Exam in MSA830 Statistical analysis and Experimental design

Friday October 23rd 2009, 8:30 – 13:30
Jour: Petter Mostad, who will be available for questions about
the formulation of exam questions at 9:30 and 11:30
Allowed during the exam: An optional calculator,
and one single page (writte on one side) of your own notes.
Number of points on the exam: 30. To pass the exam, at least 12 points are needed.

1. Alice is the manager at a chemical factory, and she is studying two different processes to manufacture a chemical. With process A she has observed the yields of 5, 7, 8, and 4, while with process B, she has observed the yields of 8, 10, 7, 11, and 9. She wants to make a hypothesis test of whether process B in general has a different yield than process A.

   a) What are the assumptions Alice has to make in order to obtain what she wants with a t-test? (1 point)

   b) Making these assumptions, choose a t-test and do a hypothesis test with a 5% significance level. Also, compute a 95% confidence interval for the differences in yields between the processes. (3 points)

   c) In fact, process A is the standard process that has been used a long time at the factory. The 4 numbers for process A above were the yields of the last four runs with this process, while the following 5 runs were made with process B, resulting in the 5 yields reported above. Is there a reason to doubt the results in b? Why? (1 point)

   d) It turns out that Alice can get access to the yields of several hundred consecutive runs of process A. Suggest a way to perform a better hypothesis test based on this data and on the 5 yields for process B. (1 points)

2. Anna has performed an experiment where she investigated the influence of three factors, A, B, and C, on some output. Each factor was investigated at 2 levels, indicated with '+' and '-' below. She performed 24 independent experiments, in a randomized order, according to the experimental plan below, with 6 experiments performed for each line of the table.

A	***B***	***C***	***Sum of 6 replications***	***Variance of 6 replications***
-	-	+	439.2	4295
-	+	-	219.7	1740
+	-	-	331.3	3855
+	+	+	192.9	873

   a) Compute the estimate for the main effect of factor B. (1 point)

   b) Compute a 95% confidence interval for the estimate in a. What assumptions do you need to make in order for this confidence interval to be valid? (2 points)

   c) Write down all pairs of main and interaction effects that are confounded in this design. (1

point)

d) Given that Anna has the time and resources to perform 24 independent experiments, write down in detail an alternative experimental plan which would allow her to estimate independently all main effects and all interaction effects. (1 point)

3. Paul wants to investigate the effect of 5 different soil types and five different tomato plant seeds on the yield of the tomato plants, in terms of the weight of tomatoes produced. Thus he would like to do 25 experiments. However, he has room for only 5 plants on his balcony, so he enlists the help of four of his friends, with the idea that each of them should try out 5 of the possible combinations of seeds and soil. Make a detailed proposal of how the different possible experiments should be divided among the friends. Also, make a comment of why your proposal is better than some other possible ways to perform the experiment. (2 points)

4. Albert has what he assumes are random samples of 10 values from a population A and 12 values from population B. The variance of the values from population A is 5.9 and the variance of the values from population B is 18.8. Albert is unsure about whether the two populations have the same variances.

   a) Make a hypothesis test for this question, and compute an approximate p-value. (1 point)

   b) What assumptions does Albert need to make in order to trust the results of this hypothesis test? Suggest a plot he can make to convince himself and others about one of these assumptions. (1 point)

   c) If Albert finds that he cannot reasonably make the assumtions you mention above, suggest an alternative test for the null hypothesis that the two sets of values come from the same population. (1 point)

5. Alan has investigated the effect of 5 different color schemes, and 4 different font types, on the click-through rates of his internet ad. Testing all 20 combinations, he has produced the results in the table below. For your convenience, the table also contains the averages of the rows and columns, and the grand average. Alan has also computed that the variance of his 20 observations is 29.47386.

	Font 1	Font 2	Font 3	Font 4	**AVERAGE**
Color A	*32*	*40*	*25*	*27*	**31**
Color B	*35*	*37*	*40*	*32*	**36**
Color C	*29*	*35*	*35*	*33*	**33**
Color D	*27*	*28*	*29*	*36*	**30**
Color E	*37*	*45*	*36*	*42*	**40**
AVERAGE	**32**	**37**	**33**	**34**	**34**

   a) Assuming that there is no interaction between the color schemes and the font types, make an ANOVA table for Alan's data, and find approximate p-values for each of the two null hypotheses that the color scheme has no effect and that the font type has no effect. (4 points)

   b) What other assumptions does Alan need to make to trust the p-values found in a? (1 point)

6. Assume Alex is planning to perform a sequence of independent experiments, each with a probability *p* of success.
   a) Write down the formula for the probability that Alex will observe the following sequence of successes and failures (1 point):

   Success, Failure, Failure, Success, Success, Failure, Success, Failure, Success, Success.

   b) Assume that in order to complete his project, Alex needs a total of 6 successful experiments. Write down the formula for the probability that he will have do endure exactly 4 unsuccessful experiments before he has conducted the sixth[1] successful experiment (2 points).

   c) Write down the formula for the probability that Alex will have to endure k failed experiments before he has finished r successful experiments (2 points).

7. For each of the three statements below, explain why you believe it is true or false:

   a) As the sample size increases, the distribution of a sample from any distribution becomes more and more like the normal distributuion, because of the central limit effect. (1 point)

   b) If you take random samples of size 100 from a uniform distribution, their means will be approximately normally distributed, because of the central limit effect. (1 point)

   c) The t statistic computed when comparing the means of two distributions is approximately normally distributed when the sample sizes are large, because of the central limit effect. (1 point)

8. Abdul wants to study the average amount of money people in a specific population spend on clothes every year. After making 50 independent measurements, he has an estimate for the population average, and a confidence interval with a width of 230 kronor. If Abdul goes on making 150 additional independent measurements, *approximately* how wide would you expect his confidence interval to be when it is based on the total of the 200 measurements? Explain why you would expect this. (1 point)

1 In the original exam, there was a misprint here: It was written ”fifth” instead of ”sixth”.