Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 16 March 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. Betty likes to play all kinds of lotteries. For one particular lottery, it is advertised that 25% of all lottery tickets give some kind of win.

   (a) If Betty buys 10 lottery tickets, what is the probability that she will win on exactly 5 of them? (1 point)

   (b) Betty decides to buy four tickets each week, one for her self and one for each of the other tree members of her family. If she continues to do this for a year (52 weeks), what is the probability that 75 or more of the tickets will give a win? (1 point)

   (c) Assuming Betty buys a total of 4 tickets each week for her family and herself as described above, what is the probability that, after 4 weeks, each of the four family members has won exactly one time? (1 point)

2. Ali is investigating an outbreak of a viral disease in cows at a farm. The virus comes in three variants, A, B, or C. The first variant is the most common, and is responsible for 70% of all outbreaks. Variants B and C are responsible for 20% and 10% of outbreaks, respectively. For this particular outbreak though, Ali suspects that variant C is responsible, because the cows show a particular symptom that variants A and B give rise to in only 10% of the the cases where they are responsible, while variant C produces this symptom 90% of the time. Given this information, what is the probability that the cows are infected with variant C? (2 points)

3. Billy is experimenting with two different designs for his bookshelfs. Among other things he wants to find out how much weight each design can take before it collapses. He has built 6 bookshelfs of each type, and has measured the weight that made them collapse:

Design A	82, 74, 64, 78, 73, 82
Design B	86, 87, 84, 72, 90, 116

   (a) Make a hypothesis test where the null hypothesis is that the values for design A and B respectively come from two normal distributions with equal standard deviations, while the alternative hypothesis is that they have different standard deviation. (2 points)

   (b) Billy would like to find a credibility interval for the difference in the expected amount of weight each design can take before it collapses. Describe a set of assumptions he can make in order to compute such an interval (i.e., a model one can use). (1 point)

(c) Compute a 95% and a 90% credibility interval for the difference mentioned in (b), using your assumptions from (b) (2 points)

(d) How can Billy correctly formulate conclusions from his experiment? (1 point)

4. Bart would like to investigate how 5 different factors influence the accuracy of his slingshot. The factors and factor levels he would like to investigate are

	Opening	Band length	Band width	Elasticity	Stick material
High (+) level	Wide	Long	Wide	High	Hard
Low (-) level	Narrow	Short	Narrow	Low	Soft

(a) Bart would like to construct 8 different slingshots, testing out combinations of the factors above. In addition to the factors themselves, he is particularly interested in investigating interaction between Band width and Elasticity, and Band length and Elasticity. Write down a fractional factorial experimental plan satisfying Bart's requirements. (2 points)

(b) Bart does randomized experiment, but then lists the data in the order of the lines of your experimental plan. First make the design matrix he should use when not considering interaction. Then make the design matrix he should use when taking the interactions that interests him into account. (2 points)

5. Alex is investigating how well his new invention, a soap-bubble machine, works. He is trying out three different mixes for the soap: A, B, and C. He is also trying out four different designs for the mechanism that blows the bubbles: X, Y, Z, and W. In each experimental run, he makes a soap mix according to A, B, or C, adjusts the machine according to X, Y, Z, or W, and measures the resulting size and life-lengths of bubbles, resulting in a single number R for that experimental run. For each combination of soap mix and mechanism design, he has made four experimental runs, resulting in a total of 48 values for R. The table below[1] lists average values for R: The averages for combinations of the factor values, the averages for each factor value, and the grand average. The variance of all his observations of R was 163.5918.

	A	B	C	Average
X	29	40.75	35.25	35
Y	30.5	29.75	36.75	32.3333
Z	45	54.5	24.75	41.4167
W	22	31.75	40.75	31.5
Average	31.625	39.1875	34.375	35.0625

(a) Alex first assumes that there is no interaction between the factors. Make an ANOVA table for this situation, and formulate conclusions. (4 points)

(b) Make a new ANOVA table, but now including interaction. Formulate conclusions. Would you recommend Alex to continue his analyses with a model with or without interaction? (2 points)

[1] In the original exam, there were two errors in the data table, which have now been removed

(c) When Alex made his 48^2 experimental runs above, he performed them in the same order as they are listed in the table above. Give an argument that could convince Alex to do further experiments in a different way. (1 point)

6. Lisa is investigating how the concentration she uses of the chemical X influences the production of some compound Y from her genetically engineered bacteria. She has made 5 experimental runs, obtaining the results below:

Chemical X	0.02	0.04	0.06	0.08	0.1
Produced compound Y	4.3	4.6	3.7	2.1	1.9

(a) Assuming Lisa would like to analyze the data using simple linear regression, compute the coefficients of the best-fitting line in this model (i.e., the slope and the intercept of the regression line). (3 points)

(b) Discuss how the simple linear regression model could be inadequate (i.e., not good for this situation). Assuming Lisa goes on to produce data from a total of 50 experimental runs, how could she check if the model is good for this data? (1 point)

(c) Use the model and results from (a) to make a prediction (a single "best prediction" value) for the amount produced of the compound Y when the concentration of chemical X is 0.2. Is this prediction realistic? (1 point)

(d) Compute the residuals and the sum of squares of the residuals. (2 points)

(e) Find a 95% credibility interval for the standard deviation parameter of the linear model. (1 point)

[2] In the original exam, this was erroneously written as 36