ST517-HW2

Nora Quick

1. The dataset `bearings.csv` contains failure times (measured in millions of cycles) of engine bearings made from five different compounds..

   (a) Read in the data. What type of data object is bearings? With ggplot2, create side-by-side boxplots of the failure times by compound.

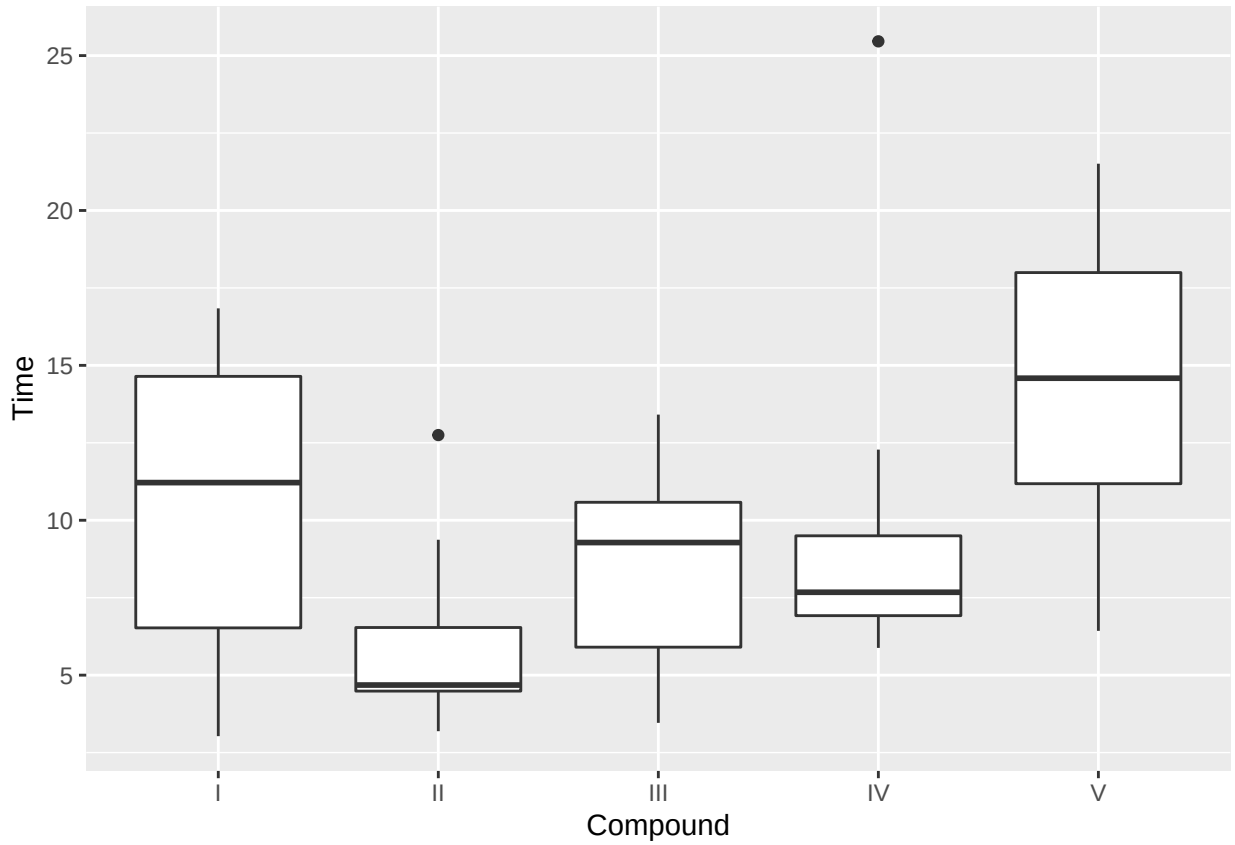
```
str(bearings)
```

```
## 'data.frame':    50 obs. of  2 variables:
##  $ Time    : num  3.03 5.53 5.6 9.3 9.92 ...
##  $ Compound: chr  "I" "I" "I" "I" ...
```

```
Compound <- bearings$Compound
Time <- bearings$Time
```

The two data objects are Time which consists of numbers and Compound which consists of characters.

```
ggplot(bearings, aes(x=Compound, y=Time)) + geom_boxplot()
```



(b) Determine the pairs of engine ball bearing compounds for which there is a significant difference in mean failure times. Present your findings in a short statistical report (abt. 4 sentences).

```
fit <- aov(Time ~ Compound, data = bearings)
summary(fit)
```

```
##             Df Sum Sq Mean Sq F value  Pr(>F)
## Compound     4  401.3  100.32    5.02 0.00197 **
## Residuals   45  899.2   19.98
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
TukeyHSD(fit)
```

```
##   Tukey multiple comparisons of means
##     95% family-wise confidence level
##
## Fit: aov(formula = Time ~ Compound, data = bearings)
##
## $Compound
##              diff         lwr       upr     p adj
## II-I   -4.6430001 -10.3234907  1.037490 0.1567083
## III-I  -2.0570003  -7.7374909  3.623490 0.8406451
## IV-I   -0.8950002  -6.5754908  4.785490 0.9914058
```

```
## V-I     4.0129998  -1.6674907  9.693490 0.2791467
## III-II  2.5859998  -3.0944908  8.266490 0.6964484
## IV-II   3.7479999  -1.9324907  9.428490 0.3453895
## V-II    8.6559999   2.9755094 14.336491 0.0007521
## IV-III  1.1620001  -4.5184905  6.842491 0.9771836
## V-III   6.0700001   0.3895095 11.750491 0.0308937
## V-IV    4.9080000  -0.7724905 10.588491 0.1195580
```

With a 95% family-wise confidence level and an Analysis of Variance test between the failure times of the ball bearings a small p-value is produced (0.00197) meaning there is at least one ball bearing failure time that is different than the others. Following the p-value rules with the adjusted p-values we can see that there are a two significant tests with small p-values. There are V-II and V-III with p-values of 0.00075 and 0.031 respectively, both of which are smaller than the significance level of 0.05. This determines that they both indicate that one of the variables within these tests is significantly different and as V is the common variable we can conclude that compound V has a signigicant difference in mean failure times.

2. Using the data from the last weeks lab and homework, (`case0501` in the `Sleuth3`), answer the question "Which diets differ in their mean lifetime?"

Looking at the data from last weeks hw and lab it appears that Diet C had the most difference in it's mean outcome from the other two based of df and the p-values outputted by the different tests.

3. A soda company is developing a new soda. They are tying to determine how much sugar to put in it to give it the best taste. In order to evaluate this, they have made samples with ten different sugar levels. Each level of sugar is a assigned to a random sample of seven people, and each person rates the soda on a scale from 0 to 10. The company would like to make inference on the difference between mean ratings between each pair of sugar level.

   (a) How many pairwise comparisons are there?

We are compairing each of the sugars to the rest of them, therefore, 10(10-1)/2 = 45 pairwise comparisons.

(b) Name 2 procedures that the company could use to control the familywise Type I error rate on the differences of means? Explain.

I. My first method would be to do the testing and then adjust the p-value to reduce/control the familywise Type I error rate. I would do this because it adjusts and reduces the rate of a Type I error rate.

II. My second method would be to adjust the test itself. I would increase the number of people testing in order to get a wider amount of opinion. For a company testing so many sugars they should want a wider sample size to find the most accurate means to work with.

4. A consumer research group names seven types of department stores. They take a random sample of six department stores for each type and record their yearly sales. They wish to find significant pairwise differences in mean yearly sales for seven types of stores.

   (a) How many pairwise comparisons are there?

We are comparing the mean of each store to the rest of the stores, therefore, 7(7-1)/2 = 21 pairwise comparisons.

(b) The group wishes to control the familywise Type I error rate at 1% using Bonferroni methods. What should be the Type I error rate of each pairwise comparison?

We want to make it so that the alpha devided by pariwise comparisons equals 0.01. Therefore, we would want to adjust the alpha level using Bonferroni's method to 0.21 to produce a familywise Type 1 error rate of 1%.