MTH 3240 Lab 8

Due Thu., Apr. 16

1 Part A: One-Factor ANOVA

1.1 Flowers Data Set

Different varieties of the tropical flower *Heliconia* are fertilized by different species of hummingbirds. Over time, the lengths of the flowers and the form of the hummingbirds' beaks have evolved to match each other.

The file **flowers.txt** contains data on the **lengths** (in millimeters) of three **varieties** (*H. bihai*, *H. caribaea red*, and *H. caribaea yellow*) of these flowers on the island of Dominica.

We're interested in determining if there are **any significant differences** among the mean flower lengths for the three **species**.

1. Save the **flowers.txt** data file onto your computer.

   The `read.table()` reads data into R from a text (.txt) data file. Among the arguments to `read.table()` are:

`file`	the name (and folder) of a text (.txt) file from which the data are to be read.
`header`	a logical (TRUE or FALSE) value indicating whether the file contains headers (variable names).

   The function `file.choose()` can be use to select the file in a dialog box:

   ```
   my.file <- file.choose()          # Select the .txt file in the dialog box.
   ```

   After selecting **flowers.txt** using `file.choose()`, use `read.table()`, with `header = TRUE`, to read the data into a data frame in R called, say, `my.data`:

   ```
   my.data <- read.table(file = my.file, header = TRUE)
   ```

2. The function `aggregate()` is used to compute a summary statistic separately from each group. It takes a *formula* argument (e.g. `Length ~ Species`), a data frame `data`, and an R function `FUN`, and applies that function to each group. Type:

   ```
   aggregate(Length ~ Species, data = my.data, FUN = mean)
   ```

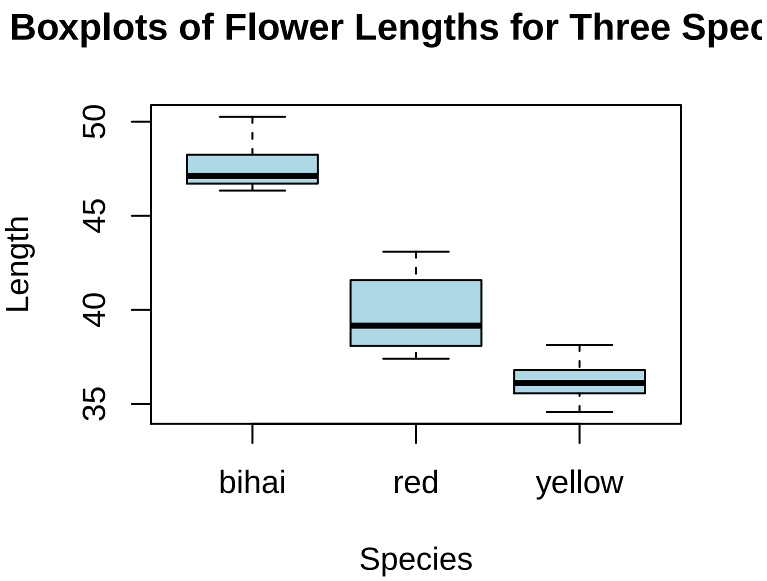
   to compute the **mean length** separately **for each** flower **species** group.

3. Now use `aggregate()` to obtain the standard deviation (`FUN = sd`) for each **species** group.

4. Use `boxplot()` (with the *formula* `Length ~ Species` and your data frame) to make side-by-side boxplots of the flower lengths for the three species, for example by typings something like this:

```
boxplot(Length ~ Species, data = my.data, col = "lightblue",
        main = "Boxplots of Flower Lengths for Three Species")
```

   Your plot should look something like the one below.



5. We'll test the hypotheses

$$H_0: \quad \mu_1 = \mu_2 = \mu_3$$
$$H_a: \quad \text{Not all } \mu_i\text{'s are equal.}$$

   The null hypothesis says there are **no differences** in the mean **flower lengths** for the three **species**. The alternative says there *are* **differences**.

   The function `aov()` will carry out a ***one-factor ANOVA*** to test the above hypotheses. It takes arguments:

   `formula` a formula specifying the model, such as y ~ x, where y is a numeric response variable and x is the factor.

   `data` a data frame from which the variables in the formula will be found.

Carry out the ***ANOVA*** and save the results in an object called `my.anova` by typing:

```
my.anova <- aov(Length ~ Species, data = my.data)
```

Then use `summary()` to look at the ***ANOVA table***:

```
summary(my.anova)
```

Make sure to look at the results of the ***ANOVA F test*** to decide if there's a **species** effect (i.e. to decide if there are differences in the mean **flower lengths** for the three **species**).

2 Part B: Checking Assumptions for the ANOVA *F* Test

2.1 Flowers Data Set (Cont'd)

The **ANOVA *F* test** of **Part A** requires that either the samples are from **normal** populations (or the sample sizes $\boldsymbol{n}$ are all **larger** than about 15), and that the **population standard deviations** are **equal**.

One way check the **normality** assumption is make a **histogram** or **normal probability plot** of the ***residuals***.

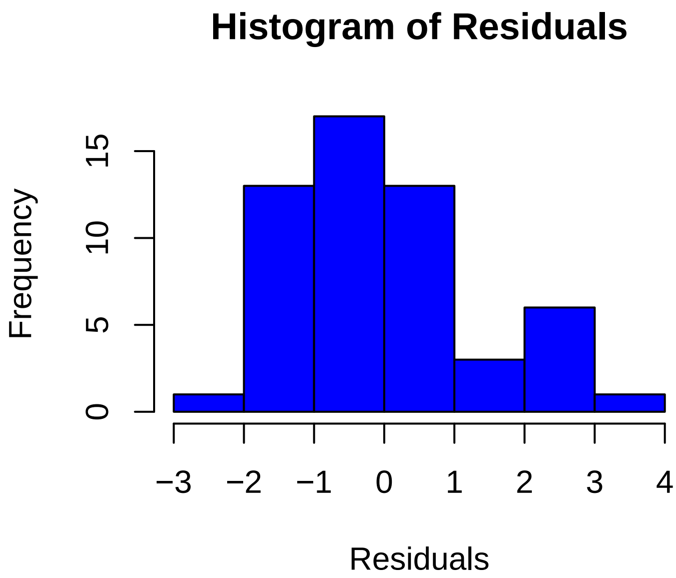
1. The object `my.anova` from **Step 5** of **Part A** is a *list* object containing several items:

   ```
   names(my.anova)
   ```

   You can get the ***residuals*** using the dollar sign operator `$`, i.e.:

   ```
   my.anova$residuals
   ```

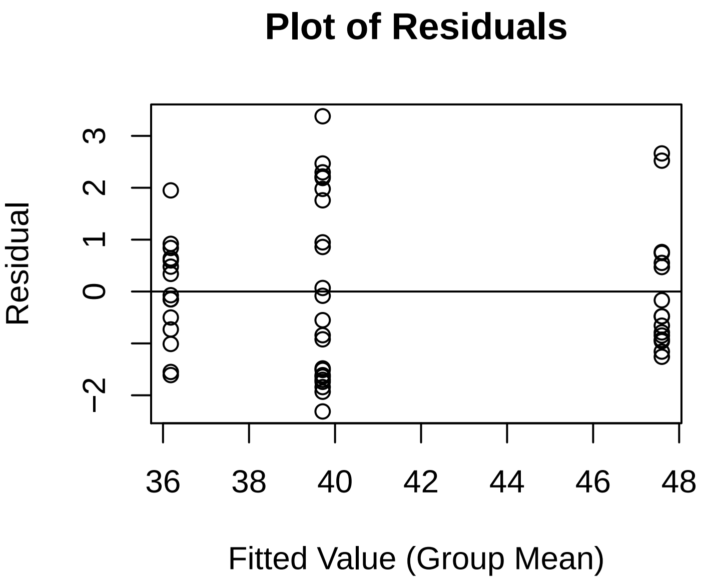
   Now check the **normality assumption** by using `hist()` to make a **histogram** of the residuals. Your histogram should look similar to the one below.



2. We check the **equal population standard deviation assumption** by plotting the ***residuals*** (y-axis) versus the ***fitted values*** (group means, x-axis) by typing something like:

```
plot(x = my.anova$fitted.values, y = my.anova$residuals,
     main = "Plot of Residuals",
     ylab = "Residual",
     xlab = "Fitted Value (Group Mean)",
     pch = 19)
abline(h = 0)
```

Your plot should look similar to the one below.


Plot of Residuals
Residual
3
2
1
0
−2
36
38
40
42
44
46
48
Fitted Value (Group Mean)