Suggested solution to

exam in MSA830: Statistical Analysis and Experimental design

December 5th 2007, 8:00 – 12:00, at Väg och Vatten

1. a) For a t-test to be valid, each observation must be an independent random sample from a normal distribution with expectation equal to the population mean for the condition. The variances of the normal distributions could be assumed equal in one type of t-test. In another type, it is only assumed that the variances within each condition is equal.

   b) We get $s^2_{pooled} = \frac{82+534}{4+6} = 61.6$ giving the test statistic $t = \frac{\bar{b}-\bar{a}}{\sqrt{s^2_{pooled}(1/5+1/7)}} = \frac{8}{\sqrt{61.6(1/5+1/7)}} = 1.74$. Looking at the table for the t-distribution with 10 degrees of freedom, we see that the probability of a value above 1.812 is 5%, so the probability of a value above 1.75 is slightly above 5%, maybe 6%. As Lisa's alternative hypothesis is one-sided, we are doing a one-sided test, so the p-value becomes approximately 6%. We conclude that there is some evidence that condition B gives a higher result than condition A, but that the difference is not significant at the 5% significance level.

   c) It is the first computation that is correct. Lisa wants to compare the difference between the averages under the two conditions with simulated numbers constructed in a similar way, which have the same distribution as the difference under the null hypothesis that all the numbers come from the same distribution.

   d) Looking at the table, we see that for the first computation, 94% of all simulations have lower values than the observed difference 17-9=8. Thus the probability of observing something more than 8 can be estimated as 100% - 94% = 6%, and this is the approximate p-value.

2. a) One possibility is to use a standard 2^3 design, with one setting of the factors for each week. It would then be sensible to use some kind of randomization for the which weeks which combinations of factors are run. A randomization could give the first four columns of:

Week number	*P1 and not P2*	*advertising board*	*music*	*Interaction*	*salesperson*
6	-	-	-	+	-
2	+	-	-	-	+
7	-	+	-	+	+
1	+	+	-	-	-
4	-	-	+	-	+
5	+	-	+	+	-
8	-	+	+	-	-
3	+	+	+	+	+

b) The main effect from the position could be computed by taking the average of the observations for the weeks where there are pluses in the column ”P1 and not P2”, and subtracting the average of the observations for the remaining weeks. The main effect of the music could be computed similarily, using the pluses and minuses from the ”music” column. The interaction effect between the two can be computed according to the ”interaction” column above, where the pluses and minuses have been obtained by multiplying the column for ”P1 and not P2” and for ”music”.

c) If the salesperson is determined according to the ”salesperson” column in the table above, then estimations for the main effects will not be influenced, as in each computation for a main effect, each possible sales person will appear an equal number of times with and without the factor. The column is produced by multiplying the signs of the three columns for the three main effects. Alternatively, they could use a 2^4 design, with each possible combination tried for half a week.

d) One way would be to randomize the order of the weekly experiments, as we have already done above. Another way would be to subdivide the weeks, so that each day represented a new experiment, although this might mean more moving and changing of equipment.

3. NOTE: There is a printing mistake in the mean square residual: It should be 1.1 (equal to 9.9/9) and not 0.9. The use of both 0.9 and 1.1 in computations has been accepted.
   a) Using 0.9, he p-value is obtained by comparing 5.1 / 0.9 = 5.67 to an F-distribution with 3 and 9 degrees of freedom. According to table D, such a distribution has a 5% chance of values above 3.86, and a 1% chance of values above 6.99. Thus our p-value is beteween 0.01 and 0.05. We conclude that tyres make a difference. Using 1.1, the p-value is obtained by comparing 5.1/1.1=4.64 to the same distribution, giving the same result.

   b) Using 0.9, the p-value is obtained by comparing 4.2 / 0.9 = 4.67 to an F-distribution with 3 and 9 degrees of freedom. Again, we get that the the p-value is between 0.01 and 0.05. We conclude that drivers make a difference. Using 1.1, the p-value is obtained by comparing 4.2/1.1 = 3.82 to the same distribution. Now, the effect of the drivers is NOT significant at the 5% level.

   c) In the new ANOVA table, the sum of squares corresponding to tyres will be the same, as will the total sum of squares. This means that the residual sum of squares will be 12.6 + 9.9 = 22.5, with 3+9=12 degrees of freedom. Thus the new mean square for residuals will be 22.5 / 12 = 1.875. The quotient of mean squares becomes 5.1 / 1.875 = 2.72, and this should be compared with an F distribution with 3 and 12 degrees of freedom. For such a distribution there is a 5% chance that values are above 3.49, and a 10% chance of values above 2.61. Thus the p-value in this case will be between 0.05 and 0.1. We can now NOT reject the null hypothesis that tyres do not influence the stopping distance, with 5% significance.

4. a) The t-distribution with 97 degrees of freedom has a 2.5% chance for values above approximately 1.98, and a 0.5% chance for values above approximately 2.617. Thus the 95% confidence interval for b_1 becomes $-0.058 \pm 1.98 \cdot 0.035$ giving the interval [-0.13, 0.01] and for b_2 we get $0.13 \pm 1.98 \cdot 0.018$ giving the interval [0.09, 0.17]. Similarly, the 99% confidence intervals become [-0.15, 0.03] and [0.08, 0.18], respectively.
   As we have as many as 100 observations, we can use a normal distribution instead of a t-distribution to find the confidence intervals. For the standard normal distribution, 1.96 is such that the probability of a value above it is 2.5%, whereas 2.58 is such that the probability of a value above it is 0.5% Thus the 95% confidence interval for b_1 becomes $-0.058 \pm 1.96 \cdot 0.035$ giving the interval [-0.13, 0.01] and for b_2 we get $0.13 \pm 1.96 \cdot 0.018$ giving the interval [0.09, 0.17]. Similarly, the 99% confidence intervals become [-0.15, 0.03] and [0.08, 0.18], respectively.

b) A confidence region for b_1 and b_2 is a region of pairs (b_1, b_2) such that, under repeated experiments with the same model, the constructed confidence region will contain the actual values of b_1 and b_2 with a given probability. Confidence regions for these models are always elliptical.

c) One would conclude that there is clear evidence that the pollutant is positively related to the humidity. However, one cannot conclude that there is any relationship between temperature and the pollutant.

d) Two problems appear in the plots. The plot of pollutant versus temperature seem to indicate that there is some relationship between temperatures and the pollutant, although not a linear one: The highest values for the pollutant are found when the temperature is lowest or highest. Thus, it seems that using a model with a linear relationship between pollutant and temperature is not adequate. Secondly, the plot of residuals versus observation days shows that there is a clear correlation between successive days. This is a violation of the assumption in such regression models that the random errors should be independent. The results of the study are infact invalidated when we observe this correlation.