Exam in MSA830 Experimental design

Friday March 13th 2009, 8:30 – 13:30

Allowed during the exam: An optional calculator, and one single page of your own notes.

Number of points on the exam: 30. To pass the exam, at least 12 points are needed.

1. Eric is comparing 4 types of plant fertilizers, named A, B, C, and D. He has a quadratic plot, subdivided into 16 quadratic subplots, in a 4x4 pattern, where he can try out how well a certain flower grows with the different fertilizers. There may be some differences in the growing conditions of the subplots in addition to the fertilizer used: For example, the amount of water in the ground could vary across the sloped plot, as could the soil quality.

   a) Eric wants to use one fertilizer on each subplot, and he wants to vary the use so that he can best estimate the differences between the fertilizers. Which of the following experimental plans would you recommend that Eric followed? Give an argument why your choice is the best (2 points):
   - The four first plots, in the first row, are grown with A fertilizer, the next four with B fertilizer, the next four with C fertilizer, and the final four with D fertilizer.
   - For each of the 16 subplots, he randomly chooses one of the 4 fertlizers to apply to the subplot.
   - He applies fertilizer A so that it is applied on exactly one subplot in each row and exactly one subplot in each column of subplots. He also applies fertilizers B, C, and D in the same ways. In this way, each fertilizer is applied exactly four times.
   - He applies fertilizer A to 4 randomly selected subplots, then fertilizer B to 4 randomly selected of the remaining subplots, then fertilizer C to 4 randomly selected of the remaining subplots, and then fertilizer D to the remaining 4 plots.

   b) Eric in fact chooses the second of the experimental plans above, and he ends up applying fertilizer A to 5 plots and fertilizer B to another 5 plots. Fertilizers C and D are quickly seen to be inferior, so he focuses on comparing the yields for A and B. These are:

   A: 40, 45, 33, 31, 41

   B: 32, 25, 27, 39, 37

   Estimate the average change in yields when using fertilizer A instead of B. Also, make a t-test of whether the change in yields is statistically significant: Choose the right t-test and estimate the corresponding p-value. (4 points)

2. Anna wants to investigate how 7 different factors influence the output in an industrial process she is studying. Each of the 7 factors, A, B, C, D, E, F, and G, is set in each experiment at a high level (denoted +) or a low level (denoted -), and the experiment is performed. Anna initially has the possibility to perform only 8 experiments, but she still wants to do it in a way so that the main effect of each of the 7 factors can be independently estimated. She has started to fill out an experimental plan below.

a) Fill in missing ”+” and ”-” symbols in the experimental plan (2 points):

A	*B*	*C*	*D*	*E*	*F*	*G*
-		-			-	
-		-			+	
-		+			-	
-		+			+	
+		-			-	
+		-			+	
+		+			-	
+		+			+	

b) She performs the experiment in the exact order of the lines in the table above, and gets the results 43, 56, 24, 98, 34, 52, 38, and 81, respectively. Compute the main effect of A and the main effect of F. (2 points)

c) Is there a potential problem with the way she performed the experiments in b? Describe this problem, and suggest a better way to perform the 8 experiments. (1 point)

d) In order to compute better estimates for various effects, Anna decides to do additional experiments. If she wants to try out all possible combinations of the 7 factors, how many experiments must she perform in total? (1 point)

e) Assume Anna instead wants to compute not only the main effect of A as in part (b) above, but also a confidence interval for this effect, and that she wants to be able to do this after performing only 8 additional experiments. Which experiments would you recommend Anna to perform? (2 points)

3. Harri investigates how different types of mice respond to several types of growth hormones. He applies 3 types of different hormones on 7 different types of mice, with 10 mice of each type receiving each type of treatment. Thus, his experiment contains a total of 210 mice, and he records the response for each mouse. He uses an ANOVA table to analyze the data.

a) Fill out the missing numbers in the white rectangles in the table below (2 points):

	Sum of squares	Degrees of freedom	Mean square	F ratio	p-value
Between hormones			60		
Between mice			30		
Residuals (errors)			15		
Total					

b) What assumptions would Harri need to make, in order for the analysis above to be meaningful? How could he check these assumptions? (1 point)

c) Making these assumptions, what can Harri conclude from the analysis above? (1 point)

4. Jenny wants to investigate the the effect of two types of running shoes on the results of athletes. 7 different athletes each try out the two shoe types, randomly choosing which type to try out first. The results are

	Athlete 1	Athlete 2	Athlete 3	Athlete 4	Athlete 5	Athlete 6	Athlete 7
Type A	103.5	111.2	121.1	101.1	112.9	119.4	102.1
Type B	102.3	110.9	122.0	100.0	110.2	116.2	102.3

Describe one possible test that could be used to analyze these data. Describe the formulas you would use in the test, what assumptions you would need to make for the test to be valid, and how you would make the computations; you do not have to do the actual computations. (3 points)

5. Selma buys a lottery ticket every day. Each lottery ticket she buys has a 15% chance of at least some kind of winning.

a) If she buys a ticket every day in a week, what is the probability that she will win on Tuesday and Thursday, but not on any of the other 5 days? (1 point)

b) What is the probability that she will win on exactly 2 out of the 7 days that week? (2 points)

c) If she plays each day a whole year (of 365 days), what is the approximate proabability that she will win on less than 20 days? (2 points)

6. Hans is recording the daily output of a pollutant from a factory. The daily output varies, but does not seem to be independent from day to day; there is a correlation between one day's pollution and the next day's pollution. One day, the factory installs a new system, which it claims reduces the average pollution. After running the new system for 30 days, Hans wants to test the claim, by somehow comparing the average of the pollution of these 30 days with the data he has recorded for every day for the last 2 years.

   a) Please give a suggestion of how Hans could perform a computation to test the claim that the new system reduces pollution. (3 points)

   b) What assumptions do you need to make in order for the test you propose above to be valid? (1 point)