Petter Mostad
Mathematical Statistics
Chalmers

MSA830 Statistical analysis and experimental design

Exam 26 October 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. Sonia is working with a machine that cleans recycled bottles. For each bottle, it has been found that there is a 4% probability that the machine will not manage to clean the bottle properly.

   (a) In a case of 20 bottles, what is the probability that one or more bottle is not properly cleaned by the machine? (1 point)

   (b) A case as above, where not all bottles are cleaned, is called a "problem case". If the machine cleans 50 cases of bottles, what is the approximate probability that 20 or more of these cases will be problem cases? (2 points)

2. In a game show, the contestant can choose between three boxes, A, B, and C. One of them contains a 1 million kronor prize, the other two are empty. The money has been placed randomly with equal probability in one of the boxes. Giving the choice a lot of thought, the contestant eventually decides that he would like have the contents of box A.

   (a) Before box A is opened, the game show host adds some exitement: She opens one of the boxes the contestant has not chosen, and shows that it is empty. She always follows the following procedure when choosing a box to open: She will not choose the box with the money (she knows which one it is), she will not choose the box the contestant has chosen, and otherwise she will choose a box randomly with equal probability. What is the probability that she will choose to show the contents of box C? (1 point)

   (b) The host opens box C and shows the contestant that it is empty. Use Bayes formula to compute the updated probability that box A now contains the 1 million kronor prize, using this new information. (1 point)

   (c) Given the information that box C is empty, what is now the probability that box B contains the prize? If the contestant is allowed to change his choice, from box A to box B, should he do so? (1 point)

3. Billy is studying how his new and improved food for farmed fish helps to increase the weight of the fish, compared to the traditional kind of food. He has asked 5 fish farms (of similar sizes) to test this out for him: Each farm gives some fish the traditional food while other fish get Billy's new and improved type, and they measure the results on each group of fish. The resulting data are given in the table below, with one line for each fish farm:

Traditional type	Billy's new and improved
659	680
613	607
501	505
337	373
222	202

Billy would like to learn about the increase in the expected weight measurements of the fish, when the traditional food is replaced with his new type.

(a) Choose a method to answer Billy's question, and argue why this method is reasonable. (1 point)

(b) Using your method (or one of the methods you think is reasonable for this data) compute a 95% credibility interval for the increase in the expected weight measurements of the fish. (3 points)

(c) Find the approximate probability outside the smallest credibility interval containing zero for the increase in the expected weight measurements of the fish. (1 point)

4. Tommy would like to see how the package coloring and package print influence the sales of his product. He his trying out 4 colors: Red, Black, Green, and Blue, and he is also trying out 4 types of print, let us call them print A, B, C, and D. He makes one sales test for each combination of color and print, and gets the sales results in the table below:

	A	B	C	D	Average
Red	43	42	43	45	43.25
Black	40	47	40	42	42.25
Green	56	47	43	54	50
Blue	49	51	47	51	49.5
Average	47	46.75	43.25	48	45.25

The table also shows the averages for different colors and different prints, and the grand average. The sample variance for all the data is 23.8. Tommy would like to analyze and conclude from these data; regarding assumptions he must make to do so, see question (d) below.

(a) Make an ANOVA table for Tommy's data. Do not include interaction. Formulate the conclusions you can make from the table. (3 points)

(b) In his continued analysis, Tommy decides to drop considering the colors, i.e., he will just ignore this factor. Set up a new ANOVA table for this situation, and draw conclusions (2 points)

(c) In the situation where colors are ignored, find a 95% credibility interval for the difference in expected sales between print A and print D. (2 points)

(d) What assumptions do you need to make in order to make the computations above? How can one check whether these assumptions are reasonable? (1 point)

5. Sally is running a shop selling flowers, and she would like find ways to keep her flowers fresh as long as possible. She would like to investigate a number of factors such as temperature, light, water nutrients, air quality etc. Let us call the factors she would like to investigate A, B, C, D, E, and F, where each of the factors are considered at two different levels. She would like to make 16 experimental runs.

   (a) Make a table containing a fractional factorial plan for Sally. Give an example of an interaction effect she will not get information about from her data, and one she will get information about, if she follows your experimental plan. (3 points)

   (b) Give advice to Sally about how she should perform her experiment, to get as scientifically reproducible conclusions as possible. (1 point)

6. Krzsysztof investigates how the strength of a new material changes with the addition to the material of a chemical that he has invented. So far he has made 8 experiments, and the results are given in the table below:

Amount added	3	5	5	6	6	7	8	9
Strength	41.3	50.0	56.6	56.7	59.1	62.6	68.0	64.2

   From the data he has computed the following values:

$$
\begin{aligned}
3 + 5 + 5 + 6 + 6 + 7 + 8 + 9 &= 49 \\
41.3 + 50.0 + 56.6 + 56.7 + 59.1 + 62.6 + 68.0 + 64.2 &= 458.5 \\
3^2 + 5^2 + 5^2 + 6^2 + 6^2 + 7^2 + 8^2 + 9^2 &= 325 \\
41.3^2 + 50.0^2 + 56.6^2 + 56.7^2 + 59.1^2 + 62.6^2 + 68.0^2 + 64.2^2 &= 26781.35 \\
3 \cdot 41.3 + 5 \cdot 50.0 + 5 \cdot 56.6 + 6 \cdot 56.7 + 6 \cdot 59.1 + 7 \cdot 62.6 + 8 \cdot 68.0 + 9 \cdot 64.2 &= 2911.7
\end{aligned}
$$

   (a) Find the least squares estimates for the parameters of a simple regression line for this data. (2 points)

   (b) There are 8 residuals. Compute two of them (any two you like). (1 point)

   (c) If one computes all the residuals, explain how one can use these to check if the simple linear regression model is appropriate for Krzysztof's data. What are the assumptions of such a model? (1 point)

   (d) The sum of the squares of the residuals is in fact 73.8612. Compute a 95% credibility interval for the distribution standard deviation e^{λ} for the ϵ_i's in the simple linear regression model. (2 points)

   (e) Krzysztof would like to use his analysis to make a prediction for the strength he will find in his next experiment. Is his credibility interval for a new strength observation where he adds 6 units of his chemical larger, smaller, or of equal length compared to his credibility interval for a new strength observation where he adds 12 units of his chemical? Why? (1 point)