DEPARTMENT OF POLITICAL SCIENCE
AND
INTERNATIONAL RELATIONS

Posc/Uapp 816

WINDOWS, MINITAB, AND INFERENCE

I. AGENDA:
- A. Small sample means test with t-distribution
- B. Confidence intervals
- C. Two-sample means tests
- D. Reading:
  1. Agresti and Finlay, ***Statistical Methods for the Social Sciences***, chapters 5 and 6 as needed.

II. SIMPLE EXAMPLE REVISITED:
- A. We are interested in knowing whether there has been a decrease in infant deaths per 1,000 live births from 1988 to 1992.
  1. If so we will infer that welfare programs are partly respsonsible.
- B. The way this problem has been set up, which is admittedly a bit simplistic and awkward, is to test the hypothesis $\mu_{1988} > 8.5$, 8.5 being the mean in 1992.
  1. As seen below the null hypothesis is H_0: $\mu_{1988} = 8.5$.
     i. If this holds our data would indicate no decrease in infant mortality.
- C. As we indicated last time we have a very small sample, N = 4 cases.

III. GENERAL STEPS IN TESTING HYPOTHESES:
- A. Covers much of the material from Class 2.
- B. At the outset note that inference procedures such as hypothesis testing involve the possibility of making an error. After all, we are guessing about unknown parameters.
  1. So part of our procedure involves thinking about the costs and probabilities of making various kinds of errors.
- C. Step 1: formulate a research hypothesis.
  1. μ, the infant death rate was higher in 1988 than in 1992 when it was 8.5.
- D. Step 2: translate it into a statistical hypothesis or hypotheses
  1. Null: H_0: $\mu = \mu_{1988} = \mu_{1992} = 8.5$
  2. Alternative hypothesis H_A: $\mu_{1988} > \mu_{1992} = 8.5$
- E. Step 3: choose an appropriate sampling distribution.
  1. Given certain conditions (independent random sampling, relatively small N's, etc.) and the null hypothesis, we expect a test statistic will have a certain distribution. If we know the properties of the distribution we can

use it to gauge the likelihood of sample results.

2. As noted several times in Class 2, a sample distribution is a function that pairs possible sample results with probabilities of occurrence under specified conditions.
   i. A sampling distribution of means tells us how likely we would observe a range of Y given that the population mean is μ and the standard deviation is σ and the sample mean is estimated with a random sample of size N.
3. In this case we are going to estimate the mean on the basis of a sample of 4 cases. Sample means based on small N's are not distributed normally but have a (Student's) t-distribution.
   i. t-distribution is a symmetric roughly bell shaped distribution centered at 0.
   ii. Its exact form is determined partly by the degrees of freedom, which are calculated as ***df*** = N - 1. Thus it is a family of distributions, the shape of each of which depends on N.
   iii. Although a t-distribution looks like a normal distribution, its "tails" (i.e., areas at each end) are somewhat larger.
      (a) This suggests that sample means quite far from the population value have a somewhat greater chance of occurrence than if they were normally distributed.
      (b) The areas under the t-curve or in the t-distribution have been tabulated. See Agresti and Finlay, Table B of the 3rd edition.
   iv. As N gets large (over 50, say) the t-distribution becomes indistinguishable from the standard normal. So if N is large we'll use the z table instead of t.

F. Step 4: Select a "critical region"
   1. Some possible sample results will be deemed so unlikely that should one of them occur we will reject the null hypothesis. Others will be considered possible and will cause us to accept the null hypothesis.
   2. So in essence divide the possible sample outcomes into two mutally exhaustive and exclusive groups: those such that if any one occurs will leads to reject the null hypothesis and those such that if any one occurs will cause us to accept the null hypothesis.
      i. This choice depends partly on how much "error" we are willing to accept.
   3. The appropriate sampling distribution helps us define these regions of acceptance and rejection.

G. Step 4: select "critical values"
   1. Because we do not know the true situation we might make an <u>error</u> in rejecting or accepting the null hypothesis.
      i. Types of inference errors

(a) **Type I error**: falsely rejecting the null hypothesis. The probability of making a type I error is alpha (α) and is called the level of significance. Usually one chooses a level of significance (an alpha level) of .05, .01, or .001.

(b) **Type II error**: falsely accepting the null hypothesis.

ii. The choice of alpha determines the critical values and hence critical region, namely those sample outcomes that will lead to rejection of the null hypothesis. Thus, if a test statistic equals or exceeds a critical value, we reject the null hypothesis.

2. Use a tabulated version of the t-distribution to find the critical value(s) that define the critical region(s).

i. See the attached figure.

ii. The critical t value is denoted t_{crit}

H. Step 5: test statistic:

1. The appropriate test statistic is determined by the data, the nature of the problem, and the assumptions that are made.

2. For this example, we are interested in testing a hypothesis about a mean. The sample size is relatively small (N = 4), and we do not know much about the population (its standard deviation, for example). These considerations suggest using the t distribution.

3. Since we are going to use it, we need to calculate a t-statistic. (After all, a t-distribution gives the distribution of t's.

4. The observed t statistic is

$$t_{obs} = \frac{(\bar{Y} - \mu)}{\hat{\sigma}_{\bar{Y}}}$$

i. The sigma in the denominator, the **estimated standard error** of the mean, has to be estimated.

(a) This is why we use the t-distribution.

(b) If σ, the population standard deviation were known, we could use the z statistic and the standard normal distribution.

ii. The formula for the estimated standard error is

$$\hat{\sigma}_{\bar{Y}} = \frac{\hat{\sigma}}{\sqrt{N}}$$

(a) Here $\hat{\sigma}$ is the sample standard deviation.

iii. In words, find the sample mean, the sample standard deviation, which is divided by the square root of N, and use the formula to

obtain an observed t statistic.

I. Step 6: make a statistical decision
   1. Compare the observed statistic (here the t) with the critical value.
      (a) If it is larger, then reject the null hypothesis; otherwise, do not reject H_0.
   2. In this case compare t_{obs} with t_{crit}: if the absolute of the observed t is equal to or larger than the critical value, reject the null hypothesis; otherwise accept it.

J. Step 7: Interpretation
   1. Ask what the statistical machinations tell you about the substantive issue.
   2. In particular, if the null hypothesis is rejected, estimate the magnitude of the parameter along with confidence intervals and report them.

IV. MINITAB:

A. Let's use MINITAB to do the analysis. The data are stored in a worksheet called "infant-crime."

B. First look up a critical values for the .05 and .01 levels in the t-table. For N - 1 = 3 degrees of freedom the t's are: $t_{crit1} = 2.353$ and $t_{crit2} = 4.541$.

C. Start MINITAB, go to **File**, then click **Open worksheet...**
   1. In the dialogue box find the file, here "infant-crime."
   2. If you have "raw" data, open **Other files** and **Special files** data.
      i. If using the student version, go to **File** and the **Import ascii**

D. Now let's draw a sample of 4 cases.
   1. Go to **Calc**, then **Random data**, then **Sample from columns**.
   2. In the dialogue box enter 4 in the sample window, choose c1 (infant) from the variables, and store the sample in c5.
   3. Press **Ok**.
   4. You can rename c5 to sample.

E. Click **Stat**, then **Basic Statistics**
   1. Choose **1-sample t**
   2. Highlight variable name infant by clicking on it and then clicking **Select**
   3. The check **Test mean**.
   4. In the window type 8.5.
   5. Then select from the **Alternative** box **greater than**.
   6. Here are the results:

```
                Test of mu =  8.50 vs mu >  8.50
Variable     N      Mean    StDev   SE Mean        T          P
Sample       4    12.625    1.195     0.598     6.90     0.0031
```

      i. The sample mean, $\bar{Y}$, (based on 4 cases) is 12.625, with a standard deviation of $\hat{\sigma}$ = **1.195**
      ii. The observed t (denoted **T** in the printout) is 6.90, which exceeds

both critical values, suggests that the null hypothesis should be rejected.

iii. But of course this test involved only four cases.

F. **Important: MINITAB like most programs reports the achieved or attained level of significance.**

1. It is the probability of observing the sample result we have in fact observed **given** or assuming the null hypothesis is true.

i. In this example, we see that the probability of obtaining $t_{obs} = 6.90$ is .0031.

2. You should always report the observed or attained probability.

G. We can of course observe that the 1988 "population" mean for the cities (the one based on all 77 cases) is 12.039 and hence considerably larger than the 1992 average of 8.5.

1. So assuming that this parameter corresponds to the 1992 figure we can see that there has in fact been a decline in infant death rates, but our small sample did not "pick it up."

H. Note that of course all of this has been a numerical exercise since the data used to compute our 1988 sample are not at all comparable to the 1992 data used in calculating the average.

1. Moreover, when you or I draw another independent sample from these data the result of the test might (probably would be quite different).

V. CONFIDENCE INTERVALS:

A. General idea: William Hays, a psychologist and statistician, likens confidence intervals to tossing rings at a fixed post. The post symbolizes an (unknown) population parameter. The rings stand for confidence intervals. Their size represents what we think we know about the parameter: the smaller the ring, the more we know. Large rings mean that we know very little. Now, the goal is to throw the rings at the post in hopes that they will slide down the shaft (that is, cover the post). If the rings have a small diameter, it will be hard to hit the post; if they are large, it will be relatively easy. Thus, there's an inverse relationship between our ability to hit the post and our level of knowledge. For example, tossing a hula hoop would allow us to hit the post almost every time, but it wouldn't be much of a challenge (that is, it wouldn't convey much information). If the rings were much smaller and we managed to hit the shaft, we could feel proud (confident) of our ability, which in this example means we would have considerable knowledge about the post.

1. Agresti and Finlay (page 126 of the 3rd edition of their book) explain confidence intervals this way: "A **confidence interval** for a parameter is a range of numbers within which the parameter is believed to fall."

2. Some intuitive idea of confidence intervals can be seen by referring to the attached figure 1 and figure 2.

B. To illustrate the idea of confidence interval in more detain let's first estimate a

population mean, **μ,** using a **large** sample. The so called "central limit theorem" indicates that the sampling distribution of $\bar{Y}$, the estimator of μ, will (under certain conditions) be normal with $E(\bar{y}) = \mu$ and standard error (deviation) $\sigma_{\bar{Y}}$.

1. Usually, $\sigma_{\bar{Y}}$ has to be estimated by:

$$\hat{\sigma}_{\bar{Y}} = \frac{\hat{\sigma}}{\sqrt{N}}$$

2. Here $\hat{\sigma}$ is the sample standard deviation.

C. Recall that 95 percent of the normal distribution lies within two 1.96 standard deviations of the mean; 99 percent falls within 2.56 standard deviations.

D. Under these (and other) circumstances, we can say that over all samples of size N, the probability is approximately .95 that

$$-1.96\hat{\sigma}_{\bar{Y}} \leq \mu - \bar{Y} \leq 1.96\hat{\sigma}_{\bar{Y}}$$

1. That is, very nearly 95 percent of all possible means calculate from samples from the population with mean **μ** will lie within 1.96 standard deviations of the true mean **μ**,. ($1.96\hat{\sigma}_{\bar{Y}}$ determines the "size" of the ring.)

E. We can restate the inequality by adding $\bar{Y}$ to each term:

$$\bar{Y} - 1.96\hat{\sigma}_{\bar{Y}} \leq \mu \leq \bar{Y} \quad 1.96\hat{\sigma}_{\bar{Y}}$$

1. That is, over all possible samples the probability is about .95 that the range between $\bar{Y} - 1.96\hat{\sigma}_{\bar{Y}}$ and $\bar{Y} \quad 1.96\hat{\sigma}_{\bar{Y}}$ will include the true mean. Stated more, crudely if we took, say, 10,000 independent samples of size N from a population with mean μ and standard deviation **σ**, about 95% of the calculated intervals would include **μ** somewhere between the upper and lower limits of the interval.

F. 99 percent confidence intervals are given by

$$\bar{Y} - 2.58\hat{\sigma}_{\bar{Y}} \leq \mu \leq \bar{Y} \; 2.58\hat{\sigma}_{\bar{Y}}$$

G. In general, (1 - α)% confidence intervals for μ can be constructed by

$$\bar{Y} \pm z\hat{\sigma}_{\bar{Y}} = \bar{Y} \pm z\left(\frac{\hat{\sigma}}{\sqrt{N}}\right)$$

1. z is selected from the standard normal table in such a fashion that we can assert with probability of (1 - α) that the random variable $(\bar{Y} - \mu)/\hat{\sigma}_{\bar{Y}})$ will lie between -z and +z.

H. When estimating means, if we were dealing with small samples and did not know the population standard deviation, which is almost always the case, we would use t's instead of z in the formulas above:

$$\bar{Y} \pm t\hat{\sigma}_{\bar{Y}} = \bar{Y} \pm t\left(\frac{\hat{\sigma}}{\sqrt{N}}\right)$$

I. More generally:

1. Given random samples from a population having a parameter q and assuming certain conditions are met the general form of (1 - a) percent confidence intervals is:

$$(\hat{\ } - \hat{\sigma}) \leq \quad \leq (\hat{\ } \quad \hat{\sigma})$$

2. where is θ the population parameter, $\hat{\ }$ is an estimator of that parameter, $\hat{\sigma}$ is the (estimated) standard error of the appropriate distribution, and ζ, which depends on the desired degree or level of confidence, is chosen so as to define or mark off (1 - α) percent of the appropriate sampling distribution.
   i. See the attached figure.

VI. MINITAB AGAIN:

A. Although the hand calculations for this problem are trivial, let's use MINITAB to construct confidence intervals for our sample mean of infant mortality.
   1. Later we'll do them by hand.

B. Go to **Stat** and **Basic statistics**.

C. Choose **1-sample t**.

D. Make sure the column containing the sample data is selected and check the

confidence interval box.

1. The default is 95 percent, so click **Ok**.

E. The results are:

```
                          Confidence Intervals
Variable     N      Mean    StDev  SE Mean       95.0 % CI
Sample       4    12.625    1.195    0.598  (  10.723,  14.527)
```

1. Interpretation: these intervals have been constructed so that we can be 95 percent confident that the true mean is between 10.723 and 14.527.

F. You can see the effect of increasing and decreasing confidence on the size of the intervals.

1. 90 percent intervals are rather narrow, but of course we can't be too confident that they contain the true value:

```
                          Confidence Intervals
Variable     N      Mean    StDev  SE Mean       90.0 % CI
Sample       4    12.625    1.195    0.598  (  11.218,  14.032)
```

i. Note the interval covers only 14.032 - 11.218 = 2.814 infant deaths.

2. 99 percent intervals are relatively wide:

```
                          Confidence Intervals
Variable     N      Mean    StDev  SE Mean       99.0 % CI
Sample       4    12.625    1.195    0.598  (   9.134,  16.116)
```

i. If we need to be 99 percent confident, then we need wide intervals.

G. There is thus a trade off: more confidence, wider intervals (less "precise" estimates); less confidence, narrower intervals.

VII. DIFFERENCE OF MEANS TEST:

A. Consider this thought experiment: we are interested in comparing California and Alabama in terms of some variable such as "the violent crime rate." Suppose we take two <u>independent</u> samples from the population of counties in each state. (Denote the sample sizes N_{CA} and N_{AL} respectively. That is, we might draw a sample of 10 counties from California and 15 from Alabama. Then, $N_{CA} = 10$ and $N_{AL} = 15$.) After finding out the <u>mean</u> or <u>average</u> rate in the two states, we then calculate:

$$\hat{} = \bar{Y}_{CA} - \bar{Y}_{AL}$$

B. Now suppose we repeat the process by drawing another sample of N_{CA} cases from California and N_{AL} from Alabama and calculate the difference of means on the crime rate. This estimated difference will probably not equal the first one because we have drawn new independent samples.

C. Let's continue in this manner, drawing samples of size 10 and 15 respectively from California and Alabama an infinite number of times. Stated differently, suppose we some how obtain sample differences of means from all possible samples from these two states. We will have a "pile" of $\hat{}$ ***s***.

D. What will be the mean, standard deviation, and shape of this collection?

1. Theory tells us that if we are have small samples and unknown population standard deviations, the appropriate distribution will the t-distribution.
2. Parameters of a sampling distribution.
   i. The mean of the sampling distribution usually equals the expected value of the sample statistic.
      (a) The mean of the difference of means, for example, will equal **Δ**, the population difference of means. In other words,

$$E(\hat{}) = \quad = \mu_{CA} - \mu_{AL}$$

   ii. The standard deviation of the sample statistics, that is the standard deviation of the sampling distribution, is called the standard error. In the case of the difference of means we would denote it $\sigma_{\hat{}}$.
3. The form of the sampling distribution.
   i. If the sample sizes are large enough and other conditions are met, the sampling distribution of a sample statistic($\hat{}$) will be **normal** with mean **Δ** and standard deviation (called the standard error) of $\sigma_{\hat{}}$.
   ii. In the case of the difference of means statistic where the two sample sizes are relatively small, $\hat{}$ will have a so-called t distribution. If the population variances (standard deviations) are the same (e.g., σ_{AR} equals σ_{VA}), then the sampling distribution will be the t distribution with (in this case):

$$df = N_{CA} \quad N_{AL} - 2$$

VIII. NEXT TIME:

A. More on inference including statistical power.

Go to Statistics page.