K300 (4392) Statistical Techniques (Fall 2007)
Assignment 5: T-test (260 points, Due November 5)
Instructor: Hun Myoung Park
kucc625@indiana.edu, (317) 274-0573

Please read the following instructions carefully. If you have any problem with any of the questions, please contact the instructor.

- Download the SPSS data set from OnCourse or the course web page at http://www.masil.org/method/statistics.html
- Conduct the t-test in question using SPSS and print out the output.
- **Write down answers on the SPSS output.** Use separate sheets if you really need or are asked to do so.
- Hand in this assignment by Monday, November 5.
- You may ask your classmates about using SPSS, but **you MAY NOT discuss with other classmates** when answering questions. Remember the Student Code of Conduct and SPEA policies. Please talk to me if you have any problem.

1. (60 points) Solve question 20 on page 423. This question appears in assignment 4. Show your work clearly. *See pages 415-417*.

> **1) (10 points)** Launch SPSS and load the data set you downloaded. Click Analyze→Compare Means→One-Sample T Test… and then choose the variable `family` (click the variable name and the right arrow so that the variable name appears on the right-hand side box). You may see the one box labeled as "Test Value" at the bottom. Now, change the value from 0 to 3.18, which is the hypothesized value in question. And then click OK to conduct the one-sample t-test. Print out the output and close the Output window.
>
> **2)** On the output, circle N, the sample mean labeled "Mean," standard deviation labeled as "Std. Deviation," and standard error labeled as "Std. Error Mean".
>
> **3)** On the output, circle the hypothesized value labeled as "Test Value," t statistic, degree of freedom (df), and p-value labeled as "Sig. (2-tailed)". Note that the label should be "p-value (2-tailed)" since the number is not the significance level or test size.
>
> **4) (5 points)** Show how to get the standard error. You need to provide the formula and plug in relevant numbers. DO NOT simply report standard error; you already did in 2).
>
> **5)** State the null hypothesis and alternative hypothesis explicitly. Is this test two-tailed or one-tailed?
>
> **6) (5 points)** Find out the critical value for the .05 significance level. Is the test statistic (t) smaller than the critical value? (ignore sign and compare magnitude only) Would you reject the null hypothesis at the .05 level?
>
> **7) (10 points)** Look at the t distribution table. Using the test statistic (t), find out the range of its p-value; you may not get the accurate p-value (that you have on the output) from the table. Just report the range in the form of *".xxx < p-value < .xxx"*. Does the p-value from the output fall in the range?

8) (10 points) Would you reject the null hypothesis from the p-value approach at the .05 level? How would you explain substantively the p-value to your boss? Note that the p-value is associated with risk taking of rejecting the true null hypothesis. Lecture note 1 will be helpful to answer correctly.

9) (10 points) Construct the 95 percent confidence interval using the critical value you got from 6). See page 358-359 if you do not have any idea. Your confidence interval is different from that SPSS produces because SPSS (strangely) reports confidence interval of difference. Add 3.18 to both lower and upper bounds to know how SPSS did; you should get the same interval as one you manually computed. Does the hypothesized value in the null hypothesis exist in the confidence interval? Would you reject the null hypothesis? Note: the mean difference is the (unstandardized) effect size, $\bar{x} - c$ (c is the hypothesized value).

10) State your conclusion **_substantively_** and specify the p-value at the end of the sentence. For example, "*The population average ... (p<.024).*" Since this is the final step of hypothesis testing, do not simply say "*reject the null hypothesis at the .05 significance level*" or "*do not reject the null hypothesis at the .01 level.*"

2. (70 points) Now, let us move on to the paired sample t-test. Question 7 on page 500 examines the effectiveness of a new grammar program. A student was asked to write a two-page essay before and after the new grammar program. The two scores (the numbers of errors a student made) are related in a sense they are mistakes made by the same student. Remember that Nicole's 90 (before the treatment) and 95 (after the treatment) should be related somehow. We just want to know what made the difference. Therefore, the paired sample t-test is widely used by (medical) scientists (IU School of Medicines) to explore the impact of a treatment (e.g., new medicine, new flue shot, new operation method, etc.).

1) (10 points) Launch SPSS and load the data set you downloaded. Click Analyze→Compare Means→Paired-Samples T Test… and then choose the variables `before` and `after`. You should see "before – after" on the right-hand side box labeled "Paired Variables:" And then click OK to conduct the paired sample t-test. Print out the output and close the Output window.

2) On the output, circle N, sample mean, and standard deviation of two variables. Skip the standard error.

3) On the output (skip the second table), circle the mean of difference (*d*), standard deviation of the difference, standard error of the difference. Show how to get the standard error by providing a formula with relevant numbers plugged in.

4) State the null hypothesis and alternative hypothesis explicitly. Let us use δ "delta" (the corresponding Greek letter for *d*) instead of µ.

5) (5 points) Find out the critical value for the .01 significance level. Look for and circle the degrees of freedom. How do you compute the degrees of freedom?

6) (5 points) Circle the test statistic (t). Is this test statistic smaller than the critical value you got in 5)? Would you like to reject the null hypothesis at the .01 level?

7) (5 points) Again look at the t distribution table. Using the test statistic (t), find out the range of its p-value; you may not get the accurate p-value (that you have on the output) from the table. Report the range in the "*.xxx < p-value < .xxx*" form. Does the p-value from the output fall in this range?

8) (10 points) Would you reject the null hypothesis from the p-value approach at the .01 level? How would you explain substantively the p-value to your boss? Would you like to reject the null hypothesis *at the .05 level*?

9) State your conclusion *substantively* and specify the p-value at the end of the sentence. See 11) in question 1. Was the new grammar program effective?

10) (25 points) Let us manually compute the test statistic for replication. Use a separate sheet and construct a table as follows. I guess this is the most easiest and efficient way to compute variance and standard deviation. If you have no idea, see example 9-12 and 9-13 on pages 493-497.

before	after	d=before-after	$(d_i - \bar{d})^2$	d_i^2
12	9	3=12-9	…	9=3×3
…	…	…	…	…
N/A	N/A	$\sum d_i =$	$\sum (d_i - \bar{d})^2 =$	$\sum d_i^2 =$

- Compute the difference between `before` and `after`. Subtract a value of `after` from its corresponding value of `before`.
- Compute the mean of the difference $\dfrac{\sum d_i}{n}$
- Compute $\sum (d_i - \bar{d})^2$ and $\sum d_i^2$
- Compute the standard deviation of the difference d using the two formulae. See the midterm exam for the formula.
- Compute standard error of the mean difference $\bar{d}$ (not the standard deviation of d). If you are confused with the standard deviation of standard error, see the lecture note 1.
- Now compute the test statistic using the formula for the paired t-test. Do you have the same t statistic that SPSS produces for you?

3. (80 points) Independent sample t-test compares the means of independent samples. Look at Question 13 on page 488. We are going to examine if there is any mean difference in the out-of-state tuitions between private and public colleges.

1) (10 points) Launch SPSS and load the data set you downloaded. Click the "Data View" tab at the left bottom of the window so that data points are listed in the spreadsheet form. Look at the variables `tuition` and `private` only. Private colleges are set 1 and 0 for public colleagues. SPSS requires this type of data arrangement for the independent sample t-test (this is NEVER flexible). Now, click Analyze→Compare Means→Independent Samples T Test… Click the variable `tuition` and move it to the right-hand side box labeled as "Test Variable(s)." Click the variable `private` and move to the box labeled "Grouping variable." You may see "private(? ?)." SPSS does not know which values to use for categorization. Click "Define Groups…" and type 1 in "Group 1" and 0 in "Group 2". Now click "Continue" and finally click OK. Print out the output and close the Output window.

2) On the output, circle N, sample mean, and standard deviation of two variables. Skip the standard error.

3) On the output (second table), circle the F statistic and its p-value under the label "Levene's Test for Equality of Variance." Would you like to reject the null hypothesis of equal variance at the .05 level? (Look at the p-value that SPSS computes for you; you do not need to look up the F distribution table. Do not forget that "Sig." is not the significance level, but the p-value for the F statistic).

4) (5 points) Conduct the (folded form) F test manually. See pages 473-476 if you have no idea. Identify the larger and smaller variances (not standard deviations) of two variables and plug appropriate numbers into the formula. What is your F statistic and degrees of freedom? Report them in a form of "*x.xxx* ~F(df_1, df_2)," which means x.xxx follows the F distribution with first and second degrees of freedom df_1 and df_2.

5) (10 points) Look up the F distribution table on page 639 and find out the critical value. Is the F statistic you computed larger than the critical value? Would you like to reject the null hypothesis of equal variance on the basis of this test statistic approach? Of course, the significance level is .05. Is your conclusion different from the result of the Levene's test?

6) State the null hypothesis and alternative hypothesis explicitly.

7) (10 points) If you do not reject the null hypothesis of equal variance, read the first line labeled as "Equal variances assumed" to get the t statistic and its p-value. If you reject the null hypothesis, read the second line labeled as "Equal variances not assumed" Circle the appropriate test statistic and p-value. Could you show me how the t statistic is computed using two sample means and standard error labeled as "Std. Error Difference?"; provide the formula with proper numbers plugged in. You do not need here to show entire computation of standard error. You will be given the opportunity to show your work in 8) and 9) below.

8) Would you like to reject the null hypothesis at the .05 level? Just look at the p-value and decide whether you reject the null hypothesis.

9) State your conclusion *substantively* and specify the p-value at the end of the sentence. Do private and public colleges have different average out-of-state tuition?

10) (10 points) Let us compute the standard error manually. Assume that two variables have the **same variance**. Show how to compute standard error by providing the formula with standard deviations and numbers of observations plugged in. Keep in mind that you need to compute the pooled variance. See page 481 to get the formula. Do not try to memorize it. However, you have to understand what is going on the standard error in order to interpret the output correctly. Compare your standard error (read the first line for equal variance) and one that SPSS produces. Are they identical? Note: using R will be much easier and safer (avoiding human errors) to do such a complex computation.

11) (5 points) Now, assume that two variables have **different variances**. Show how to compute standard error by providing the formula with standard deviations and number of observations plugged in. See page 481 to get the formula. Compare your standard error (read the second line for unequal variance) and one that SPSS produces. Are they identical?

12) (15 points) The textbook appears to ignore the approximation of degrees of freedom when variances of two variables are not equal. However, you have to

know the approximation in order to understand why SPSS reports non-integer degrees of freedom. Circle the degrees of freedom for unequal variance. The degrees of freedom you got are the Satterthwaite's approximation of degrees freedom. Get the formula from the powerpoint slide and show how the approximation was done. You need to compute *c* first from standard deviations and number of observations of two variables. Note: using R will be much easier and safer to do such a complex computation. You may take a look at an example from http://www.masil.org/documents/ttest.pdf. Is your answer the same as what SPSS reported?

4. (50 points) Now, we are going to compare two proportions. See pages 504-505 or section 7 of http://www.masil.org/documents/ttest.pdf. And then look at the example 9-16 on page 506-507. Change the sample size from 200 to 500 and test size from .01 to .05.

> **1) (15 points)** Follow all steps shown in the textbook. You have to compute the pooled probability first. Do not forget to explicitly state the null and alternative hypothesis. The textbook takes the test statistic approach.
>
> **2) (10 points)** Compare proportions using the p-value approach. You have to compute the p-value from the standard normal distributing table and compare it with the test size. 1) state the null and alternative hypothesis; 2) state the significance level; 3) compute the test statistic; 4) compute p-value and compare with alpha and then decide whether or not to reject the null hypothesis ; 5) draw a conclusion (substantive interpretation)
>
> **3) (5 points)** Launch SPSS and load the data set you downloaded. Click the "Data View" tab at the left bottom of the window so that data points are listed in the spreadsheet form. Look at the variables `control` and `surgeon` only. Surgeons are set 1 and 0 for general practitioners. Click Analyze→Compare Means→Independent Samples T Test… Click the variable `control` and move it to the right-hand side box labeled as "Test Variable(s)." Click the variable `surgeon` and move to the box labeled "Grouping variable." You may see "surgeon(? ?)." Click "Define Groups…" and type 1 in "Group 1" and 0 in "Group 2". Now click "Continue" and finally click OK. Print out the output and close the Output window. This is the independent sample t-test.
>
> **4)** On the output, circle N, sample mean, and standard deviation of two variables. Skip the standard error.
>
> **5)** On the output (second table), circle the F statistic and its p-value under the label "Levene's Test for Equality of Variance." Would you like to reject the null hypothesis of equal variance at the .05 level?
>
> **6) (5 points)** Conduct the t-test on the based of the F test in 5). Simply look at the p-value of the t statistic. Would you like to reject the null hypothesis at the .05 level?
>
> **7) (10 points)** Compare z score you got in 1) and 2). Is the t statistic different from your z score? How much? How can you interpret this result in conjunction with the normal approximation of a binomial distribution when N is large?