UNIVERSITY OF CAPE TOWN

2019

Course: STA2007F

Student details: Kyle Du Plessis (DPLKYL002)

Malikah Hardenberg (HRDMAL001)

Tanweer Beckett (BCKTAN001)

Project 2: Experimental Design

Due Date: 6th May 2019

Abstract:

The purpose of the study was to determine the effect increasing nitrogen concentration has on plant biomass and growth. We chose to measure leaf area as our biomass indicator. The plant we chose as our subject was cowpea, Vigna unguiculata, seedlings as they have a shorter germination time and would be ideal to use in such an experiment.

We chose to have 12 replicates of our treatment levels, and so had 60 experimental units. We planted 60 seeds into 60 individual plant pots. We adhered to the principles of completely randomized design and used R to randomly assign treatments. The plants were placed in the same area to limit the amount of variation we could not control such as temperature and air pressure. They were watered every day at the same time with a nutrient solution adjusted for N treatments as follows: 0, 2,8,16,24 mM NO_3. Because we could not determine the already present nitrate concentration in the soil we used, the treatment levels indicate the amount of nitrate added to the soil, and therefore not the total amount of nitrate in the experimental unit.

After 4 weeks we measured the leaf area of each of the 60 plants and used R to carry out statistical analysis. We fitted an ANOVA model with contrast parametrization to the data to determine whether or not there was a difference in leaf area between plants that received the treatments 2mM, 8mM, 16mM, 24mM and the plants that received the control, 0mM. The difference estimate between treatments 2mM and 0mM was 45.539, 120.050 between 0mM and 8mM, 144.672 between 0mM and 16mM and 195.177 between 0mM and 24mM. The p-values we obtained for these estimates were very small ($p<0.005$) which indicated strong evidence against the null hypothesis that we would observe no difference in SLA between plants that received the treatment and those that didn't.

We also wanted to determine if there were differences between all treatment levels and so we used Tukey's method in R to obtain pairwise comparisons.

Treatment pair	Difference estimate
8mM NO3-2mM NO3	74.511
16mM NO3-2mM NO3	99.133
24mM NO3-2mM NO3	149.638
16mM NO3-8mM NO3	24.622
24mM NO3-8mM NO3	75.127
24mM NO3-16mM NO3	50.505

Table 1.0: Part of the result obtained from the Tukey's pairwise comparison.

We obtained very small p-values for these estimates as well ($p<0.005$). Based on our analysis of our experimental data we concluded that there was enough evidence to suggest that increasing nitrate concentration in of the soil did increase leaf area, and therefore biomass.

Introduction:

Plant growth is a result of the amount of carbon the plant is able to fix during photosynthesis. Nutrients such as nitrogen and various leaf traits are important factors that effects the general wellbeing of the plant, based on its ability of these factors to allow growth and maximize photosynthesis. The relationship between the growth capacity and morphological and physiological leaf traits has been observed in numerous species, according to (Poorter et al. 1990). Leaf trait variation can affect both water and nitrogen use efficiency and productivity (Reich et al. 1997). Nitrogen is seen as an important macro element needed by plants to survive. A reduction in nitrogen in plants is seen to reduce chlorophyll content (which then has an effect on the ability to capture light and hence photosynthesize) and this therefore affects plant growth.

The aim of the experiment is to determine the effect of nitrogen-adjusted nutrient solutions on plant growth and specific leaf area for cowpea plants (*Vigna unguiculata*) grown in pots filled with sand for a period of approximately 3-4 weeks and watered with a nutrient solution with nitrogen adjusted as follows: 0mM NO_3^-, 2mM NO_3^-, 8mM NO_3^-, 16mM NO_3^- and 24mM NO_3^-. Specific leaf area is a leaf photosynthetic trait and is known to be a key leaf characteristic used during the study of leaf traits (Hoffmann et al. 2005).

Our objectives / research questions for this experiment are:

1. Do any of these treatments differ from each other (if not, all treatments led to the same specific leaf area)?

2. Do these treatments differ from the control in the resulting specific leaf area examined?

3. Do these treatments differ from each other in the resulting specific leaf area examined?

We hypothesized: increasing the level of nitrogen in the nutrient solutions applied to the cowpea plants increases plant growth / biomass and specific leaf area.

Methods and Data:

Single factor completely randomized design:

The treatment structure is single factor and the blocking structure used is a completely randomized design. The response is the specific leaf area - SLA $[cm2/g]$.The treatment factor is the nitrogen-adjusted nutrient solution added to the normal watering regimen. There are 5 treatment levels (or treatments): 0mM NO_3^-(plain water), 2mM NO_3^-, 8mM NO_3^-, 16mM NO_3^- and 24mM NO_3^-. The 0 mM NO_3^- treatment is the control treatment. The experimental units are the pots and the observational units are the leaf area taken from each plant. Also, we do not expect any other effects that could introduce systematic changes in the response over space or time. Therefore, no blocks are required. We suspected no spatial gradients and all cowpea plants are planted under identical conditions (same size of container, same volume of soil, etc.) so that the comparison between plants receiving the treatments is not confounded with anything else. We want to know whether increasing the level of nitrogen in the nutrient solutions applied to the cowpea plants increases plant growth and specific leaf area at all. Therefore, we require a control treatment with zero nitrogen in the nutrient solution.

Each treatment was replicated 12 times and we have 60 experimental units in total. This is a balanced experiment – each treatment has the same number of replicates.

Vigna unguiculata (cow pea) seeds were bought from a local nursey and grown in pot plants for 4 weeks. They were watered every day at the same time with a nutrient solution adjusted for N treatments as follows: 0, 2,8,16,24 mM NO_3^-. After 4 weeks, leaf trait characteristics from each pot plant was measured.

Other considerations:

Time: Watered plants every day at the same time for 4 weeks and thereafter leaf trait characteristics were measured.

Experimental setup: Plants were grown in a lab growth chamber, where all other conditions were controlled.

Cost: Seeds - R14.50 per packet (20 seeds) = R 43.50 (60 seeds)

Soil - R120 (20kg bag of soil)

Plastic pots - R5.50 (10 plastic pots) = R33 (60 plastic pots)

N treatment - R145 (per bottle of N liquid) = R435 (3 bottles)

Total cost = R631.50

Therefore, it can be seen that the experiment is feasible with respect to costs and time. The experiment is ethical, and causes no harm to any creatures or the environment.

Randomisation in R:

R output:

```
 [1] "0"  "2" "8" "16" "8" "16"  "2" "24"  "8"  "16"  "0" "0" "0"  "24"  "0"  "2" "2"
"8" "16"
[20] "8"  "16"  "24"  "2"  "16"  "16" "24" "24"  "24"  "0"  "8"  "24" "24"  "8" "0"
"16" "0"  "16"  "2"
[39] "0"  "0"  "16"  "8" "24"  "8"  "24"  "24" "0"  "2" "8"  "0"  "2" "24" "2" "8"  "2"
"2"  "8"
[58] "16" "2" "16"
```

Randomisation has been done using R. For a completely randomized design, the experimental units are not blocked, so the treatments and their replicates are assigned completely at random to all experimental units available. Each treatment is randomly assigned to experimental units and each experimental unit is equally likely to receive any of the treatments.

Pot number	Treatment [mM NO3]	SLA [cm2/g]
1	0	184.383
2	2	194.694
3	8	263.029
4	16	335.738
5	8	351.844
6	16	347.461
7	2	229.386
8	24	391.596
9	8	252.234
10	16	340.354

Table 1.1: Part of the data set of the experiment that examined the effect of nitrogen-adjusted nutrient solutions on plant growth and specific leaf area for cowpea plants (Vigna unguiculata)

Statistical Analysis:

Assumptions of the completely randomised design:

We have analysed this experiment as a completely randomised design and checked that the data matched the few key assumptions. They are:

1. **The analysis matches the design (randomisation, sample size, blocking).**

In our experiment, we have one factor with five levels (nitrogen-adjusted nutrient solutions including the control). Treatments are randomised to experimental units without restriction (i.e. no blocking had been used). We also need to check that we have one observation per experimental unit. In our experiment, the experimental units are the pot. We also have one measurement per pot, so this assumption is met.

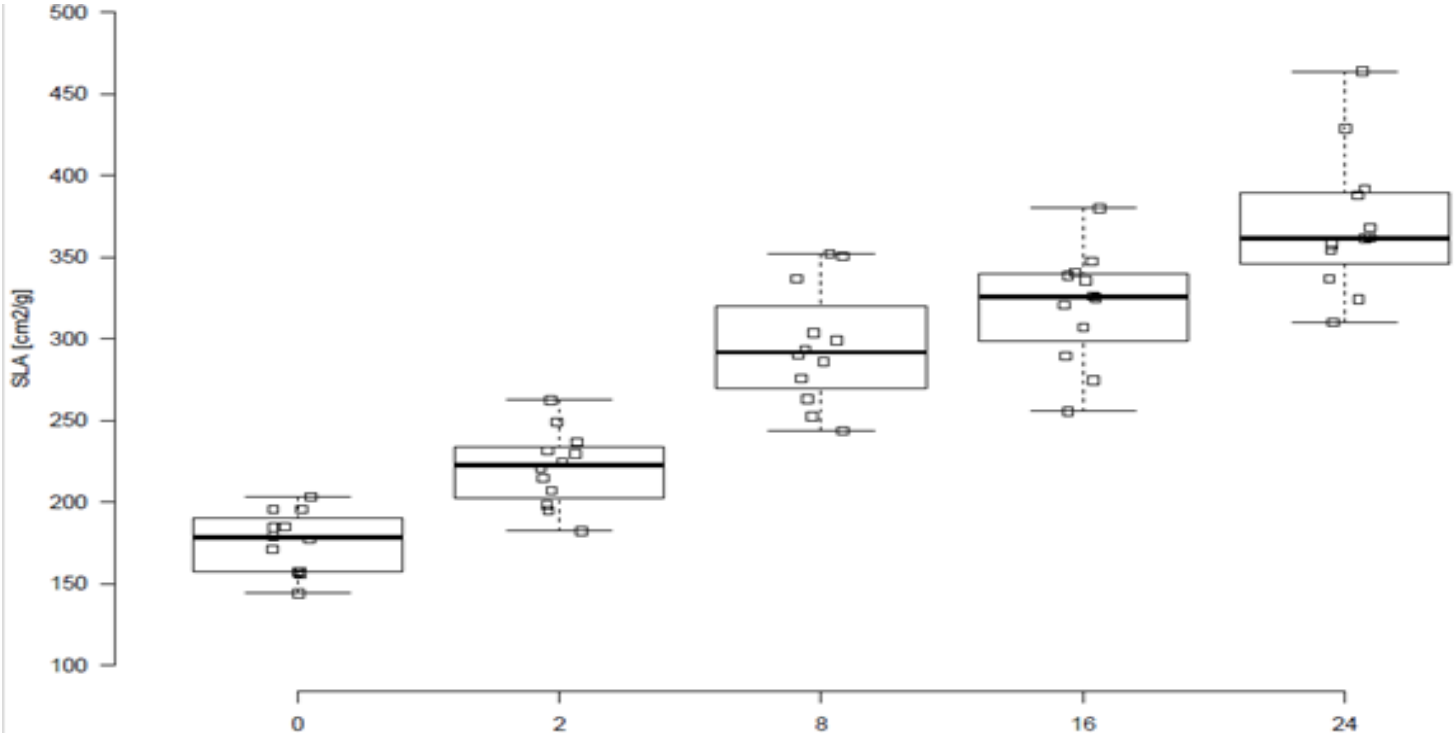

Figure 1.1: Side-by-side boxplots of leaf data.

2. **There are no outliers.**

From the side-by-side boxplots of the leaf data, it can be seen that there is no sign of any clearly worrysome outliers. This means that this assumption is met.

3. **All groups have equal population variance.**

The boxplots also suggests that there is not much variation between the treatments - the boxes do not vary widely in height, suggesting that the interquartile ranges are comparable across treatments.

R output:

0	2	16	8	24
18.66019	23.14583	34.06210	35.80480	43.02643

Table 1.3: The standard deviations of the observations per treatment sorted in increasing order.

Furthermore, in our experiment, the smallest standard deviation (observed for 0mM NO_3^-) was less than three times smaller than the largest standard deviation (observed for 24mM NO_3^-). This ratio is less than about five which means that this assumption is met.

4. **The errors are normally distributed.**

From the side-by-side boxplots of the leaf data, it can be seen that the box plots are quite symmetrical and the data points tend to cluster around the medians. We will revisit this assumption later after fitting the ANOVA model by inspecting the distribution of the residuals. In our data set, there are no signs of any problems and this means that this assumption is met for now.

5. **The errors are independent.**

We have plotted the leaf data in the order in which they were collected (i.e. order in which they appeared in the data set) to examine whether there are any obvious patterns. From the dot plot of the leaf data, it can be seen that there are no clear patterns and this means that this assumption is met.

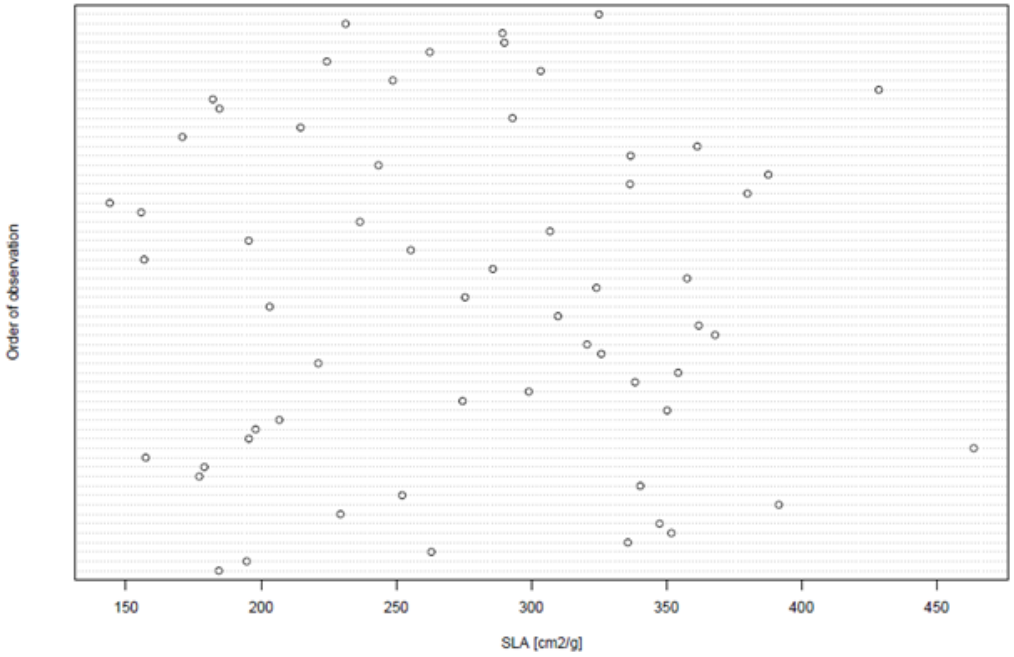

Figure 1.2: Dotplot of the leaf data.

<u>A simple Model for a Completely Randomised Design:</u>

For our first research question, we want to test whether there are differences between the treatment means. The treatment means can only be equal if all α_i are zero.

$$H_0: \alpha_{0mM\ NO3^-} = \alpha_{2mM\ NO3^-} = \alpha_{8mM\ NO3^-} = \alpha_{16mM\ NO3^-} = \alpha_{24mM\ NO3^-} = 0$$

Then we fit the ANOVA model using function aov().

> m1<-aov(SLA..cm2.g. ~ Treatment..mM.NO3.,data=leaf)

> summary(m1)

	DF	Sum Sq	Mean Sq	F-value	Pr(>F)

Treatment..mM.NO3	1	269298	269298	193.7	<2e-16 ***
Residuals	58	80650	1391		

Table 1.3: The ANOVA table for the leaf data.

From the above model output, it can be seen that the F-value is large (193.7) and the P-value is very small (<0.001). We have very strong evidence against H_0 and the answer to the first question is that the treatments did indeed differ from each other, leading to different specific leaf areas.

Checking model fit:

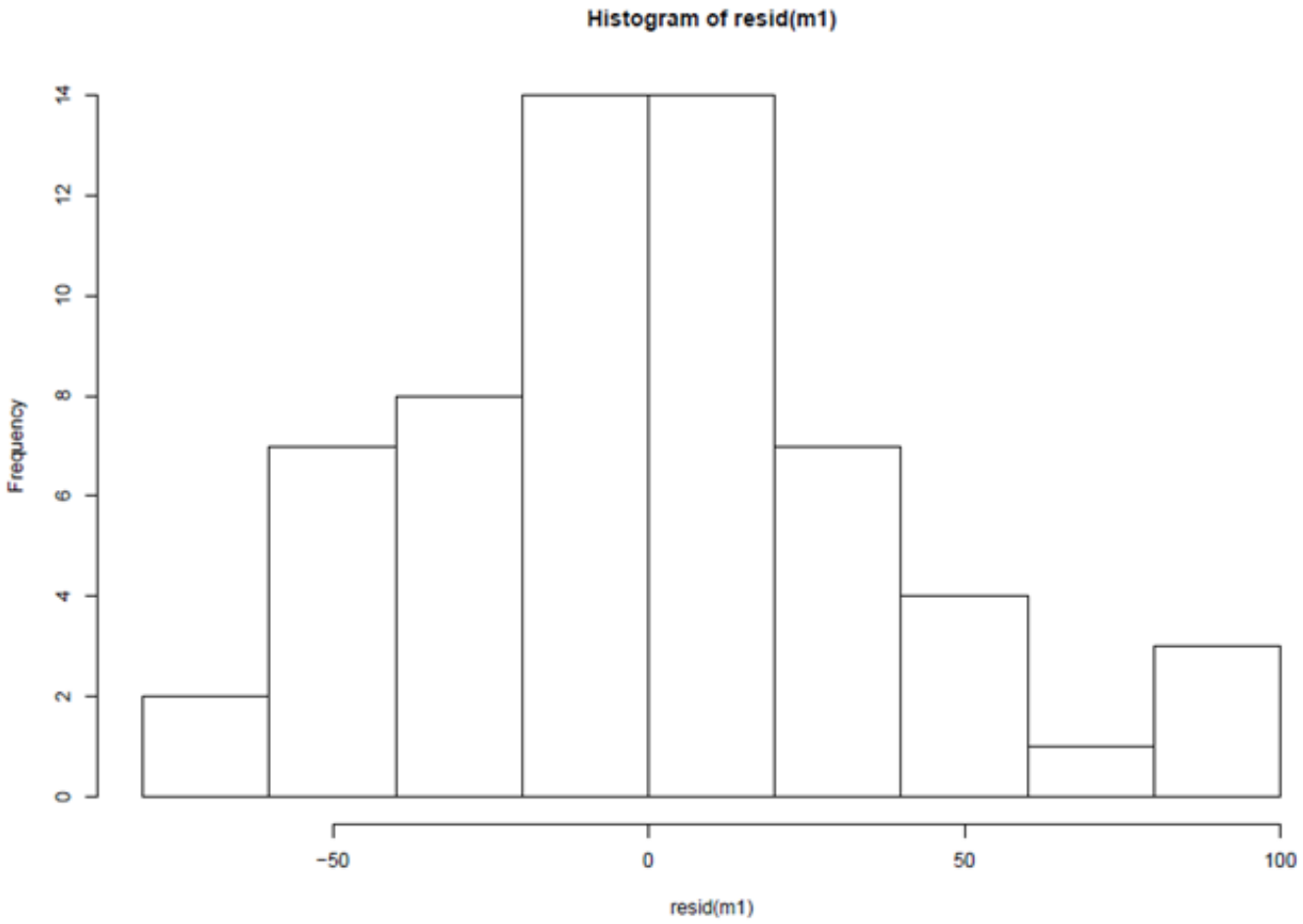

Figure 1.3: A histogram of the model residuals.

From the histogram above, it can be seen that it is quite symmetrical.

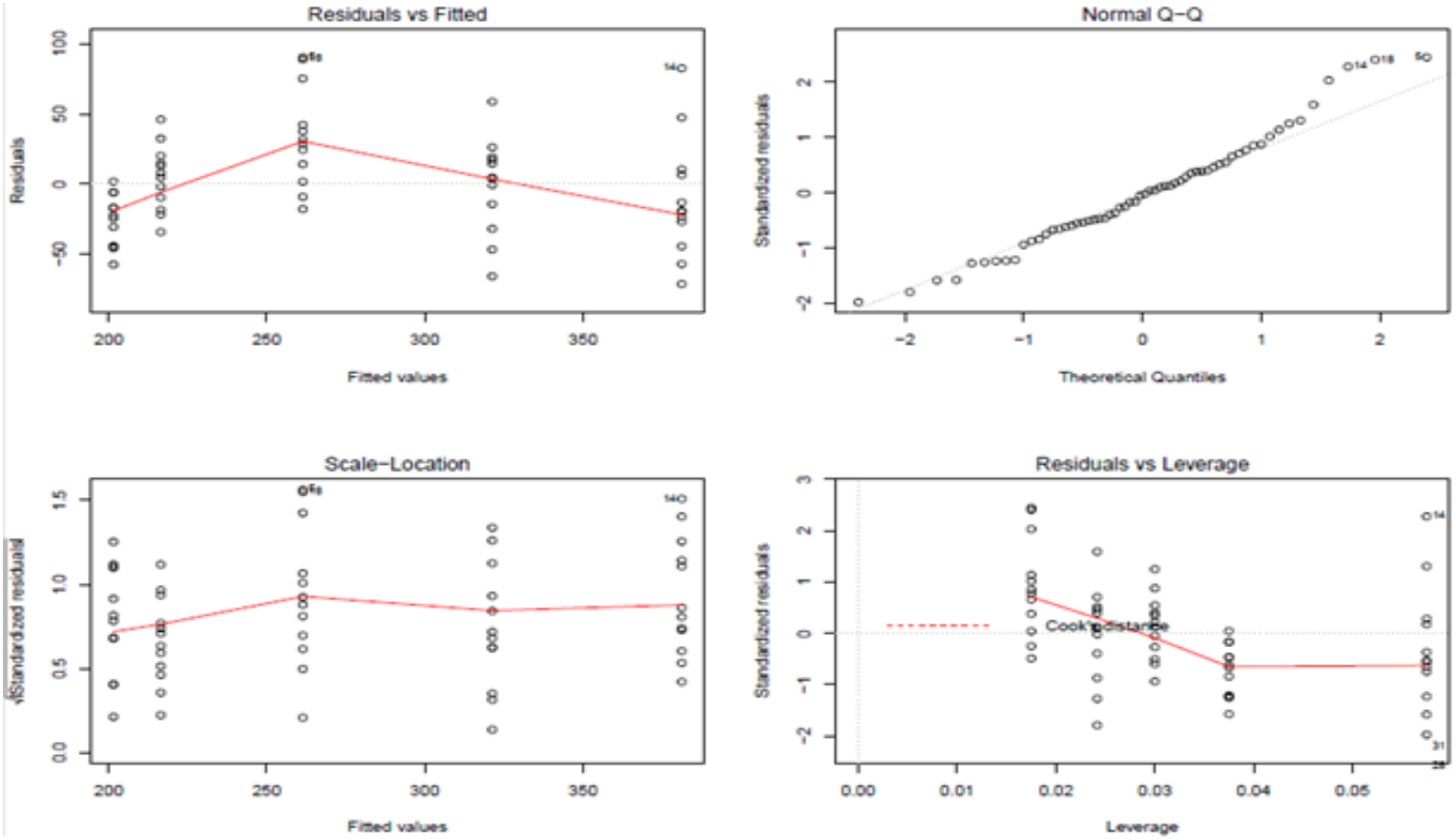

Figure 1.4: Plot of residuals against fitted values

In the plot of residuals against fitted values, there is a bit an indication of a pattern but it is not very strong. In the QQ plot, we see the points falling close to the dotted line and it is normal for the tails to veer off a bit. In the scale-location plot, again we see that there are no clear patterns. Furthermore, in the last plot of residuals against leverage values, shows that there are no clear patterns (i.e. particularly influential observations in the data). Overall, we conclude that the model fits the data reasonably well.

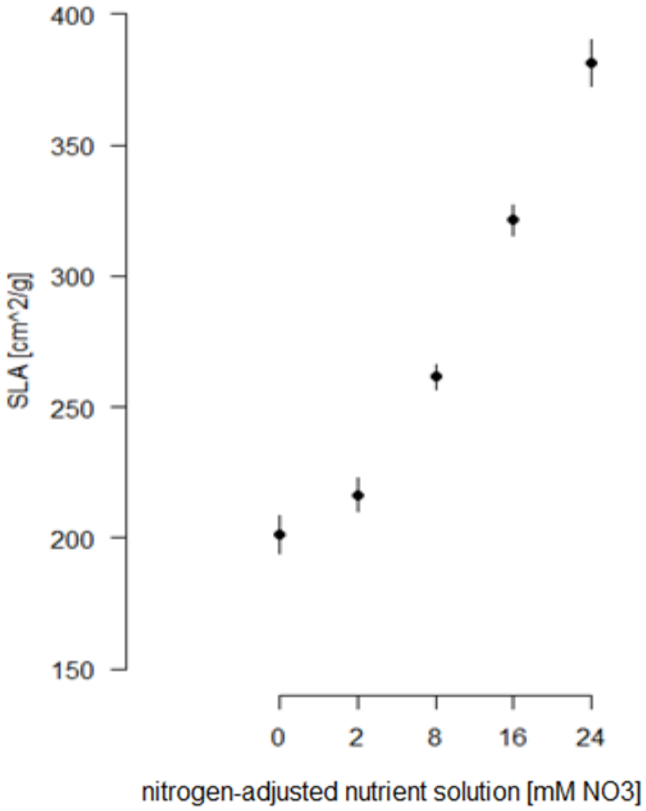

Figure 1.5: Estimated (or predicted) treatment means for the experiment - the error bars show one standard error to either side of the mean.

Contrasts:

We have fitted the ANOVA model with the treatment contrast parameterization to the leaf data.

> m1.tc <- lm(SLA..cm2.g. ~ Treatment..mM.NO3., data=leafNew)

> summary(m1.tc)

Call:

lm(formula = SLA..cm2.g. ~ Treatment..mM.NO3., data = leafNew)

Residuals:

Min	1Q	Median	3Q	Max
-64.575	-18.856	0.489	17.529	93.258

Coefficients:

	Estimate Std	Error		t value	Pr(>|t|)
(Intercept)	175.311	9.289		18.872	< 2e-16 ***

Treatment..mM.NO3.16mM NO3	144.672	13.137		11.013	1.57e-15 ***
Treatment..mM.NO3.24mM NO3	195.177	13.137		14.857	< 2e-16 ***
Treatment..mM.NO3.2mM NO3	45.539	13.137		3.466	0.00103 **
Treatment..mM.NO3.8mM NO3	120.050	13.137		9.138	1.27e-12 ***
Residual standard error			32.18 on 55 degrees of freedom		
Multiple R-squared			0.837		
Adjusted R-squared			0.825		
F-statistic			70.74 on 4 and 55 DF		
p-value			< 2.2e-16		

Table 1.3: Table showing the parameter estimates of the ANOVA model for the leaf data with the treatment contrast parameterization.

Interpretation:

The first parameter, called 'Intercept' (175.311), estimates mean SLA in the first treatment (0mM NO3 – R uses this treatment as the baseline treatment as the treatments are sorted alphabetically).

The next parameter (144.672) estimates the difference between the mean in treatment 16mM NO3 and treatment 0mM NO3.

The next parameter (195.177) estimates the difference between the mean in treatment 24mM NO3 and treatment 0mM NO3.

The next parameter (45.539) estimates the difference between the mean in treatment 2mM NO3 and treatment 0mM NO3.

The next parameter (120.050) estimates the difference between the mean in treatment 8mM NO3 and treatment 0mM NO3.

The test that R prints for each coefficient tests the null hypothesis that the parameter value is equal to zero. In other words:

H_0: the difference between 16mM NO3 and 0mM NO3 is zero (and found that $p<0.002$)

H_0: the difference between 24mM NO3 and 0mM NO3 is zero (and found that $p<0.002$)

H_0: the difference between 2mM NO3 and 0mM NO3 is zero (and found that $p<0.002$)

H_0: the difference between 8mM NO3 and 0mM NO3 is zero (and found that $p<0.002$)

From the R output it can be seen that in all cases, the p-values are very small (<0.002) and this indicates very strong evidence against the null hypothesis for each of these cases.

For the intercept:

H_0: the mean of the baseline treatment (0mM NO3) is zero (and found that $p<0.002$)

From the R output it can be seen that the p-value is very small (<0.002) and this indicates very strong evidence against the null hypothesis. In our experiment, this is certainly not an informative and useful test because if there was no SLA in these treatments, this would have been obvious to the students taking the SLA measurements.

So we have strong evidence that there were real differences between these treatments to the control and this answers our second research question: Do these treatments differ from the control in the resulting specific leaf area examined?

Multiple Comparison:

Tukey's Method:

Tukey's method is used to correct the overall type I error when making all pairwise comparisons between a treatment means. In R, we use the function TukeyHSD() to calculate the contrasts and their confidence intervals.

	diff	**lwr**	**upr**	**p-adj**
2mM NO3-0mM NO3	45.539	8.489	82.590	0.009
8mM NO3-0mM NO3	120.050	83.000	157.101	0.000
16mM NO3-0mM NO3	144.672	107.621	181.722	0.000
24mM NO3-0mM NO3	195.178	158.126	232.228	0.000
8mM NO3-2mM NO3	74.511	37.460	111.561	0.0000053
16mM NO3-2mM NO3	99.133	62.082	136.183	0.000
24mM NO3-2mM NO3	149.638	112.587	186.688	0.000
16mM NO3-8mM NO3	24.622	-12.429	61.672	0.343
24mM NO3-8mM NO3	75.127	38.076	112.178	0.0000044
24mM NO3-16mM NO3	50.505	13.455	87.556	0.0028104

Table 1.3: Table of R output showing the results of the contrasts involving all pair-wise comparisons using Tukey's HSD.

Interpretation:

The test that R prints for each coefficient tests the null hypothesis that the parameter value is equal to zero. We are interested in interpreting the last 6 lines here as we want to answer the final research question: Do these treatments differ from each other in the resulting specific leaf area examined?

H_0: the difference between 8mM NO3 and 2mM NO3 is zero (and found that p<0.003)

H_0: the difference between 16mM NO3 and 2mM NO3 is zero (and found that p<0.003)

H_0: the difference between 24mM NO3 and 2mM NO3 is zero (and found that p<0.003)

H_0: the difference between 16mM NO3 and 8mM NO3 is zero (and found that p>0.005)

H_0: the difference between 24mM NO3 and 8mM NO3 is zero (and found that p<0.003)

H_0: the difference between 24mM NO3 and 16mM NO3 is zero (and found that p<0.003)

From the R output it can be seen that in all cases except the third last case, the p-values are very small (<0.003) and this indicates very strong evidence against the null hypothesis for each of these cases. In the third last case, the p-value is large (`0.3431643 > 0.005`) which indicates that there is not much evidence the treatment means are not equal. We would conclude that all treatments differ from each other in the resulting specific leaf area examined except for the 16mM NO3 and 8mM NO3 treatments which may result in the same specific leaf area.

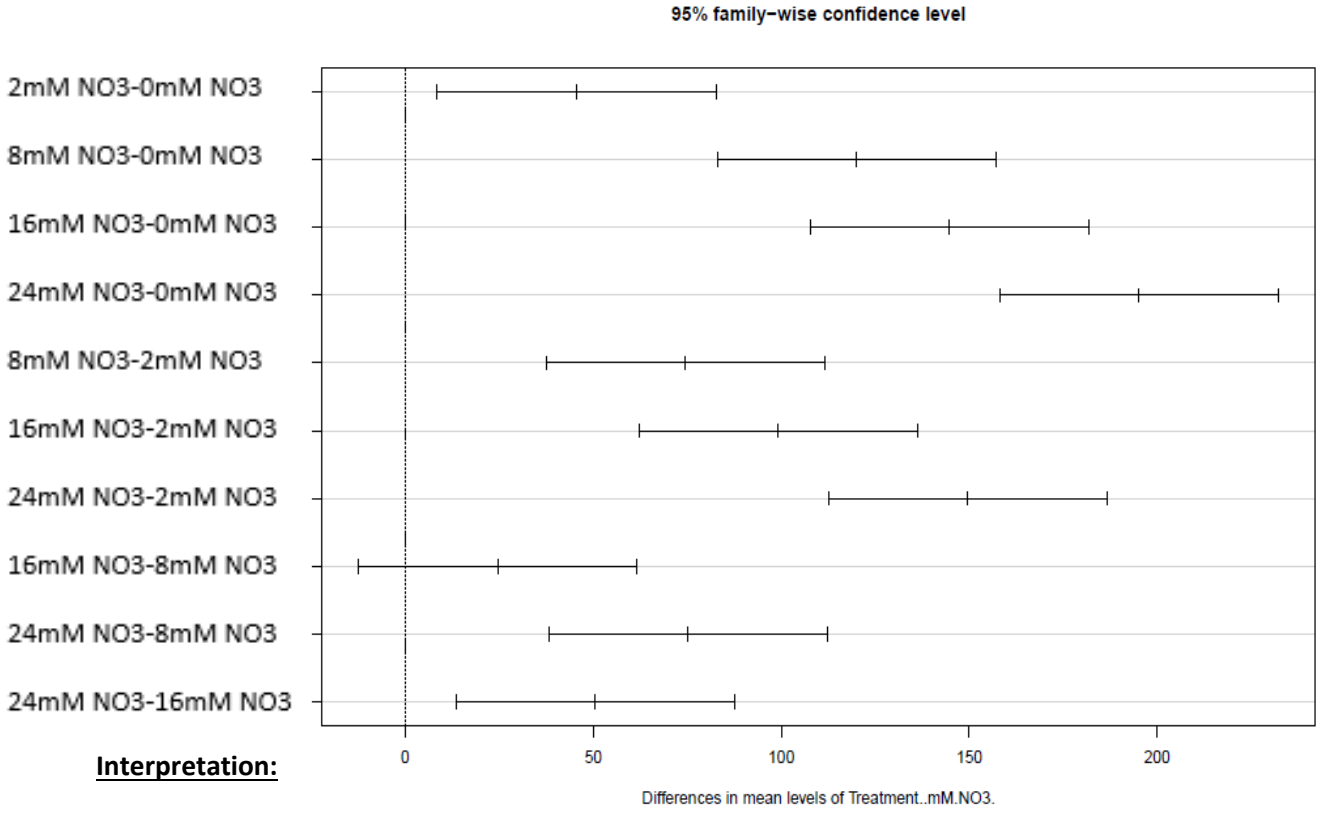

Figure 1.6: The figure shows the output of Tukey's procedure applied to the leaf data.

Interpretation:

In the figure above, each row shows one pair-wise contrast with adjusted 95% confidence interval. Those confidence intervals that don't overlap zero have an adjusted P-value < 0.05 and constitutes strong evidence against the null hypothesis, as discussed in the previous interpretation for the table of R output showing the results of the contrasts involving all pair-wise comparisons using Tukey's HSD.

Pilot experiment:

Power analysis:

This is our best estimate of the average within-treatment standard deviation for our experiment:

>[1] 32.17892

Hence, we will use a standard deviation of 32.

Firstly, we would like to know how many replicates and experimental units we require if we want to be reasonably confident that our experiment will reveal an effect that is biologically meaningful.

Two-sample t test power calculation	
n (number in each group)	26.714
Delta	25
Sd	32
Sig.level	0.05
Power	0.8
alternative	Two.sided

Table 1.4: Table Showing the output for the power calculation

So the answer is that we need 27 replicates, i.e. 54 experimental units (pots) in total for our experiment, if we want to be reasonably sure (have 80% power) to detect a difference of 25% SLA when the standard deviation within treatment is 32. Therefore, since we have already established that we have sufficient resources to run the experiment with 60 experimental units (pots) in total, we expect to be fine.

Calculating the expected power with a 60 experimental unit set-up:

> power.t.test(n=60, delta=25, sd=32, sig.level=0.05)

Two-sample t test power calculation	
n (number in each group**)**	60
Delta	25
Sd	32
Sig.level	0.05
Power	0.989
alternative	two.sided

The answer is that the power of our test would only be 98.88% in this situation and this confirms that we can be confident about the power of our experiment to detect a biologically meaningful effect.

References:

The work presented is not our own, we have borrowed from many sources:

- PoorterH, Remkes C, Lambers H.1990. Carbon and nitrogen economy of 24 wild species differing in relative growth rate. Plant Physiology 94: 621–627.
- Reich PB, MB Waters and DS Ellsworth (1997) From tropics to tundra: global convergence in plant functioning. Proceedings of the National Academy of Sciences USA 94: 13730-13734
- Hoffmann, W.A.; Franco, A.C.; Moreira, M.Z.; Haridasan, M. (2005). "Specific leaf area explains differences in leaf traits between congeneric savanna and forest trees". Functional Ecology. 19: 932–940