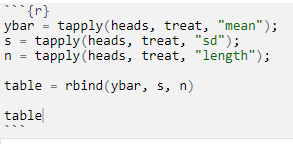
ST515-HW2b

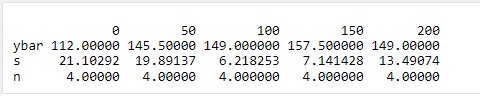
Nora Quick

2.4:

1. Construct a table of the mean number of heads of lettuce (y), the standard deviation s (square root of the sample variance), and the number of plots (n) in the five nitrogen treatments. Show your R code.



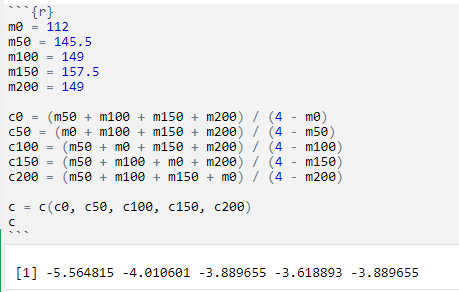
2.4.1



2.4.2

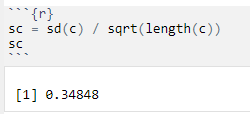
1. Consider the contrast C = ($$50 + $$100 + $$150 + $$0, where the subscripts refer to the nitrogen amounts. Show your arithmetic in answering the following questions.

(i) Estimate C using the sample means from your table. Call the estimate c.



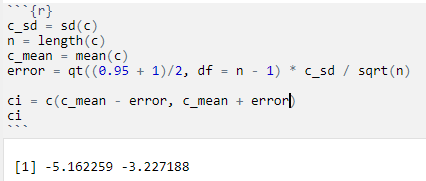
2.4.3

(ii) Estimate the standard error (square root of the estimated variance) of c. Call it sc. You will need the MSE from the single-factor ANOVA model in your calculations.



2.4.4

(iii) Use the above results to construct a 95% confidence interval for C.

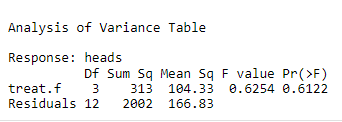


2.4.5

(iv) Interpret the confidence interval. What does it tell you about the effect of the nitrogen treatments on lettuce production?

From the confidence interval we can conclude that the more nitrogen that was added to the lettuce the more heads were produced. The less nitrogen added the less heads are produced on average somewhere between 3-5 heads less.

1. Do an analysis of variance of heads by treatment, excluding the zero-nitrogen plots, and interpret the results of the F-test.

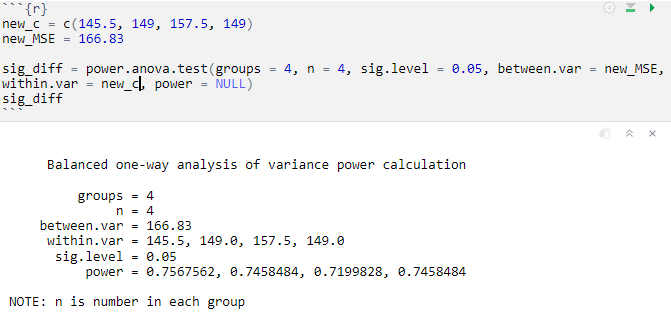


2.4.6

Based on the large f-value we can see that there is high variation between the sample means even when the zero-nitrogen plots are left out. In other words the means of each of the non-zero nitrogen plots vary in a significant way.

1. Suppose you are planning a future study of differences in lettuce production among the 50-, 100-, 150- and 200-unit nitrogen treatments. Assume that the true mean numbers of heads will be identical to those observed in the current study (145.5, 149.0, 157.5, and 149.0), and the mean square error will be 166.83. Use the R function power.anova.test to answer the following two questions. Include your code in the answers.

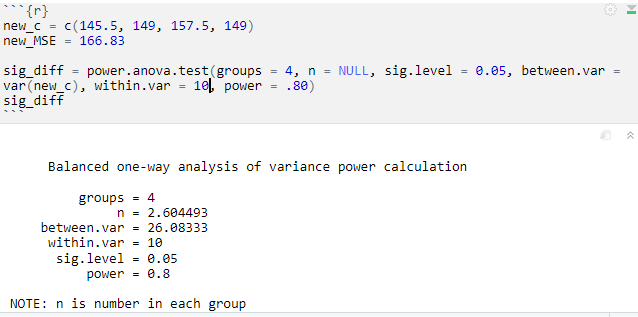
(i) How much power would this design (using 4 plots per treatment) have to find statistically significant (at the 0.05 level) difference among means?



2.4.7

From the test we can see we need about 74.5% power to find statistically siginficant differences among the means.

(ii) How many plots per treatment would we need to have 80% power to find a statistically significant difference among means?

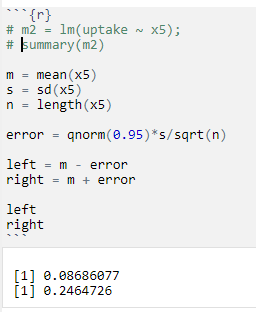


2.4.8

From the test we can see that we need about 3 (2.6 rounded up) plots per treatment to have 80% power to find a statistically significant difference among means.

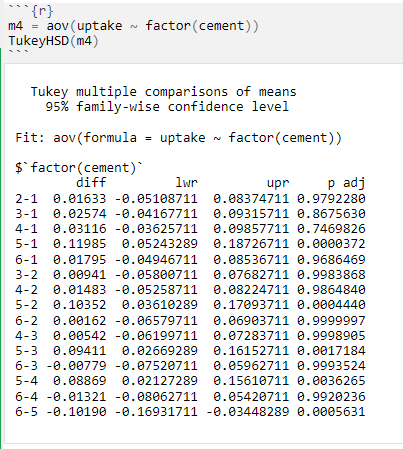
2.5:

1. Fit the regression mode y=$$0h+ $$1x1 + $$2x2 + $$3x3 + $$4x4 + $$5x5 + ei is fluoride uptake; xi = 1 for cement type i and ke; xi = 0 otherwise, for i = 1,…,5 (i.e., type 6 is the reference group); and e is random error. Use one of the estimated ’s (you have to figure out which one) and its standard error to construct the confidence interval. Interpret your result.



2.4.9

1. Use the Tukey method, as shown below. The lwr and upr columns contain the confidence limits.



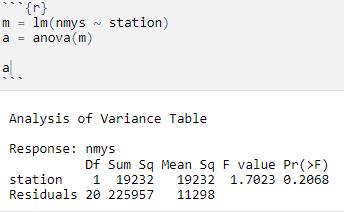
2.4.10

1. You should have found that the interval from (a) is narrower than the interval from (b). Why is that?

It is much smaller for a few reasons. The interval from (a) is smaller because we are looking at a specific cement type. In addition to that as we can see from the Tukey method that cement 5 (the one we tested in part a) is significant. If it is different from all the others than it will have a range that is smaller when not compared to the others.

2.6:

1. Do an ANOVA of nmys by station, assuming all assumptions are met. State the hypothesis being tested, show the ANOVA table, and interpret the results of the 𝐹-test.

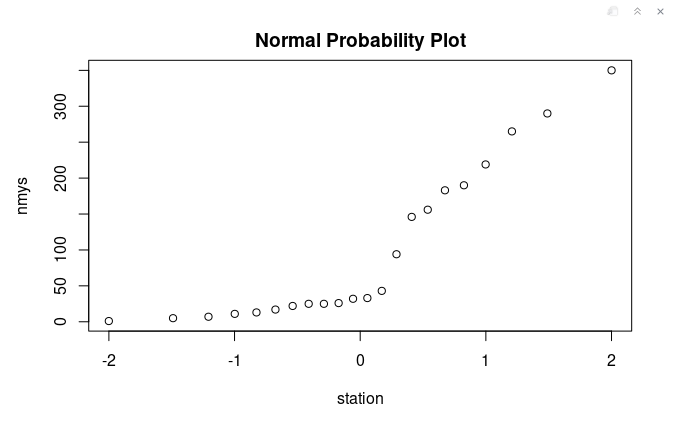


2.4.11

The hypothesis being tested is if the average number of mysids in each haul are the same for all stations.

From the anova table we can see that there is a high f-value. This variance between the averages of each group haul are larger than we would expect by chanve. In other words the mean of each station mysids haul is not the same between stations.

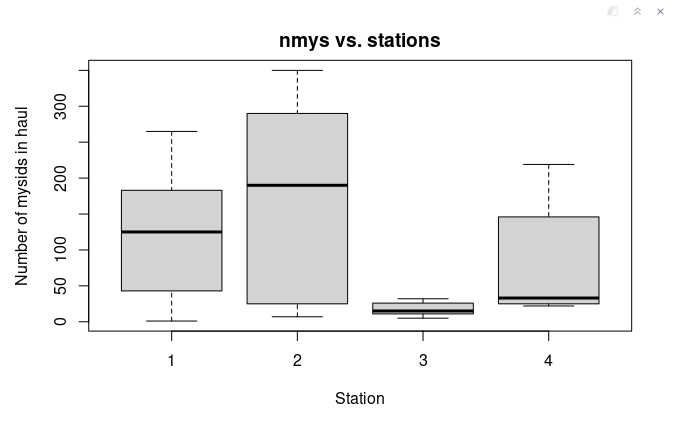
1. Using the R function qqnorm, create a normal probability plot that helps you to judge how well these data follow the normality assumption made in the analysis of variance. Comment on the graph. You don’t need to do any formal tests here.



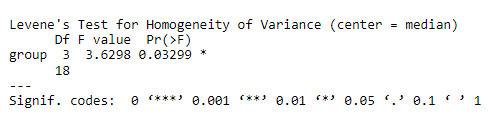
2.4.12

The graph spikes a little both on the whole the graph shows a relatively linear relationship.

1. Check the assumption of constant variance by (i) doing a multiple boxplot of nmys by station, (ii) examining a plot of residuals vs. predicted numbers of mysids, and (iii) applying Levene’s test (load the code in levene2.R into R). Interpret your results.



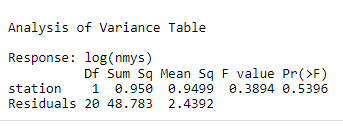
2.4.13



2.4.14

Again, we have a large f-value indicating that the variance between the groups is not the same. In other words each group has a different variance.

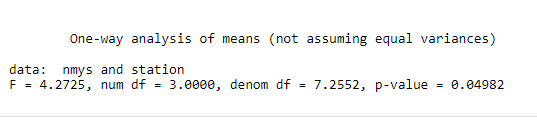
1. Re-do the ANOVA using log(nmys) as the response, and check the constant-variance assumption using the same three tools asked for in part (c). Which response—nmys or log(nmys)—is better to use in the ANOVA, in your opinion? Why?



2.4.15

In my opinion the log(nmys) is better for checking the constant-variance assumptions. I believe this because the f-value is much smaller and more reasonable for the anova test. While it is still a significant f-value it shows a better assumption of the varriance.

1. Finally, use Welch’s test (oneway.test in R) to test the association between (untransformed) numberof mysids and station. Interpret the result.



2.4.16

The Welch’s test checks if the means of populations are equal. From the large f-value we can see that there is a significant difference in the means of the populations are not equal.