

MTH 4230 Lab 6

Due Wed., Apr. 15

1 Part A: Regression with Interactions

1.1 Real Estate Data

A real estate agency provides data on rental properties for clients in a metropolitan area (from **Problem 6.18** in the book). The file **properties.txt** contains the real estate data on **rental rate**, **age**, **operating expense**, **vacancy rate**, and **square footage** for each of $n = 81$ properties.

1. Read the data into R using `read.table()`. Then **remove** the `Op.Expense` column (we won't be using it):

```
my.data$Op.Expense <- NULL
```

2. Use `lm()` to fit the **multiple regression model** to the data, with `Rent.Rate` as the response and the **three** predictors (`Age`, `Vac.Rate`, and `Sq.Ft`) and their **two-** and **three-way interactions** in the model either by listing the interactions explicitly (using the `:` operator):

```
my.reg <- lm(Rent.Rate ~ Age + Vac.Rate + Sq.Ft +  
            Age:Sq.Ft + Age:Vac.Rate + Vac.Rate:Sq.Ft +  
            Age:Vac.Rate:Sq.Ft,  
            data = my.data)
```

or by using the `*` operator:

```
my.reg <- lm(Rent.Rate ~ Age * Vac.Rate * Sq.Ft, data = my.data)
```

Then use `summary()` to look at the results.

2 Part B: Qualitative and Quantitative Predictors: Test for Equal Intercepts

2.1 Fisher's Iris Data Set

The `iris` data comes built in to R, and contains the famous data on iris flowers analyzed by the statistician R. A. Fisher. For more information, type:

```
? iris
```

We'll use a subset of the data consisting of the **lengths** and **widths** of petals of two different species of iris flowers, *Iris Versicolor* and *Iris Verginica*:

```
# Extract the rows corresponding to I. Versicolor and I. Verginica,  
# and just the 3rd through 5th columns:  
my.data <- iris[iris$Species %in% c("versicolor", "virginica"), 3:5]
```

```
# Simplify the column names:  
names(my.data) <- c("Length", "Width", "Species")
```

```
# Convert from factor to "character" to prevent headaches:  
my.data$Species <- as.character(my.data$Species)
```

1. Make a scatterplot of the data, with **Length** on the *y*-axis and **Width** on the *x*-axis, and with different **colors** representing **Species**:

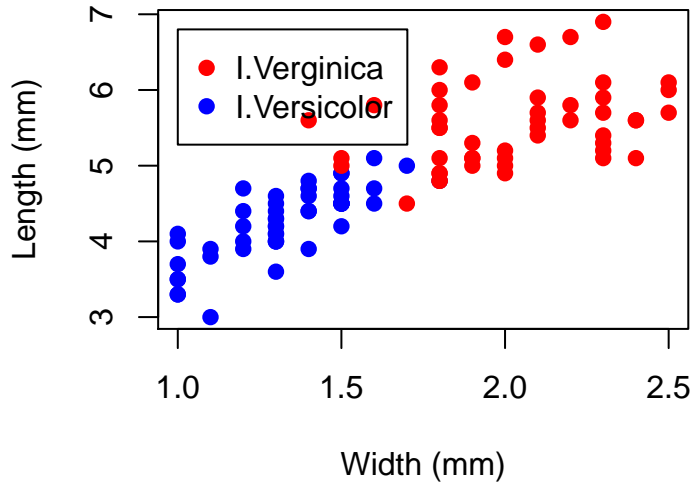
First make a "character" vector of colors corresponding to **Species**:

```
# "character" vector of colors corresponding to Species:  
my.colors <- ifelse(my.data$Species == "virginica", "red", "blue")
```

```
plot(x = my.data$Width, my.data$Length, col = my.colors, pch = 19,  
     main = "Length vs Width of Pedals for Two Iris Species",  
     xlab = "Width (mm)", ylab = "Length (mm)")  
  
legend(x = 1.0, y = 6.8,  
       pch = 19, col = c("red", "blue"),  
       legend = c("I.Verginica", "I.Versicolor"))
```

Your plot should look like this:

Length vs Width of Pedals for Two Iris Species



2. Fit the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon$$

where

$$\begin{aligned} Y &= \text{Length (mm)} \\ X_1 &= \text{Species} = \begin{cases} 0 & \text{if versicolor} \\ 1 & \text{if verginica} \end{cases} \\ X_2 &= \text{Width (mm)} \end{aligned}$$

for example by typing:

```
my.reg <- lm(Length ~ Species + Width, data = my.data)
```

then check R's *coding scheme* and look at the *design matrix* using:

```
contrasts(as.factor(my.data$Species))  
model.matrix(my.reg)
```

3. Use `summary()` to look at the results of the regression analysis from the previous step.
4. Use `abline()` to add two lines to the scatterplot of Step 1, one corresponding to *Iris Verginica* and the other to *Iris Versicolor*, for example by typing:

```

b0 <- my.reg$coefficients[1] # 2.8485
b1 <- my.reg$coefficients[2] # 0.5468
b2 <- my.reg$coefficients[3] # 1.0645

abline(a = b0 , b = b2, col = "blue")
abline(a = b0 + b1, b = b2, col = "red")

```

3 Part C: Qualitative and Quantitative Predictors: Test for Equal Slopes

3.1 Fisher's Iris Data Set (Cont'd)

In Part A, we tested whether for the lines for the two species had the same **intercepts** (assuming they had the same slopes).

Now we'll test whether they have the same **slopes** by including the **interaction** term in the model.

1. Fit the model with the **interaction** between and **Species** and **Width**,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_1 X_2 + \epsilon,$$

where (again)

$$\begin{aligned}
Y &= \text{Length (mm)} \\
X_1 = \text{Species} &= \begin{cases} 0 & \text{if versicolor} \\ 1 & \text{if virginica} \end{cases} \\
X_2 &= \text{Width (mm)}
\end{aligned}$$

and the product $X_1 X_2$ is the **interaction** term, for example by typing:

```
my.reg <- lm(Length ~ Species + Width + Species:Width, data = my.data)
```

then look at the **design matrix** using:

```
model.matrix(my.reg)
```

2. Use `summary()` to look at the results of the regression analysis from the previous step.
3. Recreate the scatterplot of Step 1 of Part B.

Then use `abline()` to add two lines to the scatterplot corresponding to *Iris Virginica* and *Iris Versicolor* by typing:

```
b0 <- my.reg$coefficients[1] # 1.7813
b1 <- my.reg$coefficients[2] # 2.4594
b2 <- my.reg$coefficients[3] # 1.8693
b3 <- my.reg$coefficients[4] # -1.2221

abline(a = b0 , b = b2, col = "blue")
abline(a = b0 + b1, b = b2 + b3, col = "red")
```