

**K300 (4392) Statistical Techniques (Fall 2007)****Assignment 6: ANOVA and Chi-Square test (190 points, Due November 14)**

Instructor: Hun Myoung Park

kucc625@indiana.edu, (317) 274-0573

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

- Download the SPSS data set `assignment6.asv` from OnCourse CL or the course web page at <http://www.masil.org/method/statistics.html>
- **Write down answers on the SPSS output.** Use separate sheets if you really need.
- Hand in this assignment by Wednesday, November 14.
- 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.

**1. (80 points)** One-way ANOVA compares means of more than two groups. Variables should be normally distributed and have equal variance. Note that the t-test is a special case of one-way ANOVA. The data set is drawn from question 28 on page 610, which examines if the average time it takes people to commute to the college is different from three groups (students, faculty, and staff). The significance level is .05. *See the powerpoint slide about ANOVA.*

**1) (10 points)** Launch SPSS and load the data set `assignment6.sav` you downloaded. Click Analyze→Compare Means→One-Way ANOVA... Choose the variable `commute` and move it to the right-hand side box labeled “Dependent List”. Choose variable `group` and move to a small box labeled “Factor”. Notice that a grouping variable is called a *factor* in ANOVA. Now, click **Options** button at the bottom and check **Descriptive** to tell SPSS to produce descriptive statistics. Click **Continue** button and click **OK** to conduct the one-way ANOVA. Print out the output and close the Output window.

**2)** On the output, circle N, mean, standard deviation of three groups. Ignore standard error, confidence interval, minimum, and maximum.

**3) (5 points)** On the output, circle SST (sum of squares of the total). This is the

numerator of variance of variable `commute`,  $s^2 = \frac{\sum (y_i - \bar{y})^2}{n - 1}$  and

$\sum (y_i - \bar{y})^2 = s^2(n - 1)$ . Show how SST is computed from descriptive statistics you obtained in 2).

**4) (5 points)** On the output, circle SSE (sum of squares due to error or within-

group sum of squares). The short-cut is  $SSE = \sum_{j=1}^t (n_j - 1)s_j^2$ , where  $n_j$  and  $s_j^2$  are

the number of observations and the sample variance of group  $j$ . Show how SSE is computed from descriptive statistics you obtained in 2).

**5)** On the output, circle SSM (sum of squares of the model or between-group sum of squares).  $SST = SSM + SSE$ . Show how SSM is computed.

- 6) (5 points)** On the output, circle degrees of freedom of SSM and SSE. Show how these degrees of freedom are computed. See the descriptive statistics you obtained in 2).
- 7) (5 points)** On the output, circle MSM (mean squares of the model or between-group mean squares) and MSE (mean squares due to error or within-group mean squares). Show how MSM and MSE are computed. See 4), 5), and 6).
- 8) (5 points)** On the output, circle F score. Show how the F score is computed.
- 9) (15 points)** Now, conduct a one-way ANOVA test. (1) state the null and alternative hypothesis; (2) find out the critical value for the .05 significance level from the F distribution table on page 639. Use degrees of freedom you obtained in 6); (3) Report F score you computed in 8); (4) compare the F score with the critical value. Would you like to reject the null hypothesis?; and (5) interpret the output substantively. You need to re-paraphrase the null hypothesis. Do not forget to report the p-value SPSS produced for you at the end of the sentence.
- 10) (25 points)** Let us manually compute SSM, SSE, and SST. See the computation table in powerpoint slide (p. 9). Note that ANOVA partitions variance components and evaluates the extent that the model (or factor) accounts for the total variance. Let us construct the following table.

<b>x</b>	<b>group</b>	$(x_i - \bar{x})^2$	$\bar{x}_j$	$(\bar{x}_j - \bar{x})^2$	$(x_i - \bar{x}_j)^2$
12	1				
...	...				
57	2				
...	...				
15	3				
...	...				

- Compute the overall mean  $\bar{x}$  and group means  $\bar{x}_j$ . Note that there are three groups (students, faculty, and staff). So you should have three group means. Fill the  $\bar{x}_j$  column with proper group means.
- Compute  $(x_i - \bar{x})^2$ . You must be familiar with this variance component. Sum them up to get SST. Is your SST the same as what SPSS computed for you? Note that you have  $n$  data points and one overall mean involved in computation; so the degrees of freedom of SST are  $n-1$ .
- Compute  $(\bar{x}_j - \bar{x})^2$ , which means subtracting a group mean from the overall mean  $\bar{x}$  and then squaring. Sum them up to get SSM. Is your SSM the same as what SPSS computed for you? Note that you have  $j$  group means and one overall mean involved in computation; so the degrees of freedom of SSM are  $t-1$ .
- Compute  $(x_i - \bar{x}_j)^2$ , which means subtracting an individual data point from its group mean and then squaring. Sum them up to get SSE. Is your SSE the same as what SPSS computed for you? Note that you have  $n$  data points and  $t$  group means involved in computation; so the degrees of freedom of SSE are  $n-t$ .

**2. (30 points)** Download the data set for assignment 5, `assignment5.sav`. This question asks you to compare the independent sample t-test and one-way ANOVA.

- 1) (5 points)** Launch SPSS and load the data set you downloaded. 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.” 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) (10 points)** Click Analyze→Compare Means→One-way ANOVA... Click the variable `tuition` and move it to the right-hand side box labeled as “Dependent List.” Click the variable `private` and move to the box labeled “Factor.” Click **OK** to conduct the ANOVA. Print out the output and close the Output window.
- 3) (15 points)** On SPSS output, circle t statistic you obtained from 1) and F statistic you obtained from 2). Circle degrees of freedom of SSM in ANOVA table from 2). Show the relationship between t statistic and F statistic when the first degrees of freedom is one (comparing two group means). See the powerpoint slides for the t-test and ANOVA.

**3. (80 points)** This question is drawn from question 19 on page 617 of Gravetter and Wallnau (2007), which examines if people who volunteer to participate in research studies tend to have higher intelligence than nonvolunteers. In other word, we want to check if volunteering is independent of intelligence. These two variables were measured in ordinal variables assuming there is some meaningful order between volunteers and nonvolunteers. A contingency table is a good tool for summarizing information of two categorical variables. Take a look at the following contingency table (cross-table).

	High IQ	Medium IQ	Low IQ	Total
Volunteers	43	73	34	150
Nonvolunteers	7	27	16	50
Total	50	100	50	200

- 1) (10 points)** Launch SPSS and load the data set you downloaded. Click Analyze→Descriptive Statistics→Crosstabs... Click the variable `volunt` and move it to the upper right-hand side box labeled as “Row(s).” Click the variable `IQ` and move to the box labeled “Column(s).” Click **Statistics...** button, check **Chi-square**, **Lambda**, and **Gamma**, and then click **Continue** button to return. Click **Cells...** button, check **Expected**, and then click **Continue**. Note that **Observed** is already checked. You are now ready to conduct the chi-square test. Click **OK** button. Print out the output and close the Output window.
- 2) (5 points)** Take a look at the second table with the title “volunt\*IQ Crosstabulation.” The row labeled as “Count” reports observed frequencies, while the “Expected Count” reports expected frequencies. Circle all expected frequencies on the SPSS output. Check if there is any cell whose expected frequency is less than 5. What is the smallest expected frequency? Keep in mind that a chi-square test is not reliable if expected frequency in any cell is less than 5.

**3) (5 points)** Take a look at the third table with the title “Chi-Square Tests.” Read the row labeled as “Pearson Chi-square.” Circle the chi-square statistic, degrees of freedom, and p-value labeled as “Asymp. Sig. (2-sided).” As I told you there are many variants of the chi-square test. Show how the degrees of freedom are computed.

**4) (15 points)** Conduct the chi-square test for independence. (1) state the null and alternative hypothesis. See the lecture note for chi-square test; (2) find out the critical value for the .05 significance level from the  $\chi^2$  distribution table on page 636. Use degrees of freedom you obtained in 3); (3) Report  $\chi^2$ ; (4) compare the  $\chi^2$  score with the critical value. Would you like to reject the null hypothesis?; and (5) interpret the output substantively. You need to re-paraphrase the null hypothesis.

**5) (5 points)** Take a look at the p-value in 3). Would you like to reject the null hypothesis in this p-value approach?

**6) (5 points)** Construct a table for expected frequency. You need to take a look at the lecture note; textbook does not mention this issue. Double check if your expected frequencies are the same as those SPSS produced in 2)

**7) (10 points)** Compute  $\chi^2$  manually using the formula provided in the lecture note or pages 577-580 of the textbook. Is your  $\chi^2$  the same as one you obtained in 3)?

**8) (5 points)** Take a look at the last table (fifth table). Read the first row labeled as “Ordinal by Ordinal Gamma.” Gamma is a measure of association of two ordinal variables. Report Gamma statistic labeled as “Value” and its p-value labeled as “Approx. Sig.” Does the Gamma score report a positive relationship between two variables? Read the p-value and determine if the measure of association is significant. Do you think the measure of association is reliable?

**9) (10 points)** Compute Gamma manually. You need to take a look at the lecture note; textbook does not mention this issue. Compute C and D first. Is your Gamma the same as one SPSS produced?

**10) (5 points)** Take a look at the fourth table. Read the third row labeled as “Nominal by Nominal Lambda—volunt dependent” Report Lambda statistic labeled as “Value” Note that if both categorical variables are not ordinal, you should compute Lambda instead of Gamma. How do you interpret the Lambda statistic with respect to the relationship of two variables? See the lecture note.

**11) (5 points)** Compute Lambda manually. You need to take a look at the lecture note; textbook does not mention this issue. Is your Lambda the same as one SPSS produced for you?