ST515-HW2

Nora Quick

2.1:

(a) Do a graphical comparison of the numbers of embryos in the two cohorts of females included in this data set, e.g., a boxplot, or a pair of histograms drawn to the same scale. Comment on the graph.

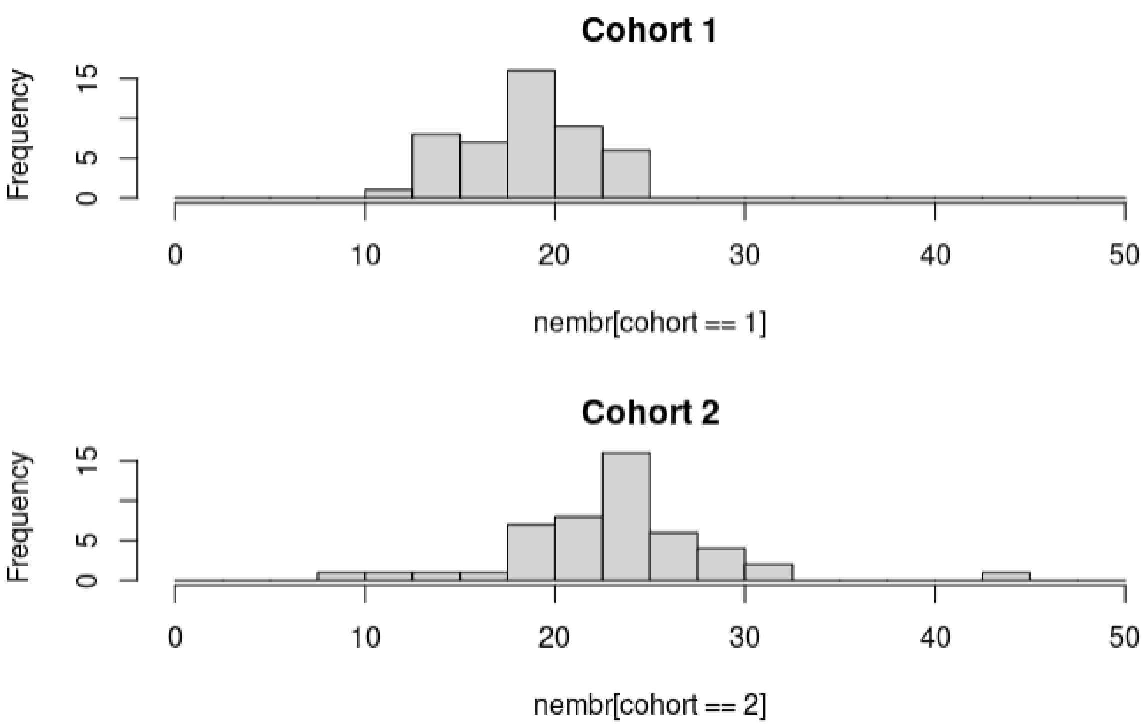

Figure 1: 2.1.1

Based on our quesion of whether there is a difference between years in the average number of embryos carried by gravid mysids the graphs show a lot of information. Based on the means of the graphs we can see that year two (cohort 2) carried more embryos than the first year (cohort 1).

(b) Do an equal-variances t-test comparing the numbers of embryos in cohort 1 vs. cohort 2. Show the output, and interpret your results in a nontechnical way.

Based on the small p-value for the two sample t-test there is strong evidence that the means of the two populations are different. This means that there is strong evidence that between the two years there was a difference in the mean number of embroys carried by gravid mysids. Additionally, we can see that in the second year there is a larger mean than the previous year.

```
        Two Sample t-test

data:  nembr[cohort == 1] and nembr[cohort == 2]
t = -4.6899, df = 93, p-value = 9.353e-06
alternative hypothesis: true difference in means is not equal to 0
95 percent confidence interval:
 -6.319571 -2.559861
sample estimates:
mean of x mean of y
 18.89362  23.33333
```

Figure 2: t-test comparing

(c) Do an analysis of variance on these data, i.e., test the hypothesis that the mean number of embryos doesn't depend on cohort. Show the output, and interpret your results.

```
Analysis of Variance Table

Response: nembr
          Df  Sum Sq Mean Sq F value    Pr(>F)
cohort     1  468.09  468.09  21.995 9.353e-06 ***
Residuals 93 1979.13   21.28
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Figure 3: aov

Again, we have a small p-value indicating that the number of embryos does depend on the cohort. Because we were trying to prove that it doesn't matter the p-value shows that we cannot conclude that so we must conclude the opposite.

(d) How does the error sum of squares in the ANOVA table relate to the pooled sample variance? How does the F-statistic from the ANOVA relate to the t-statistic from the t-test?

Both the t-value and the f-value show values testing whether the hypothesis is true and if the two standard deviations are similar respectively. Therefore, we can conclude that the relationship between the two is they are both testing if the two cohorts are giving similar data/answers to which we can conclude that they are not.

(e) The variable called size (body size of the mysid) may confound your tests of the association between embryo numbers and cohort. Can you find any evidence, graphical or otherwise, that body size is a confounding factor? Explain.

Based on both this histogram and an ANOVA test done we can see that there isn't much different in the size between the two years. Therefore, based on this graph alone we cannot find that size is a confounding factor. If we test the embryos we get largely different numbers so if we wanted to conclude that size was a confounding factor we would want to see at least a little more variance between the two cohorts.

2.2:

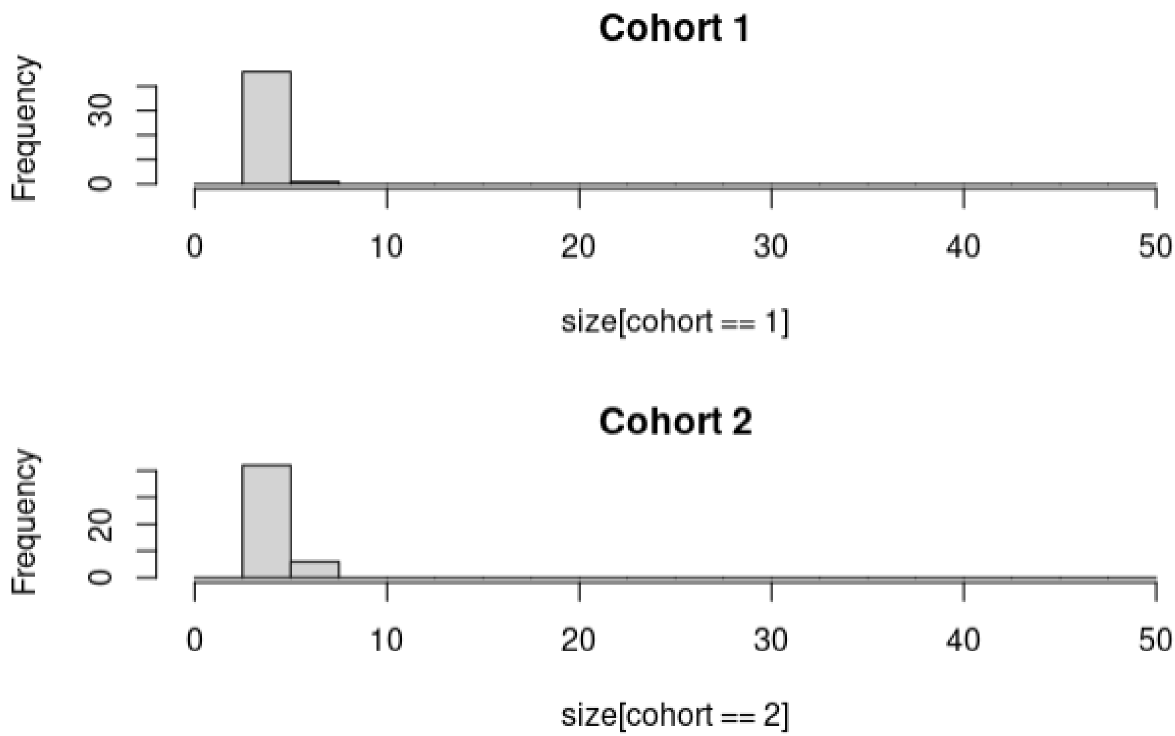

Figure 4: size

(a) According to the design described above, use R to generate one instance of randomization of the five nitrogen treatments to the 20 plots. Use a 4×5 table to show your randomization result. (Show your R code.)

```
```{r}
c1 <- c(sample(lettuce$treat, size = 5, replace = TRUE))
c2 <- sample(lettuce$treat, size = 5, replace = TRUE)
c3 <- sample(lettuce$treat, size = 5, replace = TRUE)
c4 <- sample(lettuce$treat, size = 5, replace = TRUE)

c1
c2
c3
c4
```
```

Figure 5: sample

See next page.

(b) Draw a multiple boxplot of the numbers of heads of lettuce in the five groups.

See next page.

(c) Perform an analysis of variance on heads of lettuce as a function of nitrogen treatment. (For now, don't worry about checking assumptions or considering transformations of the response variables.)

```
 50 150 150 200 150
200 200  50 100  50
  0   0 100 150   0
 50  50 200 150 100
```

Figure 6: table

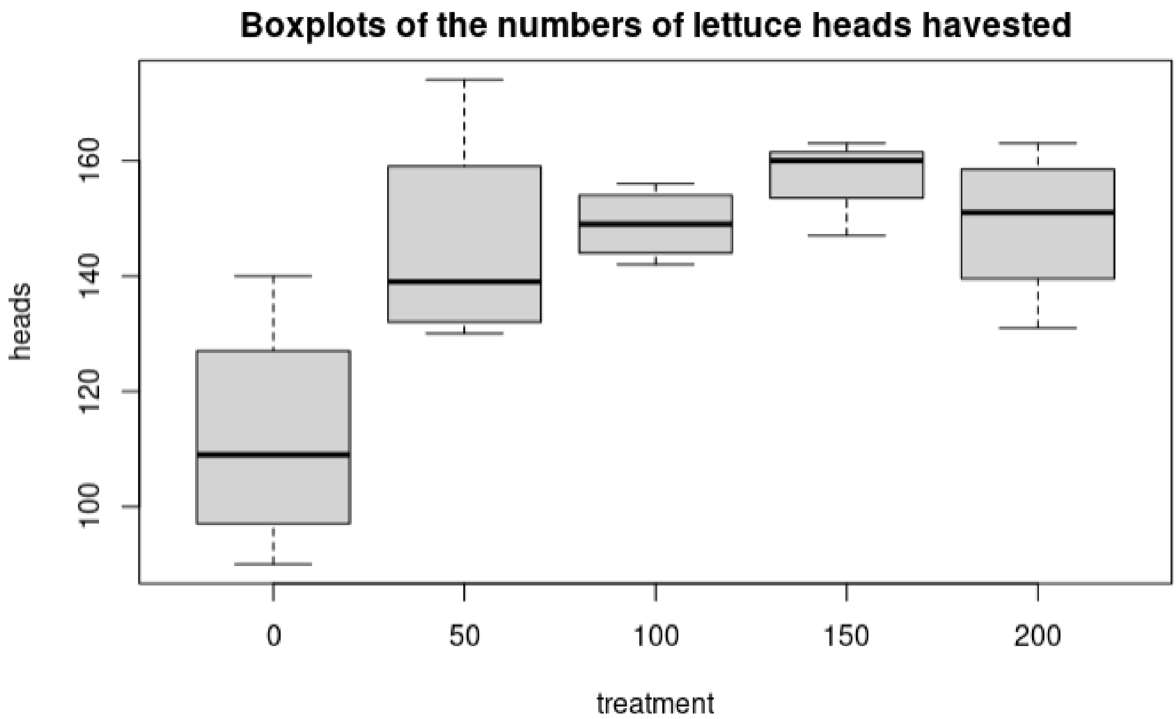

Figure 7: box

```
Analysis of Variance Table

Response: heads
          Df Sum Sq Mean Sq F value   Pr(>F)
treat.f    4 4994.8 1248.70  5.6113 0.005757 **
Residuals 15 3338.0  222.53
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

        One-way analysis of means (not assuming equal variances)

data:  heads and treat.f
F = 3.456, num df = 4.0000, denom df = 7.1589, p-value = 0.07152
```

Figure 8: aov2

(d) Provide nontechnical interpretations of the boxplot and the results of the ANOVA.

Based on both the boxplot and the ANOVA test we can see that the different levels of nitrogen fertilizer affected the mean number of heads of lettuce grown in a field. Simply looking at the boxplot we can see that there is a significant difference in no nitrogen fertilizer to some. Additionally, we can see that different amounts vary the number of heads grown. Looking at the p-value of the ANOVA test we can also see that there is not enough evidence to say that the same number of heads in grown. In other words there is a difference in the number of heads grown between the different batches.

2.3:

(a) Describe in detail how you would design and execute your experiment. Organizing your answer according to the elements of experimental design as discussed in the lecture.

For this experiment we need to observe if there is any difference in the levels of CO_2 for plant growth. We need to ask questions such as how much CO_2 will effect the plant growth? Therefore, the null hypothesis should be that they will all have the same levels of grown despite the levels of CO_2 and the alternate hypothesis is that they will not have the same levels of growth.

I would run both a two sample t-test as well as an ANOVA test to test the difference in means for the growth of the plants. To gather the data we need I would start with 12 randomly selected pots in each of the chambers. Starting at chamber 1 I would have limited/no CO_2 added to the plants. Then in increasing increments I would add CO_2 levels to each of the chambers with chamber 6 having the most. Each week I would test the weight of the plants in each chamebr and record them.

(b) Describe the statistical analysis that will be applied to the experimental results.

I would do both a two-sample t-test to create plots to determine the difference in means. I would then do an ANOVA test to check the variance of the plant growth and make another set of graphs based on that.