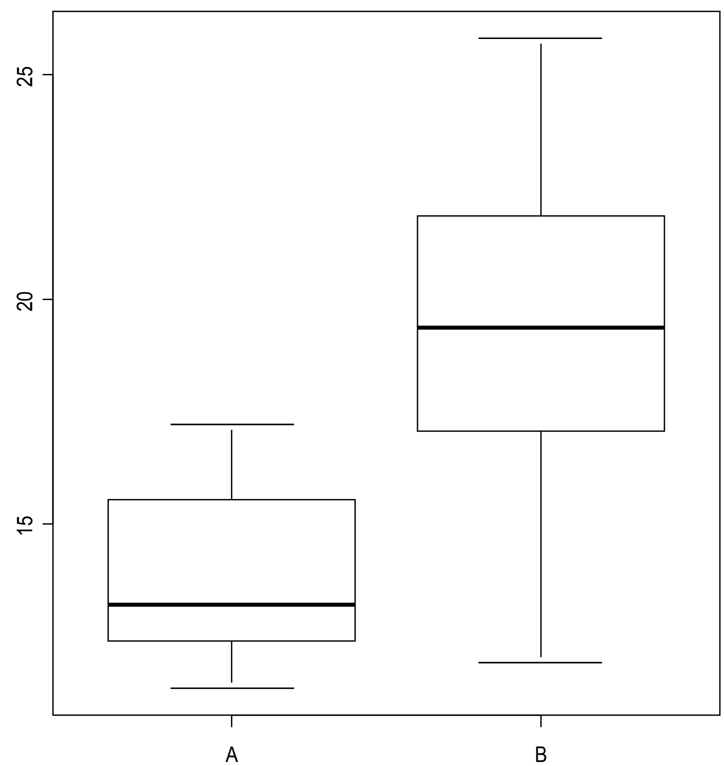
Exam in MSA830 Statistical Analysis and Experimental Design

Friday October 24th 2008, 8:30 – 13:30
Jour: Petter Mostad, who will be available for questions about
the formulations of the exam questions at 9:30 and 11:30.
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. The growth of a particular type of plant was investigated by measuring the weight after 10 weeks for 20 plants growing in condition A and 20 plants growing in condition B. The collected data are illustrated in the box plot below:



a) For each of the following statements: Which are true, which are false, and which do you not have enough information to decide about? Explain in each case. (3 points)

1. The 15 plants that grew best among those groing under condition B grew better than the 15 plants that grew worst among those growing under condition A.
2. The variance of the weights of plants growing under condition A was 148.3.
3. The condition A seems to be better for the plants than condition B.

b) In order to analyse statistically which growing condition was best for the plants, what kind of test would you use? If $a_1, a_2, ..., a_{20}$ are the weights under condition A and $b_1, b_2, .., b_{20}$ are the weights under condition B, explain in detail which formulas you would apply to these numbers, and how you would obtain a p-value in this situation (2 points).

2. a) What is meant by the central limit effect? Explain in a couple of sentences what this effect is, and why it is important in statistics (1 point).

   b) What is a normal probability plot? How is it constructed, and how can it be interpreted? Mention at least one situation in which it can be useful (2 points).

3. Hector is doing a factorial experiment with 4 factors at 2 levels. He uses + and – to indicate the levels in the plan below. He plans to use 8 different combinations of settings of the factors, and the first 6 lines of his experimental plan looks like

X	***Y***	***U***	***W***
-	-	-	-
-	-	+	+
-	+	-	+
-	+	+	-
+	-	-	+
+	-	+	-

   a) Make a proposal for the last two lines of his experimental plan, so that he kan make an efficient investigation of the effects of the different factors. What can you call his experimental plan (in the notation 2_*^{*})? Also, write down two two-way interactions that are confounded in this design (2 points)

   b) With the full experimental plan as you suggest it, Hector makes 7 replications at at each combination of settings, and obtains the following means and sample variances of the results:

Mean	***Sample variance***
14.2	2.5
21.3	2.0
4.7	0.7
10.1	1.7
6.0	0.3
14.3	1.2
5.2	0.7
19.9	1.3

   Compute the main effect of U, and the interaction effect of X and Y (2 points).

   c) Compute a 95% confidence interval for the effect of U computed above (2 points).

d) After analysing the results, Hector realizes that his experient does not allow him to estimate all interaction effects independently. He would like to extend the 56 experiments above with 56 additional experiments, again using 7 replications at 8 different combinations of factors, in such a way that he can now estimate all interaction effects independently. Make an extended experimental plan for Hector (1 point).

4. A research lab is doing a test production of a technological component. 20% of the produced components end up being defective in their experimental production process. The lab produces 10 components each day.

   a) What is the probability that none of the produced components will be defective on a given day? What assumption do you need to make in order to do this computation? (2 points)

   b) Making the same assumption, what is the approximate probability that 6 or more components will be defective that day? (1 point)

   The component becomes a success, and the production process is improved so that 10 000 components are produced each day, but still with an average of 2 defective components each day.

   c) What is the now the probability of exactly 2 defective components on a given day? (1 point)

5. Jack is investigating whether the braking distance for bikes depends on the type of tyre material on the bike wheel. He uses 4 test bikers, who each test each of 3 possible materials 10 times. He analyses his results using ANOVA, producing a table like the one below:

   a) Fill in numbers where there are dots in the table. What would be Jack's conclusion from this experiment? (2 points)

	Sums of squ.	***D. freedom***	***Mean squ.***	***F***	***p***
Tyre	.	.	.	.	.
Biker	87	.	.		
Interaction	13	.	.		
Error	93	.	.		
Total	234	.			

   b) Jack was smart and randomized the order in which the experiments were performed. Discuss shortly what kinds of problems might have occurred if he had performed all the experiments sequentially according to his experimental plan. (1 point)

   c) Jack could also have viewed his experiment as a test of 3 possible materials, with 40 replicated experiments for each tyre material. Write down the ANOVA table which would result in this case, and draw conclusions from it. (2 points)

6. Sara has a farm, and is comparing the weight gain of 4 pigs getting feed type A, with the weight gain of 4 other pigs getting feed type B. The results are 43, 20, 35, 39 for feed type A and 40, 41, 34, 49 for feed type B.

   a) Compute an estimate for the added weight gain a pig would get on feed type B instead of on feed type A, and find a 90% confidence interval for this difference. (3 points)

   b) Assume that Sara has a lot of experience with various feed types, and that she knows from other data that the standard deviation of the weight gain under any of these feed types can be assumed to be 8.0. What would change in the computation in a? What would be the confidence interval now? (1 point)

   c) Assume Sara then decided that she needed a more accurate estimate of the difference; she decides she needs a confidence interval that has only a third of the length of the interval computed in b)[1]. How many additional experiments would she need to perform? (2 points)

1 In the original exam, there was a printing error, with "d" here instead of "b".