STA 2101/442 Assignment Twelve[1]

Do these questions in preparation for the final exam. The R questions will not be on the final, but they provide some guidance about what to do with the data. So if you do them, you will be able to guess what I will do with similar data sets on the final exam.

1. In a dichotic listening experiment, subjects wear stereo headphones that allow the presentation of different sound tracks to each ear, at the same time. In this example, right-handed female university students listened to short lectures on art history in the presence of background noise. After each lecture, they answered a set of multiple choice questions.

   Two factors were varied experimentally:

   - Noise Type: The background noise consisted of either Hip-hop music, Classical music or Radio commercials. Volume was carefully held constant.
   - Presentation: Subjects heard (simultaneously) either
     - Lecture (Signal) in the left ear and distraction (Noise) in the right, or
     - Distraction (Noise) in the left ear and lecture (Signal) in the right, or
     - Both Signal and Noise in both ears

   Each subject in the experiment experienced all nine treatment combinations, in a balanced order that was different for each subject, and randomly assigned. Thus, there are nine data values for each subject: number of questions answered correctly in each experimental condition. The layout of the data is given below:

```
          Signal in          Signal in           Signal in
          Left Ear           Right Ear           Both Ears

     -------------------- -------------------- --------------------
     HipHop Classc Radio  HipHop Classc Radio  HipHop Classc Radio
     ------ ------ ------ ------ ------ ------ ------ ------ ------
     test11 test12 test13 test21 test22 test23 test31 test32 test33
1      13     12     10     15     14     14     14     13     14
2       4      8      8      6      5      8      6      3      4
3      13     15     11     11     13     15     11     13     12
```

   Data are given in the file Dichotic.data. There is a link from the course home page in case the one in this document does not work.

   (a) Test for both main effects and for the interaction. Be able to give the F statistic, the degrees of freedom and the p-value.

   (b) If a main effect is significant, calculate the marginal means and test for pairwise differences of marginal means, with a Bonferroni correction. Remember, Hotelling's T-squared for a single contrast is equivalent to a matched t-test. Be able to state conclusions (if any) in plain, non-statistical language.

   (c) If the interaction is significant at $\alpha = 0.05$ (and only then), make a two-way table of treatment means. Conduct tests of contrasts that will allow you to understand the interaction and describe it in plain, non-statistical language. Again, only do this if you reject the null hypothesis of no interaction.

[1] Copyright information is at the end of the last page.

2. In a test of how well people remember instructional materials, subjects were presented with training materials that were either in Black and White or in Colour. Their ability to recall the material was tested with both Cartoon and Realistic testing materials at two points in time – immediately after training, and several weeks later. The variables are

   - Colour versus Black and White training materials
   - Cartoon1: Recall at Time One, Cartoon testing materials
   - Real1: Recall at Time One, Realistic testing materials
   - Cartoon2: Recall at Time Two, Cartoon testing materials
   - Real2: Recall at Time Two, Realistic testing materials

   Data are given in the file `cartoon2.data`. There is a link from the course home page in case the one in this document does not work.

   (a) Think of this as a three-factor design. Which factors are between cases and which are within?

   (b) Fortunately, the within-cases factors have only two levels, so you never need to test several difference variables simultaneously. This means you don't need `manove`; you can get away with purely univariate analysis of variance. Think about this.

   (c) Your goal is to test all the main effects and interactions. In every case, you are doing a regression analysis with effect coding, and the response variable is *computed* from the data in the data file. For each effect (how many are there?) specify the response variable, and give the null hypothesis in terms of β values. Which of the tests are about the intercept?

   (d) Produce all tests of the main effects and interactions. For each one, be able to give the F statistic, the degrees of freedom and the p-value. You might as well make a table.

   (e) Interpret all results in plain, non-statistical language. You will have to calculate some treatment means to figure out what is going on.

   (f) This data set is similar to one of the data sets for the final exam. Which one?

3. The Ontario government selects a random sample of q grade schools. From each school, a random sample of k students is selected and given a reading test. School is the explanatory variable. Because the values of this variable represent a random sample from a larger populations, a *random effects* model is appropriate. A standard version that applies to this situation is

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where

μ is an unknown constant parameter.

$\tau_i \sim N(0, \sigma^2_\tau)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$.

τ_i and ϵ_{ij} are all independent, $i = 1, \ldots q$ and $j = 1, \ldots, k$.

(a) What is the distribution of Y_{ij}? Just write down the answer. You need not show any work.

(b) Are the Y_{ij} all independent? Consider two cases.

(c) What is the distribution of $\overline{Y}_i = \frac{1}{k}\sum_{j=1}^k Y_{ij}$? State your answer; the only work you need to show is your calculation of the variance.

(d) Find $Cov(\overline{Y}_i, Y_{ij} - \overline{Y}_i)$. Show your work.

(e) Define $SSTR = k\sum_{i=1}^q(\overline{Y}_i - \overline{Y}_.)^2$, where $\overline{Y}_. = \frac{1}{q}\sum_{i=1}^q \overline{Y}_i$. Find the distribution of $\frac{SSTR}{\sigma^2+k\sigma^2_\tau}$. Hint: You can make this a very easy problem. What is the joint distribution of $\overline{Y}_1, \ldots, \overline{Y}_q$?

(f) Define $SSE = \sum_{i=1}^q \sum_{j=1}^k (Y_{ij} - \overline{Y}_i)^2$. Find the distribution of $\frac{SSE}{\sigma^2}$. Again you may use a well-known fact to make the problem easier, but *do not forget your answer to 3b*.

(g) The proportion of variance in an observation Y_{ij} that is explained by School is $\frac{\sigma^2_\tau}{\sigma^2_\tau+\sigma^2}$. Give a reasonable estimator for this quantity; show some work.

(h) What null hypothesis would you use to test for the effect of school on students' reading scores? State the null hypothesis in symbolic form; that is, it's a statement in terms of Greek letters.

(i) An exact (not large-sample) test is available for this hypothesis. Give a formula for the test statistic. Also state its distribution under H_0, including the degrees of freedom. Briefly indicate why it has the distribution you claim.

(j) Show that the power of this test is based on a *central* rather than a non-central F distribution.

(k) Suppose that the government's primary interest is in testing whether School has any effect at all on average reading score. Since resources are limited, would you advise the government to spend money on sampling more schools, or more students per school? Why?
