Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 4 June 2010
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. Hans is studying the average weights of adults of animal species X and Y. In particular, he is studying whether adults of species X on average weigh more than adults of species Y. He has randomly selected 5 adults of speices X, and 6 of species Y, and measured their weights. The results are

Species X	53, 46, 54, 52, 39
Species Y	45, 30, 40, 46, 41, 33

   (a) Assuming that the weights of adults of species X are normally distributed (and assuming the standard priors) find a 95% credibility interval for the mean of this distribution. (2 points)

   (b) Making the same assumptions, find a 95% credibility interval for the precision of the distribution. (2 points)

   (c) Assuming also that the weights of adults of species Y are normally distributed, and that the precisions for the two distributions are the same (and assuming the standard priors) find a 95% credibility interval for the difference between the means of the two distributions. (2 points)

2. Tanya is selling t-shirts from a small stand at the beach at a summer resort. She would like to maximize the number of shirts sold daily. Although the sales are influenced by a number of things outside her control, such as the weather, she would also like to perform an experiment to find out how a number of factors she can control influence daily sales. She decides to investigate the impact of either doing or not doing each of the following things:

   - A: Putting up posters with nice pictures, to give a nicer atomsphere for shoppers.
   - B: Hanging up large posters advertising the low low prices.
   - C: Displaying some of the shirts on some wooden torsoes.
   - D: Putting up a sunroof of cloth, providing some extra shadow for the customers.
   - E: Playing music.
   - F: Trying to attract customers by calling out to them.

   (a) If she wanted to spend one day trying out each possible combination of either doing or not doing each of these things, how many days would she need to do her experiments? (1 point)

(b) Tanya would instead like to do 16 days of experiments, where she each day tries out a particular combination of "factor settings", where a factor is one of A, B, ..., F above, and a "factor setting" is either doing or not doing the factor. She would like to do this in such a way that the effect of each factor can be estimated from her data independently of the effect of the other factors. Propose an experimental plan for Tanya to achieve this. The plan should contain 16 rows and 6 columns, and should indicate which factor settings should be used on which days. The plan should represent an efficient way to learn about the effects of the factors. (2 points)

(c) Give advice to Tanya about how she should perform her experiments. In what order should she do her experiments? Are there other things she should think about, in order to ensure the the best possible validity of her results? (2 points)

(d) If Tanya uses your experimental plan, some interaction effects of factors may be possible to estimate, and some such interaction effects may be confounded with others. Write down exactly how two-way interaction effects are confounded with each other, or with main effects, in your experimental plan. (2 points)

3. Takashi is studying DNA sequences: These are sequences of *nucleotides*, each of which is of one of the types A, T, C, or G. So, for example, a sequence of length 6 could be written TTAGCT.

   (a) How many different sequences of length 6 are possible? (1 point)

   (b) How many sequences of length 6 contain exactly 3 nucleotides of type T? (1 point)

   (c) Takashi is studying some DNA where he can assume that the nucleotide type at different positions in his sequence are independent. He also knows that, for this type of DNA, the probability at each position is 35% for a G, 35% for a C, 15% for an A, and 15% for a T. If Takashi has a sequence of length 10, what is the probability that it contains exactly 3 nucleotides of type T? (1 point)

   (d) For the same type of DNA, what is the probability that it contains exactly 3 nucleotides of type T, 4 of type A, 3 of type G, and none of type C? (1 point)

4. Ben is trying to determine which bird species his friend Dan has observed: He has established that it is either species A or species B. In the area, there are 3 times as many birds of species A as of species B, so Ben recons that, a priori, the probability that Dan has observed species A is 75%. However, Dan also reports that the bird engaged in a particular behaviour; this behaviour is seen in 90% of all birds of species B, but only 10% of all birds of species A. Given this information, what is the probability that Dan has observed species A? (2 points)

5. Roza is investigating how her different running shoes and her different running clothes may effect her running times on the standard loop she usually runs. She has three different pairs of shoes, A, B, and C, and three different types of clothes, X, Y, and Z. Over 9 days, she tries out (in a random order) all the 9 possible combinations of these factors, and records the running times. She would like to analyze the results using an ANOVA table. She decides to assume that there is no interaction between the effects of the shoes and the clothes. She computes the following three numbers for her ANOVA table:

	Sum of squ.	Deg. freed.	Mean squ.	F value	p value
Shoes	7.8104				
Clothes	3.1599				
Residuals					
Total	13.7854				

(a) Complete the missing parts of the ANOVA table. For the p-values, find intervals containing the true p-value. (2 points)

(b) How should Roza conclude from her experiment? Could she conclude that the shoes and/or the clothes make a difference to her running times? (1 point)

(c) In fact, it later turns out that Roza made a mistake filling out the first 3 numbers of the ANOVA table. Given that Roza's running times are given in the table below, fill out the first 3 numbers in the table correctly. (For your convenience, the numbers are given in decimal minutes). (3 points)

	Shoe A	Shoe B	Shoe C
Clothes X	30.0	33.5	34.9
Clothes Y	31.1	33.5	32.3
Clothes Z	33.0	33.7	32.9

6. Lurleen is given the following information about two types of work clothes that have been tested and compared: Six people have tested type A work clothes. Each reported the number of weeks until the clothes were "worne out", as defined by the user. Six different people tested type B work clothes, and also reported the number of weeks the clothes lasted. For type A, the results were 23, 35, 32, 36, 39, 25. For type B, the results were 25, 37, 35, 46, 45, 37.

(a) Assume that the time until the clothes were "worn out" was normally distributed for both types of clothes. Assume that the precisions in the two groups may have been different. Find a 95% credibility interval for the *difference* in the average time until the clothes are worn out. Can you conclude that one type of clothes are more durable than the other? (3 points)

(b) Later Lurleen learns that she was misinformed: In fact it was the same six people that testet both types of clothes. Each person tested the types in a random order. So the results can be represented in the following table:

	Person 1	Person 2	Person 3	Person 4	Person 5	Person 6
Type A	23	35	32	36	39	25
Type B	25	37	35	46	45	37

Re-analyze the data using the new information. Can you now conclude something about whether one type is more durable than the other? (2 points)