

Directions: Complete the exercises below. When you are finished, turn in any required files online in Canvas, then check-in with the Lab TA for dismissal.

Introduction to Confidence Intervals in SAS

The same SAS code that conducts the two-sample t-test in SAS will also provide a corresponding confidence interval. As an example, consider the observational study on texting speeds from last time:

For a school statistics poster competition in 2006, students timed 15 randomly selected teenagers from the school and 15 randomly selected staff from the school over the age of 30 on how long it took each person to text the following sentence on their phone: “the quick brown fox jumps over the lazy dog.” Each subject had the sentence in front of them while they were typing. The text message had to be typed with no errors, no abbreviations, and no use of the phone directory. Time was measured using a stop watch to within 0.01 seconds. Participants were timed using two phones - their own phone and a “control” phone, which was the same for all participants. We would like to determine if teenagers were faster “texters,” on average, than adults. The data are located in the `smsspeed.csv` file and the full SAS program in `smsspeed_Lab3.sas` within our course’s shared folder in SAS Studio.

- First, load in the dataset:

```
data SMS;  
  infile ‘~/my_shared_file_links/u63538023/STAT5000_Fall2024_ISU/  
    smsspeed.csv’ dlm=‘,’ firstobs=2;  
  input Age AgeGroup $ Own Control;  
run;
```

- Then, use the `proc ttest` command to conduct a test and obtain the corresponding confidence interval for the difference in mean speed between the teenagers and adults. Use the `class` option to specify the category variable and the `var` option to specify the response variable.

```
title1 ‘T-test for Difference in Mean Times - Own Phone’;  
proc ttest data=SMS;  
  class AgeGroup;  
  var Own;  
run;
```

You’ll find the corresponding 95% confidence interval in the columns of the output table for 95% CL Mean and then look at the row for Diff (1-2) Pooled.

- You can change the confidence level using the `alpha=` parameter. For example, a 99% confidence interval can be obtained by:

```
proc ttest data=SMS alpha=0.01;  
  class AgeGroup;  
  var Own;  
run;
```

Sample Size Simulations in SAS

In lecture, we looked at an example of a randomized experiment to determine which of two treatments was the most effective at reducing bone loss in elderly women. In this experiment, we will assume equal sample sizes, equal population variances, and normally distributed response variables in both samples. We will also assume an estimate of the pooled sample variance for the response variable is available from previous studies, denoted as S_p^2 . The SAS code to calculate sample sizes is provided in the `power.Lab3.sas` file in the course's shared folder in SAS Studio.

- Suppose our research question is to determine whether or not the two treatment means are different. We will use a hypothesis test with Type I error rate of α and will want the power to detect a difference of δ units between the treatment means to be $1 - \beta$.

In lecture, our example used $\alpha = 0.05$ (`alpha`), $1 - \beta = 0.8$ (`power`), $\delta = 4$ (`meandiff`), and $S_p^2 = 25$ (take the square root and enter as `stddev`). The code that produced the result of 26 subjects in each treatment group (specified using a period for `npergroup`, meaning this is what you want SAS to solve for) is given below.

```
proc power;
  twosamplemeans test=diff
  alpha = 0.05
  meandiff = 4.0
  stddev= 5
  npergroup = .
  power = 0.80;
run;
```

- Then, use SAS to help you determine the effect of changes to the values of α , $1 - \beta$, δ , and S_p^2 on the sample size (n). To make it easier to study these changes, you can modify the SAS code to study the sample size for multiple values of an input value at the same time. For example, to study the effect of increasing power $1 - \beta$, you can change the power command to

```
power = 0.80 to 0.95 by 0.05;
```

or you can list values to study, like

```
power = 0.80, 0.9, 0.95, 0.99;
```

- Instead of the analysis above, suppose our research question is to estimate the difference between the two treatment means using a $100(1 - \alpha)\%$ confidence with width of no more than δ units.

In lecture, our example used $\alpha = 0.05$, $\delta = 4$, and $S_p^2 = 25$. From the calculation, we obtained a sample size of 50 from each sample. The code that produced this result is given below.

```
proc power;
  twosamplemeans test=diff
  alpha = 0.05
  meandiff = 4.0
  stddev= 5
  npergroup = .
  power = 0.975;
run;
```

Note: For sample size determinations using the confidence interval method, the value of `power` should always be set to the confidence level, $1 - (\alpha/2)$.

Assignment

1. Conduct the t-test for the SMS speed example in SAS and complete the following exercises:
 - (a) Using the formula from the notes, calculate by hand a 95% confidence interval for the difference in the two treatment means. Use $t_{28,0.975} = 2.0484$.
 - (b) Provide a screenshot of the SAS output and use it to verify your calculation.
 - (c) Interpret the confidence interval in the context of the problem.
2. Use SAS to explore sample size determinations for the bone loss example using the **hypothesis testing method** and complete the following exercises:
 - (a) Explore the effect of changing just the significance level - For $\alpha = 0.01, 0.05, 0.1$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (b) Explore the effect of changing just the power - For $1 - \beta = 0.99, 0.95, 0.9, 0.8, 0.7$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (c) Explore the effect of changing just the true effect size - For $\delta = 1, 2, 3, 4, 5, 6$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (d) Explore the effect of changing just the estimated population variance - For $S_p^2 = 1, 4, 9, 16, 25, 36$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
3. Use SAS to explore sample size determinations for the bone loss example using the **confidence interval method** and complete the following exercises:
 - (a) Explore the effect of changing just the significance level - For $\alpha = 0.01, 0.05, 0.1$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (b) Explore the effect of changing just the true effect size - For $\delta = 1, 2, 3, 4, 5, 6$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (c) Explore the effect of changing just the estimated population variance - For $S_p^2 = 1, 4, 9, 16, 25, 36$, what are the resulting sample sizes? Summarize your findings in one concise sentence.
 - (d) Think about how the sample size determination using the confidence interval method relates to the **standard error method**. Summarize your findings in one concise sentence.

Total: 50 points # correct: _____ %: _____