Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 21 August 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. Christine is working in a lab doing DNA paternity testing: Given DNA tests of a mother and her child, and of a putative father, Christine concludes either that the putative father is the real father, or that he is not. In 43% of last year's cases, she concluded that the putative father was not the father.

   (a) Assume that the rate of true fatherhood is constant. In 10 new cases, what is the probability that the putative father is found to not be the real father in exactly 2 of the cases? What (additional) assumption do you need to make to do your computation? (2 points)

   (b) In 100 new cases, what is the approximate probability that the putative father is found not to be the father in 50 cases or more, still assuming that the rate of true fatherhood is constant? (1 point)

   (c) In fact, in the 10 cases so far this year, the putative father was found not to be the father in 9 cases. Compute the probability that this happens first assuming that the rate of true fatherhood is 43% and then assuming that this rate has increased to the regional average of 65%. (2 points)

2. Hans is a medical doctor, and he is trying to establish whether a patient has a disease X, and if so, which of the four variants A, B, C, or D the patient has. In this population, 2% have disease X. Among those who have the disease, 80% have variant A, 10% variant B, 5% variant C, and 5% variant D. There is a test that is sensitive only to variant A: For patients with variant A, there is a 90% chance of a positive test, and a 10% chance of a false negative reslut. For patients without variant A (i.e., either healthy or with one of the other variants), there is a 10% chance of a positive test and a 90% chance for a negative test.

   (a) Given that the patient gets a positive test, what is the probability that the patient has variant A? (2 points)

   (b) Given that the patient gets a negative test, what is the probability that the patient has variant A? (1 point)

   (c) Given that the patient gets a negative test, what is the probability that the patient has one of the variants B, C, or D? (2 points)

3. Olof is studying the distribution of lengths of male students at his university. He has recorded that the lengths of 6 randomly selected male students are

$$192, 187, 173, 173, 180, 167.$$

(a) Find a 95% credibility interval for the mean of the lengths of male students at Olof's university. (1 point)

(b) Find a 95% credibility interval for the standard deviation of the lenghts of male students at Olof's university. (2 points)

(c) In a much larger investigation, it is found that the actual mean of the lengths of male students is 181. Use this mean together with Olof's data to re-compute a 95% credibility interval for the standard deviation of the lengths. (2 points)

(d) What assumptions do the computations in questions a,b,c above depend on? (1 point)

4. Susanna is studying how temperature, lighting, and smell in a grocery shop influences the total sales there. She is studying the following levels of these factors:

Temperature	Low, High
Lighting	Low, Medium, High
Smell	Smell a, smell b, smell c, smell d

For each combination of these factors, she makes 3 independent experiments, giving her a total of 72 experiments, which she performs in a randomized order. In the table below is listed the average result for each combination of temperature and lighting, the average result for each temperature, the average result for each lighting, and the grand average. In addition, Susanna has computed the average for each of the smells a, b, c, and d to be 28.39, 29.61, 29.0, and 37.06, respectively. She has also computed that the variance of all the experimental results is 61.30966.

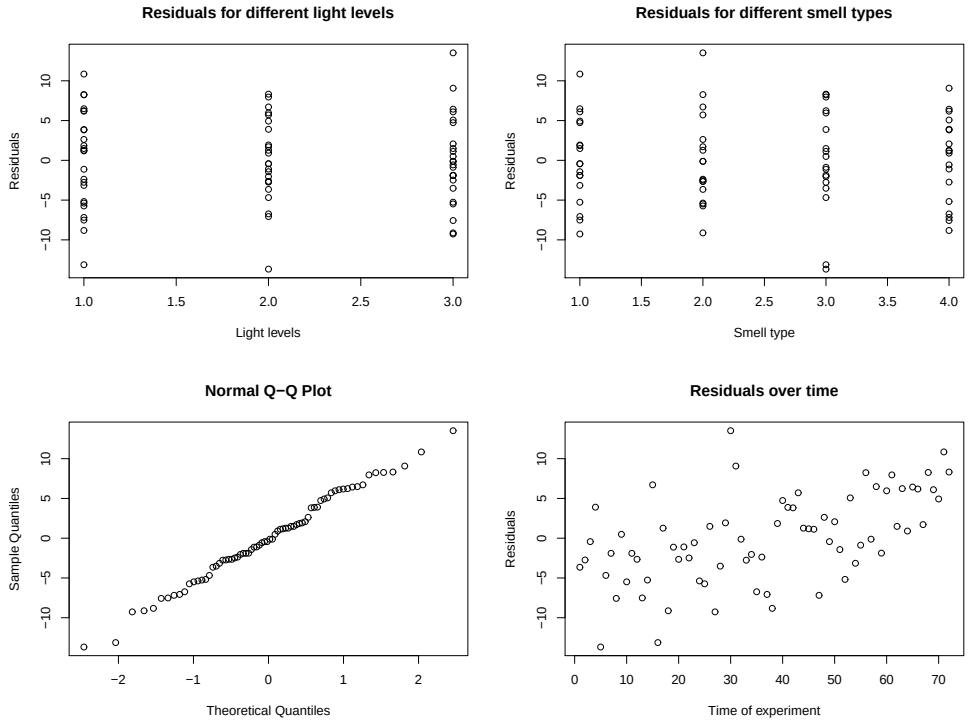
	Lighting low	Lighting medium	Lighting high	Average
Temperature low	27.92	31.42	38.75	32.69
Temperature high	27.0	27.33	33.67	29.33
Average	27.46	29.37	36.21	31.01

(a) Make an ANOVA table where the smell factor is ignored, and there are no interactions. Draw conclusions. (3 points)

(b) Make an ANOVA table including the smell factor, but with no interactions. Draw conclusions. (2 points)

(c) Make an ANOVA table including all three factors and the interaction between Temperature and Lighting. Compute the p-value for the interaction; would you based on this recommend to keep the interaction in the model or not? (2 points)

(d) Below are some plots based on the residuals in the model used in part (b) above, where all three factors are considered but there is no interaction. The top two plot the residuals for the different levels of the Lighting and Smell factors, respectively. The bottom left is a normal probability plot for the residuals, and the bottom right plots the residuals against the time ordering of the experiments. Make comments describing what these plots tell you about whether the assumptions of the model are appropriate for this data, or if not, how the model could be improved. (2 points)



5. Alizadeh is a chemist working in a factory. She is studying a process in which several chemicals are mixed together with a catalyst in a particular machine, to produce a new chemical. The amount of the new chemical (the yield) is then measured. Alizadeh wants to increase the yield, and she is investigating whether a new catalyst produces a higher average yield than the standard one. She has the possibility to do 6 runs with the standard catalyst and 6 runs with the new one.

   (a) Give general advice to Alizadeh about how she should conduct her experiment, with the aim of producing scientifically reproducible results. (2 points)

   (b) Alizadeh performs her experiments and gets the results

	Yields						Average	Variance
Stand. catalyst	1061	1046	922	1022	1046	1008	1017.5	2549.5
New catalyst	897	945	1128	1058	1193	977	1033	12925.2

   Make a hypothesis test, including conclusion, of whether the population variances are the same for yields produced with the standard catalyst and with the new catalyst. (1 point)

   (c) Assuming the population variances are the same, use the data to get a 90% credibility interval for the difference in expected yield between the new and the standard catalyst. (1 point)

   (d) Assume that, instead of the experiment above, Alizadeh is forced to do her 12 experimental runs on an old machine which is known to produce generally decreasing yields. Give advice to Alizadeh about how she should perform her experiment using this machine, and suggest possible analysis methods. (1 point)