Statistics 512: Problem Set 3
Due Tuesday, February 20, 2018 11:59 PM

Important Note – *Every graph or plot you create should have your name printed as a subtitle. Consequently, any graph with no name will result in a* **20% off** *the problem. Also, please attach your code at the end; any homework with no code provided will result in a* **50% off** *on the entire assignment.*

1. 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 rate and 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. With this mode, $\frac{1}{V}$ can be written as a linear model with explanatory variable $\frac{1}{C}$:

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

Concentration	Rate	Concentration	Rate
0.02	76	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

(a) Generate a scatterplot of V vs C. Comment on the shape.

(b) Define new variables for $\frac{1}{V}$ and $\frac{1}{C}$ in SAS, and generate a scatterplot of the new variables. Does the fit appear linear? Do any assumptions appear to be violated? The new variables can be defined as follows (if the dataset `original` contains the raw data):

```
data reaction;
     set original;
     vinv = 1/v;
     cinv = 1/c;
```

(c) Determine the least squares regression line for $\frac{1}{V}$ vs $\frac{1}{C}$. Save the residuals and predicted values. Does the residual plot suggest any problems?

(d) 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. To generate the predicted curve, simply take the predicted values from the regression model and "re-invert" them. For example, suppose `results` is the data set containing the residuals and predicted values (variable `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;
```

For the next 3 questions, use the grade point average data described in the text with Problem 1.19 .

2. Describe the distribution of the explanatory variable. Show the plots and output that were helpful in learning about this variable.

3. Run the linear regression to predict GPA from the entrance test score, and obtain the residuals (do not include a list of the residuals in your solution).

   (a) Verify that the sum of the residuals is zero by running `proc univariate` with the output from the regression.

   (b) Plot the residuals versus the explanatory variable and briefly describe the plot noting any unusual patterns or points.

   (c) Plot the residuals versus the order in which the data appear in the data file and briefly describe the plot noting any unusual patterns or points.

   (d) Examine the distribution of the residuals by getting a histogram and a normal probability plot of the residuals by using the `histogram` and `qqplot` statements in `proc univariate`. What do you conclude?

4. Change the data set by changing the value of the GPA for the last observation from 2.948 to 29.48 (e.g., a typo). You can do this in a `data` step. For example, `data a2; set a1; if _n_ eq 120 then gpa = 29.48`; an alternative is simply to edit the data file.

   (a) Make a table comparing the results of this analysis with the results of the analysis of the original data. Include in the table the following: fitted equation, t-test for the slope, with standard error and p-value, R^2, and the estimate of σ^2. Summarize the differences.

   (b) Repeat parts (b), (c), and (d) from the previous problem and explain how these plots help you to detect the unusual observation.