STAT 305 D Homework 4

Due February 21, 2012 at 12:40 PM in class

1. Vardeman and Jobe chapter 4 section 2 problem 1 (page 161). The data are available at http://will-landau.com/data/csv/polypolyols.csv.

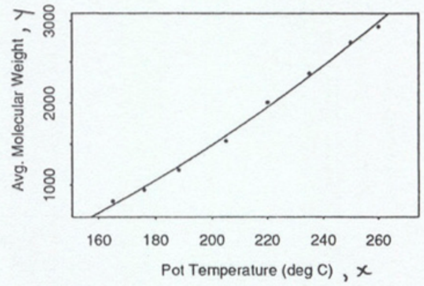
   Return to Exercise 3 of Section 4.1. Fit a quadratic relationship $y \approx \beta_0 + \beta_1 x + \beta_2 x^2$ to the data via least squares. By appropriately plotting residuals and examining R^2 values, determine the advisability of using a quadratic rather than a linear equation to describe the relationship 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?

   The relevant information from section 4.1 exercise 3 is this: The article Polyglycol Modified Poly (Ethylene Ether Carbonate) Polyols by Molecular Weight Advancement by R. Harris (Journal of Applied Polymer Science, 1990) contains some data on the effect of reaction temperature on the molecular weight of resulting poly polyols. The data for eight experimental runs at temperatures 165°C and above are as follows. Here, x is pot temperature (°) and y is average molecular weight.

x	y
165.00	808.00
176.00	940.00
188.00	1183.00
205.00	1545.00
220.00	2012.00
235.00	2362.00
250.00	2742.00
260.00	2935.00

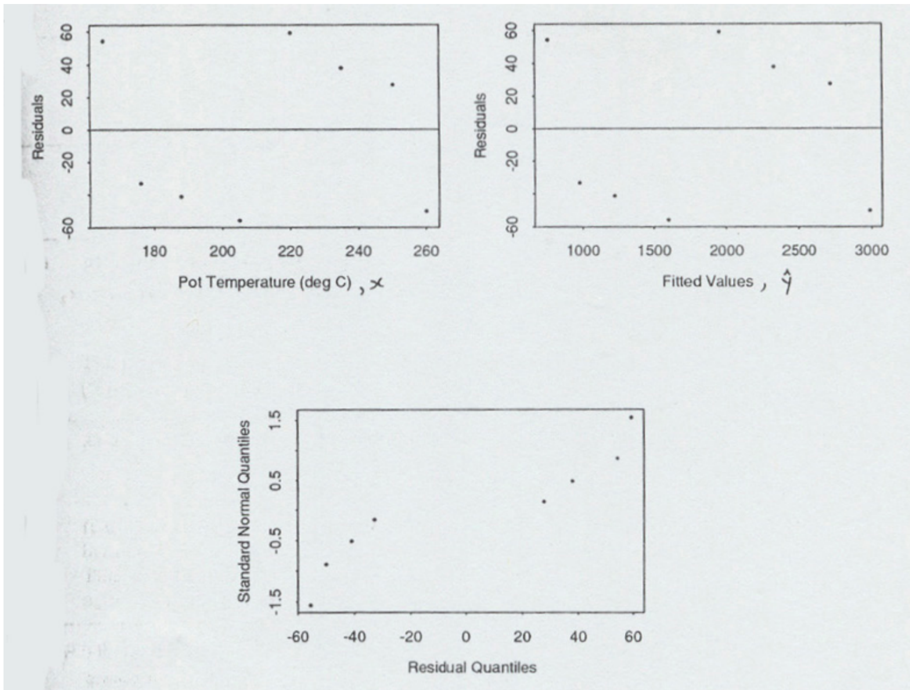
1. The least squares equation is

$$\hat{y} = -1315 + 5.59x + .04212x^2.$$



$R^2_Q = .996$, compared with $R^2_L = .994$. This is a very small improvement, at the cost of using

a more complex equation.



The residuals here are smaller, as they will always be for a more complex model. There is no noticeable improvement in the residual plots, compared to those from the straight-line model. In fact, the residual plots for the quadratic model look more patterned. The scatterplot of y versus x also indicates that the quadratic model would be "overfitting" the data. The simpler straight-line relationship seems to be adequate.

For the quadratic model, at $x = 200°$C,

$$\hat{y} = -1315 + 5.59(200) + .04212(200)^2 = 1487.2,$$

which is relatively close to 1525.1 from Ch. 4, Sec. 1, Ex. 3.

2. Vardeman and Jobe chapter 4 section 2 problem 2 parts a-c (page 161). The data are available at http://will-landau.com/stat305/data/csv/pulp.csv. Here are some data 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, x1, the pretreatment time in minutes, x2, and the resulting value of a specific surface area variable, y (with units of cm3/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
3.0	30	5.95
3.0	60	5.60
3.0	90	5.44
9.0	30	6.22
9.0	60	5.85
9.0	90	5.61
15.0	30	8.36
15.0	60	7.30
15.0	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 this 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.)

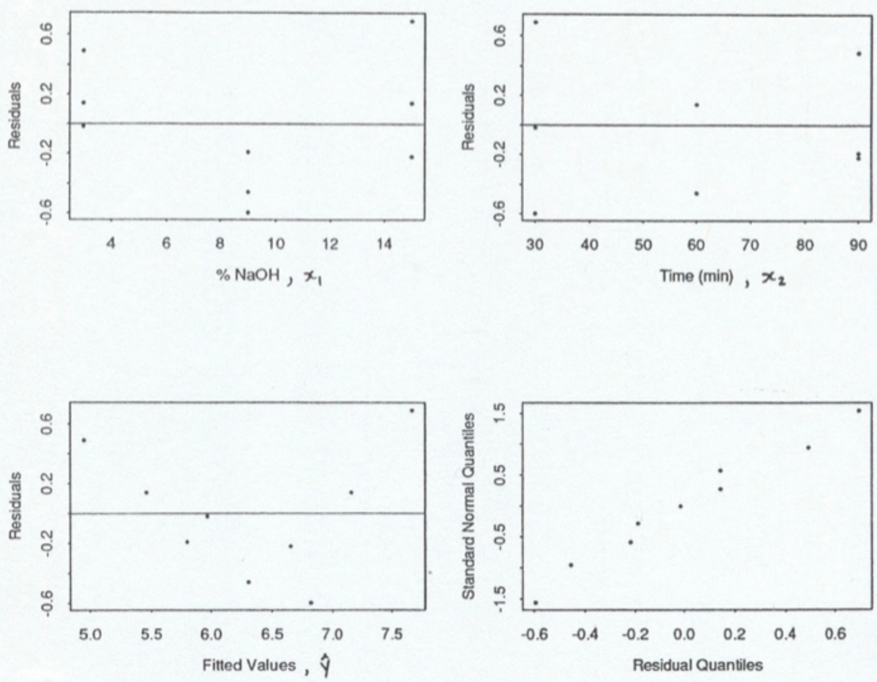
(c) Make a plot of y versus x_1 for the nine data points and sketch on that plot the three different linear functions of x_1 produced by setting x_2 first at 30, then 60, and then 90 in your fitted equation from (a). How well do fitted responses appear to match observed responses?

2. (a) The least squares equation is

$$\hat{y} = 6.0483 + .14167x_1 - .016944x_2.$$

Assuming the fitted equation is appropriate, this means that as x_1 increases by 1% (holding x_2 constant), y increases by roughly .14167 cm^3/g. As x_2 increases by 1 minute (holding x_1 constant), y decreases by roughly .016944 cm^3/g. The R^2 corresponding to this equation is .807.

(b) The residuals are $-.015$, $.143$, $.492$, $-.595$, $-.457$, $-.188$, $.695$, $.143$, $-.218$.



Both the plots of residuals versus x_1 and residuals versus $\hat{y}$ show a positive-negative-positive pattern of residuals, indicating that the relationship between x_1 and y is not completely accounted for by the current model. These plots suggest adding an x_1^2 term. The plot of residuals versus x_2 is patternless; x_2 seems to be well represented. The normal plot of residuals is fairly linear, indicating that the residuals are bell-shaped.

(c) For $x_2 = 30$, the equation is

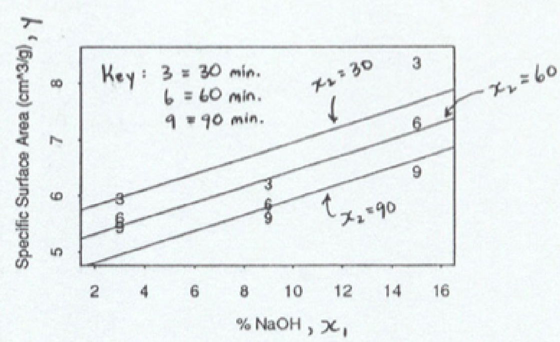
$$\begin{aligned}\hat{y} &= 6.0483 + .14167x_1 - .016944(30)\\ &= 5.53998 + .14167x_1.\end{aligned}$$

For $x_2 = 60$, the equation is

$$\begin{aligned}\hat{y} &= 6.0483 + .14167x_1 - .016944(60)\\ &= 5.03166 + .14167x_1.\end{aligned}$$

For $x_2 = 90$, the equation is

$$\begin{aligned}\hat{y} &= 6.0483 + .14167x_1 - .016944(90)\\ &= 4.52334 + .14167x_1.\end{aligned}$$



The fitted responses do not match up well, because the relationship between y and x_1 (%NaOH) is not linear for any of the x_2 values (Time).

3.

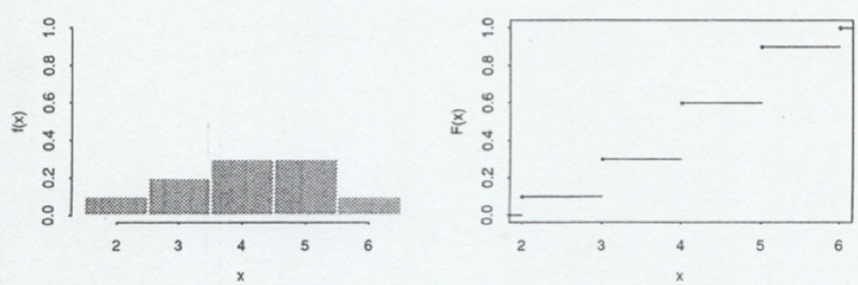
A discrete random variable X can be described using the probability function:

x	2	3	4	5	6
$f(x)$	.1	.2	.3	.3	.1

a. Plot f(x) in a "probability bar chart". Also, plot $F(t)$, the cumulative probability function of X.

b. Calculate $E(X)$ (the mean of X) and SD(X) (the standard deviation of X).

(a)



(b) Using equation (5-1),

$$EX = 2(.1) + 3(.2) + 4(.3) + 5(.3) + 6(.1) = 4.1.$$

Using equation (5-2),

$$\text{Var}X = 2^2(.1) + 3^2(.2) + 4^2(.3) + 5^2(.3) + 6^2(.1) - (4.1)^2 = 1.29,$$

so the standard deviation of X is $\sqrt{1.29} = 1.136$.

4.

Airlines sometimes overbook flights. Suppose that for a plane with 50 seats, 55 passengers have tickets. Define the random variable Y as the number of ticketed passengers who actually show up for the flight. The probability mass function of Y appears in the accompanying table.

y	45	46	47	48	49	50	51	52	53	54	55
$p(y)$	.05	.10	.12	.14	.25	.17	.06	.05	.03	.02	.01

a. What is the probability that the flight will accommodate all ticketed passengers who show up?

b. What is the probability that not all ticketed passengers who show up can be accommodated?

c. If you are the first person on the standby list (which means you will be the first one to get on the plane if there are any seats available after all ticketed passengers have been accommodated), what is the probability that you will be able to take the flight? What is this probability if you are the third person on the standby list?

a. In order for the flight to accommodate all the ticketed passengers who show up, no more than 50 can show up. We need $Y \le 50$.
$P(Y \le 50) = .05 + .10 + .12 + .14 + .25 + .17 = .83$

b. Using the information in a. above, $P(Y > 50) = 1 - P(Y \le 50) = 1 - .83 = .17$

c. For you to get on the flight, at most 49 of the ticketed passengers must show up. $P(Y \le 49) = .05 + .10 + .12 + .14 + .25 = .66$. For the 3rd person on the standby list, at most 47 of the ticketed passengers must show up. $P(Y \le 47) = .05 + .10 + .12 = .27$

5.

A contractor is required by a county planning department to submit one, two, three, four, or five forms (depending on the nature of the project) in applying for a building permit. Let Y = the number of forms required of the next applicant. The probability that y forms are required is known to be proportional to y—that is, $p(y) = ky$ for $y = 1, \ldots, 5$.

a. What is the value of k? [*Hint:* $\sum_{y=1}^{5} p(y) = 1$.]

b. What is the probability that at most three forms are required?

c. What is the probability that between two and four forms (inclusive) are required?

d. Could $p(y) = y^2/50$ for $y = 1, \ldots, 5$ be the pmf of Y?

a. $\sum_{y=1}^{5} p(y) = K[1+2+3+4+5] = 15K = 1 \Rightarrow K = \frac{1}{15}$

b. $P(Y \le 3) = p(1) + p(2) + p(3) = \frac{6}{15} = .4$

c. $P(2 \le Y \le 4) = p(2) + p(3) + p(4) = \frac{9}{15} = .6$

d. $\sum_{y=1}^{5} \left(\frac{y^2}{50}\right) = \frac{1}{50}[1+4+9+16+25] = \frac{55}{50} \ne 1$; No

6. Weekly feedback. You get full credit as long as you write something.

   1. Is there any aspect of the subject matter that you currently struggle with? If so, what specifically do you find difficult or confusing? The more detailed you are, the better I can help you.
      You got full credit as long as you wrote something.
   2. Do you have any questions or concerns about the material, class logistics, or anything else? If so, fire away. You got full credit as long

as you wrote something.