Biostats VHM 801 Course
Fall 2003, Atlantic Veterinary College, PEI
Henrik Stryhn

Solution to assignment IV

The review and discussion of methods given here is generally more advanced than in Chapter 13 of the textbook. All analyses shown used Minitab, but Stata or other programs should give the same results.

Part/Experiment 1

The Plants2 dataset of the textbook contains measurements of fresh biomass weight for 112 plants in a greenhouse/growth chamber experiment.

Question 1: The statistical design of the experiment

The 112 plants were of 4 different species (coded as 1–4) and were subjected to 7 different watering levels (coded as 1–7 but corresponding to 50, 150,...,650 mm). The statistical design is a 2-way factorial or a 2-factor (or 4×7) layout with 4 replications of each combination of species and water, and therefore balanced.

Question 2: Randomization

The 4 plants per species and watering condition should be selected randomly. Since the species is an inherent property of each plant and cannot be imposed as a treatment, the experiment is based on 28 plants of each of the 4 species. These plants should be suitably selected to make them representative for their species. Then the 28 plants of each species should be distributed randomly onto the 7 water groups, like in a completely randomized experiment (p. 237 in IPS). This could e.g. be achieved by drawing cards from a hat, or by using software to select a random reordering (permutation) of the numbers from 1 to 28, whereafter the 4 first ones will go to water level 1, the next 4 to water level 2, etc.

Question 3: 2-way ANOVA

The statistical model is

$$\begin{aligned} X_{ijk} &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad \text{where} \\ & \quad X_{ijk} \text{ is the fresh biomass of plant } k \text{ of species } i \text{ and water level } j, \\ & \quad \varepsilon_{ijk}\text{'s are "errors "assumed to be i.i.d. and } \sim N(0,\sigma), \\ & \quad i = 1,\ldots 4 \sim \text{species},\ j = 1,\ldots,7 \sim \text{water levels},\ k = 1,\ldots,4 \sim \text{plants}. \end{aligned}$$

Alternatively, the model can be written using a model formula:

```
FBIOMASS = SPECIES + WATER + SPECIES*WATER + ERROR.
```

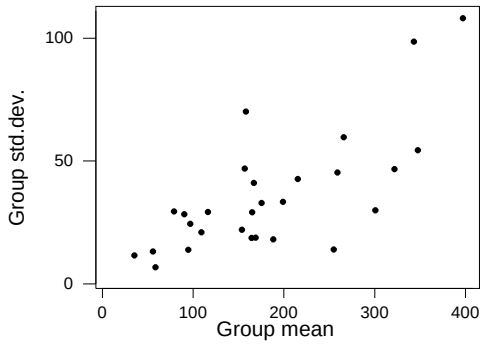
The ANOVA table is shown in the Minitab listing below.

Source	DF	SS	MS	F	P
species	3	458295	152765	81.45	0.000
water	6	491948	81991	43.71	0.000
Interaction	18	60334	3352	1.79	0.040
Error	84	157551	1876		
Total	111	1168129			

The ANOVA table shows strongly significant effects of species and water levels as well as a weakly significant interaction between the two. In a first conclusion, this means that both the species and the water levels show impacts on the fresh biomass, and by the interaction these impacts depend to some degree of the level of the other factor. To study this further, one would most naturally look at the interaction plot, but we will check the model assumptions first.

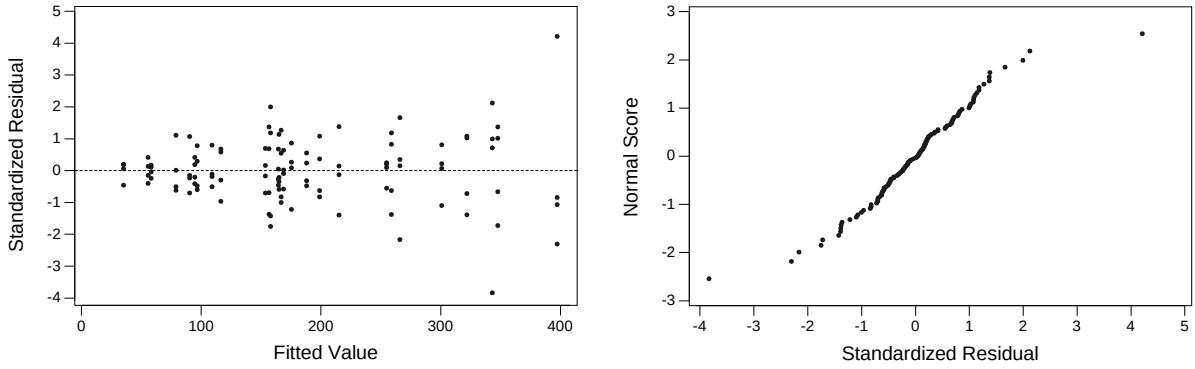
Question 4: Model validation

Two procedures of model validation are discussed here: comparison of within-group variation across groups and inspection of residuals. For the first one, we compute the means and standard deviations of the 28 groups formed by the two factors. The standard deviations range from 6.8 to 108.0, clearly violating the IPS guideline (max/min $\leq$ 2). However, when testing the hypothesis of equality of variance across the groups, one test (Levene's test) is nonsignificant whereas the other test (Bartlett's test) is clearly significant, leaving as somewhat inconclusive on this point. On the other hand, the graph below shows that the standard deviation increases with the mean.



The plot shows a clear and systematic violation of the model assumptions. A linear increase in the standard deviation with the mean makes a log-transformation is useful to stabilize the variance. This does not look unreasonable from the plot but a slightly (upwards) curved type of association seems possible as well.

A model validation from the residuals could use both the raw or standardized residuals, the latter being generally preferred because of the advantages of a reference distribution ($N(0,1)$) to assess extreme values. The next figures show the standardized residuals plotted against the fitted values and plotted in a normal probability plot.



The first plot also shows the increasing variance with the mean (a "fan" or "cone" shape). The spread of the points is lowest to the left and increases to the right. The normal plot shows a reasonably symmetric distribution with two extreme points; these values are -3.83 and 4.21. Even if the values

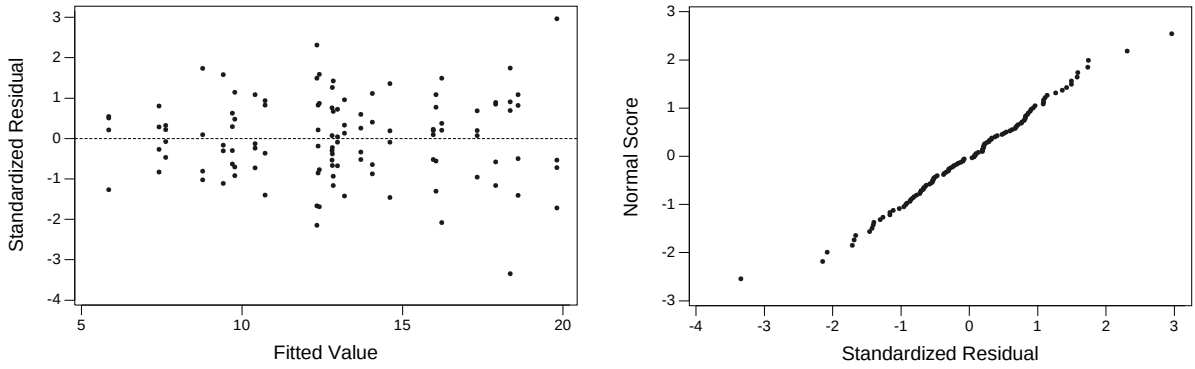
are quite extreme in a $N(0,1)$ and the normal distribution assumption might seem acceptable without them, it is not of interest to remove them. This is because even without these points there is a clear tendency of increasing variation with increasing mean (as seen most clearly in the first plot above). Therefore, removing those points would not solve all problems with the model. Also, if the variation increases with the mean, the points might be perfectly sensible because they arise from the two groups with the highest means. In conclusion, we have identified a clear violation of the model's assumption of constant variances across all groups. Whether the ANOVA method is sufficiently robust to such model violations to make our previous conclusions valid even here, is not easy to say. However, if we could come up with a better model (e.g. by transformation, see below) it would be hard to justify to stick to this clearly deficient model.

Question 5: Square-root and log transformations

We apply the same procedures as above to the square-root and log transformed data. The table and figures below give the results for the square-root transformed fresh biomass weights.

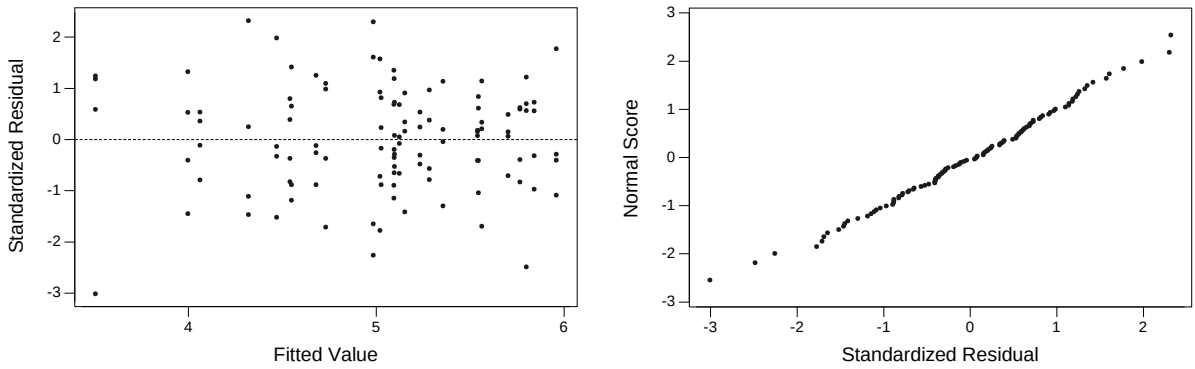
Analysis of Variance for rootfbio

Source	DF	SS	MS	F	P
species	3	638.11	212.70	99.32	0.000
water	6	753.84	125.64	58.67	0.000
Interaction	18	40.80	2.27	1.06	0.407
Error	84	179.89	2.14		
Total	111	1612.64			



The ANOVA table shows still strongly significant effects of species and watering, but the interaction is non-significant. The residual plots show some improvement relative to the ones for the untransformed data, but the pattern of increasing variation with increasing mean remains. The extreme residuals noted for the untransformed data remain visible but are less extreme in value and of little concern. The standard deviations for the 28 species $\times$ watering groups range from 0.44 to 2.88, but both tests for equal variances are non-significant. The next table and figures give the results for the log-transformed data.

```
Analysis of Variance for lnfbio
Source        DF        SS        MS        F        P
species        3   16.6932    5.5644   101.45    0.000
water          6   21.8933    3.6489    66.53    0.000
Interaction   18    1.4404    0.0800     1.46    0.127
Error         84    4.6074    0.0548
Total        111   44.6343
```



The ANOVA table looks similar to the previous one, but the P-value of the interaction has dropped to 0.13, which is still non-significant but getting close. However, no obvious patterns are seen in the interaction plot (not shown), and we could therefore disregard the interaction. The residual plot have changed remarkably; the tendency of increasing variation with the mean has certainly been eliminated, but perhaps there is now a tendency the other way around (indicating that the log transformation was "too strong"; a more detailed analysis using the so-called Box-Cox transformation method shows the optimal transform to be around $x^{0.25}$ or $\sqrt[4]{x}$). The normal plots looks almost perfect, and the standardized residuals are now within the much narrower range of -3.01 to 2.42. The standard deviations for the 28 species × watering groups range from 0.056 to 0.46, but both tests for equal variances are non-significant. In conclusion, both transformations improve the model's compliance with the ANOVA assumptions, and there is little to choose between them (and conclusions should be very similar). We choose the log transformation by the more narrow range of the standardized residuals and its easier interpretation.

Question 6: Analysis of log-transformed data

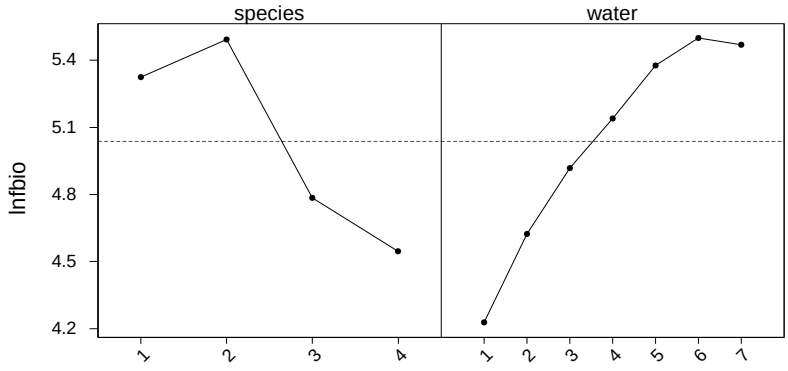
The statistical model is

$$\begin{aligned} \ln(X_{ijk}) &= \mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}, \quad \text{or} \\ X_{ijk} &= \exp\{\mu + \alpha_i + \beta_j + \gamma_{ij} + \varepsilon_{ijk}\} \\ &= \exp\{\mu\} \times \exp\{\alpha_i\} \times \exp\{\beta_j\} \times \exp\{\gamma_{ij}\} \times \exp\{\varepsilon_{ijk}\}, \end{aligned}$$

which shows that an additive model on logarithmic scale corresponds to a *multiplicative* model on original scale. This allows a particularly simple interpretation of effects on logarithmic scale. For example, if watering level 2 is estimated to be 0.4 higher than level 1 on logarithmic scale, this corresponds to level 2 being $\exp\{0.40\} = 1.49$ times the value of level 1, in other words the fresh biomass weight is almost 49% higher at 150 mm of watering than at 50 mm of watering. This type of reasoning applies to a general comparison of watering levels when there is no interaction (the $\gamma_{ij}'s$) in the model, or to a comparison for a specific species. For these data, we previously concluded the interaction to be non-significant and ignorable.

By the non-significant interaction, we present the results for the species and waterings separately. To

begin with, we plot the estimated levels of the species and watering levels. For simplicity, these are plotted on the log scale, but the plots could also be done on original scale.



The plots give a first impression of the species and watering effects which we supplement with tables of estimates, 95% confidence intervals and values to perform pairwise comparisons at individual and overall 5% error levels. Throughout we use the values $s = \sqrt{\text{MSE}} = \sqrt{0.05485} = 0.2342$ and $t^* = t_{.975,84} = 1.989$. For the Bonferroni procedure (using $\text{MSD}_{.95}$), we use $t^{**} = t_{1-.025/6,84} = 2.702$ for the species (6 comparisons). Note that the backtransformed group mean is an estimate of the *median*, not the mean for the group values on original scale.

			Species			
Scale	Statistic	Formula	1	2	3	4
natural	mean	$\overline{\ln X}_{i\cdot\cdot}$	5.324	5.493	4.785	4.545
log	95% CI	$\pm t^* s\sqrt{1/28}$		± 0.088		
	$\text{LSD}_{.95}$	$t^* s\sqrt{2/28}$		0.124		
	$\text{MSD}_{.95}$	$t^{**} s\sqrt{2/28}$		0.169		
original	median	$\exp\{\overline{\ln X}_{i\cdot\cdot}\}$	205.2	243.0	119.7	94.2
	CI lower	$\exp\{\overline{\ln X}_{i\cdot\cdot} - t^* s\sqrt{1/28}\}$	187.9	222.5	109.6	86.2
	CI upper	$\exp\{\overline{\ln X}_{i\cdot\cdot} + t^* s\sqrt{1/28}\}$	224.1	265.3	130.7	102.8

The overall F-test for species is strongly significant, and the pairwise comparisons show that all four species are significantly different as well, with a slight reservation for species 1 and 2 whose difference is exactly borderline ($P \approx 5\%$) using the Bonferroni method. Further, the species' fresh biomass weight are grouped; species 1 and 2 have high biomass weights, and species 3 and 4 have clearly lower biomass weights. We turn now to the watering, the factor of primary interest in the study. Note that the nonsignificant interaction, the effect of watering is the same for all species on the log-scale; this corresponds to the same *relative effects* of watering on fresh biomass weights for all species.

Scale	Statistic	Formula	Watering (mm) 50	150	250	350	450	550	650
natural	mean	$\overline{\ln X}_{\cdot j\cdot}$	4.228	4.624	4.918	5.140	5.378	5.500	5.470
log	95% CI	$\pm t^* s\sqrt{1/16}$				± 0.116			
	$\text{LSD}_{.95}$	$t^* s\sqrt{2/16}$				0.165			
original	median	$\exp\{\overline{\ln X}_{\cdot j\cdot}\}$	68.6	101.9	136.7	170.7	216.6	244.7	237.5
	CI lower	$\exp\{\overline{\ln X}_{\cdot j\cdot} - t^* s/\sqrt{16}\}$	61.1	90.7	121.8	152.0	192.9	217.9	211.5
	CI upper	$\exp\{\overline{\ln X}_{\cdot j\cdot} + t^* s/\sqrt{16}\}$	77.0	114.4	153.5	191.7	243.2	274.8	266.7

The overall F-test for watering is strongly significant, and the plot of group means show a yield increase with increased watering up till the last two or three levels where a plateau seems to be reached (possibly the yield drops off at further watering). As we know the actual watering levels, it would be possible to model the watering by a regression equation instead of as a factor (estimating unrelated means for each watering level); however, this is beyond the scope of the assignment. Anyway, the primary interest is in the response profile as depicted in the plot, and not in carrying out all multiple comparisons between the 7 watering levels. Therefore, we abstain from the formal (Bonferroni) comparisons, and simply use the $\text{LSD}_{.95}$ value and the confidence intervals to indicate the precision of the estimated means. It is seen that there is no evidence to distinguish the yields at the three highest watering levels, from which we conclude that the optimal yield would be achieved within a watering range of 450–650 mm. The increases in yield from the lowest (50 mm) to the highest (650 mm) watering level are considerable and highly significant. The estimated difference on log scale is $5.470 - 4.228 = 1.242$ with a margin of error of 0.165; on original this corresponds to a $\exp\{1.242\} = 3.46$ fold increase, with a 95% CI of $(\exp\{1.077\}, \exp\{1.407\}) = (2.94, 4.08)$. We previously computed the estimated yield at watering 150 mm to be 1.49 times that at watering 50 mm; the 95% CI is (1.26,1.75). By the lack of interaction, all species tolerate stress equally well, on a relative scale.

Part/Experiment 2

The Plants1 dataset of the textbook contains measurements for 252 plants in a greenhouse/growth chamber experiment of the percent of the plant that consists of nitrogen. The design is similar to the one in the Plants2 dataset, except that there are 9 instead of 4 replications of the (same) species and watering combinations.

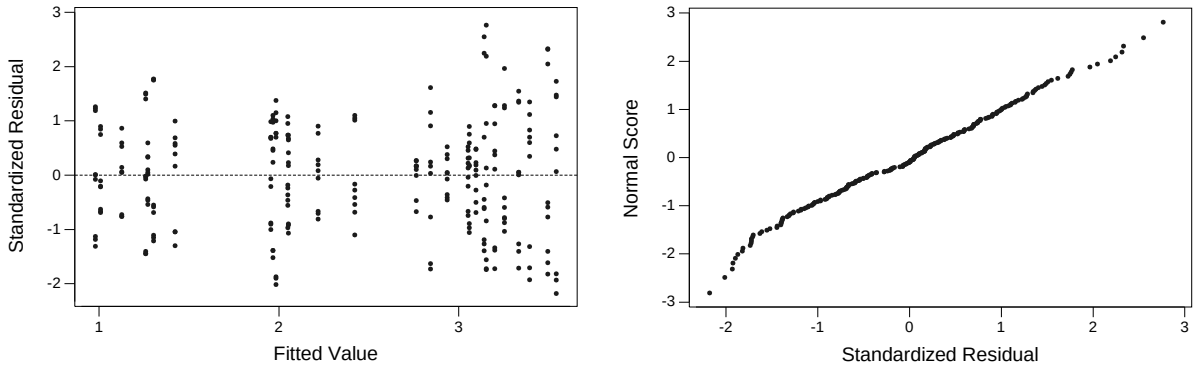
Question 1: 2-way ANOVA

The same statistical design and model as in part 1 give the ANOVA table below.

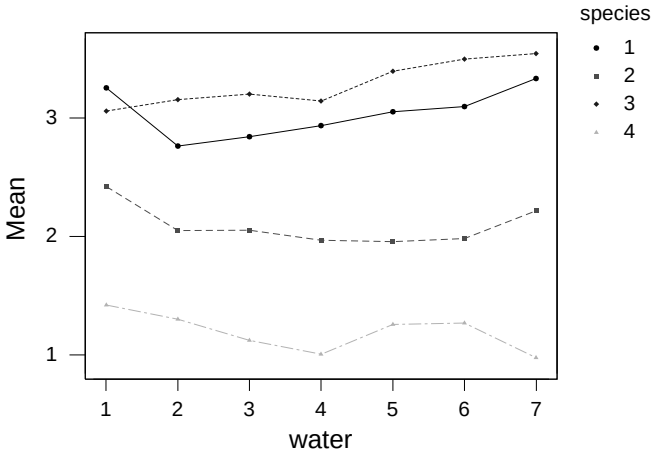
```
Analysis of Variance for pctnit
Source        DF        SS        MS        F        P
species        3  172.3916   57.4639  1301.32    0.000
water          6    2.5866    0.4311     9.76    0.000
Interaction   18    4.7446    0.2636     5.97    0.000
Error        224    9.8914    0.0442
Total        251  189.6143
```

There are strongly significant effects of the species and water levels as well as their interaction. Before proceeding towards the conclusions, we need to check the model assumptions. The within-group standard deviations range from 0.067 to 0.344 among the 28 groups, thus clearly violating the IPS guideline, and both Bartlett's and Levene's test for equal variances are significant (the former

strongly significant). Thus, there *is* evidence against the assumption of equal variances. However, a plot of group standard deviations against group means like in part 1 does not show any increase in the standard deviations with increasing mean (not shown).



The residual plot does not show the same pattern of increasing variance with increasing fitted values as in part 1, and generally looks quite ok. The normal plot looks fine as well. In summary, the invalidated variance assumption gives rise to some concern about the ANOVA results, but there is no clear pattern seen in how the variances are not constant across groups, and there is no easy way to solve the problem (e.g., the power and log-transformations do not help). We decide to continue the analysis while being a bit cautious in our conclusions. By the clearly significant interaction, the interaction plot is the natural base for analysis beyond the ANOVA table.



The plot shows 4 roughly parallel and roughly horizontal lines for the 4 species. Species 1 and 3 are clearly higher in percent nitrate levels than species 2 which again is above species 4. Species 1 and 3 show a slightly increasing nitrate levels with increasing water whereas the pattern for species 2 and 4 is more irregular. Thus, there is no clear and easily interpretable pattern in the interaction. Yet, as it is highly significant, we should base our conclusions on the 28 means in the plot. For convenience, the values are given in the table below, with the margin of error for a 95% CI and the $\text{LSD}_{.95}$ (using $s = \sqrt{\text{MSE}} = 0.21014$ and $t^* = t_{.975,224} = 1.971$).

	Watering (mm)						
Species	50	150	250	350	450	550	650
1	3.254	2.764	2.843	2.936	3.052	3.096	3.333
2	2.422	2.050	2.052	1.967	1.956	1.984	2.218
3	3.059	3.154	3.200	3.142	3.396	3.496	3.544
4	1.423	1.304	1.125	1.009	1.258	1.271	0.979

margin of error$=t^* s/\sqrt{9} = 0.137$, $\text{LSD}_{.95} = t^* s\sqrt{2/9} = 0.195$

Using the $\text{LSD}_{.95}$ for informal comparisons, we see that not only are the species 2 and 4 clearly different and lower than 1 and 3, but the latter two also differ at several of the waterings. It is therefore fair to say that there is some evidence that species 3 overall (at least for watering above 50 mm) has a higher nitrate percentage than level 1 (a precise comparison could be done using contrasts). Further, we compare the watering levels within each species; it is seen that for all species there are some differences between the watering levels. All species except no. 3 show a drop in the outcome from 50 mm to 150 mm, and it seems significant for species 1 and 2. Other interesting effects are maybe the increases with increased watering for species 1 and 3. If the question in terms of drought stress tolerance was to select the species with highest nitrate percentage at 50 mm watering, the choice would be species 1 or 3 (their difference is at most borderline significant), and the latter may be preferable because of its higher values for increased water levels.

Question 2: Separate analyses for each species

The factor with huge differences between its levels is the species. Possible reasons for preferring separate analyses for the levels of this factor are: i) less problems with the model assumptions, ii) little interest in the comparisons between species, or iii) little interest in common effects of other the factor(s) across the species (that is, in the watering main effects) because the species are not expected to show similar effects to waterings anyway. As to ii), we don't know how much interest there was in overall species comparisons, but the main interest was in the watering. As to iii), the previous analysis showed a clear interaction with no obvious interpretation (therefore, results need to be interpreted by species anyway). The table below gives selected statistics from the separate analyses, to assess item i) and to summarize the results. The models are those of a one-way ANOVA (seven watering groups and 9 observations per group).

Species	max/min std.dev.	pooled s	F-value	P-value
1	0.248/0.067=3.73	0.166	14.3	$<$0.0005
2	0.290/0.124=2.34	0.186	7.51	$<$0.0005
3	0.344/0.153=2.25	0.282	4.15	0.002
4	0.266/0.080=3.35	0.187	6.85	$<$0.0005

The comparison of group standard deviations shows that the variation between groups is somewhat smaller for species 2 and 3; in addition, for these species there is no evidence against the assumption of equal variances (results not shown). As to the differences in variation across species, the pooled standard deviation for species 3 is slightly larger than for the other species. In summary, with respect to item i) the gain by analysing the species separately is not obvious. All the one-way ANOVAs show clearly significant differences between waterings, but the evidence is least pronounced for species 3. We can now carry informal or formal (Bonferroni-adjusted) comparisons among the watering groups

for each species. As the group means are already listed in the table above, we omit the details.

Question 3: Analysis of means

A more detailed explanation of why the analysis of means might be more appropriate than the analysis of individual plant values is beyond the course (the question is whether the plants represent *subsampling* or true replicates of each watering condition). By analysing the means we effectively use the interaction as our error variation, thus measuring the effects of species and watering relative to the interaction. The model is a 2-way ANOVA without interaction:

$$\bar{X}_{ij\cdot} = \mu + \alpha_i + \beta_j + \varepsilon_{ij},$$

where still $i \sim$ species and $j \sim$ watering and the errors ε_{ij} are assumed i.i.d. from $N(0, \sigma)$. The ANOVA table becomes:

Source	DF	SS	MS	F	P
species	3	19.1546	6.3849	218.00	0.000
watering	6	0.2874	0.0479	1.64	0.195
Error	18	0.5272	0.0293		
Total	27	19.9692			

The species still show a clear statistical significance, but there are no significant differences between the waterings. An (informal) comparison of species 1 and 3 is weakly significant ($\text{LSD}_{.95} = t_{.975,18}\sqrt{\text{MSE}}\sqrt{2/7} = 0.19$), and the other comparisons are still clearly significant (with or without Bonferroni adjustment). The residual plot does not look too strange, but the normal plot shows some departure from the straight line (not shown). Nevertheless, the model can be considered acceptable. The results show that *relative to the interaction* there is no effect of watering. This is not too surprising because the overall effects of watering were not clear in the previous analyses either. It is tempting to conclude that the significant interaction represents some kind of "noise" (no matter whether it resulted from a suspect or correct statistical design) and that the useful conclusions from the experiment are the species differences and the non-significant watering effects.