2017 Spring

EDMS 646: General Linear Models I

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

Assignment 3

(Due: March 9, 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: Multiple Regression - Initial model: Use data set HSB1

For the following questions, you will use datasets HSB1. The data are on predictors of student achievement from the study called High School and Beyond. (If you have a dataset of your own that you prefer to use for this assignment, please consult the instructor for PRIOR approval. To determine whether your dataset is appropriate for this assignment, you will be asked to submit a description of the dataset and information about the variables you wish to analyze, including the measurement scales and basic descriptive data such as means and standard deviations.)

1. Carry out a multiple regression. Choose one of the following variables as your dependent (outcome) variable: read (reading score), write (writing score), math (math score), science (science score), or socst (social science score). Use all of the following variables as your independent (predictor) variables: locus (locus of control), concept (self-concept), and mot (motivation). Include all predictor variable in the analysis in the same step (that is, do not use stepwise analysis). Report the resulting prediction regression equation using the values from the unstandardized solution. Give a specific interpretation of the unstandardized regression coefficients in terms of the variables. Do not use the statistical write-up here.

2. Explain the H_0 associated with the F-test in the ANOVA table. Based on the information in the ANOVA table, write down the appropriate null and alternative hypotheses in terms of a model comparison, a decision rule, and indicate whether you reject H_0. Explain what statistical criterion you used as the basis for your answer. Do not use the statistical write-up here.

3. Explain the H_0 associated with the unstandardized regression coefficients. Based oh the t-tests, what statistical conclusions do you make and why (e.g., do you reject or retain the hypotheses and what criterion are you suing to make that decision)? Do not use the statistical write-up here. In general (not just for this example), what do the standard errors of the regression coefficients tell you?

4. Examine the assumptions of normality, linearity, and homoscedasticity. Describe the results of your examination/investigation, including any problems you found.

PART 2: Multiple Regression - Final model: Use data set HSB1

1. If any of the predictors in PART 1 are not statistically significant, carry out appropriate analyses to produce a final regression analysis in which every predictor is statistically significant. If all of the predictors in PART 1 are statistically significant, you can treat the results of your PART 1 analysis as the final model.

2. Report and interpret the results, using the statistical write-up template. Present and interpret relevant coefficients (including both unstandardized and standardized regression coefficients), p-values, and R^2. Include an APA-formatted table that presents the final results of your multiple regression analysis. (APA = American Psychological Association).

PART 3: ANOVA using Multiple Regression

An experiment is conducted comparing four instructional methods to teach children with particular disabilities to perform a specific task. The experimenter divides subjects into four groups of 10. Each group of children is assigned a method of instruction. For each child a motivation score is measured prior to group assignment and prior to treatment; after treatment a score on the desired task is observed. The results are as follows:

Treatment A		Treatment B		Treatment C		Treatment D	
score	mot.	score	mot.	score	mot.	score	mot.
94	14	80	38	92	55	80	37
96	19	84	34	96	53	82	24
98	17	90	43	99	55	79	22
100	38	97	43	100	52	100	43
102	40	97	61	102	35	110	49
105	26	98	63	104	46	98	41
109	41	130	93	99	57	97	26
110	28	89	74	110	55	113	70
111	36	120	76	115	42	115	63
130	66	120	79	118	81	104	24

To do the following part of the assignment, you need to come up with a data file using the information provided in the table above. Let's only consider the score variable.

1. Run the regression of score (dependent variable) on the independent dummy variables using Treatment D as the reference group. Write out the fitted regression equation using the values of the unstandardized regression coefficients from the output. Including the intercept, interpret each of these coefficients in terms of group means.

2. Write out the null hypotheses for testing each of the regression coefficients, in terms of group means or group mean differences. Watch your notation please. Are there significant group differences? Use appropriate statistical write-up to explain your answers.

3. Using your fitted model from question 1, find the predicted (expected) value for a child in group D, by plugging in the appropriate values for the variables. Interpret this value. Do the same for a child in group B.