Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 10 January 2012, 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. Alexander would like to know the average age of the people taking a particular cooking course. He asks 5 randomly selected participants about their age, and gets the numbers 19, 37, 33, 49, 25. Alexander assumes that the ages of the people taking the course follow a normal distribution.

   (a) Find a 95% credibility interval for the average age of the participants, based on the information Alexander has. (2 points)

   (b) Find a 95% credibility interval for the standard deviation of the distribution of the ages of the course participants, based on the information Alexander has. (2 points)

   (c) From course registration information, it is computed that the true average age is 37.9. Use this additional information to recompute a 95% credibility interval for the standard deviation of the distribution of the ages of the course participants. (2 points)

2. Anna is studying the difference in luminosity between stars A and B at a certain wavelength. She is using a method which she based on previous experience assumes produces readings that are normally distributed with the true luminosity as expectation and with a standard deviation of 0.37. She makes 7 observations of A and obtains an average of 56.2, and 4 observations of B, with an average of 61.4. Find a 90% credibility interval for the difference in luminosity. (2 points)

3. Joachim is comparing the sizes of adults of two closely related animal species (measuring size on a logged scale). He has made 15 measurements of species A, obtaining a mean of 2.31 and a variance of 0.87 He also made 18 measurements of species B, obtaining a mean of 2.95 and a variance of 0.96.

   (a) Joachim would like to assume that the two groups of observations come from two normal distributions. How can he investigate whether this is a reasonable assumption for his data? (1 point)

   (b) Joachim goes on making the assumption from (a). Make a hypothesis test of whether he can assume that the standard deviations of the two normal distributions are equal. (1 point)

   (c) Joachim bases his model choice on the test of (b). Using this model, make a hypothesis test of whether the expectations of the two normal distributions are the same. Draw a conclusion for Joachim. (2 points)

4. Sally is investigating the occurrence of car accidents resulting in fatalities in her city. Over the past 30 years, there have been a total of 137 such accidents. Sally initially assumes that all such accidents happen independently, and that there is no trend or change in the probability of accidents.

   (a) With these assumptions, what is the probability that there will be no accidents resulting in fatalities next year? (1 point)

   (b) What is the probability that there will be exactly 5 such accidents? (1 point)

   (c) What is the (approximate) probability that there will be 10 or more such accidents? (2 points)

5. Günther is studying for an exam, where he believes that the main exam question is likely to focus on either subject area A or subject area B. Based on previous exams, Günther estimates that the probability that the exam question will be about A is 40% and that it will be about B is 30%, while it is also possible that it will be about some other subject. Günther believes that if the question is about A, he has a 90% chance of passing, if it is about B, he has a 70% chance of passing, while if it is about some other subject, he only has a 10% chance of passing. What is Günthers probability for passing the exam? (1 point)

6. Peter an engineer at a factory, and he is studying the effect of two factors on the yield from his production process: He is trying out three different temperatures A, B, and C, and three different pressures X, Y and Z. In addition, he suspects that the results might depend on which machine he is using, so he is performing his experiments with three different machines, M1, M2, and M3. For each possible combination of these three factors he does two replicated experiments, so that the total number of observations is 54.

   The table below shows the average yield for each combination of temperature and pressure, the average yield for each setting of temperature and pressure separately, and the grand average yield. In addition, Peter has computed that the average yield for machines M1, M2, and M3 was 87.5, 88.78, and 87.22, respectively. He has also computed that the variance of all the data was 9.1981.

	Temperature A	Temperature B	Temperature C	Average
Pressure X	86.11	88.78	86.11	87
Pressure Y	86.61	89.28	86.61	87.5
Pressure Z	88.11	90.78	88.11	89
Average	86.94	89.61	86.94	87.83

   (a) Make an ANOVA table including the effects of all the factors, including the machine factor. Do not include interactions. Express the conclusions you can make from the p-values you compute. (4 points)

   (b) Extend the ANOVA table by including interaction between the Pressure and Temperature (but not interactions between these factors and the machines). If you should base the decision about whether or not to include interaction in your model only on the p-value you compute, what would you decide? (2 points)

   (c) The conclusions you make above depend on some assumptions you need to make in order to use this model. What are these assumptions? Mention some ways with wich you can check whether these assumptions are reasonable for your data. (1 point)

7. Anja would like to make an experimental plan for 8 experiments in which she will investigate the effects of 7 different two-level factors on some output. Write down an experimental plan for Anja, such that she may later use a balanced design matrix for the analysis of her data. (1 point)

8. Harri works for a chain of 3 bakeries, studying how to extend the time before a certain type of bread goes stale (i.e., tastes bad). In particular, he is interested in the effect of the speed of the cool-down after baking, and the effect of a certain ingredient X. He instructs bakery A to bake a bread, use a fast cool-down and then report how many days it goes before the bread is stale, then to bake another bread, use a normal cool-down and report how many days before the bread goes stale, and finally to bake a third bread, use a slow cool-down, and report the number of days before the bread goes stale. Bakery A should all the time use a small amount of ingredient X. He gives the same instructions to bakery B, but tells them to use a medium amount of ingredient X. Finally, bakery C should do the same thing, but with a high amount of ingredient X.

   Are there any problems with Harri's experimental design? Can you give him advice on how to improve it? (2 points)

9. Marge is studying how the weight of the output produced in an experiment depends on two types of input, x_1 and x_2, which are both measured on a continuous scale. She does 4 experimental runs, where the values of x_1 and x_2 are fixed according to the table below:

	x_1	x_2
Experimental run 1	1.2	1.2
Experimental run 2	1.2	3.8
Experimental run 3	3.8	1.2
Experimental run 4	3.8	3.8

   She analyzes the results using multiple linear regression with no interaction.

   (a) Which of the following sets of numbers could possibly be the least squares estimates of the analysis, and why could the other sets not be the least squares estimates? (1 point)

      i. 3.1, 1.1, -0.4, 0.2
      ii. 4.1, 1.1, -0.4
      iii. -3.1, -1.1, -0.4
      iv. 43.1, 1.1, -0.4, 0.2

   (b) Which of the following sets of numbers could possibly be the fitted values of the analysis, and why could the other sets not be the fitted values? (1 point)

      i. 4.94, 3.90, 7.80
      ii. 4.94, 3.90, 7.80, 6.76
      iii. 4.94, 4.94, 4.94, 6.76
      iv. 4.94, 3.90, 3.90

   (c) Which of the following sets of numbers could possibly be the residuals of the analysis, and why could the other sets not be the residuals? (1 point)

      i. 0.31, -0.31, -0.31, -0.31

ii. 0.31, 0.31, 0.31
iii. 0.31, -0.31, -031
iv. 0.31, -0.31, -0.31, 0.31