (s^2), sample proportion process (p), sample proportion failure (q), sample mean $(\bar{X})$ etc. statistic are also called estimator because statistic are used to estimate population parameter (Enloe, [2]). Estimated specific value of the estimator is known as estimate. The main propose of estimation is to obtain or guess population parameter from the sample statistic. There are so many examples in our daily life where we guess. For example by testing a mango we can estimate test of all the mangoes, by testing a piece of cooked food we can estimate about whole food, from the blood sample doctors can estimate content and diseases of our body, by testing density of little soil engineers can estimate density of total soil of that area, by testing a cup of juice people can estimate total juice test, from a cup of water in the sea people can estimate all the test of whole sea water etc.

2. Standard of a Statistic

Sample statistic is to be close to the population parameter. There are four conditions for good estimator as: Unbiasedness, Consistency, Efficiency, and Sufficiency. These criteria are described in (Gupta, [3]).
An estimator i.e., sample value must be representative for whole population. In other word all the expected vale of statistic is equal to parameter. That is

$$E(\bar{X}) = \mu, E(s) = S = \sigma, E\,(s^2) = \sigma^2, E(p) = P, E(q) = Q, \text{ where } E = \text{Estimation.}$$

Let us assume, if t is the values of sample and θ is the values of population, according to estimation theory; E (t) = θ, which is condition of unbiasedness. In this condition $\bar{X}$ is a good estimator of μ, p is a good estimator of P, q is a good estimator of Q but s is not a good estimator of σ but in large sample s is a good estimator of σ (Pillai and Bhagavathi, [7]) and (Sharma and Chaudhary, [9]).

If s = Sample Standard Deviation (Biased estimate) and S = Adjusted Sample Standard Deviation (Unbiased estimate), then

$$s^2 = \frac{\sum(X-\bar{X})^2}{n} \text{ and } S^2 = \frac{\sum(X-\bar{X})^2}{n-1}, \text{ which give, } S^2 = \frac{ns^2}{(n-1)}$$

For the large sample (n-1) is nearly equal to (n), so, $S^2 = s^2$ therefore $s = S = \sigma$.
So for large sample, $\hat{\sigma} = \sigma$.

In field of pure science, mathematics, engineering, medical science and natural phenomena, estimation of sample statistics is equal to population parameter. That is a sample is always representative of all population. For example a piece of blood and fruit represents whole of population of blood and fruit. But in social science, a person is lazy cannot conclude all people are lazy and vice versa. Thus **consistency** in social science refers to the effect of sample size on the accuracy of the estimator. A statistics is said to be consistent trees for population parameters if the sample size increases.

Thus mathematically (Enloe, [2]), consistency means:

$$\lim_{n\to\infty} E(\bar{X}) = \mu,\ \lim_{n\to\infty} E(s) = S = \sigma\ ;\ \lim_{n\to\infty} E(s^2) = \sigma^2;\ \lim_{n\to\infty} E(p) = P \text{ and } \lim_{n\to\infty} E(p) = P.$$

In the same way, if (t) is the statistic and θ is the parameter than ${}_{n\to\infty}^{lt}E(\text{t}) = \theta$.Thus consistency increases with increase of sample size. In other word results of different researcher in the same topic in same period of time must be equal. i.e. $t - \theta = 0 = \text{negligible amount}$.

Sample values are *efficient* if its value remains stable from sample to sample. The best estimator would be the one which would have the least variance or standard deviation from sample to sample taken randomly from same population. Less variation in sample is good sign for estimation. Let $t_1\ and\ t_2$are consistent estimators of a population parameter θ than variance of $t_1 > variance\ of\ t_2$ means consistency of t_2 is better than consistency of t_1. That is efficiency identifies no variance, no error, no deviation etc.

Sample values are sufficient *estimator* for the parameter (θ) if it is unbiased, consistent and efficient. It also uses all the information of the population by the sample is called sufficient. In other world a sample is said to be sufficient if it is representative of total population, least variance or standard deviation from sample to sample observation, less error and so on.

The theory of estimation was developed by Professor R.A. Fisher in 1930. He proposed two Theories of estimation: point estimation and interval estimation (Hentschke, [4]), which are described below:

A particular sample value (Statistic) which is used to estimate unknown population value (parameter) is called ***point estimate***. Good estimator must have four feature such as sample must be unbiased, consistent, efficient and sufficient to be true representative of population (Hunt, [5]). So point estimate is used to estimate unknown population parameter from sample statistic. This method is error free method. That is error is ignored in this method. Homogeneous data like science, mathematics, engineering, medical science apply this method. Thus Mathematically (Gupta, [3]),

$$(\bar{X}) = \mu,\ E(s) = S = \sigma,\ Es^2 = \sigma^2,\ E(p) = P \text{ and } E(q) = Q.$$

In social science and management we cannot ignore errors. Thus population parameter lies between less or more than statistics. An estimate of population parameter given by the two numbers of statistics is called *interval estimate*. In this method first of all we calculate error. Then we fixed confidence level of working as we know statistical findings are never 100 percentages. Finally interval estimate may lie less or more of the statistics. In point estimate statistics is equal to parameter. But in interval estimate population parameter lies between two ranges of statistics. To find out interval estimate fixed the confidence level which provide standard normal value (Z) and calculate the standard error. The interval estimate = point estimate ± Z × standard error,where Z = standard normal value depends on confidence level given in the table as (Pillai[7]):

Confidence level$(1-\alpha)\times 100)$%	90%	95%	96%	98%	99%	99.7%
Standard normal value (Z-value)	1.645	1.96	2.05	2.33	2.58	3

3. Interval Estimates

Population is divided in two parts such as success and failure, male and female, educated and uneducated, rich and poor and so on. But these two parts are unknown; it is difficult to compute success and failure part in the population. For this propose we draw sample and calculate sample proportion success (p) and failure (q) then we estimate population parameter. Mathematically, (James, [6]) Interval estimate = point estimate ± Z standard error. Standard Errors for infinite population and infinite population are $\sqrt{\frac{pq}{n}}$ and $\sqrt{\frac{pq}{n}(\frac{N-n}{N-1})}$ respectively. Thus Interval Estimate for Population Proportion (P) are $p\pm\sqrt{\frac{pq}{n}}$ (for infinite population) and $p\pm Z\sqrt{\frac{pq}{n}(\frac{N-n}{N-1})}$ (for finite population) respectively.

Population contents so many variables. Mean value of population cannot be calculated and unknown in various situations. It is difficult to compute population mean in the population always. So to enumerate population value we draw sample and calculate sample mean, then we estimate population mean(μ), (Gupta, [3] and Raymond,[8]) . The interval estimate for population mean (μ)= point estimate ± Z (standard error). Standard error of sample mean, for infinite population and finite population are $\frac{\hat{\sigma}}{\sqrt{n}}$ and $\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ respectively. Thus Interval Estimate for Population Mean (μ) are respectively $\bar{X}\pm\frac{\hat{\sigma}}{\sqrt{n}}$(For infinite population) and $\bar{X}\pm Z\frac{\hat{\sigma}}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$ (For finite population).

4. Concluding Remarks

Population parameters are guessed from sample statistics. In administration, there are various activities for example; economic activities, research activities, budgeting, and planning, etc. need estimation. In medical sciences, estimation theory is highly used in medical research, diagnosis of disease, treatment to patients. Total population study is not possible so, by using appropriate sampling for the study and applying estimation theory, we conduct various activities in research and administration.

References

[1] Brian, P. M. and Philip, M. N. **(2006)**. *Applied Statistics for Public Policy*, New Delhi, Prince Hall of India.

[2] C. Lon Enloe, Elizabeth Garnett, Jonathan Miles **(2000)**. *Physical Science: What the Technology Professional Needs to Know* p. 47.

[3] Gupta ,S.C. **(2010)**. *Fundamental of Statistics***.** Himalaya Publishing House, New Delhi.

[4] Hentschke, H. and Maik C. S.**(2011)**. Computation of measures of effect size for neuroscience data sets . *European Journal of Neuroscience*. 34 (12): 1887–1894. doi:10.1111/j.1460- 9568.2011.07902.x. PMID 22082031. S2CID 12505606.

[5] Hunt, M.**(1997)**. *How science takes stock: the story of meta-analysis*. The Russell Sage Foundation. New York.

[6] James T. and John S. **(2003)**. An overestimate is an estimate you know is greater than the exact answer, *Reviewing Mathematics* , page 27.

[7] Pillai R.S.N. and Bhagavathi V.**(1993)**. *Statistics*. Prince Hall of India, New Delhi.

[8] Raymond, A. K. **(2001)**. *Estimation, Data Construction and Data Analysis for Survey Research* , p. 157.

[9] Sharma, P.K. and Chaudhary, A.K. **(2015)**. *Statistical Methods*. Khanal Publication Pvt.Ltd., Kathmandu.