IOSR Journal of Computer Engineering (IOSR-JCE)
e-ISSN: 2278-0661,p-ISSN: 2278-8727, Volume 16, Issue 6, Ver. II (Nov – Dec. 2014), PP 30-35
www.iosrjournals.org

Estimation Theory

Dr. Mcchester Odoh and Dr. Ihedigbo Chinedum E.
Department Of Computer Science Michael Opara University Of Agriculture, Umudike, Abia State

I. Introduction

In discussing estimation theory in detail, it will be essential to recall the following definitions.

Statistics: A statistics is a number that describes characteristics of a sample. In other words, it is a statistical constant associated with the sample. Examples are mean, sample variance and sample standard deviation **(chukwu, 2007).**

Parameter: this is a number that describes characteristics of the population. In other words, it is a statistical constant associated with the population. Examples are population mean, population variance and population standard deviation. A statistic called an unbiased estimator of a population parameter if the mean of the static is equal to the parameter; the corresponding value of the statistic is then called an unbiased estimate of the parameter. **(Spiegel, 1987).**

Estimator: any statistic 0=0 (x1, x2, x3.......xn) used to estimate the value of a parameter 0 of the population is called estimator of 0 whereas, any observed value of the statistic 0=0 (x1, x2, x3.......xn) is known as the estimate of 0 **(Chukwu, 2007).**

II. Estimation Theory

Statistical inference can be defined as the process by which conclusions are about some measure or attribute of a population (eg mean or standard deviation) based upon analysis of **sample data**. Statistical inference can be conveniently divided into two types- **estimation** and **hypothesis testing (Lucey,2002).**

Estimation deals with the estimation of population characteristics (such as the population mean and standard deviation) from sample characteristic (such as the sample mean and standard deviation) the population characteristics are known as **population parameters** whilst to the sample characteristics are known as **sample statistics. An estimate** is therefore a statistic obtained from a sample that enables us to make a projection about its corresponding population parameter. **(Lucey, 2002). Types of estimate (choice of an estimate):** for any given parameter of a population. We have two types of estimate: a **point estimate and interval estimate:**

- **A point estimate:** A point estimate uses a single sample value to estimate the population parameter involved. For example mean x is a point estimate of the population mean u. The sample variance S^2 is a point estimate of population variance ; σ^2 (**Ewurum, 2003).**
- **Interval estimation:** consist of two numerical values which define on interval with some degree of confidence within which the value of the parameter being estimated lie. We can use mean, medium and mode as an estimate of the population (**Ugwuga, 2004**).

It is necessary to distinguish between the symbols used for sample statistics and population parameters as follows.

	Sample statistic	Population parameter
Arithmetic mean	x	μ
Standard deviation	S	σ
Number of items	n	N

Properties of good estimator

There are four four properties of good estimator

a**. unbiased**: the mean of the distribution of sample would equal the population mean.

b. **Consistency:** as the sample size increases, the precision of the estimate of the population parameter also increases.

c. **Efficiency:** an estimator is said to be more efficient than another if in repeated sampling, its variance is smaller.

d. **Sufficiency:** an estimator is said to be sufficient if it uses all the information in the sample estimating the required population parameter (**Lucey, 2002)**.

If we say that a distance is measured as 5.28meter we are giving a point estimate. If on the other hand, we say that the distance is 5.28+0.03m (ie the distance lies between 5.25 and 5.31m) we are giving the interval estimate.

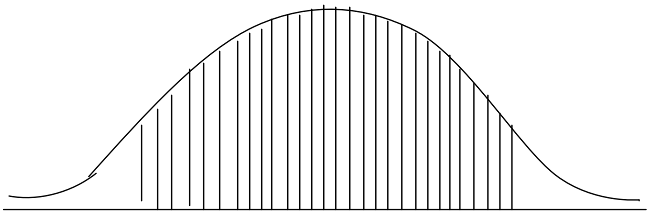
A statement of the error (or precision) of an estimate is often called its reliability (**Spiegel and Stephens, 1998**).

Generally, our focus would be centered on interval estimation. This interval has a specified confidence or probability of correcting estimating the true value of the population parameter.

Confidence interval estimation of the mean (0 known)

Confidence level	99.73%	99%	98%	96%	95.45%	95%	90%	80%	68.24%	50%
Z_c	3.00	2.58	2.33	2.05	2.00	1.96	1.645	1.28	1.00	0.67

Either from the central limits theorem or from knowledge of the population we could determine the proportion of sample means that fell within certain limits of the population mean. If the statistic S is the sample mean X, then the 95% and 99% confidence limits for estimating the population mean u are given by X ± 1.96 x and x±2.58 0x, respectively. More generally the confidence limits are given by x ± Zc 0X, where Zc (which depends on the particular level of confidence desired. If sampling is either from infinite population or with replacement from a finite population the formula is given by
u = x ± Zc σ/n (**Spiegel and Stephens, 1998**).
For example, for the distribution of tyre kilometer distance having x=23,000, 0=5,000 and a sample size of 100, when sampling with replacement given 95% confidence level.
Solution: μ=23,000 ±Zc 500/√100
μ=23,000±(1.96) (500)
μ=23,000 ±980
Therefore, the population mean interval would fall between 23,000+980=23,980 and 23,000-980=22,020.



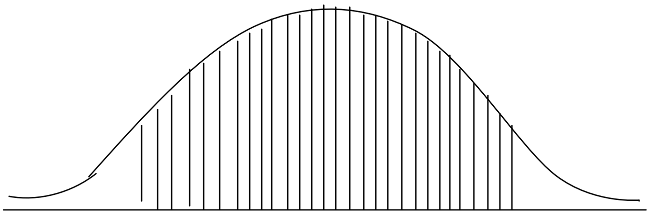
23,980 ≤μ≤ 22,020

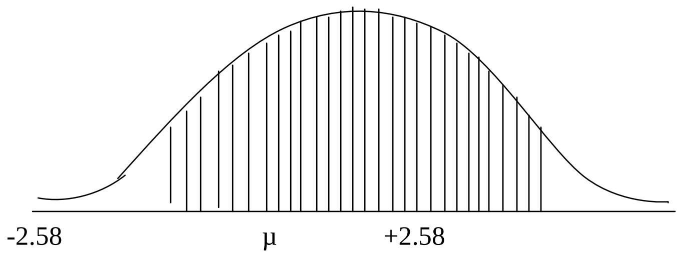
What if the sample mean was 23,500 kilometers? Then the interval would be 23,500+980 that is 24,480≤ μ ≤ 22520 observe that the population mean of 23,000 is included within this interval and hence estimate of u is a correct statement. What if the sample mean were 21,000 kilometer?
The interval would then be 21,000±980 that is 21,980
≤μ≤20000

(**Ewurum, 2003**)

We should realise that this estimate of £ would not be correct statements since the true population mean of 23000 is not included in the interval. For some samples, the estimate of £ will be correct. The population mean could be either at the upper limit of the interval or at the lower limits of the interval (**Ewurum, 2003)**.

In general a 95% confidence interval estimate can be interpreted to mean that he all possible sample size n were taken, 95% of them would include the true population mean some where within their interval, while only 5% of them would fail to estimate the true mean correctly we might desire a higher or less confidence than 95%. In some problems we might desire a high degree of assurance (such as 99%) of including the population mean with the interval. In other cases we might be willing to accept less assurance (such as 90%) of correctly estimating the true population mean. If 90% and 99% confidence were desired, the area 0.09 and 0.99 would be divided into two leaving 0.45 and 0.495 respectively between each limit and £. The z value for 90% confidence is ± 1.645 while that of 99% confidence is ± 2.58





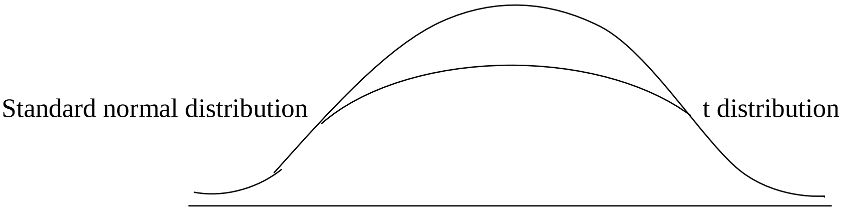
Given x =2457.92, n =72, 6=72, £=495.21, the 99% confidence interval estimate for the population mean in found to be
$X \pm Z\ \sigma/n = 2457.92 \pm (2.58)\ (495.21)/\sqrt{72}$
$= 2457.92 \pm 150.57$
$2,307.32 \leq \mu \leq 2,608.49$

III. Confidence Interval Estimation Of Mean (□ Unknown)

Just as the population mean £ is usually not known, the actual standard deviation of the population σ is also NOT likely to be known. Therefore, we need to obtain a confidence interval estimate of σ by using only sample statistic X and s. The distribution that has been developed to be applied to this situation is students t-distribution. For sample of size n>30, called large sample, the sampling distributions of many statistics are approximatly normal, the approximation becoming better with increasing n. For samples of size n<30, called small samples, this approximation is not and becomes worse with decreasing n, so that appropriate modifications must be made. A study of sampling distribution for small sample is called small sampling theory. However a more suitable name would be exact sampling theory, since the result obtained hold for large as well as for small samples (**Ewurum,2003**).

Students t -distribution

Discovered by **W.S Gosset**, who published his works under the pseudonym **'student'** during the early part of the twentieth century (**Spiegal And Stephens 1998**).
$t_{n-1}\ x-\mu/\ \sigma/\sqrt{n}$ (**Unyimadu 2001**)
The t-distribution is a symmetrical distribution similar to the normal distribution although it has more area in the tail and less in the center than does the normal distribution.



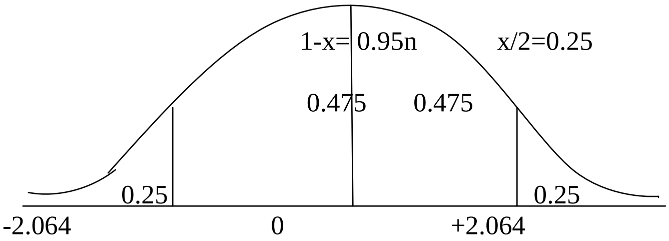
However, as the number of degree of freedom increases, the t distribution gradually approaches the normal distribution until the two are practically identical (**Ewurum 2003).**

The concept of degree of freedom: the number of degree of freedom of a statistic generally denoted by V is defined as the number n of independent observations in the sample (i.e the sample sample size) minus the number k of population parameter that must be estimated from sample observations. In symbols, v =n-k (**Ewurum, 2003**).

In the case the t statistic, the number of independent observations in the sample is n, from which we can compute x and s however, since we must estimate
μ, k=1 and v = n -1. The confidence interval for the mean with £ unknown is expressed as follows.
$x \pm t_{n-1}\ s/\sqrt{n}$
$P(x - t_{n-1}\ s/\sqrt{n} \leq \mu \leq x + t_{n-1}\ s/\sqrt{n}) = 1-\alpha$
Where 1-α level of confidence (**Ewurum, 2003**).

As long as the sample size is not too small and the population is not very skewed the t distribution can be used in estimating the mean when £ is unknown. The critical value of the t- distribution for the appropriate degree of freedom (**Ewurum, 2003).**

For example, with 24 degree of freedom, if 95% confidence were desired the appropriate value of t would be found in the following manner, the 95% confidence level indicates that there would be an area of 0.025 in each tail of the distribution. Looking in the column for an upper tail area of 0.025 and in the row corresponding to 24 degree of freedom result in value of t of 2.064. Since t is a symmetric distribution. He the upper tail value is +2.064 the value for the lower tail (lower 0.025) would be -2.064. A t value of 2.064 means that the probability that t would exceed + 2.064 is 0.025



Given, x=75, N=10,000, s=20, n=25 and t=2.064

Thus x ±tn-1 s/√n 75 ± (2.064) 20 / √ 25

=75 ± 8.256

$66.74 \leq \mu \leq 84.286$ (**Ewurum**, **2003**)

IV. Estimation For Population

Confidence Interval Estimate for The Population Total T.

An important property of the mean is that it man be used to estimate a **TOTAL AMOUNT** in a population when appropriate.

This is estimated from the following equation

Total = Nx

Where N = population size x = sample arithmetic mean. The confidence interval estimate for the population total T is x +tn -1 Ns /

Confidence Interval Estimation For The Proportion

Estimating population proportion from sample statistics follows the same pattern as outline for means. The major differences bring that the binomial distribution is involved. Statistical inference based on the binomial distribution involves complex technical difficulties caused by the discreteness of the distribution and the asymmetry do confidence intervals (**Lucey , 2003**)

Fortunately when 'n' is large and np and n (1-p) are over 5 then the binomial distribution can be approximated by the normal distribution. This greatly simplifies the analysis and the concepts outlined for the mean can be applied directly tn the proportion p from the sample proportion Ps we could set up the following confidence interval estimate for the population proportion p.

Ps ± Z √Ps (1-Ps)/n

P (Ps – Z √Ps (1-Ps)/ n≤ P ≤ Ps+ Z √Ps (1-Ps) / n= 1-α

The same proportion Ps is used as an estimate if P (in P(1-p)/n) since the true value of the P is unknown and infact is the quantity to be estimated.

A sample poll of 100voters chosen fat random from all voters in a given district indicated that 55% of them were in favor of a particular candidate, find (a) 95% (b) 99% and (c) 99.73%. confidence limits for the proportion of all the voters in favor of this candidates.

Solution (a) the 95% confidence limits for the population P are

P±1.96 σ P =P ±1.96 √P (1-P)/n

0.55 ± 1.96 √ (0.53(0.45)/100

=0.55± 0.10, where we have used the sample proportion P to estimate P

(b) the 99% confidence limit for P are

0.55±2.58 √(0.55) (0.45) 100

=0.55 ± 0.13

(c) The 99.73% confidence limits for P are

0.55 ± 3 √ (0.55) 90.45) /100

=0.55 ± 0.15 **(Spiegel and Stephens, 1998)**

Confidence Interval For Difference And Sums

If S1 and S2 are two sample statistics with approximately normal sampling distribution, confidence limit for the difference of the population parameters corresponding to s1 and s2 arwe given by

S1-s2 ±Zc σ S1-S2=s1-s2±Zc$\sqrt{\sigma^{2}_{s}}$1+σ2s2

While confidence limits for the sum of the population parameter

Are s1-s2±Zc σ S1 +s2=s1+s2±Zc$\sqrt{\sigma^{2}}$s1+σ^{2}s2

Provided the sample are independent.

For example, confidence limit for the difference of two population means, in the case where then population are infinite, are given by X1-x2+Zcσ x1-x2
$=x1-x2\pm Zc\sqrt{}\sigma^2/n_1+\sigma^2/n2$
Where x, σ, n1 and x2,σ^2,n are the respective means, standard deviations and sizes of the two samples drawn from the population.
Similarly, confidence limit for the difference of two populations, where the populations are infinit aree given by the
Ps1-Ps2 ± Zcσ Ps1- P2s=P1s-Ps2±Zc√(P1(1+P1/n1+P2(1-P2)/n_2
Where p0s1 and ps2 are the two sample proportions, n1 and n2 are the sizes of the two samples drawn from populations, and p1 and p2 are the proportion in the two population (estimated) by Ps1, ps2 (Ewurum, 2003).
The confidence interval for standard deviation σ of a normally distributed population as estimated from a sample with standard deviation, s are given by s ±Zc σ5 = s + Zc σ/2n.
In computing this confidence limits we use to estimate σ PROBABLE ERROR: the 50% confidence of the population parameters corresponding to a statistics s are given by
S ±0.6745 σs. The quantity 0.6745σ, is know as probable error of the estimate (**Ewurum, 2003**)

The voltage of 50 batteries of the same type has a mean of 18.2v and a standard deviation of 0.5v. find (a) the probable error of the mean
(b) the 50% confidence limits
Solution
(a) Probable error of the mean.
=0.674σx =o.674 σ/√n =0.6745 s/n
=0.6745 s/√n-1=0.6745 0.5/√49=0.048v
Note that if the standard deviation of 0.5v is computed as s, the probable error is 0.675 (0.3/√50)
=0.048 also, so that either estimate can be used if n is large enough.
(b) The 50% confidence limits are
18.2 ±0.048v (**Spiegel Stephen, 2003**)

Sample Size Determination For The Mean

In the business world, the determination of proper sample size is a compiled procedure that is subject to constraint of budget, time and ease of selection, in determining the sample size for estimating the mean those requirements must be kept in mind along with information about the standard deviation. If σ was known the confidence interval estimate for the population mean is obtained from the equation
x±Z □/ □n
recall the equation Z (x-μ) □ x thus we have
(x-μ)=Z □/ □n
The sampling error is equal to the difference between the estimate from our sample x and the parameter to estimate μ. The sampling error can be defined as
e=Z □/n
solving the equation for n, when have $n= Z^2 \sigma^2/e^2$
therefore to determine the sample size, three factors must be known.
1the confidence level desired, Z.
2 the sampling error permitted, e.
3 the standard deviation, σ (**Ewurum, 2003**)

Sampling size determination for a proportion

The method of sample size determination that are utilized in estimating a time proportion that are aare utilized in estimating a time proportion are similar to mthese employed in estimating the mean.
The confidence interval estimate of the true proportion P is obtained **f**rom
Ps +Z√Ps(1-Ps)/n
Recall that since Z+Ps-P/P(1-P)/n we have
Ps=P=ZP√(1-P)/n
The sampling error is equal to the difference between the estimate from the sample Ps and the parameter to be estimated P. The sampling error can be defined as
E=Z√P(1-P)/n
Solving for n we obtained n=Z2 P (1-P)/ez
In determining the sample size for estimating a proportion, three factors are needed

1. The level of confidence, Z
2. The sampling error permitted
3. The estimated true proportion of success, P (Ewurum 2003)

Estimation and ample size determination for finite populations

When sampling without replacement from finit population, finite population correction (FPC) factor searve to reduce the standard error by a factor equal to (N-n/N-1). When estimating population parameters from such samples without replacement the finite population correction factors should be used for developing confidence interval estimate for the mean would become

$x \pm tn-1\ s/\sqrt{n}\sqrt{N-n} / N-1$ (**Ewurum, 2003).**

The confidence interval estimate of the total when sampling without replacement would be

$N\ x \pm tn-1\ Ns/ \sqrt{n}\ \sqrt{N-n}/N-1$

The c onfidence interval estimate of the proportion sampling replacement would be

$Ps \pm Z\ \sqrt{Ps}\ (1-Ps)/n\ \sqrt{N-n}/N-1$

In estimating proportion the sampling would be

$e=Z\ \sqrt{P}(1-P/n\ \sqrt{N-n}/N-1)$

While in estimating means the sampling error would be

$E= Z\sigma/n\sqrt{N-n}/N-1$

In determining the sample size in estimating them mean, we would have from an earlier equation.

$No=Z^2\ \sigma^2/e^2$

Applying the correction factor to the result in

$N=n_o/n +(N-1)N$ (**Ewurum, 2003**)

V. Conclusion

Statistical inference is the process of drawing conclusion about the population. Therefore they involve random sampling which means that every item in the population has an equal chance of being chosen and that there is no bias. (**Ewurum, 2003**).

Estimation is concerned with estimating population parameter from sample statistic where the sample size is relatively large to sampling distribution of mean is a normal distribution. (**Lucey, 2003**)

There are numerous reason why sample are taken and analysed in order to draw conclusions about the whole population. Sampling is cheaper, quiker and is often only feasible method of finding information about the population. (**Lucey, 2003**)

References

[1]. Murray R. Spiegel, and Larry J. Stephens (1998) schaums outline of theory and problems of statistics

[2]. 3rd e.d., Mc Graw-Hill

[3]. Dr. U.J.F. Ewurum (2003) Module on Man 0507

[4]. Analytical techniques in busness.

[5]. T. Lucey (2002) quantitative Techniques-6 Edition Book power/ ELST edition

[6]. Uwaja. J. O.(2004)n Basic Inferential statistic for higher Education. High class quality integrated press.

[7]. Dr. Stephenson O. UnyiMadu (2007) Advanced statistics. Harmony Books.

[8]. Ben Chukwu (2007) Fundamental Business statistics. Horsethrone concept.