

MSA830 Statistical analysis and experimental design

Exam 8 June 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. Olle has received a precious stone as part of a gift, but he worries that the stone originates in an area from which export of such stones is illegal, as do 5% of all such stones. He finds out that he can check this by measuring the amount of a certain element in the stone: Among stones from the area, 3%, 15% and 82% have low, medium, or high levels of the element, respectively, while among stones from other areas, 93%, 6%, and 1% have low, medium, or high levels of the element, respectively. Olle finds that his stone has a high level of the element. What is the probability that it is from the area from which export is illegal? (2 points)
2. Betty is investigating how often people talk on a hand-held mobile phone while driving their cars. She assumes that the incidences of this are independent of each other. By making observations from a bridge over a road for a total of 100 hours during rush-hour, she records a total of 13 incidences.
 - (a) What is the probability that she will observe no new incidences, if she observes for 1 additional hour at the same spot during rush-hour? (1 point)
 - (b) What is the probability that she will observe 2 or more incidences, if she observes for 1 additional hour at the same spot during rush-hour? (1 point)
 - (c) Betty would like to observe a total of 100 or more incidents. She manages to recruit 5 persons who each are willing to make observations for 100 hours at similar places during similar times. What is the (approximate) probability that the total number of incidents observed by these persons, plus those previously observed by Betty, will give a total of 100 incidents or more? (2 points)
3. Caroline is planning to study the effect of 6 different two-level factors A,B,C,D,E,F on the oxygen output, i.e., the photo syntesis level, of her green plants. She has the possibility to do 16 experiments, and would like to have a balanced experimental plan, so that the effect of each of her factors can be estimated independently from the others. Write down an experimental plan for Caroline, specifying the setting “+” or “-” for each factor in each experiment. Also, give general advice to Caroline about how she should perform her experiment: How should she deal with other factors that she might control, and other factors which she cannot control, that might influence the results? (3 points)
4. David is trying to find systematic differences in the plumage colouring of three closely related species of birds. He is measuring the amount of a certain carotenoid, and has obtained the results below for 5, 6, and 4 birds from species A, B, and C, respectively:

	Measurements	Sample mean	Sample variance
Species A	334, 423, 570, 487, 501	463	7927.5
Species B	408, 690, 588, 823, 970, 567	689.3333	37420.67
Species C	966, 1002, 636, 1217	955.25	57578.25

To analyze the data, David first makes the assumption that the data for each species comes from a normal distribution, and that these three normal distributions may have different expectations and different standard deviations.

- (a) Find a 95% credibility interval for the expected measurement for birds of species A. (2 points)
 - (b) Find a 95% credibility interval for the standard deviation of measurements for birds of species A. (2 points)
 - (c) Find a 95% credibility interval for the difference between expectations of measurements for birds of species A and B. (2 points)
 - (d) Now, assume that the data for the three species come from normal distributions with equal standard deviations, but with different expectations. Answer each of the questions (a) and (b) again, under such assumptions. (4 points)
 - (e) Is there a way for David to analyze the data without making any assumptions that the data come from normal distributions? Make a brief explanation. (1 point)
5. Paola is a gardener, and she is trying to find the best way to grow a grass lawn in the area where she lives. She is trying out four different types of grass seeds, $G_1, G_2, G_3,$ and $G_4,$ and two different types of fertilizer, F_1 and $F_2.$ She has tried out each combination of seed and fertilizer in 5 different lawn patches, and has measured the amount and quality of the grass after a certain period of time, obtaining one response measurement for each patch and thus a total of 40 measurements. The table below lists the averages, for each combination of factors, for each factor, and the grand average. In addition, Paola has computed that the sample variance of her data is 28.47179.

	G_1	G_2	G_3	G_4	Average
F_1	55.3	57.2	59.8	55.7	57
F_2	58.9	60.8	63.4	59.3	60.6
Average	57.1	59	61.6	57.5	58.8

- (a) First, Paola would like to study only the effect of the type of seed, ignoring that different fertilizers have been used. Make an ANOVA table for this single factor, and draw conclusions from it. (2 points)
- (b) Secondly, Paola decides to study only the effect of the fertilizer, ignoring that different seeds have been used. Make an ANOVA table for this single factor. (1 point)
- (c) Assuming that the observations with fertilizers F_1 and F_2 come from two different normal distributions with the same variances, compute a 90% credibility interval for the difference in the response when changing from fertilizer F_1 to fertilizser $F_2.$ (2 points)
- (d) Make an ANOVA table for Paola's data including both the factors, and also interaction. Find a p-value for the interaction. If you should decide on whether to include interaction or not based on this p-value, what would you decide? (2 points)

- (e) No matter what the p-value was above, it is decided to use a model not including interaction for the analysis of the data. Make an ANOVA table without interaction. What can you conclude about the influence of the factors on the amount and quality of the grass? (1 point)
- (f) Finally, in order to draw conclusions in parts (a), (b), (d), and (e) above, some assumptions need to be made. What are these assumptions? Mention some ways you can study whether these assumptions are reasonable. (2 points)