STAT 420 Spring 2014

Homework 7: Due April 4 by 7:00pm

Exercise 1 (revisited)

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). Overlay the least-squares (linear) regression line, and comment on whether the line seems to fit the data.

(b) Fit a quadratic model. Overlay the least-squares (quadratic) regression line; does it seem to provide a better fit?

(c) Fit a cubic model. Overlay the least-squares (cubic) regression line; which model (linear, quadratic, or cubic) do you choose and why?

Exercise 2

A survey was conducted to study teenage gambling in Britain. (Ide-Smith & Lea, 1988, Journal of Gambling Behavior, 4, 110–118) The data is stored in the data frame `teengamb` (library `faraway`). This dataframe contains the following columns:

`sex`	0=male, 1=female
`status`	Socioeconomic status score based on parents occupation
`income`	in pounds per week
`verbal`	verbal score in words out of 12 correctly defined
`gamble`	expenditure on gambling in pounds per year

We will try to model `gamble` as the response and the other variables as predictors.

```
> library(faraway)
> data(teengamb)
```

(a) Plot `gamble` vs `status`, `gamble` vs `income`, and `gamble` vs `verbal`, using different symbols for males and females. Do these plots suggest the possible need for the interaction terms between sex and the other predictors?

(b) Fit a model with `gamble` as the response and the other variables as predictors that includes the interaction terms between `sex` and the other predictors. Determine whether this model may be reasonably simplified.

Exercise 3

Suppose a complete second-order model

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1} x_{i2} + b_4 x_{i1}^2 + b_5 x_{i2}^2 + e_i$$

was fit to $n = 24$ observations.

```
> sum( lm( y ~ 1 )$residuals^2 )
[1] 360
> sum( lm(y ~ x1+x2)$residuals^2 )
[1] 126
> sum( lm(y ~ x1+x2+I(x1*x2))$residuals^2 )
[1] 100
> sum( lm(y ~ x1+x2+I(x1*x2)+I(x1^2)+I(x2^2))$residuals^2 )
[1] 72
```

(a) Perform the "significance of the regression" test at a 5% level of significance.

(b) Test whether the second-order terms are significant at a 5% level of significance. What is the p-value of the test?

Exercise 4

Can a corporation's annual profit be predicted from information about the company's chief executive officer (CEO)? Forbes (May, 1999) presented data on company profit (y), (in \$ millions), CEO's annual income (x_1) (in \$ thousands), and percentage of the company's stock owned by the CEO (x_2).

Company	Profit (y, \$)	CEO	Income (x_1, \$)	Stock (x_2, %)
Gap	824.5	Drexler	3,743	1.71
Intel	6,068.0	Grove	52,598	0.13
Gateway 2000	346.4	Waitt	855	43.93
HJ Heinz	746.9	O'Reilly	2,916	1.63
Conseco	630.7	Hilbert	124,579	3.64
Citicorp	5,807.0	Reed	6,200	0.22
Cisco Systems	1,362.3	Chambers	560	0.06
General Electric	9,296.0	Welch	40,626	0.03
America Online	254.0	Case	26,917	0.54
Computer Associates	570.0	Wang	10,614	3.79
Lockheed Martin	1,001.0	Augustine	2,533	0.01
Bear Stearns	538.6	Cayne	23,215	3.44

(a) Fit the interaction model

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1} x_{i2} + e_i$$

Give the least squares prediction equation and determine whether the overall model is statistically useful for predicting company profit at $\alpha = 0.10$.

(b) Is there evidence to indicate that CEO income x_1 and stock percentage x_2 interact? Use a level of $\alpha = 0.05$.

(c) Based on the least squares estimates of the b_j parameters, give the estimate of the change in profit for every one thousand dollar increase in a CEO's income when a CEO owns 2% of the company's stock.

Exercise 5

Suppose the interaction model

$$y_i = b_0 + b_1 x_{i1} + b_2 x_{i2} + b_3 x_{i1} x_{i2} + e_i$$

was fit to n = 20 data points, and the following results were obtained:

```
> sum( lm( y ~ 1 )$residuals^2 )
[1] 57
> sum( lm( y ~ x1 )$residuals^2 )
[1] 40
> sum( lm( y ~ x2 )$residuals^2 )
[1] 45
> sum( lm( y ~ x1 + x2 )$residuals^2 )
[1] 36
> sum( lm( y ~ x1 + x2 + I(x1*x2) )$residuals^2 )
[1] 30
> lm( y ~ x1 + x2 + I(x1*x2) )$coefficients
(Intercept)          x1          x2     I(x1 * x2)
         10           5          -2              3
```

(a) Perform the significance of the regression test at $\alpha = 0.05$.

(b) Do x_1 and x_2 interact? Perform the appropriate test at $\alpha = 0.05$.

(c) Is there sufficient evidence to indicate that x_2 contributes information for the prediction of y? Perform the appropriate test at $\alpha = 0.05$. What is the p-value of this test?

(d) Estimate the change in $E(Y)$ for every 1-unit increase in x_1 when $x_2 = 2$.

(e) Estimate the change in $E(Y)$ for every 1-unit increase in x_2 when $x_1 = 3$.