

STA441 Assignment 9

Quiz in Tutorial on Tuesday March 20th

$$\mu = \begin{pmatrix} \mu_1 \\ \mu_2 \\ \vdots \\ \mu_k \end{pmatrix} = \begin{pmatrix} E[y_1|\mathbf{X}=\mathbf{x}] \\ E[y_2|\mathbf{X}=\mathbf{x}] \\ \vdots \\ E[y_k|\mathbf{X}=\mathbf{x}] \end{pmatrix} = \begin{pmatrix} \beta_{0,1} + \beta_{1,1}x_1 + \cdots + \beta_{p-1,1}x_{p-1} \\ \beta_{0,2} + \beta_{1,2}x_1 + \cdots + \beta_{p-1,2}x_{p-1} \\ \vdots \\ \beta_{0,k} + \beta_{1,k}x_1 + \cdots + \beta_{p-1,k}x_{p-1} \end{pmatrix}$$

This assignment uses 2 different data sets, so **please bring 2 separate sets of output**, the log file and the results file for each data set. You may not hand both of them in.

1. An IQ (Intelligence Quotient) is a score on a test that supposedly measures intelligence. A score of 100 is average. IQ tests are controversial, but it's fair to say that an IQ test is an inexact measure of one kind of intelligence -- the kind that helps you do well in school. In the Longitudinal IQ data, adopted children had their IQs tested at ages 2, 4, 8 and 13. Birth mother's IQ and adoptive mother's education are also available for each child.

The data are available in the file [origIQ.data.txt](#). The variables are

- Adoptive mother's education in years
 - Birth mother's IQ
 - Child's IQ at age 2
 - Child's IQ at age 4
 - Child's IQ at age 8
 - Child's IQ at age 13
- a. Use `proc means` to produce the default descriptive statistics on all the variables.
 - b. Just to get a feel for the relationships among variables, produce a correlation matrix with `proc corr` and take a look. Be able to interpret all the tests. Is there evidence that adoptive mother's education is related to birth mother's IQ?
 - c. Now treat this as a multivariate regression with two explanatory variables (no interaction) and four response variables.
 - i. Carry out multivariate tests of adoptive mother's education controlling for birth mother's IQ, and birth mother's IQ controlling for adoptive mother's education. What do you conclude?
 - ii. You have four univariate tests of adoptive mother's education controlling for birth mother's IQ, and four univariate tests of birth mother's IQ controlling for adoptive mother's education. Protecting these 8 tests with a Bonferroni correction is a good idea, and makes you wonder why you bothered with the multivariate test. With the correction, what do you conclude? Be able to state your conclusions in plain, non-statistical language.
 - iii. I can't ask this on the quiz, but do you see why these results could make some people uneasy? Think of the political implications.

Bring your log file and results file to the quiz.

2. A random sample of male and female university students is weighed midway through year 1, 2, 3 and 4.
 - a. What are the cases in this study?

- b. How many numbers (observations) do you have for each student?
- c. This is a factorial experiment. What are the factors? (Let's say that case is not a factor.)
- d. Classify each factor as between-cases or within-cases.
- e. Make a 2 by 4 table. Draw an oval or ovals on the table, indicating the crossing or nesting of cases within experimental conditions. See the lecture slides for some examples.
- f. In the "multivariate" approach to within-cases analysis, you set up effect coding dummy variables for the between-cases factors (if any), and calculate response variables that are linear combinations of the variables that are recorded for each case. You can then obtain tests for all the main effects and interactions by testing null hypotheses about the β values in the regression model. Sometimes the model has more than one response variable (linear combination). In this case it really is multivariate, and the second subscript on the β s refers to the response variable.

For this sex by weight example, denote the four weights for student i by $y_{i1}, y_{i2}, y_{i3}, y_{i4}$. The response variables will be linear combinations of these values. First consider the main effect for sex of student.

- i. Give a formula (or formulas) for the linear combination (or combinations) that you would use as the response variable (or variables).
 - ii. Write the regression model -- just the expected value(s).
 - iii. What is the null hypothesis in terms of the β values from your model?
 - g. A single model applies to Year and Gender by Year.
 - i. Give a formula (or formulas) for the linear combination (or combinations) that you would use as the response variable (or variables).
 - ii. Write the regression model -- just the expected value(s).
 - iii. In terms of β values from your model, what is the null hypothesis for testing the main effect of Year?
 - iv. In terms of β values from your model, what is the null hypothesis for testing the Gender by Year interaction?
3. In an experiment on anxiety medications, volunteer patients took a pill in the morning every day and in the evening they rated how anxious they had felt on average during that day. What was in the pill was unknown to the patient, and came in a different random order for every patient. The pill contained Drug A (Yes or No) and Drug B (Yes or No), in all four combinations. The four numbers for each patient are actually average ratings over 10 cycles, so the experiment took 40 days. The four numbers for each patient are
- y_{11} : Average rating in the No, No condition. $E(y_{11}) = \mu_{11}$
 - y_{12} : Average rating in the No, Yes condition. $E(y_{12}) = \mu_{12}$
 - y_{21} : Average rating in the Yes, No condition. $E(y_{21}) = \mu_{21}$
 - y_{22} : Average rating in the Yes, Yes condition. $E(y_{22}) = \mu_{22}$
- a. What are the cases in this study?
 - b. This is a factorial experiment. What are the factors? (Let's say that case is not a factor.)
 - c. Classify each factor as between-cases or within-cases.
 - d. Make a 2 by 2 table and write expected values in the cells.
 - e. Make an oval or ovals on the table, indicating the crossing or nesting of cases within experimental conditions. See the lecture slides for some examples.
 - f. You can test the main effects and interactions in this study with one-sample t-tests, testing whether the means of certain linear combinations of the y_{ij} variables equal zero. Give the linear combination you would use to test for each of the following. In each case, the answer is a formula, a function of the y_{ij} .
 - i. Main effect for Drug A .

- ii. Main effect for Drug *B*.
- iii. Drug *A* by Drug *B* interaction.
- g. Psychoactive drugs can have very different effects depending on the age of the person taking them. So consider independent samples of patients aged 5-12, 13-18, 19-29, 30-64 and 65+.
 - i. Write a regression model in which the expected value of the response variable depends on age group. You don't have to specify how your dummy variables are defined. You will do that in the next part.
 - ii. Make a table with 5 rows, showing how your dummy variables for age group are defined. Add another column for the expected value of the response variable.
 - iii. Why is your dummy variable coding scheme a good choice for testing whether the *average* expected value of some linear combination is equal to zero?
 - iv. We now have a 3-factor design, with one between-cases factor and 2 within-cases factors. What is the between-cases factor?
 - v. For each of the effects in your 3-way design, give the linear combination of y_{ij} , you would use as the response variable, and the null hypothesis you would test in terms of β values from your regression model.

| Effect | Linear combination | Null hypothesis |
|--------------------------------|--------------------|-----------------|
| Drug <i>A</i> | | |
| Drug <i>B</i> | | |
| Age | | |
| $A \times B$ | | |
| $A \times \text{Age}$ | | |
| $B \times \text{Age}$ | | |
| $A \times B \times \text{Age}$ | | |

4. In a study of the psychology of attention, subjects attempted to solve word problems while listening to distracting background noise. The distracting material was either music, or spoken words related to the problem they were trying to solve. The distracting material was presented at three different levels of loudness. Each subject attempted 10 problems at each combination of loudness and type of distraction, for a total of 60 problems. Order of presentation was randomized. Data for each subject are number correct in each of the six treatment combinations, and are available in [distract.data.txt](#).

- a. How many factors are there in this study? Classify each one as between cases or within cases. Let's say that case is not a factor.
- b. Make a 2 by 3 table. Draw an oval or ovals on the table, indicating the crossing or nesting of cases within experimental conditions.
- c. Using the multivariate approach to repeated measures, carry out the appropriate analysis, obtaining F-tests for all main effects and interaction(s).
- d. Prepare a plot of the means. You can do it by hand, or use Excel or a similar program. Either way, bring your plot to the quiz. You may hand it in.
- e. Your plot suggests that the interaction and one of the main effects (not the other one) should be interpreted. Which test for main effects is more interpretable and why?
- f. Describe the more interpretable effect in plain, non-statistical language.
- g. Based on the interaction plot, let's test for a volume effect (low-medium-high) just within the musical distraction condition. This is a multivariate test. Is there evidence of an effect?
- h. Just for completeness, test all pairwise volume differences just within the **voice** condition, even though we know how it's going to come out. Use matched t-tests. There are several ways to do this. The results are so strong that correction for multiple testing is not an issue. In plain, non-statistical language, what do you conclude?

Bring your log file and results file to the quiz.

Please **bring both log files and both results files to the quiz. *Also bring the plot for Question 4.*** As usual, answers to the questions are not to be handed in. They are just practice for the quiz. Please do not write anything on your printouts except your name and student number. It is okay to highlight the results file, but do not write interpretations on your results file, or cause them to appear in any way (including comment statements) on your log file.