Homework 6

Psychology 311
Spring, 2015

1. (25 points). The file *reading.csv* contains data for boys and girls (Factor A) participating in a reading readiness testing program. The design is between-subjects (cross-sectional), rather than (within-subjects) longitudinal, so different groups of students were tested at each age. Students tested were in the 5th, 6th, 7th, 8th, and 9th grades (Factor B), and reading scores were presented in terms of "completed grade reading level" For example, a score of 6.7 indicates a student is reading at a level appropriate for students midway through the 7th grade. On the basis of these data:

   (a) Perform a standard 2-way ANOVA on the data. Is there a significant main effect for sex? Is there a significant main effect for grade? Is there a significant interaction?

   (b) Perform a trend analysis on the data by grade. Is there a significant linear effect? Is there a significant quadratic effect?

   (c) Is there a significant $A \times \text{Linear}(B)$ interaction?

   (d) Is there a significant $A \times \text{Quadratic}(B)$ interaction?

   (e) Do an interaction plot of the data, with *Grade* on the horizontal axis, and two lines representing the boys and girls. How would you characterize the plot?

2. (30 points). The file *corrdata.csv* contains 150 observations on 10 variables. Read it into a data frame called *my.data*.

   (a) Compute the 10×10 correlation matrix on the 10 variables. There will be 45 non-redundant correlations in the correlation matrix. Extract these 45 correlations into a 45×3 matrix, with column 1 being the row index of the correlation, column 2 the column index, and column 3 the correlation. Extract only the nonredundant correlations for which the row index is larger than the column index. (*Hint*: You can do this easily with a double `for` loop using an *indexing variable*. I show you how, below. Study this code carefully, as it presents a standard approach used by programmers to extract information and recast it.

```
> my.data <- read.csv("corrdata.csv")
> R <- cor(my.data)
> X <- matrix(rep(NA, 270), 45, 6)
```

```r
> colnames(X) <- c("i", "j", "r[i,j]", "p-value",
+     "Holm", "FDR")
> index.var <- 0
> for (i in 2:10) for (j in 1:(i - 1)) {
+     index.var <- index.var + 1
+     X[index.var, 1] <- i
+     X[index.var, 2] <- j
+     X[index.var, 3] <- R[i, j]
+ }
```

(b) Compute the *two-sided* p-values for the 45 correlations, i.e., the test that $\rho = 0$. Compute the significance level based on the classic t-statistic that $rho = 0$, or its square F, which tests the hypothesis that $\rho^2 = 0$ against the 1-sided alternative $\rho > 0$. You can verify yourself that a 1-sided p-value for the F test is the same as the 2-sided p-value for the t-test. When you compute the p-values, place them in the 4th column of the data frame. You might want to write a function that takes n and r_{ij} as inputs, and returns the p-value. The F statistic is

$$F_{1,n-2} = (n-2)\frac{r^2}{1-r^2} \tag{1}$$

(c) Test the correlations for significance, using Holm's test, controlling FWER at 0.05, *while keeping track of which correlations are significant*. *Hint*: To answer this question and its follow-up, try sorting the data by p-value, using the method shown on *www.statmethods.net*, at the following page: $http://www.statmethods.net/management/sorting.html$. If you do this correctly, the row and column indices will "follow" the p-values, so you can keep track of which correlations are significant. Put the decision in the 5th column of the data.

(d) Test the correlations for significance again, this time controlling FDR at 0.05, *again while keeping track of which correlations are significant*. Put the result in the 6th column of the data frame.

(e) Once you have determined significance, re-sort the data frame, so that the data are sorted first by row, second by column, so that the correlations are back in their natural order.

(f) Now that you have completed this task, you have all you need to create a function that, on demand, provides FWER or FDR control for testing tables of correlations. I'm not requiring you to do that, but I suggest that you create that function, save it, and apply it should you ever decide to compute a correlation matrix and test all its elements for "significance." How does it feel to be able to do something in a few hours that nobody at SPSS managed to do for 25 years? (Larzelere and Mulaik had a 1977 article in *Psychological Bulletin* describing how to do Holm's test on elements of a correlation matrix.)

3. (15 points). RDASA3 10.2

4. (15 points). RDASA3 10.6

5. (15 points). RDASA3 10.10