Notes 3: Statistical Inference: Sampling, Sampling Distributions
Confidence Intervals, and Hypothesis Testing

1. Purpose of statistical inference

Statistical inference provides a means of generalizing from sample data to population data; from sample findings to population relationships.

2. Population and Samples; Parameters and Statistics

- Population: Target group for our inferences, the group we to which we wish to generalize; each population must have defining characteristics. Target population is that group to which we generalize, accessible population is that group from which sample is actually drawn.
- Sample (n) : A subgroup, subset, or part of the population; the actual group used in the research.
- Parameter: Descriptive measure calculated for the population; usually denoted by Greek letters (e.g., σ, μ, β)
- Statistic: Descriptive measure calculated for the sample; usually denoted by Roman letters (e.g., s, M, $\bar{X}$)
- Estimate: Any statistic calculated for a sample is an estimate of its corresponding parameter; an estimator is some function of the scores that leads to an estimate

3. Randomness, Random Sampling, and Sampling

- Sampling: Process of selecting a subgroup from the defined population.
- Representativeness: The method of sampling is very important in judging the validity of inferences made from the sample to the population. At issue with the validity of inferences is the representativeness of the sample. Note that the size of a sample never compensates for its lack of representativeness of the population.

  * it is very important that sample be representative of the population for statistics to work properly in estimating parameters
  * no sampling procedure guarantees a single representative sample; however, some sampling procedures do guarantee representative samples over the long run--over many samples

Type of samples:

- accidental or convenience samples: these samples do not involve any type of controlled or systematically applied random selection; these samples are frequently biased, and are therefore not representative of the population.
- Probability Sampling: selection in which some form of randomized selection occurs and the chances of groups of individuals can be calculated (thus elements of the population—individuals within population—are known and can be identified [a list of individuals can be formed]). Probability sampling includes simple random samples, stratified samples, and cluster samples. Sometimes systematic samples are included in this list.
- Simple Random Sample, (SRS): a SRS involves the selection of a subgroup from the population in a manner that includes identifying all elements (e.g., people) in the population (e.g., US citizens), assigning unique identification to all elements, and selecting elements based upon some randomized procedure (e.g., using table of random numbers).

  * SRS, in the long run, will produce representative samples of the population
  * if SRS are used, any discrepancy (sampling error) between sample and population is due to chance, not to bias

- <u>sampling error</u>: difference between a statistic and its corresponding parameter for a given sample (i.e., M - μ) with any difference due to chance and not due to systematic influences
- <u>bias</u>: difference between a statistic and its corresponding parameter for a given sample (i.e., M - μ) with difference due systematic influences (e.g., selecting too many females from population due to selection strategy); a statistic is biased if it's expected value does not equal the corresponding population parameter.

  - ∗ if SRS is used, the degree of sampling error (error due to chance) can be estimated, and this allows us to determine accuracy for statistical inference
  - ∗ in short, the use of SRS eliminates systematic error (or bias) from existing in the population
  - ∗ if SRS is used, then independence of observations is certain (in the long run)

- <u>independence</u>: observations in the sample are not correlated; knowledge of one observation will not enable us to predict another

<u>example of non-independence (i.e., dependence)</u>
We wish to determine the average age in a community. We randomly select five people and ask of them their ages and their spouses ages.

True random sampling was not used since spouses were also used in the sample. Spouses were not randomly selected. The problem with this sample is that spouses' ages are not independent—that is, often people marry others with similar ages, so spouses' ages will have some degree of positive correlation. In short, if we know your age, we can probably predict your spouse's age, so spouses' ages are not independent.

<u>example of independence</u>
We wish to determine the average age in a community. We randomly select ten people and ask of them their ages.

With this selection process, we will not be able to predict the next person's age simply by looking at the previously selected person's age—ages are not likely to be dependent.

- <u>Stratified Sampling</u>: similar to SRS except that population is first divided based upon some stratification variable, then SRS is performed within each category of the stratification variable (e.g., divide population of teachers according to the district in which they work, then randomly select 5 teachers from each and every district; in this example district is the stratification variable).
- <u>Cluster Sampling</u>: random selection of groups rather than individuals (e.g., randomly selecting 10 schools from all schools in state and then surveying teachers in those 10 schools selected)
- <u>Systematic Sampling</u>: with this sampling procedure, every element in the population is placed in some order (e.g., on a list), and every n^{th} element is selected after randomly selecting a starting place near the beginning of the list.

  - ∗ systematic sampling, while it can lead to biased results, usually is representative, thus sampling results will be generalizable
  - ∗ when systematic sampling does result in biased samples, it is usually because of patterns in the lists of elements (e.g., listed created or sorted by some variable such as race, sex, IQ)

4. Point and Interval Estimates

- point estimate: a descriptive measure calculated from a sample, such as M, s, etc.
- interval estimate: a range or band within which the parameter is presumed to be included; e.g., the point estimate is M = 5, the interval estimate is 5 ± 2, so the interval would be 3 to 7; so one would expect the parameter μ to be within this interval with a given level of confidence
- interval estimates convey the degree of accuracy of the estimate, therefore it is more informative than the point estimate
- note that the parameter one is trying to estimate does not vary—it is fixed; what varies across samples are the point estimates and the intervals estimates

5. Types of Distributions

- population distribution: distribution of raw scores (Xs) in the population, such as all IQ scores in population of U.S.

Note that when we make inferences from sample to population, what we are really doing in inferential statistics is making inferences about population descriptive characteristics, parameters (e.g., μ or σ), based upon sample summary characteristics, statistics (e.g., M or s).

- sample distribution: distribution of raw scores in the sample, such as IQ scores in our sample of students, Xs.
- sampling distribution: distribution of all possible values of a statistic, such as the mean, M, median, mode, SD, variance, etc..

How is a sampling distribution created? If one were to randomly select a sample, of size n, from the population over and over again, we would get a slightly different M for each sample; these different Ms form a frequency distribution which is called the sampling distribution of M.

Note the emphasis on sample size of n; this means that the samples must all be of the same size when constructing the sampling distribution; if the sample size changes, the sampling distribution will also change.

All statistics (e.g., $\overline{X}$, s, s^2) have sampling distributions.

6. The Sampling Distribution of M ($\overline{X}$):
As noted above, all statistics (e.g., mean, variance, standard deviation, correlation) have sampling distributions. The most commonly used, however, is the sampling distribution of the mean, M (or $\overline{X}$).

If the population distribution of Xs is normally distributed with a mean of μ and variance of σ^2, then the sampling distribution of M will be normally distributed with a mean of μ and a variance of σ^2/n.

That the sampling distribution of M is normal in shape will play an important role in inferential tests to be discussed later. The shape of the sampling distribution is important because it will allow us to find probabilities based upon the distribution, just like finding probabilities for z scores.

Some common properties of the sampling distribution of $\overline{X}$ are:

(a) The mean of the sampling distribution of $\overline{X}$, which may be denoted as $\overline{X}_{\overline{X}}$, is μ. This is the same as saying the expected value of $\overline{X}$ is μ:

$E(\overline{X}) = \overline{X}_{\overline{X}} = \mu$

In short, the mean of the sample means equals the population mean.

(b) As just noted above, the shape of the sampling distribution of $\overline{X}$ is normal <u>if</u> the population distribution is normal in shape. If the population distribution is not normal, then $\overline{X}$ will not be normal, although it will approach normality rapidly as n increases.

(c) The variance of the sampling distribution of $\overline{X}$ is

$\sigma^2_{\overline{X}} = \sigma^2/n$.

and this variance is referred to as the <u>variance error of the mean</u>.

(d) The <u>standard deviation</u> of the sampling distribution of $\overline{X}$ is calculated as

$$\sigma_{\overline{X}} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}}$$

This is called the <u>standard error of the mean</u> and is symbolized $\sigma_{\overline{X}}$.

Note that $\sigma_{\overline{X}}$ is just the standard deviation of the means, $\overline{X}s$. As the sample size increases (n ↑), the standard error of the mean will become smaller, which means that we are getting more precise information about the population from the sample.

For example, if a SRS was taken of IQ scores with μ = 100 and σ = 15, and a sample size of

(a) n = 9, then $\sigma_{\overline{X}} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{9}} = 5.00$

(b) n = 16, then $\sigma_{\overline{X}} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{16}} = 3.75$

(c) n = 25, then $\sigma_{\overline{X}} = \sqrt{\sigma^2/n} = \frac{\sigma}{\sqrt{n}} = \frac{15}{\sqrt{25}} = 3.00$

So as sample size increases, the standard error of the mean will decrease.

<u>An Additional Example Calculation of</u> $\sigma_{\overline{X}}$:
Suppose we have the following SAT scores: 450, 375, 550, 675, 300, 330. We know that SAT is suppose to have a population standard deviation, σ, equal to 100. The mean for the sample, $\overline{X}$, is 446.67. What is the standard error of the mean for this data?

The formula is $\frac{\sigma}{\sqrt{n}}$, so the standard error of the mean is

$$\sigma_{\overline{X}} = \frac{100}{\sqrt{6}} = \frac{100}{2.45} = 40.82.$$

Note that averages (e.g., means) have less variability than individual observations/scores. Thus, sampling distributions will be "tighter" or demonstrate less spread than will their corresponding population distributions.

7. Central Limit Theorem

As explained earlier, the sampling distribution of $\overline{X}$ will be normally distributed only if the population distribution of Xs is normally distributed.

It is very important that the sampling distribution of the $\overline{X}$ s be normally distributed so we can calculate probabilities of the $\overline{X}$ for inferential purposes.

Fortunately, the central limit theorem is invoked for population distributions that are not normally distributed.

The central limit theorem states:

> If one is sampling independently from a population that has mean of μ and variance of σ^2, then as the sample size n approaches infinity, the sampling distribution of the sample mean $\overline{X}$ approaches normality without regard to the shape of the sampled population.

In practice, this means that the larger the sample size used, the better the sampling distribution of M will approximate a normal distribution. Note, however, that the Central Limit Theorem applies only the $\overline{X}$.

(Question: Which is it, the number of samples or the size of the sample which enables us to say that the sampling distribution of $\overline{X}$ approaches normality?)

An illustration of the central limit theorem appears in the document entitled "Central Limit Theorem Illustrated":

http://www.bwgriffin.com/gsu/courses/edur8131/content/EDUR_8131_CLT_illustrated_one_page.pdf

Why is it important that the sampling distribution of the mean be normally distributed? Later we will see that z scores can be calculated for sample means, and these z scores will aid in making inferential statements.

8. Confidence Intervals(CI)

Recall the discussion of interval estimates. CIs represent the method used to develop interval estimates.

Since M is only a point estimate of μ, we know that M varies from sample to sample. The CI will give us a better look at the possible values that could represent μ.

Some logic:

(a) From studying z scores we know that about 95% of all scores in a normal distribution lie between ± 1.96 standard deviations from the mean.

If, due to the central limit theorem, we know that the sampling distribution of $\overline{X}$ is approximately normal in shape, then we know that the probability is about .95 that $\overline{X}$ will be within

$\pm 1.96\ \sigma_{\overline{X}}$

of the population mean.

(b) To say that $\overline{X}$ is within $\pm 1.96\ \sigma_{\overline{X}}$ of μ is that same as to say μ is within $\pm 1.96\ \sigma_{\overline{X}}$ of $\overline{X}$.

(c) So 95% of all samples will capture the true μ in the interval of

$\overline{X} - 1.96\ \sigma_{\overline{X}}$ to $\overline{X} + 1.96\ \sigma_{\overline{X}}$

For example, suppose we randomly select n = 25 grade 8 students and collect their IQ scores with $\overline{X}$ = 103. Since, in theory, σ = 15 for the population distribution of IQ scores, then the standard error for this mean is

$$\sigma_{\overline{X}} = \frac{15}{\sqrt{25}} = \frac{15}{5} = 3.$$

We can anticipate 95% of all such selected samples will produce a .95 confidence interval that will capture the population value of μ. For this particular example, we are 95% confident that μ lies within the interval

$\overline{X} \pm 1.96\ \sigma_{\overline{X}}$,

$\overline{X} - 1.96\ \sigma_{\overline{X}}$ to $\overline{X} + 1.96\ \sigma_{\overline{X}}$

103 – 1.96 (3) to 103 + 1.96 (3)

103 – 5.88 to 103 + 5.88

97.12 to 108.88.

We are 95% confident that the “unknown” population mean of IQ for all grade 8 students lies between 97.12 and 108.88.

Only 5% of randomly selected samples will not provide a .95 confidence interval that captures μ.

In constructing the CI, we need also to consider the degree of accuracy of our interval estimate. The most common degrees of accuracy used in interval estimation are .95 and .99.

As illustrated above, a .95 CI is calculated as:

.95CI = $\overline{X} \pm \mathbf{1.96}\,\sigma_{\overline{X}}$

As a second example, the SAT scores presented earlier demonstrated $\overline{X}$ = 446.67 with σ = 100. The standard error of the mean was $\sigma_{\overline{X}}$ = 40.82, so the 95% CI is

$.95CI = \overline{X} \pm 1.96\,\sigma_{\overline{X}}$

$.95CI = 446.67 \pm (1.96)(40.82)$

$.95CI = 446.67 \pm 80.01$

The limits are:

lower limit (LL) is 446.67- 80.01 = 366.66
upper limit (UL) is 446.67 + 80.01 = 526.68

This CI can be interpreted as follows:

> We are 95% confident that the true population mean lies somewhere within the interval of 366.66 to 526.68.

A 99% CI can be calculated using the following formula:

.99CI = $\overline{X} \pm \mathbf{2.576}\,\sigma_{\overline{X}}$ **(or, to keep it simpler, round ± 2.576 to ± 2.58)**

Confirm that a value of ± 2.58 produces approximately 99% of observations about the mean in a normal distribution using the Z table.

What would be the corresponding value if we sought a 90% CI?

The CI constructed above only hold when σ is known. When the population SD is not known, other formulas, presented later, can be used.

- it is important to note that CIs vary from sample to sample; what does not vary is the parameter
- if .95CIs were constructed for 100 SRS, then on average the population parameter would lie within 95 of the CIs, and would not be within 5 of the CIs

9. Margin of Error

All CI are constructed from three components:

(a) An estimate of a population parameter (such as $\bar{X}$)
(b) Standard error of the estimate (e.g., $\sigma_{\bar{X}}$)
(c) And a test statistics (e.g., z = 1.96), together, then we have this example:

$\bar{X} \pm 1.96\ \sigma_{\bar{X}}$

where (1.96 $\sigma_{\bar{X}}$) is known as the margin of error.

If we believe the margin of error is too large, thus providing a large CI, we can reduce it by

- using a lower level of confidence (say from .99 to .95 reduces the test statistics from 2.576 to 1.96)
- reduce σ (which is unlikely to be something we can adjust easily)
- increase sample size (larger n will reduce the standard error $\sigma_{\bar{X}}$)

10. Properties of Estimators

(a) Unbiasedness

An estimator of a parameter is said to be unbiased if the expected value of the estimator equals the parameter, i.e.,

$E(\bar{X}) = \bar{X}_{\bar{X}} = \mu$

In other words, if the mean of the sampling distribution of the statistic equals the parameter, then the statistic is unbiased.

The term unbiased applies to all statistics, but not all statistics are unbiased. For example,

$E(s) \neq \sigma$ or $\bar{X}_s \neq \sigma$

This means that the expected value of the sample standard deviation (i.e., the mean of the sampling distribution of s) does not equal the population standard deviation, σ.

Positive bias occurs when the expected value of a statistic is greater than the parameter,
i.e., $E(Mdn) > \mu$

Negative bias results when the expected value of a statistic is less than the parameter,
i.e., $E(r) < \rho$

(b) Consistency
An estimator is said to be consistent if, as the sample size increases, the expected value of the statistic approaches (get closer) to the parameter.

As n ↑, $E(s) \rightarrow \sigma$

(c) Relative Efficiency
Efficiency refers to the precision of an estimator; the variability of an estimator from sample to sample; the degree of sampling error associated with an estimator. The variance error and standard error of statistics are measures of efficiency.

Efficiency is usually thought to be more important than either the biasness or consistency of an estimator.

For example, suppose we wish to determine which of two unbiased, equally consistent estimators is better to use, such as M and Mdn for a normally distributed population of IQ scores.

* the standard error of the mean is $\sigma_{\overline{X}} = 2$
* the standard error of the median is $\sigma_{Mdn} = 3$

Since $\sigma_{\overline{X}}$ is smaller than σ_{Mdn} ($\sigma_{\overline{X}} < \sigma_{Mdn}$), the mean is declared to be more efficient than the median for estimating μ.

Note that the assessment of efficiency is relative since it requires that the variance errors of the sampling distributions of two or more statistics be compared.