PSYC602 - Homework #4

Complete this homework as a .Rmd file and submit the knitted .html file via ELMS.

NO GRADE WILL BE ASSIGNED FOR THIS PROJECT. In completing this assignment, you may consult any text or other written materials you wish, and you **MAY** discuss the project with your classmates.

Scenario - Part I

The individual who consulted with you regarding the oneway ANOVA has now become interested in conducting a regression analysis of his data, since he has heard such an approach would be 'better'. He brings you the data set below, and indicates that it has been collected in the column order in which it is listed.
He wants to know if group is a significant predictor of the criterion variable (the numbers). After a moment or two of confusion, you decide to go ahead and accede to his wishes.

Tasks - Part I

In your desire to answer such a question, you decide to:

- State the model under investigation
- State H_0, both formally (in terms of population parameters) and in words
- State H_1, both formally (in terms of population parameters) and in words
- Select an error rate
- Conduct a regression on these data using **lm()** and defining the group as a categorical predictor (Hint: *You will need to melt the data set to get in in the correct form for this regression*).
- Conduct a regression on these data using **lm()**, with the intercept as the only predictor (hint: **lm(y~1, data)**) and compare this model to the one using group as a categorical predictor (hint: **anova(fit0, fit1)**). Is the model that includes the categorical better? How do you know?
- Plot y vs. X, y vs. $\hat{y}$, y vs. e, $\hat{y}$ vs. X, e vs. X, and $\hat{y}$ vs. e, where X is the *Group* variable (*if you add 2 variables to your data frame, you can do all these plots in one* ** gpairs() ** *call*)
- Print a data frame that includes y, $\hat{y}$, and e for each observation
- Obtain and print the correlations among y, $\hat{y}$, and e
- Explain the intercorrelations found above
- Interpret the results so far

Scenario - Part II

Your researcher 'friend' comes back to you a final time, and indicates that his real interest all along has been to compare the first two groups to the last two groups, which he claims is a comparison of two treatment groups [1 and 2] to two control groups [3 and 4]. Furthermore, he is also interested in determining whether the two treatment groups differ from one another, and whether the two control groups differ from one another.

Tasks - Part II

- Code the predictor variables such that these three questions could be answered in a single regression analysis

The purpose of last week's lab was to demonstrate the equivalence of ANOVA and regression 'contrasts'. In order to do that, we went through the incredibly confusing transformation of contrasts in **R** that involved taking the inverse of the contrast matrix. **You do not need to do that here** Simply create a matrix where each column is a contrast using integers. An example for a 4-level categorical variable, comparing the first two and last 2 groups, the first and last group, and the second and third group, is included below:

```
contrasts(df1$Group) <- cbind(":12v34" = c(-1,-1,1,1),
                              ":1v4"   = c(-1,0,0,1),
                              ":2v3"   = c(0,-1,1,0))
```

- Conduct this analysis to prove your point
- Inform your researcher regarding what conclusions he may draw regarding his "real interest", including a statement about what model you believe gave rise to the data

Data

*These data are included as **hw4.csv** on ELMS*

y	Group
1	1
2	1
3	1
2	1
3	2
2	2
3	2
4	2
4	3
3	3
5	3
4	3
4	4
3	4
2	4
3	4