Petter Mostad
Mathematical Statistics
Chalmers

MSG830 Statistisk analys och experimentplanering

Exam 20 August 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. Sandra is selling t-shirts. Among her last 20 customers, 4 bought white t-shirts without print, 7 bought white t-shirts with print, and 9 bought colored t-shirts.

   (a) Using the empirical probability distribution, what is the probability that the next customer will buy a colored t-shirt? (1 point)

   (b) Given the information that the next customer bought a white t-shirt, what is the conditional probability that it was a t-shirt with print? (1 point)

2. Maria is studying reneissance painters A and B, who both specialized in portraits; A is known to have produced about 150 portraits and B about 400. She has a painting she believes was made by one of the tho painters, but she doesn't know which one of them. However, she has discovered that a very special blue pigment has been used in the painting. According to previous research, A used this pigment in about half his paintings, while B used it only rarely, in 1 out of 20 paintings. Given this information, whith what probability should Maria believe that A made the portrait? (2 points)

3. Oliver is has an olive orchard. He has developed a classification system where black olives are classified into four different types or qualities, A, B, C, or D, according to taste. Picking olives from his orchard in general, he found for 40 olives that 13 were of type A, 5 of type B, 11 of type C, and 11 of type D. Picking 20 olives from his newest three, he found that 4 were of type A, 4 of type B, 9 of type C, and 3 of type D. Do a goodness-of-fit test of whether the olives from the new tree taste different from the olives from his orchard in general. Formulate your conclusions (2 points).

4. Bart is prospecting for gold in the old-fashioned way, using a pan. However, he believes he has found a new pan which works better than the traditional one. In 6 creeks he as tried out both types of pan for an equal number of days. The table belows shows how much gold he has found with each pan type in each creek.

	Cr. 1	Cr. 2	Cr. 3	Cr. 4	Cr. 5	Cr.6
Trad. pan	43	7	133	81	33	0
New pan	39	13	151	94	61	0

(a) How much more gold, on average, does Bart find with the new pan type? Find a 95% confidence interval for the added amount of gold Bart should expect to find using the new pan compared to the old. Make the assumptions you need in order to make the computations. (3 points)

(b) Formulate the assumptions you needed to make in order to do the calculations in (a). Are they reasonable? Are there reasons to doubt these assumptions? (2 points)

(c) Bart wants to make a hypothesis test concerning whether the new pans are better or not. State the null hypothesis for such a hypothesis test, and the alternative hypothesis. What would be the conclusion of this hypothesis test (keeping any assumptions you made in part a)? (1 point)

(d) It it possible, with the same data, to make a hypothesis test with a different set of assumptions than those you may have made in (a)? Explain how, and what kind of test this would be. (1 point)

5. Andrea is a researcher who would like to optimize the output of a certain biochemical process. She has four ideas for increased output: To raise the temperature, to replace enzyme X with enzyme Y, to increase the speed of her mixer, and to irradiate the process with light. Unfortunately, she has only 8 experiments to test out her ideas. She decides on the following experimental plan

Temperature	Enzyme	Mixerspeed	Light
Normal	X	Normal	No
Normal	X	Normal	No
High	X	Normal	No
High	X	Normal	No
Normal	Y	Normal	No
Normal	Y	Normal	No
Normal	X	High	No
Normal	X	Normal	Yes

(a) Replace Andreas experimental plan with an 8 experiment fractional factorial experimental plan, using the same type of table to describe it. (2 points)

(b) Compare the two experimental plans. What are the advantages and disadvantages of each plan? Which one do you recommend? (2 points)

6. Explain in your own words what a dendrogram is, and draw an example. (2 points)

7. Three bivariate observations are given: They are listed in the table below, and shown in Figure 1
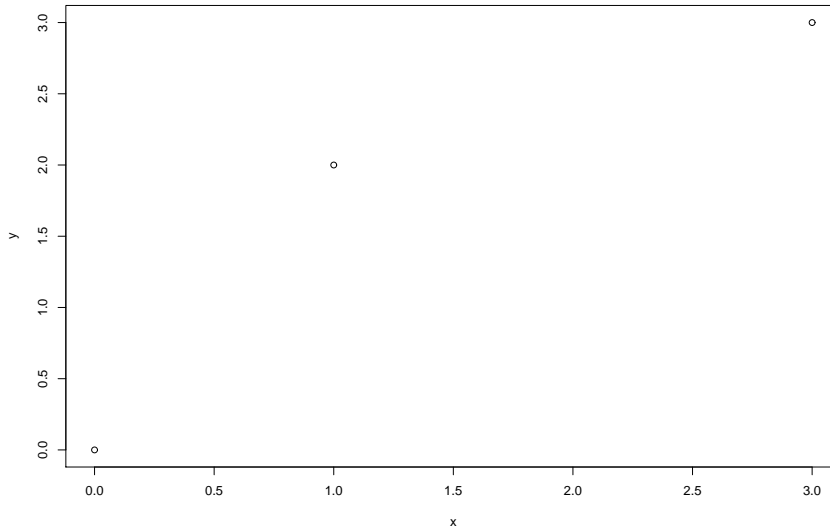
x	y
0	0
1	2
3	3

Figure 1: Plot of the three observations

(a) Compute the parameters of the simple linear regression line. (2 points)

(b) When simple linear regression is used (usually with more data) one can make a hypothesis test of whether the slope of the regression line is zero. What assumptions does one need to make in order to trust the outcome of such a hypothesis test? (1 point)

8. Describe what a permutation test is, and give an example. (3 points)

9. An experiment has been made where the yield of a production process was to be optimized in relation to temperature and pressure. Four different temperatures were tried out, and 5 different pressures. For each combination of temperature and pressure, three experimental runs were made. The mean and variance of the resulting yields were 42.2 and 23.7.

   The ANOVA table below is partially filled out. Fill out the parts of it that is possible to fill out given the information available to you (3 points):

	Sum of squares	Deg. freedom	Mean squ.	F	p
Temperature					
Pressure	235.8				
Residuals	949.6				
Total					

10. Given a dataset, describe how to construct a boxplot for it. (2 points)