Petter Mostad
Mathematical Statistics
Chalmers

MSG830 Statistisk analys och experimentplanering

Exam 15 March 2013, 8:30 - 12:30
Examiner: Petter Mostad, phone 0707163235,
visits the exam at 9:30 and at 11:30
Allowed to use during the exam: Pocket calculator
Number of points on the exam: 30. To pass the exam, at least 12 points are needed

1. Figure 1 contains boxplots comparing two datasets, A and B. For each of the statements below, select one of the following: (i) You can see from the plot that the statement is true. (ii) You can see from the plot that the statement is false. (iii) You cannot determine from the plot whether the statement is true or false. (3 points)

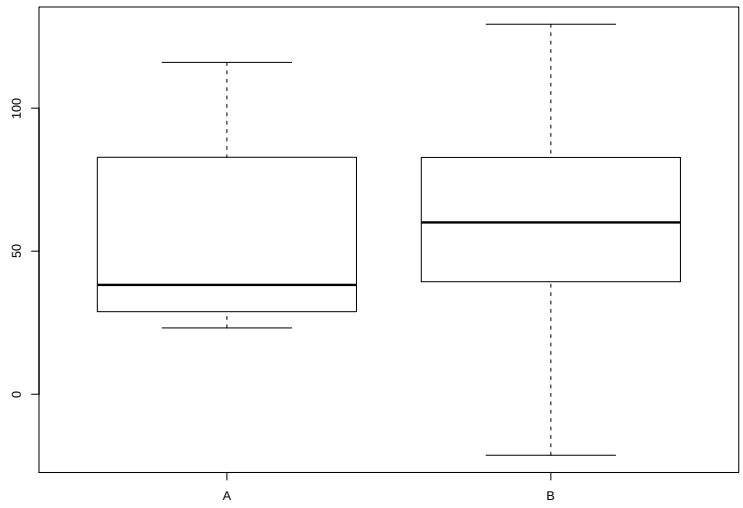

Figure 1: The datasets for question 1.

(a) The median for A is smaller than the median for B.
(b) The mean for A is smaller than the mean for B.
(c) The sum of the values in A is larger than the sum of the values in B.
(d) The smallest value in A is smaller than the smallest value in B.
(e) The variance for A is smaller than the variance for B.
(f) The median for A is smaller than 75% of the values in B.

2. Andrew is trying to find out if people in Gothenburg have a particular sense of humor. He has constructed 3 jokes, and in several Swedish cities in other parts of the country he asked randomly selected people which one they think is most funny. Out of those who selected one joke, 43% answered joke A, 12% answered joke B, and the remaining joke C. Andrew then asked 20 people in Gothenburg the same question, and out out these, 5 answered joke A, 6 joke B, and the remaining joke C. Do people in Gothenburg have a significantly different sense of humor than the rest of Sweden? Do a goodness of fit test: Compute the test statistic, and then determine whether the p-value is above or below 0.05. Describe the exact conclusions you can draw (include what your null hypothesis is). (3 points)

3. Charlie is participating in a national lottery game, where one out of every 10 tickets gives some kind of win.

   (a) If Charlie buys 11 tickets, what is the probability that he will win on exactly 3 of these? (1 point)

   (b) If he buys 120 tickets, what is the approximate probability that he will win on 20 or more tickets? (2 points)

4. Jacob is trying to determine the origin of his rock sample. He knows it is either from location A or from location B. He also knows that 30% these rock samples are from location A, while 70% are from location B. However, he has detected traces of copper in the sample. He knows that 8% of samples from location A have such copper traces, while only 0.6% of samples from location B have such traces. Based on this, what is the probability that Jacob's rock sample is from location A? (2 points)

5. Sara has a new sushi-place that will be open for lunch on weekdays, and she would like to maximize her sales. She would like to investigate the influence of prices and advertizing, but also of more subtle things like the amount of light in her restaurant, and the temperature. She decides to follow the plan below every week for 4 months, starting the 1st of April:

   Mondays: Better light, colder temperature. Tuesdays: Better light, colder temperature. Wednesdays: Normal light, normal temperature. Thursdays: Normal light, normal temperature. Fridays: Crazy Fridays: Two-for-one sale on all her sushi types (while keeping the light and temperature normal).

   In the middle of her 4-month experimental period, she plans to place several ads in the local newspaper. At the end, she will draw conclusions from the daily sales numbers from the 4 month period.

   Are there any problems with Sara's experimental plan? Explain in detail. Do you have advice for how she might improve it? (3 points)

6. Write down a fractional factorial experimental plan where 5 two-level factors are investigated in 8 experiments. (4 points)

7. What is type 1 error and what is type 2 error? (1 point)

8. Enver is trying to compare the yield from his catalytic converter with and without his secret ingredient. Without the ingredient, he obtained in 16 runs an average of 14.3 and a sample variance of 4.2. With the ingredient, he obtained in 11 runs an average of 15.7 and a sample variance of 3.8. Compute a 95% confidence interval for the increase in yield obtained with the ingredient, choosing a set of assumptions that make your computations possible. Remember to specify the assumptions you make. (3 points)

9. In Figure 2, there are scatterplots of 4 bivariate data sets: A, B, C, and D.

   (a) For each of the data sets A, B, C, and D, can you determine the correlation in the data just by looking at the plot, i.e., without computation? If so, what is the correlation? (We are here talking about the "ordinary", "Pearson" correlation). (2 points)

   (b) For each of the data sets A, B, C, and D, can you determine the Spearman rank correlation in the data just by looking at the plot, i.e., without computation? If so, what is the Spearman rank correlation? (2 points)

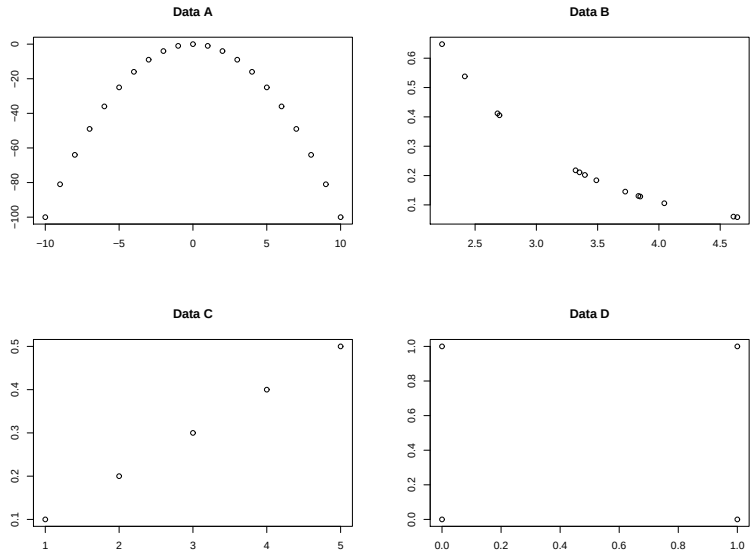

Figure 2: The datasets for question 1.

10. In the plot in Figure 3, five points are shown. Which of the 4 plots in Figure 4, X, Y, Z, or W, show a clustering of the data in Figure 3? (2 points)

11. Figures 5 and 6 each show 4 points and a line. In one of these plots, the line represents the linear regression line.

    (a) Is it Plot A in Figure 5 or Plot B in Figure 6 that represents a linear regression? Explain. (1 point)

    (b) For the plot that represents a linear regression, write down the residuals. (1 point)

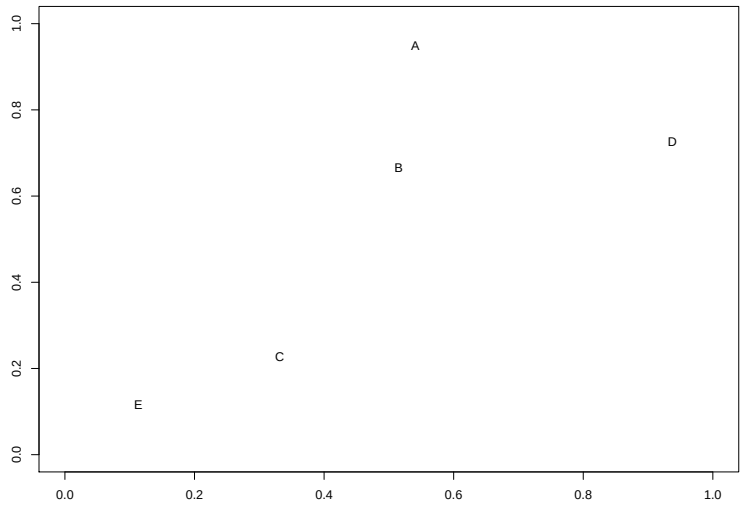

Figure 3: Five data points.

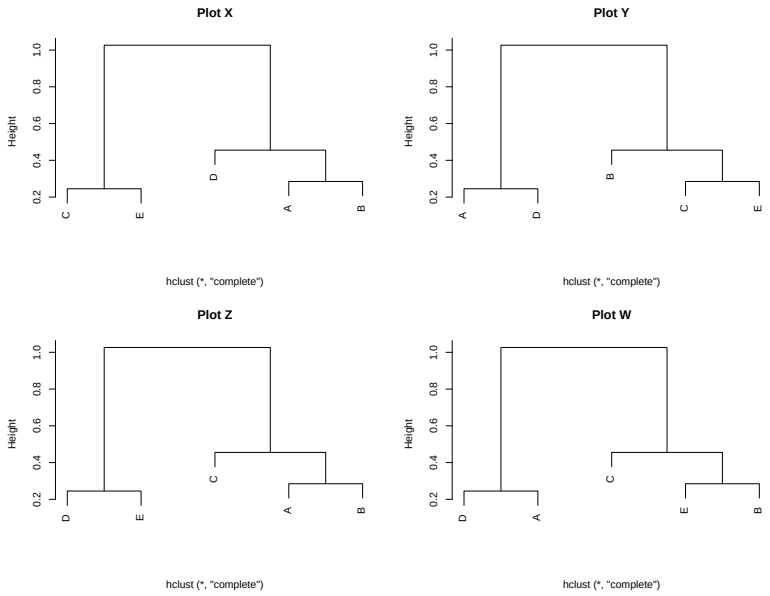

Figure 4: Clusterings of the points in Figure3?

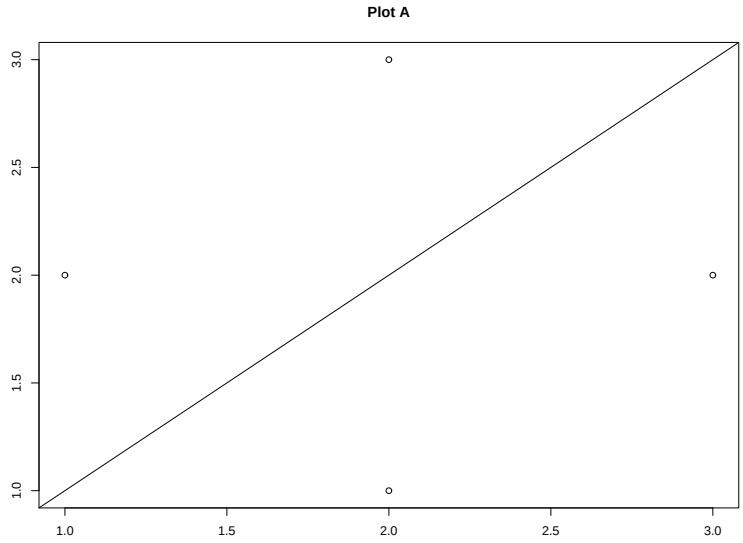

Figure 5: Possible linear regression A.

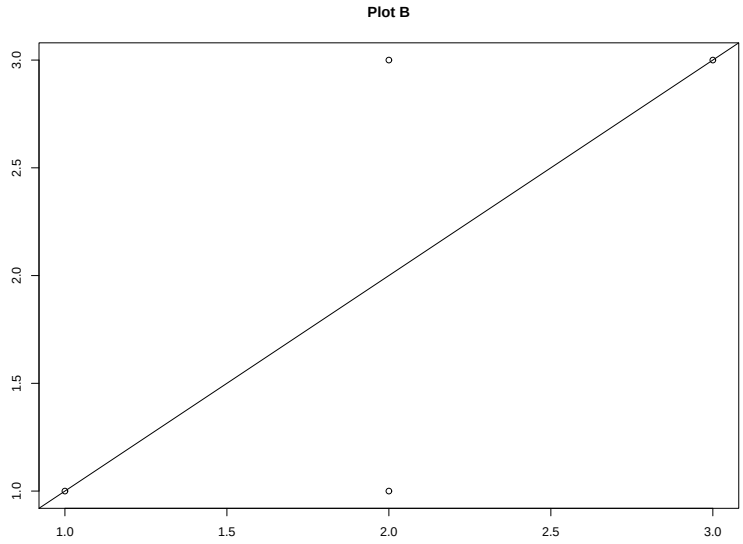

Figure 6: Possible linear regression B.