STAT 512

Homework 3 (23 pts. + 1.5 pt. BONUS) due Jan. 31

A reminder – Please do not hand in any unlabeled or unedited SAS output. Include in your write-up only those results that are necessary to present a complete solution (what you want the grader to grade). In particular, questions must be answered in order (including graphs), and all graphs must be fully labeled (main title should include the question number, and all axes should be labeled). Don't forget to put all necessary information (see course policies) on the first page. Include the SAS input for all questions at the very end of your homework; this could be important even though it won't be graded. You will often be asked to continue problems on successive homework assignments so save all your SAS code.

1. (10 pts.) Consider the following data set that describes the relationship between the rate of an enzymatic reaction (V) and the substrate concentration (C). A common model used to describe the relationship between the rate and the concentration is the Michaelis-Menten model $V = \frac{\theta_1 C}{\theta_2 + C}$, where θ_1 is the maximum rate of the reaction and θ_2 describes how quickly the reaction will reach its maximum rate. The equation can be rearranged so that $\frac{1}{v}$ can be written as a linear model with explanatory variable $\frac{1}{c}$:

$$\frac{1}{V} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1} \frac{1}{C}$$

The data set is as follows:

Concentration	Rate	Concentration	Rate
0.02	49	0.22	159
0.02	47	0.22	152
0.06	97	0.56	191
0.06	107	0.56	201
0.11	123	1.10	207
0.11	139	1.10	200

Even though we know theoretically what the final equation 'should' be, let's see if we can generate this.

(a) Since we know the theoretical equation, we will use a transformation of Y' = $\frac{1}{\gamma}$. Note: The Box-Cox procedure does NOT help in determining the transformation in this case. We will define new variables for the Y in SAS. The new variable can be defined as follows (if the raw data set used the variable v with a dataset name of original:

```
data transformY;
    set original;
    vinv = 1/v;
```

Now, we need to do the diagnostics on the data

i. Plot the inverse dependent variable versus the explanatory variable and comment on the shape and any unusual points.

STAT 512

- ii. Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points (Hint: This plot is generated after a least squares regression on $\frac{1}{v}$ vs. C is performed.) The only output that should be included in this part are plots, no listing of data or other numbers.
- iii. Examine the distribution of the residuals by generating a histogram and a normal probability plot of the residuals by using the histogram and qqplot statements in proc univariate. What do you conclude?
- iv. Do we need to do a transformation on X? Why or why not?
- (b) Since the plot still isn't linear, and we have a shape similar to Fig. 3.13c), a transformation of X' = $\frac{1}{X}$ is appropriate (since it matches the theory). Perform the data transformation similarly as before and repeat the diagnostics.
 - i. Plot the inverse dependent variable versus the inverse explanatory variable and comment on the shape and any unusual points.
 - ii. Plot the residuals versus the inverse explanatory variable and briefly describe the plot noting any unusual patterns or points
 - iii. Examine the distribution of the residuals by generating a histogram and a normal probability plot of the residuals. What do you conclude?
 - iv. Is the linear regression a valid model now? Why or why not?
- (c) Convert this regression line back into the original nonlinear model and plot the predicted curve on a scatterplot of V vs. C. Comment on the fit. Do you think that any of the points are influencing the line more than others? To generate the predicted curve, simply take the predicted values from the regression model and "re-invert" them (Hint: These values need to be outputted when the regression is performed). For example, suppose results is the data set containing the residuals and predicted values (pred)

```
data invert;
    set results;
    predv = 1/pred;
symbol1 v = circle i = none c = black
symbol2 v = plus i = sm5 c = red
proc gplot data=invert;
    plot v*c predv*c /overlay;
```

- (8.5 pts.) Using the Solution Concentration data (described problem 3.15 on p.150, CH03PR15.DAT), we want to run a linear regression to predict concentration at a particular time.
 - (a) First, we need to perform the diagnostics on the data
 - i. Plot the dependent variable versus the explanatory variable and comment on the shape and any unusual points.
 - ii. Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points.
 - iii. Examine the distribution of the residuals by getting a histogram and a normal probability plot of the residuals. What do you conclude?
 - iv. Do we need to do a transformation on X or Y? Why or why not?
 - (b) Using the automated Box-Cox Procedure, determine which transformation of Y would be appropriate (if any)?

- (c) Using the appropriate transformation, convert the Y variable in SAS as in problem 1. Repeat the diagnostics in part i.
 - i. Plot the transformed dependent variable versus the explanatory variable and comment on the shape and any unusual points.
 - ii. Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points
 - iii. Examine the distribution of the residuals by getting a histogram and a normal probability plot of the residuals. What do you conclude?
 - iv. Is the linear regression a valid model now? Why or why not?
- (d) (BONUS: 1.5 pt.) If you want, you may plot the predicted curve on the original data as in Problem 1. Comment on the fit.
- 3. (4.5 pts.) This is a continuation of the "plastic hardness", Question 1 on HW 2. The context is described in problem 1.22 with the data set of (CH01PR22.DAT).
 - (a) Calculate and interpret the 90% confidence intervals for β_0 and β_1 . Note: these are the separate confidence intervals.
 - (b) Obtain the Bonferroni joint confidence intervals for β_0 and β_1 using an $\alpha = 0.10$. Interpret your confidence intervals. Note: This should be done by hand using output from SAS.
 - (c) Compare your answers in part a) and b). That is, which one of the parts has a larger confidence interval. Why?