STA 2201S 2001 Assignment 5

Quiz on Thursday March 1st. Bring printout to the quiz.

1. Suppose a social scientist designs a factorial experiment with all combinations of 5 factors. To make this question a bit more convenient for you to answer, let it be a 2x2x2x2x2 design, with **n=5** per cell. The scientist will do a standard analysis using a linear model with normal independent errors. Naturally, he will look at all five main effects, all the two-factor interactions, all the 3-factor interactions, and so on. If even one test is significant at the 0.05 level, he will figure out what the results mean and submit a paper for publication.

   a. How many tests is the scientist planning to examine?

   b. Suppose that in reality, all of the null hypotheses are true; that is, the experiment was a perfect failure. Assuming the model assumptions of normal independent errors and equal variances to be satisfied, what would the probability be of rejecting at least one hypothesis just by chance, *if* all the tests were independent?

   c. But the tests are not quite independent, are they? The numerators are all independent, but they have the same denominator. To see this, consider just a 2x2 design with n=2 per cell, and realize that in this case, like the big one we are really interested in, we can test all the hypotheses with t-tests, provided we set up the dummy variables using effect coding (the scheme with the –1s). The numerator of each t-test is the least-squares regression coefficient, an element of $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$.

      i. What is the distribution of $\hat{\boldsymbol{\beta}}$ -- for any regression model, not just for this particular one?

      ii. Now for the particular case of this 2x2 design with n=2 per cell, write down the $\mathbf{X}'\mathbf{X}$ matrix. Why does this tell you the numerators are independent? Are you convinced about the bigger case?

      iii. What is the denominator?

   d. Since we can't easily compute the exact probability of rejecting at least one null hypothesis, we will get it by Monte Carlo. All the treatment means are equal, and since the null hypotheses all refer to differences between means, it does not matter what the common mean is. The variance does not matter either (why?). So do a simulation; choose any mean and variance you want. Let the Monte Carlo sample size m = 1,000. Give a 99% confidence interval for the probability of interest.

2. In the production of an industrial chemical, the yield (Y) depends upon cooking time (A) and cooling time (B). Two cooking times and two cooling times are to be compared in a two–way factorial design. The cell means model for this analysis would be $Y_{ijk} = \mu_{ij} + \varepsilon_{ijk}$, $i = 1, 2$; $j = 1, 2$, where μ_{ij} are unknown constants and ε_{ikj} are independent $N(0,\sigma^2)$ random variables.

a. Write this as a multiple regression model with **effect coding** dummy variables; include an intercept and the interaction. You do not need to define how the dummy variables are coded; you are asked to do that in the next part of the question. In fact, all you need to do is complete this:

$$E[Y] =$$

b. (6 pts) Now complete the table below; use **effect coding**.

Cooking Time	Cooling time			
1	1			
1	2			
2	1			
2	2			

c. (9 pts) Now fill in the table below, expressing μ's in terms of the β's of your multiple regression model. $\mu_{\cdot 1}$
is the mean of μ_{11} and μ_{21}, etc..

	Cooling Time = 1	Cooling Time = 2	
Cooking Time = 1	$\mu_{11} =$	$\mu_{12} =$	$\mu_{1\cdot} =$
Cooking Time = 2	$\mu_{21} =$	$\mu_{22} =$	$\mu_{2\cdot} =$
	$\mu_{\cdot 1} =$	$\mu_{\cdot 2} =$	$\mu_{\cdot\cdot} =$

3. The table below shows the <u>whole data file</u> for a hypothetical experiment including two factors and a blocking variable; that is, it is a randomized block design. Please fill in and label columns containing dummy variables for the blocking variable and also for main effects and interactions between A and B. By "label" I mean write variable names in the empty boxes in the top row. Write numbers in the other boxes. Use effect coding (the 0, 1, –1 scheme). Remember that in a randomized block design, there are no interactions between factors and blocks. You are NOT being asked for E{Y} yet.

Case #	Y	Block	A	B	
1	Y_1	1	1	1	
2	Y_2	1	1	2	
3	Y_3	1	1	3	
4	Y_4	1	2	1	
5	Y_5	1	2	2	
6	Y_6	1	2	3	
7	Y_7	2	1	1	
8	Y_8	2	1	2	
9	Y_9	2	1	3	
10	Y_{10}	2	2	1	
11	Y_{11}	2	2	2	
12	Y_{12}	2	2	3	

Now write E[Y] in each cell below, in terms of your β values.

	Block One		Block Two	
	A=1	A=2	A=1	A=2
B=1				
B=2				
B=3				

In terms of your β values, what null hypothesis corresponds to the claim that, averaging across A and blocks, the mean response to the three levels of B is the same?