Biostats VHM 802 Course
Winter 2011, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to additional exercise 7.6

The data stem from a greenhouse experiment with pots of cucumber plants to compare 12 lots of sphagnum moss with regard to the volume (air and water content) in the pots. The experiment was carried out using 6 watering troughs in the greenhouse. The troughs can be considered as blocks, because they clearly define a grouping of pots into sets with similar experimental conditions; on the other hand, there does not seem to be any interesting hypotheses associated with the troughs themselves. The sphagnum moss lots would most naturally be considered as treatments. Thus, the design is a block design with one treatment factor. The blocks are incomplete as there are only 8 pots in each block (and 12 treatments). Each treatment occurs 4 times in total, but the number of times two treatments "meet" in the design is not the same for all pairs of treatments. For example, treatments 11 and 12 meet 3 times (in troughs 2, 5 and 6), whereas treatments 9 and 10 only meet 2 times (in troughs 3 and 6). Therefore, the design is not a balanced incomplete block design. We use the notation:

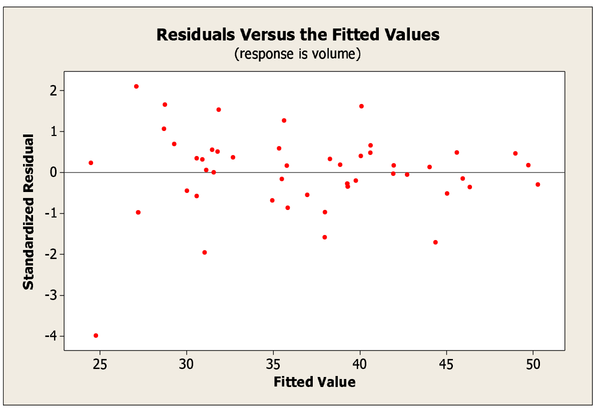
$$\begin{aligned} y_{ij} &= \text{volume \% for pot of sphagnum moss lot } i \text{ in trough } j, \\ i &= 1, \ldots, 12 \sim \text{sphagnum lots (treatments)}, \\ j &= 1, 2, 3, 4 \sim \text{troughs (blocks)}, \end{aligned}$$

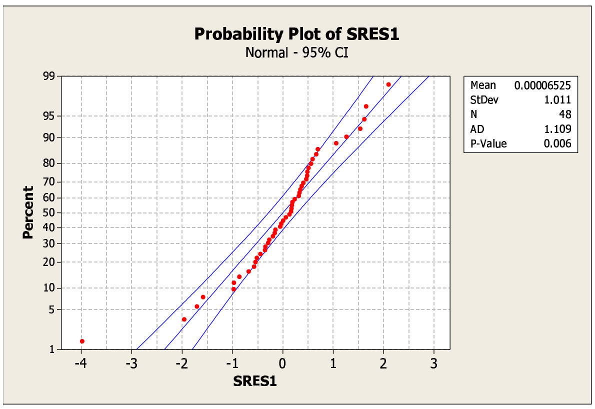
where not all combinations (i, j) occur in the design. For these data, the full model is an additive model in treatments and blocks, and with the possible reductions of this model, the following models can be considered,

$$\begin{aligned} &\text{(A)} \quad y_{ij} = \mu + \alpha_i + \beta_j + \varepsilon_{ij}, \\ &\text{(B)} \quad y_{ij} = \mu + \beta_j + \varepsilon_{ij}, \\ &\text{(C)} \quad y_{ij} = \mu + \alpha_i + \varepsilon_{ij}, \\ &\text{(D)} \quad y_{ij} = \mu + \varepsilon_{ij}, \end{aligned}$$

where the errors ε_{ij} are as usual assumed to be i.i.d. and $\sim \mathrm{N}(0,\sigma^2)$.

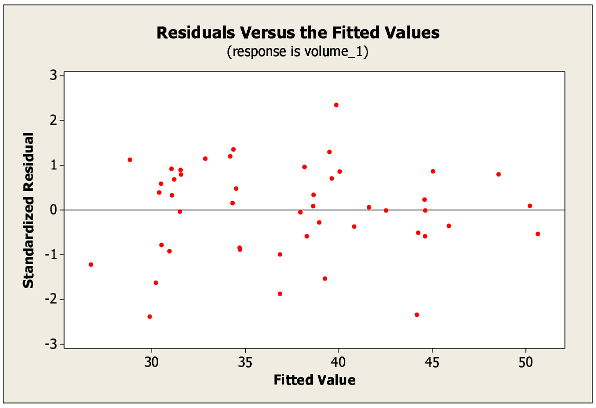
Before embarking on model comparisons we evaluate the residuals of the full model (A). The residual and normal plots (below) show one outlying observation, for lot 12 in trough 2.

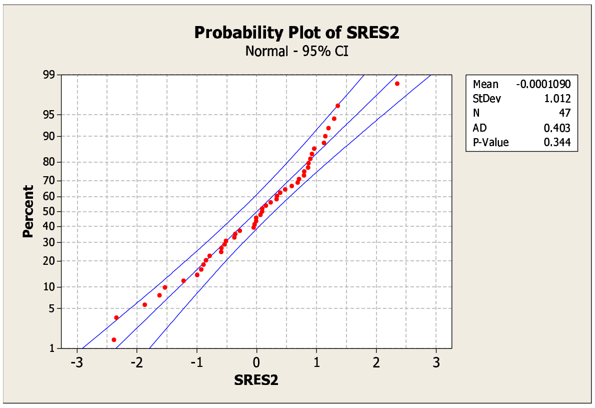




The normality test is clearly significant which is most likely due to this outlying observation. The standardised residual is -3.98 and the deletion residual is -5.61. The corresponding P-value for a significance test based on the deletion residual and the Bonferroni correction is $2 \cdot \mathrm{P}(t(30) < -5.61) \cdot 48 = 0.0002$. Therefore, this observation must be considered a gross outlier; in particular, it is by far the smallest volume recorded. We decide to remove it from the dataset before proceeding with the analysis. Note that by the strong evidence that the observation does not belong in there with the others, some kind of action is necessary. If one is uncomfortable with removing the outlier right away, one option is to carry out the analysis with and without the outlier, and compare the results.

When reanalysing model (A) without the outlier, the residual plots look reasonably well, and the normal probability plot shows no systematic disagreement with a normal sample (as judged by a P-value above 0.10 for the tests of a normal distribution). In this analysis, the most extreme deletion residual is -2.60 which is no cause of concern. We therefore proceed with the analysis of model (A).





The table below gives error sum of squares and degrees of freedom for models (A)–(D), as well as F-test statistics for model reductions. Minitab's ANOVA table for model (A) gives the sum of squares for model reductions (A)→(B) and (A)→(C) as adjusted sum of squares, as well as one of the model reductions (B)→(D) and (C)→(D) as sequential sum of squares, depending of the order of the two model terms.

Model	SSE	DFE	Submodel of	Change in SSE	Change in DFE	F-test	ref.dist.	P
(A)	439.27	30	-	-	-	-	-	-
(B)	2004.28	41	(A)	1565.01	11	9.72	$F(11,30)$	$< 10^{-6}$
(C)	551.65	35	(A)	112.38	5	1.53	$F(5,30)$	0.21
(D)	2146.29	46	(B)	142.01	5	0.58	$F(5,41)$	0.71
(D)	2146.29	46	(C)	1594.64	11	9.20	$F(11,35)$	$< 10^{-6}$

The table shows the treatments (sphagnum moss lots) to be highly significant, whereas there are no significant differences between blocks (troughs). This would allow the model reduction (A)→(C), but certainly not the model reduction (A)→(B). Thus, the further reduction (B)→(D) is of no interest. As to whether the block effect should be removed from the model, one may argue both in favour of that and against it. In favour, because we generally prefer simpler models to complex models if the latter do not add something to the model fit. Against, because the blocks were part of the experimental design, and a quite high P-value is still no "proof" of no block differences. Also, it may be said that the gain in model simplicity by eliminating the blocks is not large, and that the estimate of error variance should be sufficiently precise in model (A) with 30 degrees of freedom. The general consensus among statisticians would be *not* to eliminate the blocks in a situation like this.

Before proceeding a few remarks on partial vs. sequential sum of squares (SS). It cannot be said that one is generally preferable over the other because both have their valid uses. Partial SS are used to test every effect in a given model, on the assumption that all other effects than the one under consideration should be *in* the model. They correspond to the t- and F-tests computed from a given model within the regression framework. These tests are useful for assessing significance when all (other) effects in the model are "interesting" (e.g., if all effects are significant). If several effects in the model are of no interest, they should be removed before assessing the significance of the interesting effects. That is, the model needs to be rerun with a smaller number of effects.
While partial SS are independent on the order of the effects, the sequential SS assess the effects in a given order, usually from the bottom and upwards in the ANOVA table. This implies that an effect is assessed in a model where the effects below it are *out* of the model. One advantage of sequential SS is that the variations attributed to different effects add up to the total variation (which is not the case for partial SS, unless they are equal to sequential SS). On the other hand, sequential SS become little useful for effects listed above other effects that are significant. That is because we usually do not remove significant effects, and therefore the particular sequential SS refers to a model of no real interest. One example of this for the present data is the sequential SS for troughs (142.01) if lots are listed below troughs.

In model (A), the sphagnum lots should be compared by least squares means which take into account that treatments occur in different blocks. By the weak block effect in these data, the difference between least squares means and ordinary means is not large, but in general it could make a huge difference (and would be a big mistake to use the ordinary means). The following table gives least squares sphagnum lot means and standard errors, and an indication by letters of which lots are statistically different by pairwise comparisons using the Bonferroni method. (Two treatments with the same letter cannot be considered statistically different, at the simultaneous 5% error level. The letter coding was constructed by ordering the treatments from lowest to highest and working the way along the sequence, introducing a new letter whenever a new relation (two treatments not being statistically different) occurred. For comparison, the table also lists the simple means, which should *not* be used to represent the levels corresponding to model (A) when the design is unbalanced. The simple means correspond to model (C) without block effects.

Sphagnum moss lot	Simple mean	Least squares means mean	stand.error	comparison
1	42.80	43.21	1.962	cd
2	48.83	49.23	1.960	d
3	39.23	38.63	1.959	abc
4	44.63	43.62	1.959	cd
5	33.55	33.74	1.961	abc
6	31.05	30.57	1.958	ab
7	30.20	29.52	1.959	a
8	37.48	37.66	1.961	abc
9	30.10	30.09	1.959	ab
10	39.70	40.19	1.960	bcd
11	37.05	37.54	1.960	abc
12	32.53	33.55	2.289	abc

The table shows that lot 2 has a higher volume than 8 other lots, and that this is the most remarkable difference among lots. A few differences exist as well between the other lots, but the lot with lowest level differs only significantly from 4 other lots. The simple and least squares means are quite close in value for most of the lots, due to the weak trough effects and the approximate balancedness of the design.

Appendix: tables for analysis of full dataset

Model	SSE	DFE	Submodel of	Change in SSE	DFE	F-test	ref.dist.	P
(A)	900.03	31	-	-	-	-	-	-
(B)	2821.30	42	(A)	1921.27	11	6.02	$F(11,31)$	$< .0005$
(C)	1021.65	36	(A)	121.62	5	0.84	$F(5,31)$	0.53
(D)	3019.43	47	(B)	198.13	5	0.59	$F(5,42)$	0.71
(D)	3019.43	47	(C)	1997.79	11	6.40	$F(11,36)$	$< .0005$

Sphagnum moss lot	Simple mean	Least squares means mean	stand.error	comparison
1	42.80	42.51	2.757	bcd
2	48.83	48.76	2.757	d
3	39.23	39.08	2.756	abcd
4	44.63	44.07	2.757	cd
5	33.55	33.98	2.760	abc
6	31.05	30.83	2.757	abc
7	30.20	29.06	2.756	ab
8	37.48	37.91	2.760	abcd
9	30.10	29.62	2.756	abc
10	39.70	40.40	2.760	abcd
11	37.05	37.75	2.760	abcd
12	26.28	26.89	2.756	a