ESTIMATION

A. Armed with the central limit theorem and the normal distribution, one can obtain the probability that one's ESTIMATE of a population parameter is in error.

1. If sample means (i.e., $\bar{Y}$s) are normally distributed about a population parameter, $\mu$, then we know that 95% of all $\bar{Y}$s take values on the interval, $\mu \pm 1.96 \frac{\sigma_Y}{\sqrt{n}}$.

2. What we have in practice, however, are not $\mu$ and $\sigma_Y$, but $\bar{Y}$ and $\hat{\sigma}_Y$ which are used as point estimates of $\mu$ and $\sigma_Y$.

3. In estimating these parameters we want estimates that are both UNBIASED (centered around the parameter) and EFFICIENT (having a small degree of sampling error relative to other estimators).

For example, $\bar{Y} + 10$ is a biased estimator of $\mu$ (although it is an unbiased estimator of $\mu + 10$). Also $Y_1$ is not as efficient as $\bar{Y}$ in estimating $\mu$, because the variance of $Y_1$ is $\sigma_Y^2$, whereas the variance of $\bar{Y}$ is $\frac{\sigma_Y^2}{n}$.

4. As you might expect, $\bar{Y}$ is an efficient, unbiased estimator of $\mu$.

However, the sample variance, $s_Y^2$, is NOT an UNbiased estimator of $\sigma_Y^2$:

To demonstrate this, it must first be acknowledged that $\frac{1}{n} \sum (Y_i - \mu)^2$ is an unbiased estimator of $\sigma_Y^2$. Recall that $\sigma_Y^2$ is the average squared deviation from the population mean. Like averages for other random variables (e.g., $\bar{Y}$ for $Y$), a sample average of squared deviations from $\mu$ is an unbiased estimator of the population average of such squared deviations. By showing that this unbiased estimator (namely,
\( \frac{1}{n} \sum (Y_i - \mu)^2 \) is consistently larger than \( s_Y^2 \), it can be demonstrated that \( s_Y^2 \) is biased as an estimator of \( \sigma_Y^2 \). Thus, the thesis is that for any sample

\[
\frac{1}{n} \sum_{i=1}^{n} (Y_i - \mu)^2 \geq \frac{1}{n} \sum_{i=1}^{n} (Y_i - \bar{Y})^2 = s^2 .
\]

We begin by noting that

\[
\sum_{i=1}^{n} (Y_i - \mu)^2 = \sum [(Y_i - \bar{Y}) + (\bar{Y} - \mu)]^2
\]

\[
= \sum [(Y_i - \bar{Y})^2 + (\bar{Y} - \mu)^2 + 2(Y_i - \bar{Y})(\bar{Y} - \mu)]
\]

\[
= n(Y_i - \bar{Y})^2 + n(\bar{Y} - \mu)^2 + 2n(Y_i - \bar{Y})(\bar{Y} - \mu)
\]

\[
= n(Y_i - \bar{Y})^2 + n(\bar{Y} - \mu)^2
\]

Since \( n(\bar{Y} - \mu)^2 \geq 0 \), the proof is complete.

Thus "\( s_Y^2 \)" consistently UNDERestimates \( \sigma_Y \). We can adjust for this bias by changing the denominator of \( s_Y^2 \) from \( n \) to \( n-1 \). This is why

\[
\hat{\sigma}_Y^2 = \frac{\sum(Y_i - \bar{Y})^2}{n-1}
\]

and not \( s_Y^2 \) is used as the POINT ESTIMATE of \( \sigma_Y^2 \).

(REMEMBER: The "hat" notation indicates that \( \hat{\sigma}_Y^2 \) is an estimator of \( \sigma_Y^2 \).)

5. WARNING: When estimating the variance of \( X \), divide by "n-1"; when estimating the variance of \( \bar{X} \), divide by "n". Regarding the latter warning, recall that according to the central limit theorem,

\[
\sigma_X^2 = \frac{\sigma_Y^2}{n} \quad \text{NOT} \quad \frac{\sigma_Y^2}{n-1} !!!
\]

B. We have already talked about INTERVAL ESTIMATION in which a CONFIDENCE INTERVAL is estimated. But intervals are only meaningful along some
continuum of values. Thus it seems peculiar to speak of constructing an interval around a "mean" for NOMINAL data. You CANNOT, of course, just add up the numbers and divide by "n." (Reminiscent of our earlier discussion on LEVELS OF MEASUREMENT, this would be "like adding apples and oranges.")

Our solution will be to consider each category separately. Taking a variable such as religious affiliation, for example, a new variable (call it "D") could be constructed such that

$$D = \begin{cases} 
0 & \text{when not Catholic} \\
1 & \text{when Catholic} 
\end{cases}$$

NOTE: Variables like this (namely, that take the value, one, when subjects have an attribute and zero otherwise) are called dummy variables.

C. ESTIMATING PROPORTIONS:

1. Applying the formula for the mean to a dummy variable always produces a number with a value somewhere between zero and one. For example, if the average of the Catholic affiliation variable were .6 , this would mean
that 60% of the Catholics in your sample are Catholic. This follows because the mean of the Ds is

\[ \hat{\pi} = \frac{1}{n} \sum_{i=1}^{n} D_i \]. But since D = 0 when D ≠ Catholic and D = 1 when D = Catholic then

\[ \frac{\# \text{ Catholics}}{n} \] = proportion of Catholics in the sample.

2. Now let's consider the variance of this statistic, \( \hat{\pi} \). You will recall that the peg-board illustration (used when we introduced the concept of normality) was a demonstration of the sampling distribution of a proportion. Applying the lesson from that illustration to this dummy variable, we know that if one were to draw a lot of random samples of "n" residents from the city of Antwerp, one would expect that their respective Catholic-proportions would be distributed normally (for large n) around the true proportion of Catholics in the population.

To be able to make probabilistic inferences about the proportion of Catholics that we find in our sample, we must be more specific about the exact nature of the sampling distribution of these sample proportions:

a. Applying the formula for the variance to the dummy variable, D:

\[ \hat{\sigma}_D^2 = \frac{1}{n-1} \left( \sum_{i=1}^{n} (D_i - \hat{\pi})^2 \right) = \frac{1}{n-1} \left[ \sum (D_i^2 - 2\hat{\pi}D_i + \hat{\pi}^2) \right] \]

\[ = \frac{1}{n-1} \left[ \Sigma D_i^2 + \Sigma ( -2\hat{\pi}D_i ) + n\hat{\pi}^2 \right] \]

\[ = \frac{1}{n-1} \left[ \Sigma D_i^2 - 2\hat{\pi} \Sigma D_i + n\hat{\pi}^2 \right] \]

\[ = \frac{n}{n-1} \left( \frac{\Sigma D_i^2}{n} - 2\hat{\pi} \frac{\Sigma D_i}{n} + n\hat{\pi}^2 \right) \]
Noting that when \( D = 0 \) or \( 1 \), \( D = D^2 \),

\[
\hat{\sigma}_D^2 = \frac{n}{n-1} \left( \frac{\sum D_i^2}{n} - 2\hat{\pi} \frac{\sum D_i}{n} + \frac{n\hat{\pi}^2}{n} \right).
\]

And since \( \hat{\pi} = \frac{\sum D_i}{n} \),

\[
\hat{\sigma}_D^2 = \frac{n}{n-1} \left( \hat{\pi} - 2\hat{\pi}^2 + \hat{\pi}^2 \right) = \frac{n}{n-1} \left( \hat{\pi} - \hat{\pi}^2 \right)
\]

\[
= \frac{n}{n-1} \hat{\pi}(1 - \hat{\pi}).
\]

Finally, note that when the sample size is large, the ratio

\[
\frac{n}{n-1},
\]

can (for all practical purposes) be considered equal to 1.

Thus \( \hat{\sigma}_D^2 = \hat{\pi}(1 - \hat{\pi}) \) for large samples.

b. But this is the large-sample estimate of the variance of a dummy variable, not of a proportion. The central limit theorem is applied to \( \hat{\pi} \) nearly identically as it was applied to \( \bar{X} \). That is, the variance of a proportion equals the variance of its associated dummy variable divided by the sample size. In other words, the variance of a proportion is estimated as follows:

\[
\hat{\sigma}_\pi^2 = \frac{\hat{\sigma}_D^2}{n} = \frac{\hat{\pi}(1 - \hat{\pi})}{n}
\]

Accordingly, \( \hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n}) \) is what is called the "normal approximation of the binomial distribution." We shall return to the binomial distribution shortly.
c. One more comment about estimating proportions:  

The following is a three-part "rule of thumb" for deciding when your sample is "large enough" for you to assume that

\[ \hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n}) \] :

1) Take whichever is SMALLER of \( \hat{\pi} \) and \( 1 - \hat{\pi} \) and

2) multiply it by \( n \).

3) If this number is greater than five, you may assume that \( \hat{\pi} \) is normally distributed.

D. Sample size required in estimating proportions

1. CBS/N.Y. Times telephone pollers claim to have estimates of public opinion that are in error by only ± 3% points. Consider the question, How big a sample would you need to estimate the proportion of Americans that would answer "Yes" to the question, "Do you agree with the President's economic policies?"

2. Put differently, you want your estimate of the proportion to have a PRECISION (i.e., the substantive error one is willing to tolerate) of ± 3%! NOTE that saying that you want your estimate to be within 3% of the true amount SAYS NOTHING about the significance level (or \( \alpha \)).

a. We know that a 100(1-\( \alpha \))% confidence interval around \( \hat{\pi} \) would be

---

b. It is the last part of this expression that is the precision that the CBS/New York Times pollers set at 3 percent. In accordance with my earlier discussion of the Big Picture (Lecture Notes, p. 36), the Greek capital letter delta, Δ, refers to precision. Note that precision is one-half the width of a confidence interval. That is,

\[ \text{PRECISION} = \Delta = Z_{\alpha/2} \sqrt{\frac{\hat{\pi}(1 - \hat{\pi})}{n}}. \]

REMEMBER: The larger your precision is, the less precise (i.e., the more open to deviation from the true value) your estimate will be!

c. After selecting a precision of \( \Delta = 3\% \) and a significance level of, let's say, \( \alpha = .05 \), this formula requires that we also have a value for \( \hat{\pi} \) to determine how large a sample is needed.

The most conservative choice when choosing a value for \( \hat{\pi} \) is a value that yields the largest value of \( \hat{\pi}(1 - \hat{\pi}) \). This allows one to estimate the largest possible confidence interval (and, equivalently, the largest possible precision) that could result once one's data are collected. It is easy to verify that this most conservative choice is \( \hat{\pi} = .50 \). One need merely consider, for example, the values of \( \hat{\pi}(1 - \hat{\pi}) \) as \( \hat{\pi} \) ranges from 0 to 1:
Thus, we find that \( .03 = 1.96 \sqrt{\frac{.50(1-.50)}{n}} \), which can be rewritten as \( n = .25 \left( \frac{1.96}{.03} \right)^2 = 1067.11 \), or 1068.

Thus, \( n = 1068 \) is the size of CBS/NYT samples!

THREE COMMENTS:

1) As always, round up to whole numbers in estimating sample sizes for a given level of precision. (I.e., no half people, please.) By rounding down, a larger (i.e., less precise) precision is attained than the one that is sought. (A REMINDER: In all other cases, round to the nearest decimal value as per p. 31 of these Lecture Notes.)

2) Notice that this number (i.e., 1068) is the same NO MATTER HOW LARGE THE POPULATION!!! That is, to obtain a precision of \( \Delta = 3\% \) (at \( \alpha = .05 \)), you would need to sample 1068 people no matter if they were drawn from a town of 20,000 people or from all people in the world.
a) Of course, the reason why few surveys of the world are done is due to the difficulty in obtaining a list of all humans from which such a sample could be drawn.

b) Although a larger sample is not required for large populations, a smaller sample is allowed when one's sample size is more than one tenth the size of one's population. (For more details, see "finite population correction" in Agresti and Finlay [1986, p. 87].)

3) A few words on estimating $\sigma_D^2$ (or, when to assume $\sigma_D^2 = .25$):

a) If you have collected your data, you will be able to calculate $\hat{\pi}$. In such cases use $\hat{\sigma}_D^2 = \hat{\pi}(1 - \hat{\pi})$. (As a point of future reference: You will always use $\hat{\sigma}_D^2 = \hat{\pi}(1 - \hat{\pi})$ when finding confidence intervals.)

b) If you are testing a hypothesis, you will have decided on a "hypothesized value" of $\pi$ (call it $\pi_0$). In such cases, use $\hat{\sigma}_D^2 = \pi_0(1 - \pi_0)$. NOTE: We shall talk a lot more about $\pi_0$ when we discuss hypothesis testing. (ALSO, as a point of future reference: You will always use $\hat{\sigma}_D^2 = \pi_0(1 - \pi_0)$ in calculating p-values and Type II errors.)

c) When you have neither $\hat{\pi}$ nor $\pi_0$ (as is often the case when choosing n), use the most conservative estimate of $\sigma_D^2$, namely $\hat{\sigma}_D^2 = .25$.

d) When you have both $\hat{\pi}$ and $\pi_0$ (as often happens when you are writing up your findings), use the more conservative estimate of $\hat{\sigma}_D^2$, namely whichever is the larger of $\hat{\sigma}_D^2 = \hat{\pi}(1 - \hat{\pi})$ or
\[ \frac{\hat{\sigma}_D^2}{\sigma_D^2} = \pi_0 (1 - \pi_0) \]. (WARNING: As mentioned in "a" and "b" above, there are exceptions to this rule when finding confidence intervals, p-values, and Type II errors.)

E. You may recall from the second section of these lecture notes that neither the Central Limit Theorem nor the Big Picture holds for estimates based on small samples. The binomial distribution is a case in point. Let us consider what to do if we find that the sample is NOT large enough to assume that

\[ \hat{\pi} \sim N(\pi, \frac{\pi(1 - \pi)}{n}) \].

That is, what should be done if we find that \( n \times \hat{\pi} \leq 5 \) or \( n \times (1 - \hat{\pi}) \leq 5 \)?

To answer this, we need to think a little more about the distribution of

\[ D = \begin{cases} 
0 & \text{if not Catholic} \\
1 & \text{if Catholic} 
\end{cases} \]

Notice that this random variable acts much like flipping a coin. Assume that you have a perfectly symmetric coin (i.e., a coin with equal chances of getting a "heads" or a "tails," OR with a probability of getting "heads" of \( \pi = .5 \)). With a sample of size \( n = 2 \) coin tosses, note that the distribution already begins to take on a somewhat normal shape with

\[ \Pr(\hat{\pi} = 0) = .25, \quad \Pr(\hat{\pi} = .5) = .5, \quad \text{and} \quad \Pr(\hat{\pi} = 1) = .25. \]

This probability distribution can be expressed as that of a BINOMIAL RANDOM VARIABLE, \( Y \), by multiplying each \( \hat{\pi} \) value by the sample size. The resulting distribution is then

\[ \Pr( Y = 0 ) = .25, \quad \Pr( Y = 1 ) = .50, \quad \text{and} \quad \Pr( Y = 2 ) = .25. \]

NOTE: The BINOMIAL RANDOM VARIABLE, \( Y \), is the sum of the values taken by
each of \( n \) independently sampled observations on a dummy variable. The sampling distribution of a binomial random variable is called the BINOMIAL PROBABILITY DISTRIBUTION. The general form of this binomial distribution is as follows:

\[
\Pr(Y = y) = \binom{n}{y} \pi^y (1 - \pi)^{n-y}, \quad \text{where} \quad \binom{n}{y} = \frac{n!}{y!(n-y)!}
\]

No. Placing an exclamation point after a number does not indicate the statistician’s excitement about that particular number. When an exclamation point is placed after an integer, this refers to a “factorial.” It means that you multiply that number by all integers between it and zero. For example, "five factorial" (5!) is calculated as \( 5 \times 4 \times 3 \times 2 \times 1 = 120 \).

Moreover, by definition, \( 0! = 1 \).

Thus \( \Pr(Y = 0) = \binom{2}{0} \pi^0 (1 - \pi)^{2-0} = \frac{2!}{0!2!} (.5)^0 (1 - .5)^2 = (.5)^2 = .25 \).

This seems a lot of trouble until you consider \( \Pr(Y < 3) \) when \( \pi = .8 \) and \( n = 32 \)!

Let’s see how we would calculate this:

\[
\begin{align*}
\Pr(Y = 0) &= \binom{32}{0} (.8)^0 (.2)^{32-0} \\
\Pr(Y = 1) &= \binom{32}{1} (.8)^1 (.2)^{32-1} \\
\Pr(Y = 2) &= \binom{32}{2} (.8)^2 (.2)^{32-2}
\end{align*}
\]

After calculating these numbers, \( \Pr(Y < 3) \) equals their sum.

CONCLUSION: When your sample size is too small for you to assume that the sampling distribution of \( \hat{\pi} \) is normally distributed, you should use the binomial distribution instead of the standard normal distribution to
determine if your results are significantly different from what you would assume would happen by chance alone. Moreover, note that although the binomial distribution may be computationally impractical for very large samples, it ALWAYS APPLIES for proportions. In this last example, for instance, \( r \times n = .8 \times 32 = 25.6 > 5 \) and \( (1-r) \times n = .2 \times 32 = 6.5 > 5 \), allowing one to use either the binomial distribution or its normal approximation in drawing inferences about population proportions.

F. t-DISTRIBUTION: If you are making an interval estimate of the mean of an interval- or ratio-level variable and if your sample size is small (say, less than about \( n = 30 \)), you will need to use the t-distribution (see Table B) instead of the standard normal distribution.

1. All that is being said here is that (for the purposes hand-calculated problems during this course) when \( n > 30 \), you should use your standard normal table in estimating probabilities associated with sample means. When \( n \leq 30 \), use the t-distribution table. (Statistical software like SPSS will always use the t-distribution internally.)

2. There is an important restriction when it comes to using the t-distribution, however. If in one's population the distribution of a variable is not normal, it is legitimate (given the central limit theorem) to assume that the sampling distribution of one's statistic is normal. In contrast, when your sample is small it is NOT legitimate to assume that the sampling distribution of your statistic is that of a t-distribution, unless you have reason to believe that your statistic's underlying variable is normally distributed in the population.

3. Notice that the table for the t-distribution is organized similarly to the chi-square table, with degrees of freedom down the left column and
probabilities at the top of the columns. Like the chi-square table, the
t-table presents values on a statistic (here, t-scores instead of chi-
square scores) for a large number of distributions—each distribution
defined according to the statistic's degrees of freedom. For example,
when estimating a confidence interval for a mean, the t-distribution
with n-1 degrees of freedom should be used. (NOTE: The shape of the t-
distribution gets "flatter," the fewer its degrees of freedom.)

4. AN EXAMPLE: Imagine that a government official wishes to appropriate
funds for flood relief. As part of his planning, he wishes to estimate
(at the .10 level of significance) how many of the people who reside on
the Mississippi flood plain move off of the flood plain within a year of
the river’s flooding. He only has data for the last 30 years and during
these years only six floods have occurred. He assumes his six measures
of moved residents to have been drawn from a normally distributed
population of resident-moves during past floods. Thus, his goal is to
find an interval estimate of the number of people who move after
Mississippi floods. His data are

People who moved: 21, 31, 19, 38, 60, 47.

Clearly, this sample is smaller than n = 30. Note that if we use the
normal distribution, the 90% confidence interval is

\[ \bar{X} \pm Z_{a/2} \sqrt{\frac{\hat{\sigma}^2}{n}} \]

which since \( \bar{X} = 36 \) and \( \hat{\sigma}^2 = 248 \) \([\hat{\sigma} = 15.75]\)

yields a 90% confidence interval of \( 36 \pm 1.645 \times \frac{15.75}{\sqrt{6}} \) or an interval
between 25.4 and 46.6.
However, since \( n = 6 < 30 \), we must use \( t_{\alpha/2} \) and not \( Z_{\alpha/2} \) in this expression. Referring to Table B we find that \( t_{5,.05} = 2.015 \). Thus our 90% confidence interval becomes

\[
36 \pm 2.015 \times \frac{15.75}{\sqrt{6}} \quad \text{or an interval between 23.0 and 49.0 .}
\]

NOTE how in comparison to the z-score, the t-score leads you to make a more conservative (i.e., a 'wider') interval estimate.

5. Now let's imagine that you could obtain data on an additional fourteen Mississippi floods if you were to go to the considerable extra time and expense of obtaining data from local newspapers on floods that took place during the previous century. Making use of the data from the six cases that you do have, you assume that the population mean is in fact 36 flood victims, and that the population variance equals 248 (squared victims).

Here's a question for you: "How far from (the assumed population mean of) 36 would the mean from a random sample of 20 floods have to be for it to have a probability of less than .05 of having been sampled from this assumed population of Mississippi River floods?" Because "how far" requires that we consider deviations both larger and smaller than 36, the answer to this question (again using Table B) is \( t_{19,.025} \hat{\sigma}_x/\sqrt{n} = 2.093 \sqrt{\frac{248}{20}} = 7.37 \) moved residents "far from" 36.

6. What a complicated way to speak of probabilities! The language of HYPOTHESIS TESTING was developed to simplify probabilistic statements such as this.