STAT 420
2GR, 2UG

Spring 2017

Homework 10

(Due Friday, April 21, by 4:00 p.m.)

Please submit your assignment *on paper*, following the Guidelines for Homework Write-Ups and Submissions. Please include your name (with your last name underlined), and your NetID at the top of the first page.

Note: For this entire Homework 10, you are to use R / R Studio for all computations and plots. Please copy and paste your results and printouts into Word or other word processor.

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 table on the right gives the concentrations in ppm and the voltage in mV for calcium ISE.
The data are also stored in ISEs.dat

ppm	mV
0	1.72
0	1.68
0	1.74
50	2.04
50	2.11
50	2.17
75	2.40
75	2.32
75	2.33
100	2.91
100	3.00
100	2.89
150	4.47
150	4.51
150	4.43
200	6.67
200	6.66
200	6.57

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

(b) Fit a quadratic model with mV as the response. Does it seem to provide a better fit? Explain using an appropriate test. Use $\alpha = 0.05$. State the null hypothesis, report the value of the test statistic, the distribution of the test statistic under the null hypothesis, the *p*-value, and the decision.

(c) Add the fit from part (b) (best-fit parabola) to the plot from part (a).

2. Use the `odor` data from the `faraway` package.

```
> library(faraway)
> data(odor)
> #View(odor)
> #?odor
```

(a) Fit a model with `odor` as the response, and the first and second orders of the three other variables as predictors by including linear and quadratic terms without any interaction terms. (Note: this leads to a total of six predictor terms, with three corresponding to the linear terms and three corresponding to the quadratic terms.) Perform the significance of the regression test. Use $\alpha = 0.05$. Report the value of the test statistic, the distribution of the test statistic under the null hypothesis, the *p*-value, and the decision.

(b) What is the proportion of the observed variation of `odor` explained by the model in part (a)?

(c) Find the value of Adjusted R^2 and the value of AIC for the model in part (a).

3. For the `prostate` data from the `faraway` package, fit a model with `lpsa` as the response and the other variables as predictors.

(a) Implement the backward elimination (i.e. backward selection) method to determine the "best" model for each of the following two criteria:

i) AIC; ii) BIC.

(b) Implement the forward selection method to determine the "best" model for each of the following two criteria:

i) AIC; ii) BIC.

(c) Use the `leaps` package for model selection to determine the "best" model for each of the following three criteria:

i) AIC; ii) BIC; iii) Adjusted R^2.

(d) Among all the methods and criteria employed above in parts (a), (b), and (c), identify which model seems to be the best choice, and provide your justification.