This document describes how to use R to carry out parts of Exercises 11.31–35 on pages 497–498 in the textbook. Carry out all of the steps, but include in your homework write-up only answers to questions

1. Download the data set in `11-31.txt` to a textfile.

2. Read this data into R. Attach the data set so that individual variables may be referred to by name. Under Windows, you may need to specify a complete path name for the data set, something like `"C:/My Documents/11-31.txt"` if you put the downloaded text file in that location.

```
> beans <- read.table("11-31.txt", header = T)
> attach(beans)
```

3. Make side-by-side boxplots by group (which is the result of a two-way classification).

```
> boxplot(split(yield, group))
```

Question 1. Do the groups have similar centers? Do the groups have similar amounts of variability?

4. Calculate sample sizes, sample means and standard deviations for each group. (The function `split` partitions the first variable by the categorical levels of the second variable and stores the results in a list. The function `lapply` applies a function to each element of a list.)

```
> lapply(split(yield, group), length)
> lapply(split(yield, group), mean)
> lapply(split(yield, group), sd)
```

Question 2. Record these values in a table.

5. Carry out a one-way ANOVA with `yield` as the response variable and `group` as the explanatory variable. Show the ANOVA table. (The command `lm` fits a linear model of which ANOVA is an example.)

```
> fit1 <- lm(yield ~ group, data = beans)
> anova(fit1)
```

Question 3. Summarize the results of this test in the context of the problem. Is the test significant at the $\alpha = 0.05$ level?

6. Make a plot of the residuals versus the fitted values.

```
> plot(fit1$fitted, fit1$resid, xlab = "Fitted Values", ylab = "Residuals")
> abline(h = 0)
```

Question 4. Refer to the boxplots made previously. Does this plot indicate that the assumption of normality might be suspect? Does this plot indicate that the assumption of equal variances might be suspect?

Question 5. Refer to the residual plot. Does this plot indicate that variability is related to the mean value?

7. Make a normal probability plot of the residuals.

```
> qqnorm(fit1$resid)
```

Question 6. Do the residuals look normally distributed?

8. Transformations of the response variable often fit the assumption better than data in the original scale. Two common transformations that help when the variance seems to be a function of the mean with larger spread in populations with larger means are logarithms and square roots. (Exercise 11.35 refers to a reciprocal transformation which can make sense in some settings.) Carry out a one-way ANOVA with log `yield` as the response variable.

```
> fit2 <- lm(log(yield) ~ group, data = beans)
> anova(fit2)
```

9. Make a plot of the residuals versus fitted values, side-by-side boxplots of log `yield`, and normal probability plots of the residuals.

```
> plot(fit2$fitted, fit2$resid, xlab = "Fitted Values", ylab = "Residuals")
> abline(h = 0)
```

```
> boxplot(split(log(yield), group))
```

```
> qqnorm(fit2$resid)
```

Question 7. Do these plots indicate that the variablity within each sample are more equal for the transformed data?