Petter Mostad
Matematisk Statistik
Chalmers

Suggested solution for exam in
MSA830: Statistical Analysis and Experimental Design
22 October 2010

1. (a) The data is paired, and so we need to consider the five differences

$$\begin{aligned} 114 - 104 &= 10 \\ 107 - 93 &= 14 \\ 123 - 121 &= 2 \\ 111 - 88 &= 23 \\ 92 - 73 &= 19 \end{aligned}$$

Assuming these differences are from a normal distribution with an expectation μ, assuming that this distribution represents increases in the fitness score produced by the training program, and using standard priors for μ and the unknown logged scale λ, we get that the posterior for μ is

$$\mathrm{t}(\overline{y}, n-1, \log\sqrt{s^2/n})$$

where $\overline{y} = 13.6$ is the average, $n = 5$ is the number of observations, and $s^2 = 66.3$ is the variance of the data. The distribution thus becomes

$$\mathrm{t}(13.6, 4, \log\sqrt{13.26})$$

As $t_{0.025,4} = 2.776$ and $\sqrt{13.26} = 3.6414$, the 95% credibility interval thus becomes

$$[13.6 - 2.776 \cdot 3.6414, 13.6 + 2.776 \cdot 3.6414] = [3.5, 23.7]$$

(b) One major assumption is that the differences come from a normal distribution. Given the data of 25 persons, Ralph could see if this assumption was OK for his data by visualizing the data, or possibly by performing a hypothesis test of normality. The best visual way for checking the normality of the data would be to produce a normal probability plot, and judge whether the plot was sufficiently close to a straight line.

(c) Ralph could do a permutation test, where the null hypothesis would be that it was random which of the two measurements for each person what before or after the training program. The test statistic could be the average of the differences, $\overline{y}$. In a large number of simulations, Ralph would multiply each difference randomly with $+1$ or -1, and compute the average of the resulting differences. The sample produced could then be compared with the actual average $\overline{y}$ to see if the actual average was among the smallest 2.5% of the simulated values (or smaller) or among the largest 2.5% of the simulated values (or larger). If so, the test would reject the null hypothesis.

(d) The biggest weakness with Ralphs experiment is that changes in the fitness of the test persons could also be due to other factors than Ralph's program. For example, persons who have just joined a fitness club might increase their fitness (also) in other ways than the specific program. A better design would be to recruit a larger group of test persons from the same population (possibly from persons who have recently joined the fitness club) and then randomly select half of the persons to go through the fitness program, while the other half did not. The differences in fitness scores in the two groups should then be compared in the final analysis.

2. (a) The (empirical) probability is $34/1000 = 0.034$.

   (b) This probability is given by the Binomial distribution, calculating the probability of 2 "successes" in 10 trials when the probability of success is $128/1000 = 0.128$. This is given by

$$\binom{10}{2} 0.128^2 (1-0.128)^{10-2} = \frac{10!}{2!8!} 0.128^2 0.872^8 = \frac{10 \cdot 9}{1 \cdot 2} 0.005477 = 0.2465.$$

   (c) We use a normal approximation for the probability: The Binomial distribution with 100 trials and probability $327/1000 = 0.327$ for success in each trial has expectation $100 \cdot 0.327 = 32.7$ and variance $100 \cdot 0.327 \cdot (1 - 0.327) = 22.0$. Thus the probability we want is the probability for a standard normal distribution to be above

$$\frac{49.5 - 32.7}{\sqrt{22.0}} = 3.58$$

   According to the appropriate table, this probability is $1 - 0.99983 = 0.00017$. Using 49 or 50 instead of 49.5 gives approximately the same result.

   (d) We use Bayes formula. We can write

$$\begin{aligned} \pi(\text{buy} \mid A) &= 0.1 \\ \pi(\text{buy} \mid B) &= 0.05 \\ \pi(\text{buy} \mid C) &= 0.02 \end{aligned}$$

   Using the probabilities found earlier, $\pi(A) = 0.327$, $\pi(B) = 0.034$, and $\pi(C) = 0.128$, we get that

$$\begin{aligned} \pi(A \mid \text{buy}) &= \frac{\pi(\text{buy} \mid A)\pi(A)}{\pi(\text{buy} \mid A)\pi(A) + \pi(\text{buy} \mid B)\pi(B) + \pi(\text{buy} \mid C)\pi(C)} \\ &= \frac{0.1 \cdot 0.327}{0.1 \cdot 0.327 + 0.05 \cdot 0.034 + 0.02 \cdot 0.128} \\ &= 0.8847 \end{aligned}$$

   So the answer is 88%.

3. (a) The ANOVA table becomes

	SS	D.f.	M.sq.	F	p
Additive	96.9024	2	48.4512	4.55	$0.025 < p < 0.05$
Temperature	84.5424	2	42.2712	3.97	$0.025 < p < 0.05$
Residuals	138.3252	13	10.6404		
Total	319.77	17			

The sum of squares were computed with

$$SS_{Temperature} = 6\cdot\left((36.17 - 36.89)^2 + (34.67 - 36.89)^2 + (39.83 - 36.89)^2\right) = 84.5424$$

and

$$SS_{Additive} = 6 \cdot \left((35.17 - 36.89)^2 + (40.17 - 36.89)^2 + (35.33 - 36.89)^2\right) = 96.9024$$

while

$$SS_{Total} = 17 \cdot 18.81 = 319.77$$

and

$$SS_{Residuals} = SS_{Total} - SS_{Additive} - SS_{Temperature} = 319.77 - 96.9024 - 84.5424 = 138.3252$$

(b) The analysis shows that there is a clear ("significant") effect of the additive, and a clear ("significant") effect of the temperature. This conclusion is made under the assumption that there is no interaction between the temperature and the additive, in their effect on the durability. Different additives or different temperatures were not compared pairwise, so the analysis so far does not tell to what extent there is a clear ("significant") difference between specific pairs of additives or temperatures. However, as the highest average durability was observed for additive Z, and for temperature B, this would be the best combination to recommend based on the study and the analysis so far.

(c) The start of the new ANOVA table would be

	SS	D.f.	M.sq.	F	p
Additive	96.9024	2			
Temperature	84.5424	2			
Interaction		4			
Residuals		9			
Total	319.77	17	18.81		

The last line would not change, and the SS and degrees of freedom for the additive and the temperature would not change. However, the residual SS and degrees of freedom values from the previous table is split into the Interaction and Residuals line in this table. As there are a total of 9 combination of classes, the degrees of freedom values for the first three lines sum to 8. The $SS_{Residuals}$ could be computed as the sum of the squares of all data values minus the mean value in each of the 9 combination of classes.

(d) Ulla could for example make plots of the residuals versus each of the predictors (temperature and additive), against the order in which the experiments were done, and against the resulting value of the durability. In each plot, she could investigate whether the residuals seemed approximately independent, and with a constant variance.

4. (a) For example, with the experimental plan

A	B	C	D	E
-	-	-	+	-
-	-	+	-	+
-	+	-	+	+
-	+	+	-	-
+	-	-	-	+
+	-	+	+	-
+	+	-	-	-
+	+	+	+	+

all effects of the factors A, B, C, D, and E can be independently estimated, while the interaction between A and B can also be independently estimated (assuming there is no interaction between C and E)[1]. Here, *D* is constructed as the product between the *A* and *C* columns, while *E* is constructed as a product between all *A*, *B*, and *C* columns.

(b) It matters. If he includes the possible interaction effects in his analysis, there will not be enough data to also estimate the variance (or precision) in the linear model. With no estimate of the variance, Karim cannot compute a credibility interval. Put another way, if the number of parameters in the model equals 8, the posterior for the parameter corresponding to B will not be a proper distribution, and no credibility interval can be found. However, if Karim does not include interaction in his analysis, he will be able to compute credibility intervals.

(c) One can call the design above a 2^{5-2}_{III} design. The *III* indicates that the design has resolution 3: Threre are three columns (for example *A*, *C*, and *D*) that multiply to the identity column (of only +), but there is no pair of columns with this property.

(d) He should perform his 8 experiments in a random order, to avoid that time effects are confounded with the *A* factor. He should try to keep all other factors but the 5 he is varying as constant as possible, to reduce the influence of such factors.

MORE....

5. (a) In this case, the difference between the expectations for measurements at A and C has distribution

$$\mathrm{t}\left(\overline{y}_A - \overline{y}_C, n-k, \log\sqrt{\frac{SS}{n-k}\left(\frac{1}{n_A}+\frac{1}{n_C}\right)}\right),$$

where $\overline{y}_A = 13.93$ and $\overline{y}_C = 17.59$ are the averages of measurements at A and C, respectively, $n = 31$ is the total number of observations, $k = 4$ is the number of groups, $n_A = 7$ and $n_C = 9$ are the number of observations in grous A and C, respectively, and SS is the sum of squares of residuals for the linear model with four groups of observations. As this sum of squares can be computed from the variances in each group, we get that

$$SS = (7-1)\cdot 2.94 + (5-1)\cdot 9.01 + (9-1)\cdot 0.79 + (10-1)\cdot 5.66 = 110.94$$

[1] In the original exam, it was required that Karim should also be able to independently estimate the interaction between D and E. However, no fractional factorial experimental plan exists satisfying all these requirements

Thus the distribution becomes

$$\mathrm{t}\left(13.93 - 17.59, 31 - 4, \log\sqrt{\frac{110.94}{31-4}\left(\frac{1}{7}+\frac{1}{9}\right)}\right) = \mathrm{t}\left(-3.66, 27, \log\sqrt{1.0435}\right)$$

As $t_{0.025,27} = 2.052$ and $\sqrt{1.0435} = 1.0215$, we get the 95% credibility interval

$$[-3.66 - 2.052 \cdot 1.0215, -3.66 + 2.052 \cdot 1.0215] = [-5.76, -1.56].$$

(b) The logged scale λ has distribution

$$\mathrm{ExpGamma}\left(\frac{n-k}{2}, \frac{1}{2}SS, -2\right)$$

with the same definitions as in (a). Thus we get the distribution

$$\mathrm{ExpGamma}\left(\frac{27}{2}, \frac{1}{2}110.94, -2\right) = \mathrm{ExpGamma}(13.5, 55.47, -2)$$

As $\chi^2_{0.975,27} \approx 14$ and $\chi^2_{0.025,27} \approx 43$, we get the 95% credibility interval for e^λ

$$\left[\sqrt{\frac{110.94}{43}}, \sqrt{\frac{43}{110.94}14}\right] = [1.6062, 2.8150]$$

This gives the 95% credibility interval for the precision $1/(e^\lambda)^2$

$$[1/2.8150^2, 1/1.6062^2] = [0.13, 0.39]$$

(c) The difference now follows the distribution

$$\mathrm{t}\left(\bar{y}_A - \bar{y}_C, n_A + n_C - 2, \log\sqrt{s_p^2\left(\frac{1}{n_A}+\frac{1}{n_B}\right)}\right),$$

where s_p^2 is the pooled variance, so that

$$s_p^2 = \frac{(n_A-1)s_A^2 + (n_C-1)s_C^2}{n_A + n_C + 2} = \frac{6 \cdot 2.94 + 8 \cdot 0.79}{14} = 1.7114$$

We get the distribution

$$\mathrm{t}\left(13.93 - 17.59, 14, \log\sqrt{1.7114\left(\frac{1}{7}+\frac{1}{9}\right)}\right) = \mathrm{t}(-3.66, 14, \log\sqrt{0.4346})$$

As $t_{0.025,14} = 2.145$ and $\sqrt{0.4346} = 0.6592$ we get the 95% credibility interval

$$[-3.66 - 2.145 \cdot 0.6592, -3.66 + 2.145 \cdot 0.6592] = [-5.07, -2.24]$$

(d) The difference now has an approximate distribution

$$\mathrm{t}\left(\bar{y}_A - \bar{y}_C, \nu, \log \sqrt{\frac{s_A^2}{n_A} + \frac{s_C^2}{n_C}}\right)$$

where

$$\nu = \frac{\left(\frac{s_A^2}{n_A} + \frac{s_C^2}{n_C}\right)^2}{\frac{(s_A^2/n_A)^2}{n_A-1} + \frac{(s_C^2/n_C)^2}{n_C-1}} = 8.49$$

so that the distribution becomes

$$\mathrm{t}\left(13.93 - 17.59, 8.49, \log \sqrt{0.5078}\right) = \mathrm{t}\left(-3.66, 8.49, \log \sqrt{0.5078}\right)$$

As $t_{0.025,8.48} \approx 2.284$ and $\sqrt{0.5078} = 0.7126$, we get the 95% credibility interval

$$[-3.66 - 2.284 \cdot 0.7126, -3.66 + 2.284 \cdot 0.7126] = [-5.29, -2.03]$$