Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 10 March 2010
Examiner: Petter Mostad, phone 0707163235,
visits the exam at 9.30 and at 11.30.
Allowed to use during the exam: Pocket calculator, books, copies, and notes
Number of points on the exam: 30. To pass the exam, at least 12 points are needed

1. Susan is investigating whether her patient X has a certain gene that may cause hypertension. In fact, it is known that 18% of patients without the gene have hypertension, while 89% of those with the gene has hypertension. In general, 12% of the population carries the gene. What Susan knows about her patient X is that he has hypertension. What is the probability that he is carrying the particular gene? (2 points)

2. Morten is trying out a new hand control that he thinks will make it easier to play a certain type of videogame. To investigate whether it is better or not, he recruits 6 friends who do not have experience with either the old or the new hand control. Each friend plays for one hour with each of the two controls, and Morten chooses randomly whether they use the old or the new control the first hour. The resulting scores are:

	Friend 1	Friend 2	Friend 3	Friend 4	Friend 5	Friend 6
Old control	713	421	616	1002	568	729
New control	778	569	655	845	700	695

   (a) Compute a 95% credibility interval for the number of points the new control is expected to increase any person's score. (Assume that the increases in scores for each person is normally distributed[1].) (2 points)

   (b) After analyzing the results, Morten realizes that there is quite a strong learning effect: His friends tend to get higher scores during the second hour of testing. In fact, by chance, three of his friends tested the new control first, and three tested the old control first. Describe an alternative way to analyze the data, which would take the learning effect into account. (You don't have to make any computations) (2 points)

3. Emelie is trying to assess the evidence for whether large earthquakes are more common now than before. She has determined that, between 1950 and 1990, there were an average of 14 earthquakes in the world over a certain magnitude. Se initially decides to model earthquakes as independent, rare events.

   (a) If the true rate of large earthquakes is 14 every year, what is the probability for exactly 5 large earthquakes next year? (1 point)

   (b) If the true rate of large earthquakes is 14 every year, what is the approximate probability of 20 or more earthquakes next year? (2 points)

[1] This sentence has been slightly changed from the original exam

4. In a class on experimental design, all students are going to do the same experiment, investigating the flying properties of paper airplanes. Each student will do the following:

   The student will investigate two factors: The first factor is paper type, which can be thick and thin. The second factor is folding design, which can be type A, B, or C. For each of the 6 possible combinations of these factors, each student constructs 2 planes. For each constructed plane, the student measures the average of 10 flying distances. This gives the student 12 values to analyze. The values are to be analyzed with an ANOVA table.

   (a) Laila is quick with her experiments, and with the first part of her ANOVA table, given below.

	Sum of squ.	Deg. freed.	Mean squ.	F value	p value
Thickness	11.8				
Fold design	116.6				
Interaction					
Residuals	56.0				
Total	195.6				

   Fill out the missing parts of Laila's ANOVA table. (2 points)

   (b) Interpret the results from the ANOVA table, and make the conclusions you can make. Is it possible from the information in the ANOVA table to conclude which folding design (A, B, or C) gives the longest-flying paper airplane? (2 points)

   (c) Hans is a bit slower than Laila, but he finally gets the 12 observed values given in the table below:

	Thick	Thin
Folding design A	13.0	12.8
	8.0	13.3
Folding design B	16.0	12.8
	11.0	12.6
Folding design C	11.4	19.9
	10.4	17.2

   Compute the values in the Sum of squares column of the ANOVA table for Hans' data. You do NOT have to compute more than the first column of this table. (4 points)

5. Dmitrii would like to find out how people like his new recipe for kebab-pizza, compared to a standard one. He is considering baking 5 pizzas of each type, and inviting 20 friends to a party for pizza tasting. Make a detailed set of recommendations about how he should perform the testing, in order for his conclusions to be as reproducible and as valid as possible. Consider such things as how he should bake the pizzas, how he should organize the tasting, how he should collect information from his friends, and exactly what information he should give his friends (3 points).

6. Patrik would like to investigate the influence of four factors, A, B, C, and D, on his measured output. Each factor has two possible levels or "settings". Patrik is planning to do a total of 8 experiments, and would like to construct an experimental plan so that the effect of each factor can be estimated separately.

   (a) Make an experimental plan for Patrik, listing for each of his 8 experiments the choices for the four factors. (1 point)

(b) When Patrik does the analysis of his data, there will be several interactions of effects that will be *confounded* with each other, i.e., they cannot both be independently estimated from his data. Write down at least one pair of interaction effects that are confounded if Patrik uses the design you propose. (1 point).

7. Emma is an engineer at a food processing plant, and is comparing the daily yields of two machines: X and Y. During 10 days, she tries out the two machines, using one for five randomly selected days and the other for the other days. Her results are given in the table below. For the analysis, she assumes that the yields from each machine follow a normal distribution.

Machine X	65, 78, 80, 77, 77
Machine Y	80, 104, 81, 86, 78

   (a) Find the expecation *and* a 95% credibility interval for the difference between the expectations of the two distributions, assuming that their precisions are the same. (2 points)

   (b) Find a 95% credibility interval for the *variance* in the two groups, under the same assumption. (2 points)

   (c) Having completed her computations, Emma starts to doubt her assumption that the precisions in the two distributions are the same. Make a hypothesis test testing whether the variances of the two distributions are the same, and estimate its p value [2]. Based on this, decide whether or not you would recommend Emma that the computations in (a) and (b) should be re-done under different assumptions. (2 points)

8. Mohammad is making experiments to determine the properties of a complex machine. He is controlling the values of three inputs, x_1, x_2, and x_3, and for each experiment, he is measuring the output y. He has done 100 experiments, and is trying to analyze the results with a multiple regression model, using x_1, x_2, and x_3 as predictors. A normal probability plot for the residuals from his initial model is given in Figure 1, and plots of the residuals against the predictors x_1, x_2 and x_3 are also shown in that figure. Which of the plots indicate a problem with the analysis? How would you describe this problem? Do you have a suggestion for an improved analysis? (2 points).

[2] This sentence has been slightly changed from the original exam, to make it clearer

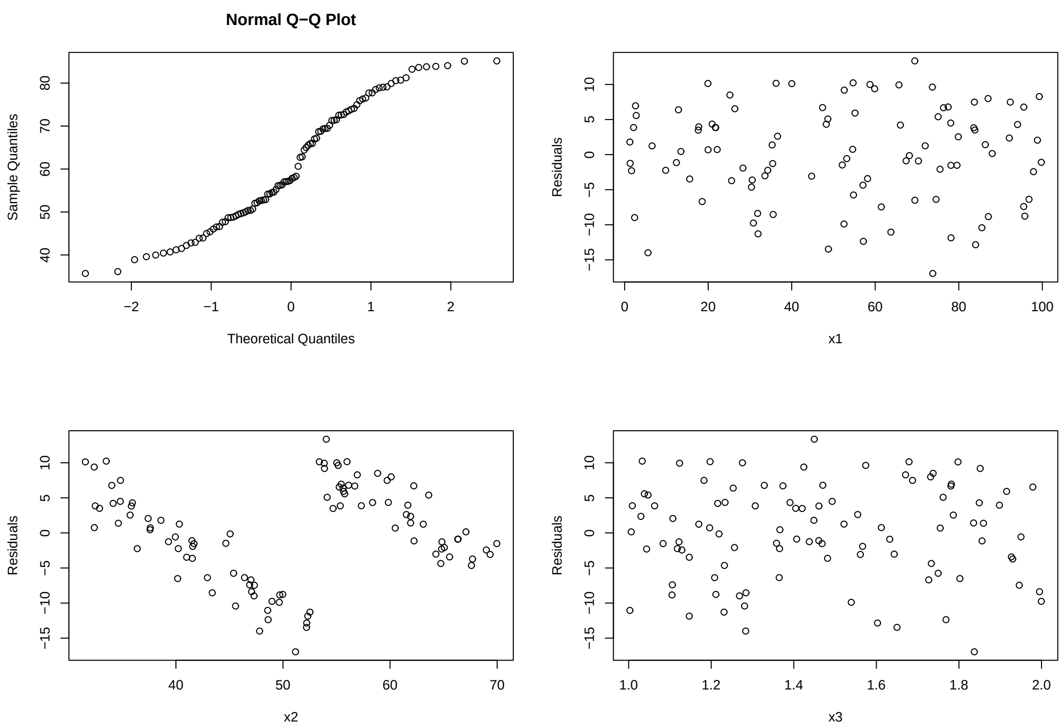

Figure 1: Plots of residuals from Mohammads analysis