MTH 3240 Lab 9

Due Thu., Apr. 23

1 Part A: Two-Factor ANOVA

1.1 Macroinvertebrates Data Set

Macroinvertebrates are larger-than-microscopic invertebrate animals. Freshwater macroinvertebrates include aquatic insects, worms, clams, snails, and crustaceans. Because of their abundance and their sensitivity to environmental impacts, they're widely used as bioindicators of water quality in rivers and streams.

A study gathered data on the change in abundance of two freshwater macroinvertebrates, mayflies (which, in their immature stages, are aquatic) and aquatic worms, in three New Zealand **streams** over two different annual **seasons**. Mayflies are pollution-sensitive, so by measuring the **ratio** of mayflies to aquatic worms, stream pollution can be assessed.

The **ratios** are given in the data file **macroinvertebrates.txt**. The two factors are stream and season. Observations were made on three days for each **stream** by **season** combination.

We're interested in examining the ***main effects*** of **season** and **stream** on the macroinvertebrate ratio, and the effect of their ***interaction***.

1. Use `read.table()` to read the data into a *data frame* called, say, `my.data`.

2. Use `aov()` to carry out the ***two-factor ANOVA***, save the result as, say `my.anova`, for example by typing:

   ```
   my.anova <- aov(Ratio ~ Season + Stream + Season:Stream, data = my.data)
   ```

3. Now use `summary()` to look at the **ANOVA table**.

4. Make an ***interaction plot*** of the mean ratio, with season on the horizontal axis and stream represented by separate lines, for example by typing:

   ```
   interaction.plot(x.factor = my.data$Season,
                    trace.factor = my.data$Stream,
                    response = my.data$Ratio,
                    trace.label = "Stream",
                    xlab = "Season", ylab = "Ratio")
   ```

   Your plot should look like this:
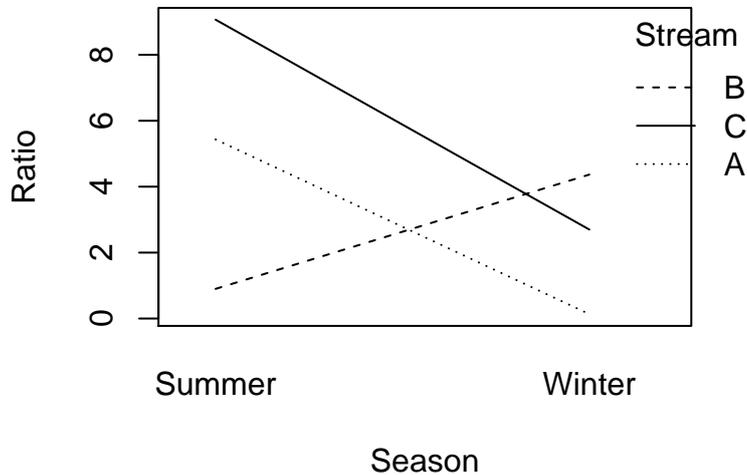
5. Use `qqnorm()` and `qqline()` to make a **normal probability plot** of the **residuals** to check the normality assumption:

```
qqnorm(my.anova$residuals, pch  = 19)
qqline(my.anova$residuals, col = "blue")
```

6. Use `plot()` to make a plot of the **residuals versus the fitted values** to check the assumption that there's a common standard deviation across groups:

```
plot(x = my.anova$fitted.values, y = my.anova$residuals, pch = 19)
abline(h = 0)
```

2 Part B: Two-Factor ANOVA

2.1 Prince William Sound Hydrocarbons Data Set

In order to establish baseline levels prior to the start of oil tanker movement through the Prince William Sound, Alaska, various hydrocarbons were measured in sediment at **seven stations** in the sound in **three months** (May, June, and August, 1978). At each station, **two observations** of several hydrocarbons were made each month. The file **prince_william_sound.txt** contains the **pristane** and **phytane** concentrations (ng/g).

We want to carry out a *two-factor ANOVA*, with **pristane** as the response and **month** and **station** as the factors.

1. Use **read.table()** to read the data from **prince_william_sound.txt** into a *data frame* in R called, say, **my.data**.

2. Use **aov()** to carry out a ***two-factor ANOVA*** (with additive factor effects) by typing:

```r
my.anova <- aov(Pristane ~ Month + Station + Month:Station, data = my.data)
```

3. Now look at the **ANOVA table** by typing:

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

4. Make an ***interaction plot*** of the pristane concentrations, with **month** on the horizontal axis and **stations** represented by separate lines, by typing:

```r
interaction.plot(x.factor = my.data$Month,
                 trace.factor = my.data$Station,
                 response = my.data$Pristane,
                 trace.label = "Stations",
                 xlab = "Month", ylab = "Mean Pristane")
```