Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 29 October 2011, 8:30 - 13:30
Examiner: Petter Mostad, phone 0707163235,
visits the exam at 9.30 and at 12.30.
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. Lisa often plays a card game with four of her friends. Lisa is quite good compared to the others, and they estimate that Lisa has a 40% chance of winning each game, while her friends each have a 15% chance of winning. Assume below that these probabilities are correct.

   (a) What is the probability that Lisa wins 8 of the 10 next games? (1 point)

   (b) What is the probability that she wins at least 8 of the next 10 games? (1 point)

   (c) What is the probability that the next 4 games are all won by different persons, non of whom is Lisa? (1 point)

2. Alexandra is a materials scientist, and she is comparing the strengths of 5 new experimental materials. For each she has made some repeated strenth measurements and computed summary statistics, with the results given in the table below:

	Sample size	Sample mean	Sample standard deviation
Material A	14	3.21	0.11
Material B	12	3.31	0.08
Material C	9	3.35	0.14
Material D	15	3.37	0.15
Material E	10	3.26	0.09

   Alexandra assumes that the observations for each material come from its own normal distribution; she does not assume anything about the parameters of these distributions.

   (a) Find a 95% credibility interval for the expected change in strength when changing from material A to material C. (2 points)

   (b) Find a 95% credibility interval for the standard deviation of the measured strength of material B. (2 points)

   (c) If Alexandra had assumed that the standard deviations for all the normal distribution were the same, what would be a 95% credibility interval for this standard deviation? (2 points)

3. Takashi is trying to determine the quality of animall feed, and in particular, he is trying to determine for each batch of feed whether the concentration of a certain bacterium is low, medium, or high. According to experience, about 0.7% of all batches have high

a concentration of the bacterium. As a quick test, Takashi measures the presence of a certain chemical. The chemical is present in 5% of all baches with low concentration of the bacterium, 35% in all batches with medium concentration, and 90% of all batches with high concentration. Given that Takashi detects the chemical in a batch, what is the probability that the batch has a high concentration of the bacterium? (2 point)

4. Mariann would like to compare two different skin lotions, X and Y. She does this by asking 7 persons (P1,. . . ,P7) to try out both lotions and give both a grade, on a scale from 0 to 100 points. The results are

	P1	P2	P3	P4	P5	P6	P7
Lotion X	85	30	95	50	83	90	44
Lotion Y	60	20	77	55	50	75	33

   (a) Compute a 90% credibility interval for the difference in customer satisfaction between lotion X and lotion Y. (2 points)

   (b) What assumption or assumptions did you need to make in order to do your computations in (a)? (1 point)

   (c) It is possible to make a hypothesis test where the null hypothesis is that there is no expected difference in customer satisfaction between the lotions, and where one does not make the assumption or assumptions discussed in (b). Describe how this hypothesis test works, and compute the test statistic on the data above. (2 points)

5. George is experimenting with a new procedure for measuring the activity of genes. It involves a large number of laboratory steps, and at the end George obtains a measurement of the quality of the result, among other things. George would like to investigate the influence of six different factors on the quality, where each factor represents one of the steps in the laboratory procedure. For each of these factors, George would like to test out two possible levels, i.e., ways to do that step in the procedure. He calls the factors A, B, C, D, E, F, and the levels for each factor '+' and '-'. He has the time and resources to do 32 experiments.

   (a) Write down a fractional factorial experimental plan (i.e., a table for how to set each of the factors in each of the 32 experiments). For full points, the plan should have the property that the inference from the experiment will not be influenced by which factor is assigned to which column in the plan (2 points).

   (b) Mention at least two important things George needs to remember when he performs his experiment, for his conclusions to be as reproducible as possible (2 points).

6. Axel is investigating how three different factors influence the flight lengths of his paper airplanes. The factors and their possible levels are listed below:

Factor	Levels
Design type	Type1, Type2, Type3, Type4
Paper thickness	Thick, Thin
Plane constructor	Axel, Eric

For each possible combination of these factors, 5 planes are constructed, and a flight distance for each of these 5 planes is found in an experiment. Thus Axel has a total of 80 values to analyze.

In the table below, the *average flight distance* for each of the 5 planes with the same factor settings are listed. The table also lists the averages for each of the design types, and for thick and thin paper. It was also computed that the average flight distance for all planes constructed by Eric was 8.55, while the average for Axel was 8.325. Finally, it was computed that the variance of all the 80 observations was 17.4644.

	Thick		Thin		Average
	Axel	Eric	Axel	Eric	
Type1	8.2	7.6	10.2	7.2	8.3
Type2	8.6	12.2	10.4	13.4	11.15
Type3	8.2	5.2	10.0	6.6	7.5
Type4	2.6	6.8	8.4	9.4	6.8
Average	7.425		9.45		8.4375

(a) Compute the following sums of squares: $SS_{\text{DesignType}}$, $SS_{\text{PaperThickness}}$, $SS_{\text{PlaneConstructor}}$, and SS_{Total}. (2 points)

(b) Make an ANOVA table without interaction. For each of the three factors, make a conclusion based on the ANOVA table and the table given above. (2 points)

(c) Make an ANOVA table including interaction. If you base a decision about whether or not to include interaction in the model solely on the p-value computed in this table, will you recommend to include interaction in the model? (2 points)

(d) Assume now that Axel had used the 16 averages in the table above as his observations, ignoring that they are computed as averages over several observations. Make an ANOVA table not including interaction for this situation, and draw conclusions about the three factors. (2 points)

7. Ingela is comparing the sizes of individuals of two different species. She has measured 200 individuals from each species. She has made a normal probability plot of all her 400 observations, shown in Figure 1.

(a) Which of the four histograms shown in Figure 2 are of the 400 observations shown in Figure 1: Is it dataA, dataB, dataC, or dataD? Why? (1 point)

(b) Ingela plans to analyze her observations by assuming that the observations for each species is normally distributed. Does Figure 1 indicate that this is an unsuitable model? Why, or why not? (1 point)

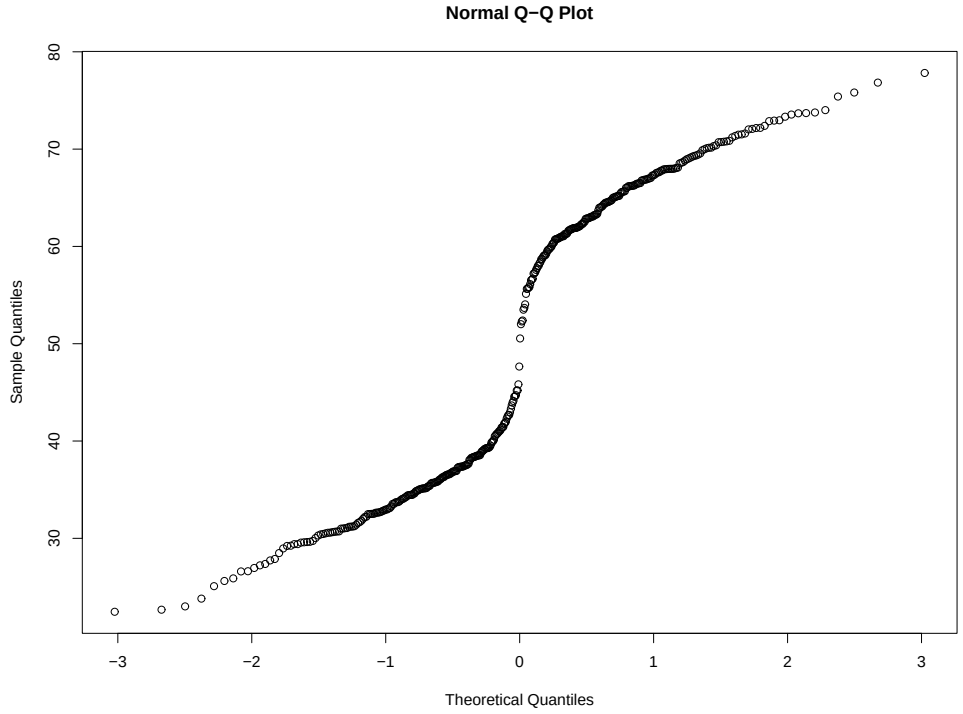

Figure 1: Normal probability plot showing all of Ingelas 400 observations

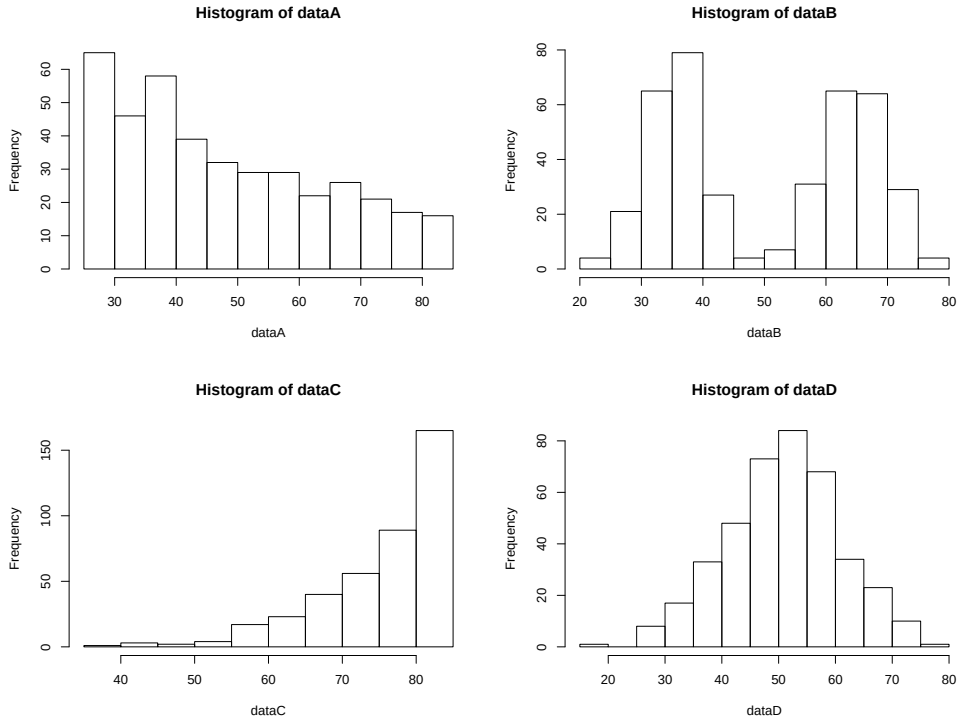

Figure 2: Histograms of the four data sets dataA, dataB, dataC, and dataD