MTH 4230 Lab 2

Due Wed., Feb. 12

1 Part A: Lack of Fit Test

1.1 Chemistry Experiment Data Set

A chemist studied the concentration of a solution (Y) over time (X). Fifteen identical solutions were prepared. The 15 solutions were randomly divided into five sets of three, and the five sets were measured, respectively, after 1, 3, 5, 7, and 9 hours. The results are in the file **chemistry.txt**.

1. Use `read.table()` (with `header=TRUE`) to read the data into an R data frame called, say, `my.data`.

2. Make a scatterplot of the **concentration** (Y) versus **time** (X) using `plot()`, for example by typing:

```
plot(x = my.data$Hours, y = my.data$Conc, pch = 19,
  main = "Concentration vs Time", xlab = "Time", ylab = "Concentration")
```

3. Use `lm()` to fit a ***simple linear regression model*** with **concentration** as the response and **time** as the predictor by typing:

```
my.reg <- lm(Conc ~ Hours, data = my.data)
```

   Then look at the results by typing:

```
summary(my.reg)
```

4. Add the **fitted regression line** to the scatterplot created in Step 2 by typing:

```
abline(my.reg)
```

5. Plot the **residuals** (y-axis) versus the **fitted values** (x-axis) using `plot()` and `abline()`:

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

6. We want to carry out a ***lack of fit test***. First fit the **full model**

$$Y_{ij} = \mu_i + \epsilon_{ij},$$

   where μ_i is the true mean concentration for the ith time point, by typing:

```
my.full.reg <- lm(Conc ~ factor(Hours), data = my.data)
```

7. Now fit the **reduced model** (the usual linear regression model)

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$$

by typing:

```
my.reduced.reg <- lm(Conc ~ Hours, data = my.data)
```

8. Lastly, carry out the ***F test for lack of fit*** by typing:

```
anova(my.reduced.reg, my.full.reg)
```

2 Part B: Transformations

2.1 Chemistry Experiment Data Set (Cont'd)

Notice from the scatterplot of Steps 2 and 4 of Part A that the relationship between **concentration** and **time** is nonlinear.

One possible remedy is to make a ***transformation*** of the **concentrations** so that their relationship to **time** is more linear.

1. Perform the (base-10) ***log transformation***

$$Y' = \log_{10}(Y) = \log_{10}(\text{concentration})$$

of the **concentrations** using something like:

```
log.Conc <- log10(my.data$Conc)
```

Then add the **log concentrations** to the *data frame* `my.data` by typing:

```
my.data$log.Conc <- log.Conc
```

2. Make a scatterplot of **log concentration** versus **time**.
3. Use `lm()` to perform a ***linear regression analysis*** with **log concentration** as the response and **time** as the predictor. Then use `summary()` to view the results.
4. Add the **fitted regression line** to the scatterplot of Step 2 using `abline()`.
5. Check the **normality assumption** for the error term ϵ in the regression model by making a **histogram** and **normal probability plot** of the residuals (`my.reg$residuals`) using `hist()`, `qqnorm()`, and `qqline()`.

6. Plot the **residuals** (y-axis) versus the **fitted values** (x-axis) to check the **constant standard deviation assumption** using `plot()` and `abline()`.

7. The ***fitted model*** has the form:

$$\hat{Y}' = b_0 + b_1 X\,.$$

where

$$Y' = \log_{10}(\text{concentration}) \qquad \text{and} \qquad X = \text{time}\,.$$

We can express the **fitted model** in the **original units** as:

$$\hat{Y} = 10^{b_0 + b_1 X}\,.$$

Plot the data on the **original scale** by typing:

```
plot(x = my.data$Hours, y = my.data$Conc, pch = 19)
```

then add the **fitted values** on the **original scale** to the plot, **connected by lines**, by typing:

```
lines(x = my.data$Hours, y = 10^my.reg$fitted.values, col = "blue")
```

3 Part C: General Linear *F* Test

3.1 Chemistry Experiment Data Set (Cont'd)

1. We want to carry out a ***general linear F test*** to compare the **full model**

$$Y_i' = \beta_0 + \beta_1 X_i + \epsilon_i$$

to the reduced model

$$Y_i' = \beta_0 + \epsilon_i$$

(both models using **log concentration** Y' as the response).

Fit the **full model**, for example by typing

```
my.full.reg <- lm(log.Conc ~ Hours, data = my.data)
```

2. Now fit the **reduced model** by typing

```
my.reduced.reg <- lm(log.Conc ~ 1, data = my.data)
```

3. Finally, carry out the ***general linear F test*** by typing:

```
anova(my.reduced.reg, my.full.reg)
```