Solution to Exercise 4.2 (Version 1, 22/09/14)

from Statistical Methods in Biology: Design & Analysis of Experiments and Regression (2014) S.J. Welham, S.A. Gezan, S.J. Clark & A. Mead. Chapman & Hall/CRC Press, Boca Raton, Florida. ISBN: 978-1-4398-0878-8

© S J Welham, S A Gezan, S J Clark & A Mead, 2014.

Exercise 4.2 (Data: courtesy K. Hammond-Kosack, Rothamsted Research)

A laboratory experiment investigated the effect of different treatments on grain production in wheat ears infected with *Fusarium graminearum* (Baldwin et al., 2010). Single wheat ears on 30 separate plants were inoculated with *F. graminearum*. Four treatments (labelled A–D) and a negative (untreated) control were then allocated to the inoculated ears as a CRD. The number of grains in the region above the inoculation position of each ear was counted. The file GRAINS.DAT contains the unit number (*DEar*), the treatment applied (factor Treatment) and the number of grains (variate *Grains*) for each ear.

a) Write down a mathematical model for the numbers of grains.
b) Write down the null and alternative hypotheses associated with this experiment.
c) Construct an ANOVA table by calculating the total, treatment and residual sums of squares and df and then deriving the other columns. Is there any evidence that grain production is affected by the treatments?
d) Calculate the predicted mean for each treatment group and the SED and LSD for treatment comparisons.
e) State your conclusions from this analysis.

(We re-visit these data in Exercises 5.1 and 5.2.)

Data 4.2 (GRAINS.DAT)

Number of grains recorded in wheat ears to investigate the effects of treatments on progress of Fusarium infection:

Ear	Treatment A	Treatment B	Treatment C	Treatment D	Control
1	11	11	9	8	9
2	7	8	7	5	7
3	10	4	5	4	5
4	10	1	3	5	5
5	13	3	5	6	3
6	14	6	3	5	6

Solution 4.2

a) We first label the $t = 5$ treatments; we label treatments A–D with levels $j = 1\ldots4$, respectively, and the control with level $j = 5$. The single factor model for this experiment can be written as

$$Grains_{jk} = Treatment_j + e_{jk},$$

where $Grains_{jk}$ represents the number of grains for the kth replicate ($k = 1\ldots6$) of the jth treatment ($j = 1\ldots5$), $Treatment_j$ is the true (but unknown) population mean for the jth treatment, and e_{jk} is the deviation from the group population mean for the kth replicate of the jth treatment. There are six replicates of each treatment ($n=6$) and there are 30 observations in total, i.e. $N = 30$.

For alternative forms of the model, see the additional material at the end of this solution.

b) The null hypothesis is H_0: $Treatment_1 = Treatment_2 = Treatment_3 = Treatment_4 = Treatment_5$, i.e. that the population means for the five treatments are all equal. The general alternative hypothesis, H_1, is that the treatment population means are not all equal.

c) The total, treatment and residual sums of squares, respectively, can be written as

$$TotSS = \sum_{j=1}^{5}\sum_{k=1}^{6}\left(Grains_{jk} - \overline{Grains}\right)^2,$$

$$TrtSS = \sum_{k=1}^{6}\sum_{j=1}^{5}\left(\overline{Grains}_{j\bullet} - \overline{Grains}\right)^2 = 6\sum_{j=1}^{5}\left(\overline{Grains}_{j\bullet} - \overline{Grains}\right)^2,$$

$$ResSS = \sum_{j=1}^{5}\sum_{k=1}^{6}\left(Grains_{jk} - \overline{Grains}_{j\bullet}\right)^2,$$

where $\overline{Grains}$ is the sample grand mean, calculated as

$$\overline{Grains} = \frac{1}{N}\sum_{j=1}^{5}\sum_{k=1}^{6}Grains_{jk} = \frac{198}{30} = 6.6,$$

and $\overline{Grains}_{j\bullet}$ is the sample mean for the jth treatment (see Table S4.2.1). In practice, the ResSS is most easily calculated by subtraction using the relationship $ResSS = TotSS - TrtSS$.

Table S4.2.1 Calculation of the mean number of grains recorded for each treatment.

Ear	Treatment A	Treatment B	Treatment C	Treatment D	Control
1	11	11	9	8	9
2	7	8	7	5	7
3	10	4	5	4	5
4	10	1	3	5	5
5	13	3	5	6	3
6	14	6	3	5	6
Total	65	33	32	33	35
Mean	10.8	5.5	5.3	5.5	5.8

Table S4.2.2 Calculation of the total sum of squares for grain numbers.

Ear	Treatment	j	k	$Grains_{jk}$	$Grains_{jk} - \overline{Grains}$	$(Grains_{jk} - \overline{Grains})^2$
1	A	1	1	11	4.4	19.36
2	A	1	2	7	0.4	0.16
3	A	1	3	10	3.4	11.56
4	A	1	4	10	3.4	11.56
5	A	1	5	13	6.4	40.96
6	A	1	6	14	7.4	54.76
7	B	2	1	11	4.4	19.36
8	B	2	2	8	1.4	1.96
9	B	2	3	4	−2.6	6.76
10	B	2	4	1	−5.6	31.36
11	B	2	5	3	−3.6	12.96
12	B	2	6	6	−0.6	0.36
13	C	3	1	9	2.4	5.76
14	C	3	2	7	0.4	0.16
15	C	3	3	5	−1.6	2.56
16	C	3	4	3	−3.6	12.96
17	C	3	5	5	−1.6	2.56
18	C	3	6	3	−3.6	12.96
19	D	4	1	8	1.4	1.96
20	D	4	2	5	−1.6	2.56
21	D	4	3	4	−2.6	6.76
22	D	4	4	5	−1.6	2.56
23	D	4	5	6	−0.6	0.36
24	D	4	6	5	−1.6	2.56
25	Control	5	1	9	2.4	5.76
26	Control	5	2	7	0.4	0.16
27	Control	5	3	5	−1.6	2.56
28	Control	5	4	5	−1.6	2.56
29	Control	5	5	3	−3.6	12.96
30	Control	5	6	6	−0.6	0.36
Total	–	–	–	198	0.0	289.20

We draw up tables to aid in the calculation of the total and treatment sums of squares, as in Example 4.1. The calculation of the TotSS is illustrated in Table S4.2.2, giving TotSS = 289.20, equal to the sum of all 30 squared differences between the data values and the sample grand mean (final column of Table S4.2.2).

The calculation of the TrtSS is illustrated in Table S4.2.3. Hence, TrtSS = 6 × 22.53 = 135.20 (equal to 6 times the sum of the squared differences between the treatment sample means and the sample grand mean, as in the final column of Table S4.2.3). Finally,

$$\text{ResSS} = \text{TotSS} - \text{TrtSS} = 289.20 - 135.20 = 154.00.$$

Table S4.2.3 Calculation of the treatment sum of squares for grain numbers.

Treatment	j	n_j	$\overline{Grains}_{j\bullet}$	$\overline{Grains}_{j\bullet} - \overline{Grains}$	$\left(\overline{Grains}_{j\bullet} - \overline{Grains}\right)^2$
A	1	6	10.83	4.23	17.92
B	2	6	5.50	−1.10	1.21
C	3	6	5.33	−1.27	1.60
D	4	6	5.50	−1.10	1.21
Control	5	6	5.83	−0.77	0.59
Total	−	30	−	0.00	22.53

The degrees of freedom associated with the sums of squares are

$$TotDF = N - 1 = 30 - 1 = 29,$$
$$TrtDF = t - 1 = 5 - 1 = 4,$$
$$ResDF = N - t = 30 - 5 = 25.$$

The treatment and residual mean squares are obtained by dividing the treatment and residual sums of squares by their df, i.e.

$$TrtMS = TrtSS / TrtDF = 135.20 / 4 = 33.80,$$
$$ResMS = ResSS / ResDF = 154.00 / 25 = 6.16.$$

The observed variance ratio is then calculated as F = TrtMS / ResSS = 5.487, and this has 4 numerator df (equal to the TrtDF) and 25 denominator df (equal to the ResDF). These results can be combined together in the ANOVA table shown in Table S4.2.4.

Table S4.2.4 ANOVA table for number of grains.

Source of variation	df	Sum of squares	Mean square	Variance ratio	P
Treatment	4	135.20	33.80	5.487	0.003
Residual	25	154.00	6.16		
Total	29	289.20			

The 5%, 1% and 0.1% critical values of the F-distribution with 4 and 25 df are $F_{4,25}^{[0.05]} = 2.759$, $F_{4,25}^{[0.01]} = 4.177$ and $F_{4,25}^{[0.001]} = 6.493$. The 5% and 1% values are smaller than the observed variance ratio of F = 5.487 so we reject the null hypothesis with $P < 0.01$. The observed significance level is $P = 0.003$, which lies between 0.01 and 0.001 as we expect from the critical values. We reject the null hypothesis and conclude that there is strong evidence that population mean grain numbers differ between the treatments.

d) The estimated treatment population means (predicted means) are equal to the sample means given in Table S4.2.1. The SED between the predicted means for the jth and ith treatments, both with replication n, is estimated as

$$SED = \sqrt{\frac{2s^2}{n}} \, .$$

Here we have $s^2 = \text{ResMS} = 6.16$ and $n=6$, so the SED for all treatment comparisons is

$$SED = \sqrt{\frac{2 \times 6.16}{6}} = 1.43 \, .$$

The LSD between any pair of treatments is then

$$LSD = t_{25}^{[0.025]} \times SED = 2.060 \times 1.43 = 2.95 \, ,$$

where $t_{25}^{[0.025]} = 2.060$ is the 97.5$^{\text{th}}$ percentile of the Student's t distribution with 25 df (df equal to ResDF from the ANOVA table).

e) We can conclude that there was strong evidence that the population mean (as number of grains per ear) differed among these treatments ($F_{4,25} = 5.487$, $P < 0.01$). The predicted mean number of grains for Treatment A was roughly double that for the other treatments, which were all similar to the control (A: 10.8, B: 5.5, C: 5.3, D: 5.5, Control: 5.8 grains; SED = 1.43, LSD = 2.95 with 25 df). Treatment A appears to increase grain number on average, but the other treatments appear no different to the control (no treatment).

Additional material

Using the material in Section 4.5, we can set up the single factor model in terms of an overall mean plus the effects of each treatment. This model is over-parameterized and we can impose constraints in several different ways. The nature of the constraints changes the estimates but gives the same ANOVA table and fitted values in each case.

The model can be written in terms of treatment effects with sum-to-zero constraints as

$$Grains_{jk} = \mu + Trt_j + e_{jk} \, ,$$

where μ is the overall population mean across all treatments, and Trt_j is the unknown population treatment effect for the jth group, i.e. the difference between the population mean for the jth treatment and the overall population mean. We impose the sum-to-zero constraint $\Sigma_j \, Trt_j = 0$ to avoid issues with over-parameterization. Using this parameterization, the null hypothesis is still that all population means are equal, now expressed as H$_0$: $Trt_j = 0$ for $j = 1 \ldots 5$. The estimated parameters for this model are given in Table S4.2.5. Estimates of treatment means are calculated as

$$\hat{\mu} + Trt_j \, , \quad \text{(note: the } Trt \text{ symbol should have a hat over it!)}$$

and it is straightforward to verify that this gives the same predicted means as reported in part (d). The SED and LSD for predicted means from this model are also the same as in part (d).

Table S4.2.5. Parameter estimates for single factor model for grain numbers with sum-to-zero constraints.

Parameter	μ	Trt_1	Trt_2	Trt_3	Trt_4	Trt_5
Estimate	6.6	4.2	-1.1	-1.3	-1.1	-0.8

Alternatively, the model can be written using first-level-zero constraints as

$$Grains_{jk} = Treatment_1 + v_j + e_{jk}.$$

Here, $Treatment_1$ represents the population mean for the first treatment (A), the effect of treatment A is constrained to be zero (i.e. $v_1 = 0$), and the effects v_2 to v_5 represent the differences between the population means for the second to fifth treatments (B, C, D and Control, respectively) and that of the first treatment (A). The null hypothesis is still that all population means are equal, now expressed as H_0: $v_j = 0$ for $j = 1\ldots5$. Estimates of treatment means are calculated as

$$Treatment_1 + \hat{v}_j, \quad \text{(note: the } Treatment \text{ symbol should have a hat over it!)}$$

and it is straightforward to verify that this gives the same predicted means as reported in part (d). The SED and LSD for predicted means from this model are also the same as in part (d).

Table S4.2.6. Parameter estimates for single factor model for grain numbers with first-level-zero constraints.

Parameter	$Treatment_1$	v_1	v_2	v_3	v_4	v_5
Estimate	10.8	0.0	-5.3	-5.5	-5.3	-5.0

Finally, the model can be written using last-level zero constraints as

$$Grains_{jk} = Treatment_5 + \omega_j + e_{jk}.$$

Here, $Treatment_5$ represents the population mean for the last treatment (Control), the effect of the control treatment is constrained to be zero (i.e. $\omega_5 = 0$), and the effects ω_1 to ω_4 represent the differences between the population means for the first to fourth treatments (A–D) and that of the last treatment (Control). The null hypothesis is still that all population means are equal, now expressed as H_0: $\omega_j = 0$ for $j = 1\ldots5$. Estimates of treatment means are calculated as

$$Treatment_5 + \hat{\omega}_j, \quad \text{(note: the } Treatment \text{ symbol should have a hat over it!)}$$

and it is straightforward to verify that this gives the same predicted means as reported in part (d). The SED and LSD for predicted means from this model are also the same as in part (d).

Table S4.2.7. Parameter estimates for single factor model for grain numbers with last-level-zero constraints.

Parameter	$Treatment_5$	ω_1	ω_2	ω_3	ω_4	ω_5
Estimate	5.8	5.0	-0.3	-0.5	-0.3	0.0