STAT 420 Spring 2014

Homework 6: Due March 19 by 7:00pm

Exercise 1

Chemists often use ion-sensitive electrodes (ISEs) to measure the ion concentration of aqueous solutions. These devices measure the migration of the charge of these ions and give a reading in millivolts (mV). A standard curve is produced by measuring known concentrations (in ppm) and fitting a line to the millivolt data. The below table gives the concentrations in ppm and the voltage in mV for calcium ISE.

i	ppm	mV	i	ppm	mV	i	ppm	mV
1	0	1.72	7	75	2.40	13	150	4.47
2	0	1.68	8	75	2.32	14	150	4.51
3	0	1.74	9	75	2.33	15	150	4.43
4	50	2.04	10	100	2.91	16	200	6.67
5	50	2.11	11	100	3.00	17	200	6.66
6	50	2.17	12	100	2.89	18	200	6.57

(a) Plot the points mV (y) versus ppm (x). Does a linear model seem appropriate here?

(b) Use the Box-Cox method to determine the best transformation on the response mV.

Exercise 2

The dataset `uswages` is drawn as a sample from the Current Population Survey in 1988. Fit a model with weekly wages as the response and years of education and experience as predictors. Report and give a simple interpretation to the regression coefficient for years of education. Now fit the same model but with logged weekly wages. Give an interpretation to the regression coefficient for years of education. Which interpretation is more natural?

```
> library(faraway)
> data(uswages)
```

Exercise 3

Data set `mammals` contains the average body weight in kg (x) and the average brain weight in g (y) for 62 species of land mammals.

```
> library(MASS)
> data(mammals)
```

Researchers such as Sprent (1972) and Gould (1996) have noted that the following relationship seems to work well:

$$\text{brain weight} = b_0(\text{body weight})^{b_1}(e)$$

This model asserts that brain weight is proportional to body weight raised to the b_1 power, with a multiplicative error e. Obviously, this model can be linearized if we take the logarithm of both x and y. That is,

$$\ln(\text{brain weight}) = \ln(b_0) + b_1 \ln(\text{body weight}) + \ln(e)$$

(a) Plot the average brain weight (y) versus the average body weight (x).

(b) Use the Box-Cox method to verify that ln(brain weight) is a "recommended" transformation of the response variable. That is, verify that $\lambda = 0$ is among the "recommended" values of λ when considering

$$g_\lambda(\text{brain weight}) = \ln(b_0) + b_1 \ln(\text{body weight}) + \ln(e)$$

where $g_\lambda(\cdot)$ denotes the Box-Cox transformation.

(c) Plot ln(brain weight) versus ln(body weight). Does a linear relationship seem to be appropriate here? Fit the model

$$\ln(\text{brain weight}) = \ln(b_0) + b_1 \ln(\text{body weight}) + \ln(e)$$

and use it to predict the average brain weight of a Siberian tiger (average body weight 227 kg). Construct a 95% prediction interval.

Exercise 4

Consider the GPA data from http://onlinestatbook.com/2/regression/intro.html. The data file `gpa.csv` is available on Compass.

You can use the `read.csv` function to read the data file in to R:

```
gpa=read.csv("/Users/Nate/Desktop/gpa.csv",header=TRUE)
```

You will have to change the file path to the location of the `gpa.csv` file on your machine.

(a) Fit the simple linear regression model

$$\texttt{univ\_GPA} = b_0 + b_1(\texttt{high\_GPA}) + e_i$$

assuming that $e_i \overset{\text{iid}}{\sim} \mathrm{N}(0, \sigma^2)$. Plot the residuals $\hat{e}_i$ (y axis) against `high_GPA` (x axis). Do you notice any pattern that may suggest a problem?

(b) Test the assumption that the residuals are normally distributed using $\alpha = 0.01$. Report the null and alternative hypothesis of your chosen test. Also, report the observed test statistic, p-value, and your decision.

(c) Test the assumption of homogeneity of variance using $\alpha = 0.01$. Report the null and alternative hypothesis of your chosen test. Also, report the observed test statistic, p-value, and your decision.