Homework 7

Assigned Wednesday, October 12
Due Wednesday, October 19 at classtime

Two-way ANOVA with balanced designs

Turn-in homework with hand-written or typed responses and include any relevant plots which you are describing.

1. “Prisoner’s Dilemma” data:
   In this study, there were twenty 4-person groups. Each group played “prisoner’s dilemma” 30 times. For each game, each subject selected either a competitive or a cooperative choice. The value reported for each group is the total number of cooperative choices (out of the 120) made by subjects in the that group. Ten of the groups recorded their choices anonymously, while the remaining 10 groups made public choices (i.e. their choice was made know to the other members of the group). Ten groups were all males, the other 10 were all females. The study was set-up as a balanced two-way ANOVA with the response of cooperation and factors of condition and sex.

   The experimenters expected to observe a higher level of cooperation in the public-choice condition, but did not make predictions about sex effects or sex-by-condition interaction.

   The Guyer data set in the car library contains the data.

   (a) Make a scatterplot of cooperation versus sex (so, sex is on the x-axis). Use different colors or symbols to represent condition in this plot. You can do a hand drawn plot, but this might be a good time to play around with a more sophisticated plot in R.
       *Comment on how sex and condition appear to affect cooperation based on your plot.
   
   (b) Get the 4 cell means.

   (c) After coding the factor dummy regressors using sum-to-zero constraints, fit a model that includes interaction and provide the Anova table.

   (d) State your conclusions to the questions of interest for this model. Use full sentences and write in terms of the research at hand(i.e. in the context of the problem), not just whether an effect is ‘significant’ or not.

More on next page...
2. Brewer’s malt is produced from germinating barley, and they’re interested in knowing under what conditions they should germinate their barley. This problem relates to data collected on germination rate of barley (a continuous variable) at 2-levels of ‘water used’ (4 ml or 8 ml) and at 5-levels of ‘weeks after planting’ (1, 3, 6, 9, or 12).

(a) Get the data set called Barley.csv from the class website under ‘Data sets’.

As Water and Timing are numeric, change them into factors for the analysis:

```r
> Water=as.factor(Water)
> Timing=as.factor(Timing)
```

(b) Provide an interaction plot that places Timing on the x-axis:

```r
> interaction.plot(Timing,Water,Germination)
```

Does it look like there is interaction between Timing and Water based on the plot? Provide the plot and give a relevant comment.

(c) Fit the model that includes interaction and test for significant interaction, using the sum-to-zero constraints and the ‘Anova’ function.

Provide the ‘Anova’ output, and give the conclusion of the test for interaction.

(d) Fit the simpler ‘additive’ model. Provide the Anova output, test for main effects, and give the conclusions.

(e) Use either the estimated coefficients from the model, or the fitted values to provide the following:

The estimated mean germination rate for barley at...

1) a Water level of 4 and a Timing of 1 week after harvest
2) a Water level of 6 and a Timing of 3 week after harvest

**NOTE**: Timing of 3 weeks is the Timing factor level 2 (not 3).

(f) As this is a balanced design, we can use the TukeyHSD function to do pairwise comparisons of different factor levels with a multiple comparison adjustment. Use the following code to determine which of the 5 levels of Timing are significantly different from each other. Provide output and your conclusion.

```r
> TukeyHSD(aov(lm(Germination ~ Timing + Water)),"Timing")
```

More on next page...
3. Many local newspapers sell advertising. At our class website, there is a data set called `newspaper_advertising.csv` showing the relationship between the number of Inquiries coming from ads based on Day of the week (Monday through Friday) and Section of the paper (business or news) that the ad appeared for a sampling of newspapers.

(a) SET-UP THE DATA IN R:
* Re-order the factor levels so they coincide with the days of the week.
* Change the dummy regression coding to *sum-to-zero* constraints.
* Load the *car* library because we will use some of its functions.

(b) Provide the interaction plot. Place the Day factor on the horizontal axis, and let Section be the trace factor in the plot. Based on the plot alone, does it look like there is interaction between these two factors?

(c) Test to see if there is significant interaction. Provide the relevant anova table, and give your conclusion to the test.

(d) Use the process we used in notes (two-way ANOVA part 3 slides 7-10) to test if there is a significant difference in the number of inquiries between the two sections of the paper on a monday. Give the R output, a p-value, and a conclusion to the test.