Homework 5

Due June 2, 2017 in class

Please show all work for full credit. Print and staple your assignment together and submit by end of class the due date. If you cannot attend class on the due date, please arrange to submit your homework prior to the due date.

1. [Ch. 4.2, Exercise 1, pg. 161] Return to Homework 2, #5 (Exercise 3 of Section 4.1 on pg. 140 of the textbook). Fit a quadratic relationship $y \approx \beta_0 + \beta_1 x + \beta_2 x^2$ to the data via least squares. By appropriately plotting residuals and examining R^2 values, determine the advisability of using a quadratic rather than a linear equation to describe the relationship between x and y. If a quadratic fitted equation is used, how does the predicted mean molecular weight at 200°C compare to that obtained in part e) of the earlier exercise?

2. [Ch. 4.2, Exercise 2, pg. 161] Here are some data (also on website as `pulp.csv`) taken from the article "Chemithermomechanical Pulp from Mixed High Density Hardwoods" by Miller, Shankar, and Peterson (*Tappi Journal*, 1988). Given are the percent NaOH used as a pretreatment chemical, x_1, the pretreatment time in minutes, x_2, and the resulting value of a specific surface area variable, y (with units of cm^2/g), for nine batches of pulp produced from a mixture of hardwoods at a treatment temperature of 75°C in mechanical pulping.

% NaOH, x_1	Time, x_2	Specific Surface Area, y
3	30	5.95
3	60	5.60
3	90	5.44
9	30	6.22
9	60	5.85
9	90	5.61
15	30	8.36
15	60	7.30
15	90	6.43

   a) Fit the approximate relationship $y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2$ to these data via least squares. Interpret the coefficients b_1 and b_2 in the fitted equation. What fraction of the observed raw variation in y is accounted for using the equation?

   b) Compute and plot residuals for your fitted equation from a). Discuss what these plots indicate about the adequacy of your fitted equation. (At a minimum, you should plot residuals against all of x_1, x_2, and $\hat{y}$ and normal-plot the residuals.)

   c) Make a plot of y versus x_1 for the nine data points and sketch on that plot the three different linear functions of x_1 produced by setting x_2 first at 30, then 60, and then 90 in your fitted equation from a). How well to fitted responses appear to match observed responses?

   d) What specific surface area would you predict for an additional batch of pulp of this type produced using a 10% NaOH treatment for a time of 70 minutes? Would you be willing to make a similar prediction for 10% NaOH used for 120 minutes based on your fitted equation? Why or why not?

   e) There are many other possible approximate relationships that might be fitted to these data via least squares, one of which is $y \approx \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1 x_2$. Fit this equation to the preceding data and compare the resulting coefficient of determination to the one found in a). On these basis of these alone, does the use of a more complicated equation seem necessary?

   f) For the equation fit in part e), repeat the steps of part c) and compare the plot made here to the one made earlier.

g) What is an intrinsic weakness of this real published data set?

h) What terminology (for data structures) introduced in section 1.2 describes this data set? It turns out that since the data set has this special structure and all nine sample sizes are the same (i.e., all are 1), some special relationships hold between the equation fit in a) and what you get by separately fitting linear equations in x_1 and x_2 to the y data. Fit such one-variable linear equations and compare coefficients and R^2 values to what you obtained in a). What relationships exist between these?

3. [Ch. 4, Exercise 12, pg. 208] The article referred to in Homework 2, #6 (Ch. 4.1 Exercise 4, pg. 140) actually considers the effects of both cutting speed and feed rate on tool life. The whole data set from the article follows and is on the website as `tool_life_full.csv`. (The data for the previous assignment are the $x_2 = .01725$ data only.)

Cutting speed, x_1 (sfpm)	Feed, x_2 (ipr)	Tool life, y (min)
400	0.01725	21.5, 24.5, 26, 33
450	0.02200	4, 4.7, 5.3, 6
500	0.01570	8.8, 11, 11.75, 19
500	0.01725	6.4, 7.8, 9.8, 16.5
600	0.01725	2.35, 2.65, 3, 3.6
600	0.02200	1.2, 1.5, 1.6, 1.6
700	0.01570	1.75, 1.85, 2, 2.2
700	0.01725	1, 1.2, 1.5, 1.6
800	0.01725	1, 0.9, 0.74, 0.66

a) Taylor's expanded tool life equation is $yx_1^{\alpha_1}x_2^{\alpha_2} = C$. This relationship suggests that $\ln(y)$ may well be approximately linear in both $\ln(x_1)$ and $\ln(x_2)$. Use a multiple linear regression program to fit the relationship

$$\ln(y) \approx \beta_0 + \beta_1 \ln(x_1) + \beta_2 \ln(x_2)$$

to these data. What fraction of the raw variability in $\ln(y)$ is accounted for in the fitting process? What estimates of the parameters α_1, α_2, and C follow from your fitted equation?

b) Compute and plot residuals (continuing to work on log scales) for the equation you fit in part a). Make at least plots of residuals versus fitted $\ln(y)$ and both $\ln(x_1)$ and $\ln(x_1)$ and make a normal plot of these residuals. Do these plots reveal any particular problems with the fitted equation?

c) Use your fitted equation to predict first a log tool life and then a tool life, if in this machining application a cutting speed of 550 and a feed of .01650 is used.

d) Plot the ordered pairs appearing in the data set in the (x_1, x_2)-plane. Outline a region in the plane where you would feel reasonably safe using the equation you fit in part a) to predict tool life.