Confidence Intervals and Tests of Hypothesis: A Minitab Demonstration

1. The Data

The data for this demonstration are taken from Ott: *Introduction to Statistical Methods and Data Analysis, 4th ed.,* (Duxbury), Problem 5.107, page 256. We are told there that the observations are the proportions of patients with a particular kind of insurance coverage at 40 hospitals selected at random from a particular population of hospitals.

The main issue here is to make inferences about the proportion of patients with this kind of insurance in the *population* of hospitals from which our sample of 40 was taken. Although this demonstration can be read without using Minitab, we recommend that you use your browser to print it out and follow through the steps using Minitab as you read.

2. Putting the Data Into a Minitab Worksheet

1.1. If you are working in the CSU Hayward Statistics Lab. Start Minitab, select the Session window, in **Files** open the Minitab worksheet CH05-107.MTW located on the server in I:\COURSWRK\STAT\3502\OTTDAT. Check the worksheet to make sure you have retrieved the correct data.

The data are as follows:

```
0.67    0.74    0.68    0.63    0.91    0.81    0.79    0.73
0.82    0.93    0.92    0.59    0.90    0.75    0.76    0.88
0.85    0.90    0.77    0.51    0.67    0.67    0.92    0.72
0.69    0.73    0.71    0.76    0.84    0.74    0.54    0.79
0.71    0.75    0.70    0.82    0.93    0.83    0.58    0.84
```

1.2. If you do not have access to the server. Open Minitab, go to the Data window, and type the 40 observations into C1 of the worksheet, proofreading carefully. *Alternatively,* once you have opened Minitab, you can switch to your browser and highlight the data, then cut and paste into row 1/ col. 1 of a blank Minitab worksheet (spaces as delimiters), and finally, "stack" all of the data into C1 (menu path: **MANIP > stack**). This

method will not preserve the order of the observations, but their order is unimportant in this situation.

2. Descriptive Statistics

2.1. Numerical Summary. Before doing any formal inference, it is always a good idea to use descriptive methods to understand the data. Here we begin by finding some numerical descriptive statistics.

```
MTB > desc c1

               N      MEAN    MEDIAN    TRMEAN     STDEV    SEMEAN
C1            40    0.7620    0.7550    0.7658    0.1086    0.0172

             MIN       MAX        Q1        Q3
C1        0.5100    0.9300    0.6925    0.8400
```
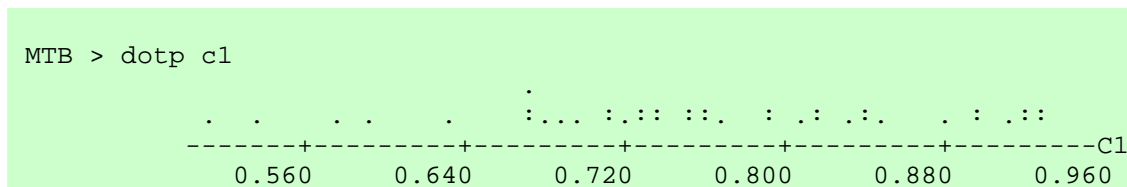
Questions:

- The mean and the median are very nearly equal. What clue does this give you about the possible skewness of the sample?
- The mean is very nearly equal to the trimmed mean (mean of the middle 90% of the observations). What clue does this give you about the possible presence of outliers?
- Make a boxplot of these data (command: `boxp`) and describe what you see. What five numbers from above are needed for making the boxplot?
- The values that will prove crucial for formal inference below are sample size 40, sample mean 0.7620, and the sample standard deviation 0.1086. The (estimated) standard error of the mean is also useful. Show how it is computed from the sample size and the sample standard deviation.

2.2. Dotplot. Next, make a dotplot as shown below. The data appear to be nearly, but perhaps not exactly, normal. In any case they are not severely skewed and there are no outliers. So with a sample size as large as 40, t-procedures for inference are OK.

```
MTB > dotp c1
                                 .
        .   .       .  .       .     :...  :.::  ::.   :  .:  .:.    . :  .::
        -------+---------+---------+---------+---------+---------+---------C1
           0.560     0.640     0.720     0.800     0.880     0.960
```

Questions:

- From the dotplot, try to judge what interval of values of the population mean μ might be believable, considering the inevitable sampling error. Notice that the plot seems to "balance" at about 0.76 (the sample mean).
- In particular, which of the following values are believable values of the population mean μ: 0.64? 0.70? 0.73? 0.75? 0.88?

3. Confidence Intervals

3.1. A 90% Confidence interval with known population standard deviation. Assume that the population standard deviation is known to be σ = 0.1.

In most real-life applications, σ is not known. Here is a scenario in which it would be reasonable to assume that the population standard deviation is known to be σ = 0.1: For several past years a state agency has collected data for all of the hospitals in the population. The population standard deviation for each of these past years was computed and has held steady at about σ = 0.1. Furthermore, we know of no reason that the population dispersion should have changed this year.

A Minitab printout of the required confidence interval is shown below. The command `zint` stands for a confidence interval based on the standard normal distribution, which is often represented by the letter *Z.*

```
MTB > zint 90 .1 c1

THE ASSUMED SIGMA =0.100

              N       MEAN     STDEV   SE MEAN    90.0 PERCENT C.I.
C1           40     0.7620    0.1086    0.0158  (  0.7360,   0.7880)
```

Things to notice.

- The command line (or the menu dialog box) must say what confidence level is desired. (Do not type a percent sign.)
- The command line (or the menu dialog box) must also contain the known value of the population standard deviation.
- Notice that the value for SE MEAN in this output differs from the (estimated) standard error of the

mean shown among the descriptive statistics. Here
SE MEAN is computed using the *known* population
standard deviation σ.

Questions:

- For a standard normal random variable *Z* what cutoff value *c* gives
  $P(Z < c) = 0.05$? (Use tables of the standard normal distribution.
  Alternatively, use the Minitab command invcdf or the Minitab
  menu path **CALC > Probability> Normal**. Preferably, use both
  methods and compare the results.)
- What is the formula for the *margin of error* in this confidence
  interval? (Hint: It involves the cutoff value of the previous
  question and SE MEAN.) Give the numerical value resulting from
  this formula.
- What is the *total length* of the confidence interval just computed?
  How is the total length related to the value of the margin of error?
- How is the margin of error in the previous question used to find the
  endpoints of the 90% confidence interval shown in the printout
  above?
- Recall that we are computing a 90% confidence interval here. Why
  was the cutoff value *c* in the first question of this group calculated
  using 5% instead of 100% – 90% = 10%?

**3.2. A 90% confidence interval for the usual case where the
population standard deviation is NOT known.** Here the confidence
interval is based on the t-distribution with 39 degrees of freedom. The
command name tint reflects the use of the t-distribution. Of course, the
command line does not include the value of the population standard
deviation because it is unknown.

```
MTB > tint 90 c1

            N      MEAN     STDEV   SE MEAN    90.0 PERCENT C.I.
C1         40    0.7620    0.1086    0.0172   (  0.7331,   0.7909)
```

Questions:

- For a random variable *T* distributed according to the t-distribution
  with 39 degrees of freedom, what cutoff value *c* gives $P(T <
  c) = 0.05$? (Use tables of the t-distribution for an approximate
  value. Alternatively, for an exact value, use the Minitab command
  invcdf or the Minitab menu path **CALC > Probability > t**.)
- Why does the value 0.0172 of SE MEAN given just above differ
  from the value 0.0158 given in the previous printout?

- What is the formula for the margin of error in this confidence interval? [Hint: It involves the cutoff value of *T* from above and `SE MEAN`.]
- How is the margin of error in the previous question used to find the endpoints of the 90% t-confidence interval shown in the printout above?

3.3. A 95% confidence interval based on the t-distribution. Because 95% is the *default* confidence level for Minitab, the command *does not require* you to state that the confidence level is 95%. (But it does no damage to include `95` in the command line; try it both ways.) Notice that the increase in confidence level from 90% to 95% increases the margin of error and thus makes the interval longer.

```
MTB > tint c1

              N       MEAN     STDEV   SE MEAN     95.0 PERCENT C.I.
C1           40     0.7620    0.1086    0.0172   (  0.7273,   0.7967)
```

Questions:

- What was the total length of the 90% t-confidence interval computed previously? What is the total length of this 95% confidence interval? Give an *intuitive* explanation why the 95% C.I. must be longer.
- What cutoff value for the t-distribution did Minitab use to compute this 95% confidence interval?
- Suppose you are wondering whether $\mu = 0.74$ is a "believable" value for the population mean. Suppose also that the population standard deviation σ is unknown. What does the 90% confidence interval have to say about this issue? Adjusting your concept of "believable" somewhat, what does the 95% confidence interval you just found have to say?
- Answer the previous question again, but this time you are wondering whether $\mu = 0.73$ is a believable value.

3.4. Which distribution do I use for a confidence interval when estimating the population mean? The correct choice between the the standard normal distribution and the t-distribution is very easily made:

- If the population standard deviation is known, always use the standard normal distribution.
- If the population standard deviation is NOT known, always use the t-distribution.

If the sample size is larger than 30, the cutoff values for 95% confidence intervals are *about* the same regardless of whether the standard normal or the t-distribution is used. (And the same can be said for 90% or 99% confidence intervals.) Before statisticians knew about the t-distribution, it was common to use the standard normal distribution to get *approximate* confidence intervals when σ was unknown. However, the t-distribution always gives more accurate results when σ is unknown. Notice that there is never any doubt which distribution is correct when you are using Minitab: `zint` requires you to input a known value of σ, `tint` does not.

This same distinction, depending on whether or not σ is known also holds for tests of hypothesis which we consider next.

4. Hypothesis Testing

4.1. Tests With Two-Sided Alternatives. Now suppose that comprehensive data from all relevant hospitals last year showed $\mu = 0.73$. Since then there have been many changes in the health insurance industry. Using the random sample of 40 we want to test whether the population mean has changed this year. In this situation:

- The null hypothesis is that $\mu = 0.73$ (no change from last year).
- The alternative hypothesis is that μ is not equal to 0.73 (some change).

Because we have comprehensive data, we could compute last year's population standard deviation σ, but suppose we are concerned that σ may have changed also, and so we are not willing to use last year's σ for our test of hypothesis. Thus we regard σ as unknown.

Notice that the statement of the null hypothesis involves the *population* mean, not the *sample* mean. We already know that the sample mean of our data is 0.7620. Because the sample mean is a good estimate of the population mean, we would not expect the population mean to be "very much" different from 0.7620. The question is whether the sample mean and the hypothetical value 0.73 of the population mean are "significantly different." How big a difference is "significant," that is to say more different than we would expect on the basis of sampling error?

In this situation a one-sample t-test is the appropriate statistical procedure to judge whether the sample mean differs significantly from the hypothetical value 0.73 of the population mean. The results of this test are shown in the Minitab printout below:

```
MTB > ttest .73 c1

TEST OF MU = 0.7300 VS MU N.E. 0.7300

              N      MEAN     STDEV    SE MEAN         T    P VALUE
C1           40    0.7620    0.1086    0.0172      1.86      0.070
```

Some notes on Minitab:

- You must specify the hypothetical value of μ in the command (or in the menu dialog box). Otherwise Minitab doesn't "know" what value you have in mind. Minitab's default hypothetical value is 0, which would make no sense at all in this problem!
- Notice that you do not need to specify a fixed significance level for your test when using Minitab. This is because Minitab prints a `P-VALUE` from which you can tell whether or not to reject the null hypothesis at any significance level of your choosing. (More on this below.)
- Minitab does not print Greek letters; `MU` stands for μ.
- In printouts, some versions of Minitab use `VS` (abbreviation for Latin *versus,* meaning *against*) to indicate that a statement of the alternative hypothesis is coming next.
- In printouts, some versions of Minitab use the abbreviations `N.E.` (not equal), `L.T.` (less than), and `G.T.` (greater than) in specifying alternative hypotheses.

Questions:

- Find the formula for the t-statistic in your text. Plug in the values of the sample size, sample mean, and sample standard deviation from this Minitab printout and verify the value of T given in the printout.
- For a 10% fixed significance level find the *critical values* of the t-distribution with 39 degrees of freedom. Critical values are those that separate the rejection regions in the tails of the distribution from the non-rejection region ("acceptance" region) in the center. At the 10% level do you reject the null hypothesis? At the 5% level?
- Notice that the formula for the test statistic T involves the difference between the sample mean and the hypothetical value of the population mean. But it *also* involves the sample size and a

measure of dispersion; explain intuitively why. [Hint: What if you converted the data from proportions to percents: 67, 74, ...? Then the difference would be 100 times as big, but would the difference be "100 times as convincing"? Returning to proportions, what if you had only 4 observations with a sample mean of 0.762? Would the difference be as convincing as with 40 observations?]

4.2. Using P-values. First, here is how to *interpret* a P-value. Suppose you have a fixed significance level α in mind for your test of hypothesis.

- If the P-value is *less than or equal* to α, then REJECT the null hypothesis.
- If the P-value is *greater than* α, then DO NOT REJECT the null hypothesis.

The smaller the P-value the stronger the evidence *against* the null hypothesis.

> Example: In the above printout the P-value is 0.07: This means that the null hypothesis cannot be rejected at the 5% significance level. However, it would be rejected at the 10% level.

Second, here is how to *compute* a P-value. Assuming the null hypothesis to be true, the P-value is the probability of observing (by chance alone) a more extreme value of the test statistic than the one actually obtained. (Some books use the terminology "observed significance level" to mean the same thing as P-value. This terminology seems not to be used very much any more.)

> Sample computation: In the above printout the test statistic has the value $T = 1.86$. It is distributed according to the t-distribution with 39 df. Values larger than 1.86 or smaller than -1.86 are judged to be "more extreme" (farther from the 0 center of the distribution) than 1.86. Thus, the P-value is

$$P(T > 1.86) + P(T < -1.86) = 2P(T > 1.86) = 0.07.$$

Questions:

- For our data, would the null hypothesis that $\mu = 0.73$ be rejected at the fixed significance level $\alpha = 1\%$?
- Generally speaking, tables of the t-distribution that appear in textbooks are not adequate to find exact P-values, because they provide cutoff values of t-distributions corresponding only to a

very few probabilities (for example perhaps: .1, .05, .025, .01, .005, .001). Furthermore, tables do not provide information for all degrees of freedom (for example, skipping from 30 df to 40 df to 60 df, etc.). Can you figure out how to use the t-tables in your text to "bracket" the P-value for the present test of hypothesis? [Typical answer: 39 df is between 30 df and 40 df. In either case (30 df or 40 df) $T = 1.86$ lies between the one-tailed cutoffs for 5% and 2.5%. Thus, the P-value is between $2(5\%) = 10\%$ and $2(2.5\%) = 5\%$. These values bracket the true P-value which is 7%.]

- Use Minitab's probability calculation capabilities to find the P-value corresponding to $T = 1.86$ for a two-sided test (39 df). [Hint: It is probably easiest to use the menu path **CALC > Probability > t**.]

4.3. Connection of two-sided test with confidence interval. The value $\mu = 0.73$ is contained in the 95% C.I. Thus, 0.73 is a "believable" value of μ with $\alpha = 5\%$. Logically enough, using the same data, the test of hypothesis "accepts" 0.73 as a believable value of μ at the 5% level.

Similar connections exist between 90% confidence intervals and tests with $\alpha = 5\%$, and between 99% confidence intervals and tests with $\alpha = 1\%$.

Caution: Below we consider tests with *one-sided* alternatives. Minitab's confidence intervals for a population mean are fundamentally two-sided, based on the sample mean *plus or minus* a margin of error. The connection stated above between confidence intervals and tests of hypothesis holds for *two-sided alternatives only.* Do not try to use it for tests with one-sided alternatives.

4.4. One-sided alternative. Now suppose we have not seen any data yet. But we know that the insurance carrier in question has conducted a large public relations campaign and has increased coverage while keeping prices about the same. If there has been any change in the choice of this coverage, we are "sure" it must have been an increase.

In this situation, the null hypothesis is still that $\mu = 0.73$, but here the sensible alternative is that $\mu > 0.73$. To perform the "one-sided test" (also called a "one-tailed test") you must select this option in menus or use a subcommand ALTERN 1 (for a left-sided test, ALTERN -1).

```
MTB > ttest .73 c1;
SUBC> altern 1.

TEST OF MU = 0.7300 VS MU G.T. 0.7300

             N      MEAN    STDEV   SE MEAN        T    P VALUE
C1          40    0.7620   0.1086   0.0172      1.86      0.035
```

Notice that the P-value here is half what it was for the two-sided test. Thus, at the 5% level we REJECT the null hypothesis $\mu = 0.73$ against the right-sided alternative. If we are working at a fixed significance level of 5%, the difference between a one-sided and a two-sided alternative makes the makes the difference whether or not to reject the null hypothesis!

In some cases, the choice whether to use a one-sided or a two-sided alternative can be controversial. Sometimes the controversy centers on the reliability of background information, sometimes on the purpose of the experiment, and sometimes on philosophical issues about hypothesis testing. Our view is that two-sided alternatives should be used unless there is very strong reason to support the use of a one-sided alternative. In any case, there is no controversy about the following two statements:

- The choice whether to use a one- or two-sided alternative must always be made *before* the data are collected. After one sees the data, it would be too easy to invent a rationale why the data "had" to fall in the direction they did. Electing a one-sided alternative *after* seeing the data is deceptive and unethical. For example, it may amount to claiming a fixed significance level of only 5%, when the real significance level is 10%.
- If a one-sided alternative is chosen, and the data should turn out (embarrassingly) to be in the opposite direction from the selected alternative, then the null hypothesis cannot be rejected, no matter how extreme the result. For example, when one chooses a right-tailed alternative, one is declaring in advance of data collection that values of the test statistic in the left tail will be attributed to faulty procedure or bad luck.

Questions:

- For the present data, the null hypothesis that $\mu = 0.73$, and a *left-tailed* alternative, what is the P-value? What is your conclusion (at the 5% level)?
- For the present data, the null hypothesis that $\mu = 0.73$, and a right-tailed alternative, and the *assumption* that $\sigma = 0.1$, what is the P-value? What is the conclusion (5% level)? [Hint: The command is `ztest` and the known value of the population standard deviation must be specified on the command line; alternatively, use menus-- or do the computation by hand and use tables of the standard normal distribution.]

Additional Things to Try:

- Is a confidence interval based on the standard normal distribution always shorter than a confidence interval (same confidence level)

based on the t-distribution? [Hint: Remember that the assumed value of σ is part of the computations. What if the sample standard deviation is quite different from the assumed population standard deviation? For these data, what happens to the z-interval if we assume σ = 0.15?]

- Even though our data appear to be nearly normal, some statisticians may prefer to do a "nonparametric" test, which does not depend at all on the assumption that the data are normal. One such test is a *sign test.* Here is how to apply it to the present situation:

    The null hypothesis is that the population *median* is 0.73. Use the two-sided alternative.

    Each observation greater than 0.73 is called a PLUS; each observation smaller than 0.73 is called a MINUS. This test is only for continuous data: if (due to rounding) any observation is recorded as 0.73, it is thrown out and the sample size is reduced accordingly. For our data, how many PLUSes and how many MINUSes are there?

    If the null hypothesis is true, then PLUS and MINUS each have probability 1/2 (due to the definition of the median). Then, the sign test becomes a test of whether the true probability of PLUS is really 1/2. (Just like judging whether a coin is fair.)

    Use the normal approximation to the binomial. Assuming, that the probability of PLUS is 1/2, standardize the observed number of PLUSes. (Remember to use the adjusted sample size when you find the mean and standard deviation of the relevant binomial distribution.)

    If this Z-score is less than -1.96 or greater than +1.96, then reject the null hypothesis against the two-sided alternative at the 5% level of significance.

    What is the P-value of this sign test? Again, use the normal approximation, and compare your result with the Minitab printout from the procedure obtained via the menu path **STAT > Nonparametrics > 1-sample sign.** [Answer: 14 MINUSes, 24 PLUSes, 2 eliminated. The P-value is about 14%; not even significant at the 10% level. Because this test looks only at PLUSes and MINUSes it loses some of the information in the data, and thus is less powerful (less able to detect when the null hypothesis is false).]