2017 Spring

EDMS 646: General Linear Models I

1121 Benjamin Building, Thursdays: 4:15-7:00pm

Assignment 4

(Due: April 20, beginning of the class)

Answer each question synthesizing the information from the lecture slides, textbook, and other resources. When appropriate, insert statistical output (and input if need be) to justify each of your answers. Answers to the questions must be word-processed using Word and Microsoft Equation Editor (an object to be inserted within Word) for statistical/mathematical notation. Answers to the homework questions should appear on 8.5" x 11" paper (not computer output) and must be legible. Students with clarifying questions about the homework should contact Dr.Yang directly. The homework grading scheme was explained during the first class and also written in the course syllabus.

PART 1: One-Way ANOVA

Mireault (1990) conducted a study of grief and loss with college students. The primary focus was examining the effect of parental loss on the perceived vulnerability to future loss (i.e., does parental loss make one feel more vulnerable to potential future loss?). Mireault collected data from 381 college students, some of who had lost a parent by death during their childhood, some of whom had not. There were a total of three groups in the study. Group 1 was composed of people who had lost a parent. Group 2 was composed of people who did not lose a parent and their parents were married to each other. Group 3 was composed of people who did not lose a parent but were divorced. Use the Mireault.sav data set to investigate group differences in perceived vulnerability using ANOVA. The grouping variable or factor is called "group" and the dependent variable is called "pv-total," which is the total score of a measure that reflects perceived vulnerability to future loss (higher scores indicate more vulnerability).

1. Generate side-by-side boxplots and descriptive statistics for the three groups. What do the sample means and boxplots indicate about perceived loss among the groups? Explain.
2. In symbols, write out the omnibus null hypothesis for this analysis.
3. State the omnibus alternative hypothesis for this analysis.
4. Generate a means plot in SPSS or R. According to the means plot, are the sample means compatible with the omnibus null hypothesis? Explain.

5. Based on the side-by-side boxplots and other pertinent output, is there evidence that the scores within the groups are all equal (i.e., is there evidence the scores are constant within groups)? Explain.

6. Based on the means and sample sizes in the "descriptive?" table (in other words, using grand mean and group means and sample sizes), compute $MS_{between}$.
   The following hints are for SPSS users.
   Hint #1: the grand mean is in the row labeled "total."
   Hint #2: you do not have equal sample sizes, so beware the formula you use. Verify this value is within rounding error of the value in the ANOVA table.

7. What does the value of $MS_{between}$ imply about the equality of the sample means? Explain.

8. Can we use $MS_{between}$ alone (by itself) to test H_0? Explain.

9. Compute R^2 (estimated eta-squared) based on the sum of squares.

10. Interpret R^2 using the names of variables.

11. Assuming $\alpha = .05$, what is the F-critical value?

12. Based on the F-statistic and the F-critical value, what decision do you make about H_0 and why do you make this decision?

13. What does your decision in the last question suggest about the population means?

14. Interpret the p-value from the output (i.e., define the p-value for this example).

15. Does your decision about H_0 leave any ambiguity about the specific relationships between the population means? Explain.

16. Conduct an appropriate post-hoc multiple comparison analysis to address the ambiguity.

17. Test the assumptions of homogeneity of variance and normality. Use graphical evidence as well as descriptive statistics and statistical tests in your justification

Suppose you are a researcher investigating gender differences in personal income. You want to examine whether respondent's income in 1991 (y = rincom91) showed a mean difference for men and women (x = sex). Being a savvy researcher you reason that perhaps men and women differ in mean income because of a difference in number of years of education (c = educ). Therefore, you want to exam the mean difference for men and women on rincom91 while controlling for the covariate education. Perform an ANOVA (PART 2) and an ANCOVA (PART 3) using and R or SPSS. If you use SPSS, please utilize General Linear Model Option for PARTS 2 and 3, using the data set GSS93.sav.

PART 2: Analysis of Variance (No covariate)

1. Explore the data and report the results.
2. Run the ANOVA and report the ANOVA table.
3. Write up the results of ANOVA using the APA format (Don't forget to include effect size measures and observed power).
4. Check if any serious model assumption violation is observed and report the results.

PART 3: Analysis of Covariance

1. Write an ANCOVA model equation.
2. Run the ANCOVA without interaction in which education is used as a covariate, and report the ANOVA table. Make sure you choose the proper type for SS decomposition.
3. Write out two null and alternative hypotheses being tested in the ANOVA table.
4. Do you think that the education variable is a useful covariate? Explain your reasons based on your analysis result.
5. Report adjusted means for two groups and explain how these are different from original means.
6. Conduct the ANCOVA with interaction and explain if homogeneity of slopes assumption is reasonably met or not based on your analysis results
7. If you have to choose a model between the ANOVA model and the ANCOVA model to describe the data, what would be your choice? Explain why.

PART 4: Analysis of Covariance through Multiple Regression

Now, perform an ANOVA and an ANCOVA using regression analyses, using the data set GSS93.sav. Before running regression analyses, you will need to recode your sex variable and generate an interaction term as follows. The following instruction is for SPSS users, and R users do not need this process according to the lab materials.

Step 1. Recode the sex variable.
a. Open the data file GSS93.sav.
b. Sex is coded 1=male and 2=female but it is preferable to have 0=male and 1=female. Therefore, recode sex by using Transform > Recode > Into Same Variable, select sex then click on Old and New Values plugging in the old and new values and clicking Add each time.

Step 2. Compute the interaction term.
a. Compute the interaction term using Transform > Compute.
b. For the Target Variable use inter, and for the Numeric Expression use sex*educ (the asterisk denotes multiplication).

Now, run the following three models and answer the following questions.

Model 1: Regress rincom91 on your recoded sex variable (ANOVA model)
Model 2: Regress rincom91 on your recoded sex variable and education variable (ANCOVA model)
Model 3: Regress rincom91 on your recoded sex variable, education variable, and the interaction term (ANCOVA model with interaction term)

1. First consider the results of Model 1 (ANOVA model). Using the names of the variables, interpret (define) the values of the intercept and the unstandardized slope in terms of means.

2. In terms of means, state in symbols the null hypotheses associated with the two t-tests of the ANOVA analysis. Based on the t-test associated with the slope of the ANOVA model, what conclusion do you make about the population?

3. Using the names of the variables, interpret the R^2 for the ANOVA model.

4. Consider the interaction model (Model 3). Is the interaction term required? Based on information on the output, explain why or why not.

5. Consider the ANCOVA model (Model 2). Using the names of the variables, interpret (define) the value of the unstandardized regression coefficient for sex in terms of means.

6. Consider the interaction model (Model 3). Based on the sample estimates solve the regression equation for each value of x (covariate, education in this example). Based on your answer; compute the group predicted values for 6 years of education and 20 years of education. Explain the nature of the sample interaction.