Biostats VHM 802 Course
Winter 2008, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to Question 1 of final exam

Question 1.

We use the following notation,

$$y_{ijk} = \text{mean test score for exam } i \text{ at university } k \text{ in a section of the course based on text } j,$$

where $i = 1, 2, 3 \sim$ E1,E2,E3; $j = 1, 2 \sim$ T1,T2; $k =$A,B,C.

A)

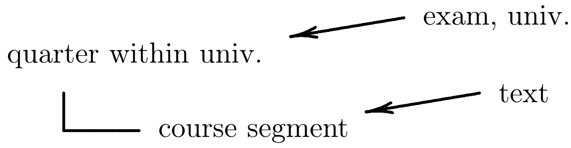
The design is a split-plot layout where the whole plots are the quarters at each university, and the subplots are the course segments taught with different textbooks. The universities would be considered as blocks; within each university there are three whole plots (quarters), and within each whole plot there are two subplots (course segments with different texts). The whole plot factor is exam type, and the subplot factor is text. The randomisation is carried out first for the whole plots and then for the subplots, as described in the study description. The corresponding statistical model is

$$y_{ijk} = \mu + \alpha_i + \beta_j + (\alpha\beta)_{ij} + \gamma_k + A_{ik} + \varepsilon_{ijk},$$

where

- ε_{ijk}'s are the subplot errors, assumed i.i.d. and $\sim \mathrm{N}(0, \sigma^2)$,
- A_{ik}'s are the whole plot random effects, assumed i.i.d. and $\sim \mathrm{N}(0, \sigma_A^2)$,
- α_i's, β_j's, and $(\alpha\beta)_{ij}$'s are the main effects of exam type, text, and their interaction, respectively,
- γ_k's are the university effects, which can be taken either as fixed effects or as random effects, in which case they are assumed i.i.d. and $\sim \mathrm{N}(0, \sigma_C^2)$.

The hierarchical structure of the design has two levels: the whole plots as the upper level, and the subplots as the lower level, as shown below. In the data table, the whole plots can be identified as the combinations of university and exam type.



B)

Among the statistical analyses presented in the listings, the second model is the split-plot model described above. The model has been run in Minitab with random effects of blocks (universities), but this has minimal impact on the analysis. The estimated variance components: $\hat{\sigma}_A^2 = 3.14$ and $\hat{\sigma}^2 = 14.28$, show that the whole plot variation is the smaller of the two, accounting for only $3.14/(3.14 + 14.28) = 18\%$ of the unexplained variation. From the ANOVA table the following conclusions can be drawn:

- no interaction between the effects of exam and text; this means that the expectation of the faculty that the exams would work differently depending on the text used is not supported by the data,
- no significant differences between the exams ($P = 0.30$); with this fairly large P-value one would conclude that the exams indeed evaluate the students consistently, as claimed by the company,
- a significant difference between the texts ($P = 0.03$); although only moderately significant there is evidence to say that the two texts lead to different exam scores,
- a close to significant difference between universities; with $P = 0.051$ it is fair to say that the data offer an indication of differences in the scores of students from the different universities.

C)

The exam effects showed no significance so there is no need for further discussion. The significant difference between the two texts gave evidence that one performed better than the other; the mean score for T2 is 5 units higher than for T1. The estimated standard error of the score difference is $\hat{\sigma}\sqrt{2/9} = \sqrt{14.28 \cdot 2/9} = 1.78$, based on the model with fixed block effects. Therefore, the 95% confidence interval is: $5.0 \pm t(6, .975)\text{SE} = 5.0 \pm 2.45 \cdot 1.78 = 5.0 \pm 4.4$.

The university effect was close to significant, and by inspection of the means it is seen that university C has the highest average score with universities A and B at a similar, lower level. Comparisons between universities are based on the whole plot mean square (MS) and degrees of freedom. Therefore, the LSD value for uncorrected comparisons between universities equals $\text{LSD}_{.95} = t(.975, 4)\sqrt{\text{MSAC} \cdot 2/6} = 2.78\sqrt{20.56 \cdot 2/6} = 7.3$. The difference in mean scores between universities A and C exceeds this value, therefore these would be considered statistically different if this particular comparison had been pre-planned (and no correction was to be employed for multiple comparisons). No other differences attain statistical significance using the LSD method. In summary, there is weak evidence to state that university C has higher scores than university A.