

Lecture 6b

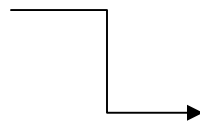
Estimation of Population parameters

Prepared by Grace S. Thomson

INTRODUCTION TO ESTIMATION

We have discussed before how effective is to work with a sample instead of a large population, for both economic and logistic reasons; but once you have your sample, **new questions arise**:

1. **Is the sample Mean equal to the population mean** $\bar{X} = \mu$?
2. If, $\bar{X} \neq \mu$? $(\bar{X} \Rightarrow \mu)$ **how close is sample mean** to the actual population?
3. **Is a sample Mean of size (n) a good estimate** of the population mean?



- Samples are different
- There are many combinations
- Sample mean may be different
- Sample % may be different

4. **Do you need to increase n** to make sample mean closer to population mean?

The **Objective of Sampling** is to gather data that mirrors a population. In this process, we will always deal with a **Sampling error**.

Each time you take a sample out of a population you will obtain a mean of the sample that is normally different from the mean of the population. Most samples taken will have a different mean from the other samples as well as the population. This is why you have sampling error and why the sampling errors are different for different samples. The ramifications of sampling error are that you do not know how close your sample mean is to the true population mean and you do not know whether it is actually lower or higher. This could lead to incorrect decisions.

Sampling error → Difference between Sample value (statistic) and Population value (parameter)

$$\bar{X} - \mu$$

Sample mean: changes depending on the sample we take.

Population mean: always the same, no matter how many times we calculate it

The potential for an extreme sampling error is great when smaller samples are used. The larger the sample, the smaller will be the error.

CENTRAL LIMIT THEOREM

In order to perform estimations of parameters of the population, we will use a Theorem called “**The Central Limit Theorem**”. **This Theorem simply states:**

Any population with μ, σ ; will result in a sample with mean $\mu_{\bar{x}} = \mu$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

When n is sufficiently large to approximate a Normal Distribution. For symmetric distribution a sufficient sample is $n=2$ or 3 . For skewed distributions a sufficient sample must be $n=25$ or 30 .

In the case of Proportions, the Sampling Distribution is expressed by $Z = \frac{\bar{p} - p}{\sigma_{\bar{p}}}$ where

$$\sigma_{\bar{p}} = \sqrt{\frac{p(1-p)}{n}}$$

Why is this information important?

When we say “**Estimation**” we are implying that through the information about samples we will be able to find out information about the population. Therefore, we can always assume normality of any event and immediately we can assume that the sample that we will use is normal, too. This is a very strong assumption in Statistics, but theoretically, that is the grounds of everything performed in this course.

ESTIMATING POPULATION PARAMETERS

There are two ways of estimating the possible mean of the “true” population; through a point estimate or an Interval estimate.

Point Estimate

We assume that the sample mean is a perfect representation of the population mean.

Ex. 62% of people that visit your hospital in a daily basis is 18 and older. (62% is your point estimate).

When you make a point estimate you will have a very high SAMPLING ERROR. To overcome this weakness it is recommended to work with confidence intervals.

Confidence Interval

Instead of coming up with a single number as an estimation of the population, it is recommended to build an interval calculated around this possible point estimate. We assume that at some point within its range the “true” population parameter will be included.

When computing confidence intervals we will need the information about the sample statistic, the confidence level, the standard deviation of the population and the size of the sample. Most of the time we don't have enough information about the population standard deviation; in these cases we make modifications to the formula.

A confidence interval is computed by adding a margin of error to the sample mean:

$$\text{Sample Mean} \pm \text{margin of error}$$

The margin of error is computed $e = Z_{\sigma/2} \frac{\sigma}{\sqrt{n}}$. In real life the margin of error is sometimes given by

the researcher or the decision maker in terms of percentage, e.g. margin of error not greater than 3%, etc.

Confidence Interval estimate for Population Values	σ known	<ul style="list-style-type: none"> Normally distributed with 	$\bar{X} \pm Z_{\sigma/2} \frac{\sigma}{\sqrt{n}}$ Confidence interval
		$\mu, \sigma \rightarrow \mu, \frac{\sigma}{\sqrt{N}}$	
	σ unknown	<ul style="list-style-type: none"> Not normally distributed, but $n > 30$; apply C.L.Theorem and assume: 	$e = Z_{\sigma/2} \frac{\sigma}{\sqrt{n}}$ margin of error
		$\mu, \frac{\sigma}{\sqrt{N}}$	
		Now there are 2 unknown, μ and σ . This is very common!! <ul style="list-style-type: none"> Assume this is a normal distribution if $n \geq 30$ Use t-value instead of Z, Use sample standard deviation (s) instead of σ 	$\bar{X} \pm t_{\sigma/2} \frac{s}{\sqrt{n}}$ Confidence interval $e = t_{\sigma/2} \frac{s}{\sqrt{n}}$ margin of error

Quick tips about confidence levels

Z value

99% confidence	$Z_{\alpha/2}$	=NORMSINV(0.01/2)=	-2.575829304
95% confidence	$Z_{\alpha/2}$	=NORMSINV(0.05/2)=	-1.959963985
90% confidence	$Z_{\alpha/2}$	=NORMSINV(0.10/2)=	-1.644853627

student t

99% confidence	$t_{\alpha/2}$	=TINV(0.01,n-1)	will vary depending on degrees of freedom
95% confidence	$t_{\alpha/2}$	=TINV(0.05,n-1)	will vary depending on degrees of freedom
90% confidence	$t_{\alpha/2}$	=TINV(0.10,n-1)	will vary depending on degrees of freedom

Application of confidence interval concepts to business

The J.R. Simplot Company produces frozen French fries that are then sold to customers such as McDonald's. The "prime" line of fries has an average length of 6.00 inches with a standard deviation of 0.50 inches. To make sure that Simplot continues to meet the quality standard for "prime" fries, they plan to select

a random sample of $n = 100$ fries each day. The quality analysts will compute the mean length for the sample. They want to establish limits on either side of the 6.00 inch mean so that the chance of the sample mean falling within the limits is 0.99. What should these limits be?

R: Within the approximate range of 5.87 inches to 6.13 inches

$$\bar{X} \pm Z_{\sigma/2} \frac{\sigma}{\sqrt{n}} = 6 \pm Z_{(1-0.99)/2} \frac{0.50}{\sqrt{100}} = 6 \pm 2.57 * 0.05 = 6 \pm 0.1285$$

$$= \text{NORMSINV}(0.01/2) = 2.57$$

Upper limit: 6.13

Lower limit: 5.87

Required sample size for estimating population mean

The determination of the optimal sample size to perform a study is one of the most important issues in research and statistics. The decision about the sample size (# of observations) depends on 2 factors: budget and the cost of sampling.

If budget is \$2,000 and the cost of sampling is \$10/observation, the maximum # of observations in my sample size will be 200.

There is always an interrelationship between the sample size, high confidence level and low margin of error.

When Standard Deviation is known, formula for Sample size is:

$$n = \left(\frac{Z_{\sigma/2} \sigma}{e} \right)^2 = \frac{Z^2_{\sigma/2} \sigma^2}{e^2}$$

When Standard Deviation is unknown, formula for sample size is:

We use a pilot sample, taken explicitly to estimate population Standard Deviation σ . Once we obtain n , we will need to subtract the pilot sample size taken previously.

$$n = \left(\frac{Z_{\sigma/2} \sigma}{e} \right)^2 = \frac{Z_{\sigma/2}^2 \sigma^2}{e^2}$$

SOCR Confidence Interval Experiment

http://wiki.stat.ucla.edu/socr/index.php/SOCR_EduMaterials_Activities_CoinfidenceIntervalExperiment

socr.ucla.edu/htmls/SOCR_Experiments.html

Description

This experiment randomly samples from a $N(\mu=0, \sigma^2=1)$ distribution and constructs a confidence interval for the mean (μ). Note how confidence interval widens as the confidence level increases. This is counterbalanced by the sample-size increase. Also, see the effect of the confidence level, $100(1-\alpha)\%$, on the frequency of intervals missing the estimated parameter ($\mu=0$, in this case).

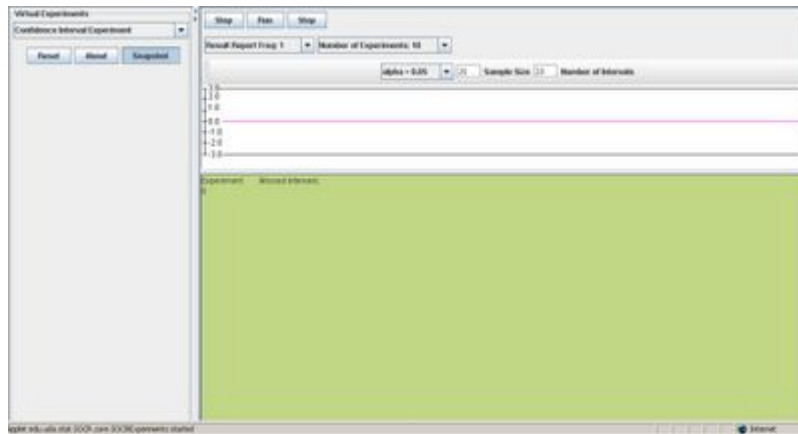
Goal

To provide a simple simulation demonstrating the effects of confidence level and sample-size on the size of the constructed confidence interval, and utilizing this experiment when sampling from almost all symmetric unimodal distributions.

Experiments

Go to the [SOCR Experiments](http://socr.ucla.edu/htmls/SOCR_Experiments.html) (socr.ucla.edu/htmls/SOCR_Experiments.html) and select the **Confidence**

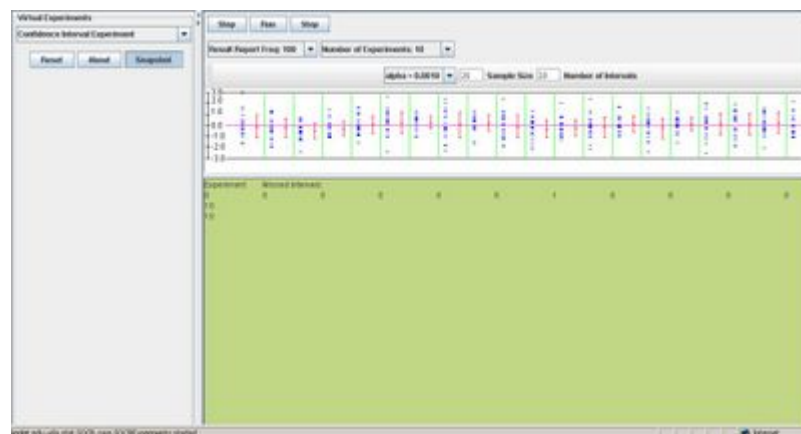
Interval Experiment from the drop-down list of experiments on the top left. The image below shows the initial view of this experiment:



The control buttons on the top allow the user when to begin and end the experiment. “Step” runs one trial, “Run” executes a number of trials depending on the settings of the list boxes below, and “Stop” ends the trials.

The sample size and number of intervals may be modified by the user’s discretion.

Notice that when alpha has a very small value, the results are also small in value and deviation. The image below shows this information:



Applications

The Confidence Interval Experiment may be applied to any situation regarding a population parameter associated with a probability that is generated from a random sample.

For example, adjusting a machine that is to fill cans with soda so that the mean content of the cups is approximately 8 ounces. Since it is not possible for every cup to be filled at exactly 8 ounces, setting the random variable X as the weight of the filling can be applied to this applet. The distribution of X is assumed to be of a normal distribution with an expected mean value and standard deviation.

Suppose engineers want to check for the amount of pollution in the waters of Los Angeles and discover the average amount with standard deviation from the past couple of years. By using this applet, they draw generalizations of the average amount of pollution in Los Angeles with a 95% confidence level.

