Biostats VHM 802 Course
Winter 2004, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Question 3 of final exam

Question 3.

The experiment comprised 20 observations and involved the two factors:

Disk drive type	levels A–E
Computer	levels 1–5

Denote by y_{ij} the average access time for the disk drive of type i, i = A,...,E, when measured in computer j, $j = 1, \ldots, 5$. Not all combinations of i and j were tested in the experiment.

A)

In this two-factor design the disk types can be considered as treatments and the computers as blocks. As the blocks do not hold all treatments, it is an *incomplete block design*. It is furthermore a balanced incomplete block design (BIB or BIBD), because any pair of treatments occur together in a block exactly 3 times. The BIBD parameters are: $t = 5$, $b = 5$, $r = 4$, $k = 4$ and $\lambda = 3$. The use of an incomplete block design became necessary because the computers were unable to hold more than 4 disk drives (thus, the blocks could not be complete). The reason to choose a BIB design was that it compares all treatments with the same accuracy — which is generally desirable and particularly appropriate when there are no preferences for specific treatment comparisons (as there would not seem to be in this case). The statistical model is that of a standard block design:

$$y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \tag{1}$$

where the errors ε_{ij} are assumed i.i.d. and normally distributed $\mathrm{N}(0, \sigma^2)$.

B)

The ANOVA table shows that in model (4) the effect of blocks is non-significant ($P = 0.22$) and the effect of treatments (disk types) is weakly significant ($P = 0.042$). No residuals or diagnostics from the model is given, so we cannot check the validity of the model. To rerun the model with residuals and diagnostics would be an obvious addition to the analysis shown.

Before continuing with comparisons of the different disk types, we need to decide whether to reduce the model by removing the non-significant computer effect. The sequential sum of squares for treatments enables us to compute a significance test for types in a reduced model: $F = [168.7/4]/[(106.8 + 65.70)/(11 + 4)] = 3.67$. This value exceeds $F(4, 15, 0.95) = 3.06$, so the effect is still significant ($P < 0.05$) in the reduced model. When conclusions are roughly the same, the choice between the two models (full and reduced) is not definitive. Generally speaking, the full model has the advantage of more correctly reflecting the sample design, and the least squares means do take into account the (moderate) computer effects. Therefore, the continued analysis here will be based on the full model.

The first step in the comparisons of treatment effects is to compute the margin of error for a 95% confidence interval for treatment means. Using the reference value $t(11, 0.975) = 2.201$ and the given SE, we compute a margin of error of $2.201 \cdot 1.599 = 3.52$. Not surprisingly (with the quite high

P-value), most of the confidence intervals overlap. Next, we compute the LSD-value for comparison of two treatments without correcting for multiple testing. Due to the unbalanced nature of the design, the least squares means are not simple means, and the usual formula for computing the SE of their difference *does not work*. However, the effective sample size for each treatment is $\lambda t/k = 3 \cdot 5/4 = 3.75$ (textbook, p. 369). This value is substituted into usual formula, to give $\text{LSD}_{0.95} = t(11, 0.975)\sqrt{\text{MSE} \cdot 2/3.75} = 5.01$ (alternatively, the SE for treatment differences is shown directly in the Stata listing). With this value, treatments A, B, and E would be different from treatment D. The calculation unrealistically assumes these comparisons to be preplanned; instead, we should correct for a total of $5 \cdot 4/2 = 10$ pairwise comparisons. Using the Bonferroni method, we replace the reference value by $t(11, 1 - 0.025/10) = t(11, 0.9975)$. Table B in the textbook tells us only that the value lies somewhere between 3.106 and 4.025. The corresponding BSD-values are 7.07 and 9.16, respectively, and their mean is 8.12. It is clear from these values that only the treatment difference between B and D can be declared significant at an overall 5% error level. It would be desirable to compute the exact BSD-value (it is 7.96). We conclude that there is a weak, but significant difference between the disk drives with respect to the access times, and that drive B has a significantly higher access time than drive D; the difference is 8.4 ms (or a 20% reduction of the access time for drive B). By the non-significant overall computer effect, we do not carry out any pairwise comparisons between computers. The observed differences in the least squares means could very well be contributed to pure noise alone. We conclude that there is no evidence to indicate that the disk drives perform differently in the five computers tested.