Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Reexam 15 August 2011, 8:30 - 13:30
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. Muhammad is comparing the life lengths of two types of batteries: X and Y. He has measured the life length of 6 batteries of each type, and he has observed the following numbers:

Batteries of type X	4.1, 5.9, 4.6, 3.9, 5.1, 5.3
Batteries of type Y	5.6, 7.2, 4.9, 6.3, 5.7, 6.7

   (a) Assuming that the observations of life lengths for batteries of type X come from a normal distribution, find a 95% credibility interval for the expected value of this distribution. (2 points)

   (b) Under the same assumption, find a 95% credibility interval for the precision of the distribution of part (a). From this, find also a 95% credibility interval for the standard deviation of the distribution. (2 points)

   (c) Assuming the observations of life lengths for batteries of type Y also come from some normal distribution, and assuming that the variances of the two distributions are the same, compute a 95% credibility interval for the difference in expected values for the two types of batteries. Which battery would you recommend, based on life length? (2 points)

   (d) The decision of whether or not to assume that the two distributions have the same variance can be aided by a hypothesis test. Perform the test, and find an interval for its p-value. Does the p-value support the decision you made in (c)? (2 points)

   (e) If Mohammad did not want to assume that the observations come from normal distributions, how could he analyze the data without this assumption? Explain; you do not have to make computations. (2 points)

2. Ingrid is working in a company mining for gold, and she is calculating the probability that a certain location contains minable amounts of gold. According to her earlier information, she believes that the probability for gold at the location is 4%. She has now received information that traces of another metal has been found at the location, and she knows that such traces are found in 10% of all locations that also contain gold, while it is only found in 0.3% of all locations not containing gold. What should now be her estimate for the probability of gold at the location? (2 points)

3. Patrik is investigating how the amount of a chemical in the leaves of a particular plant depends on two different factors: The habitat in which the plant grows, and the time of the year. He has collected the following data:

	Spring	Autumn	Average
Habitat A	39 47	54 58	49.5
Habitat B	46 44	59 63	53
Habitat C	36 40	52 50	44.5
Habitat D	59 65	88 88	75
Average	47	64	55.5

(a) Construct an ANOVA table for the data above: Do not include interaction. Find intervals for the relevant p-values, and draw conclusions. What assumptions do your conclusions depend on? (3 points)

(b) Change the ANOVA table so that it includes interaction between the two factors. Draw conclusions from the new table. (3 points)

4. Sally has a small shop, where she has estimated that the probability for each customer to be female is 69%.

   (a) Out of the next 6 customers, what is the probability that exactly 5 are female? (1 point)

   (b) What is the probability that 4 or less are female? (1 point)

   (c) Which assumption underlies the computations in a and b? (1 point)

5. Lisa works at a kennel, and is training dogs to do agility tasks. She is planning 16 separate experiments where she would like to investigate the effects of the following two-level factors on the learning time for the dogs:

   - Training in the morning or in the afternoon.
   - Training dogs that are hungry or less hungry.
   - Subdividing the training time into longer or shorter periods.
   - Training in an enclosed location or in an open location.
   - Training larger or smaller groups of dogs together.
   - Training with a loud or soft voice.
   - Male or female trainer.

   For each of the factors, Lisa would like to use her observations to find a distribution for a parameter β_i representing the effect on learning time of changing the factor.

   (a) Write down a fractional factorial experimental design that Lisa could use. (1 point)

   (b) Write down the name of your experimental design, in the form 2^{*}_{*}. (1 point)

   (c) Give advice to Lisa on how she should perform her experiment in order for the results to be as scientifically reproducible as possible. (2 points)

(d) The sum of all training times when using an enclosed location is 543, while the sum of all training times when using an open location is 619. What is the expected value of the effect of the parameter representing the effect of location? (1 point)

(e) In what situations will the 95% credibility interval for the parameter for male or female trainer be larger than the 95% credibility interval for the parameter for location? (1 point)

6. Harry is studying how a response, measured with a continuous variable y, depends on some predictors $x_1, x_2, \ldots, x_k$. He has observed both the predictors and the response in n cases, and he would like to analyze the data using multiple regression, thus finding the posterior distribution for the parameters $(\beta_1, \beta_2, \ldots, \beta_k, \tau)$ given the observed data.

   (a) What type of distribution is the joint posterior distribution for $(\beta_1, \beta_2, \ldots, \beta_k, \tau)$? (1 point)

   (b) What type of distribution is the marginal posterior distribution for β_2? (1 point)

   (c) Explain why it may be useful to study the residuals when investigating whether the chosen multiple regression model is appropriate for the data. (1 point)