STAT525 HOMEWORK#4

1. Consider the following data set that describes the relationship between "velocity" of an enzymatic reaction (V) and the substrate concentration (C).

Concentration	Velocity
0.02	76
0.02	47
0.06	97
0.06	107
0.11	123
0.11	139
0.22	159
0.22	152
0.56	191
0.56	201
1.10	207
1.10	200

A common model used to describe the relationship between "velocity" and concentration is the Michaelis-Menten model

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

where θ_1 is the maximum velocity of the reaction and θ_2 describes how quickly (in terms of increasing concentration) the reaction will reach maximum velocity. With this model, $1/V$ can be written as a linear model with explanatory variable $1/C$.

$$\frac{1}{V} = \frac{1}{\theta_1} + \frac{\theta_2}{\theta_1}\left(\frac{1}{C}\right) = \beta_0 + \beta_1\left(\frac{1}{C}\right)$$

You are asked to investigate whether this linear transformation results in a good or poor fit by doing the following steps:

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

b. Define new variables for $1/V$ and $1/C$ in SAS and generate a scatterplot. These new variables can be defined as follows:

```
DATA NEW;
    INPUT c v;
    vinv = 1/v;
    cinv = 1/c;
    CARDS;
    [PUT DATA HERE]
;
```

Does the fit appear linear? Does there appear to be any violation of assumptions?

c. Is the distribution of $1/C$ different than C? Are there any points that may be more influential in determining the fit?

d. Determine the least squares regression line for $1/V$ vs $1/C$. Save the residuals and predicted values. Does the residual plot suggest any problems?

e. 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-inverse" them. For example, consider new1 to be the data set containing the residuals and predicted values.

```
DATA new1;
    SET new1;
    pred1 = 1/pred;

SYMBOL1 V=CIRCLE I=NONE C=BLACK;
SYMBOL2 V=PLUS I=SM5 C=BLACK;
PROC GPLOT;
    PLOT v*c=1 pred1*c=2 / OVERLAY;
RUN;
```

2. KNNL Problem 4.12. This can be done in SAS or by hand. To fit a model with no intercept in SAS, you need to add the `NOINT` option on `MODEL` statement line.

3. KNNL Problem 4.21

4. KNNL Problem 4.25

5. KNNL Problem 6.2

6. (SAS Exercise) Use the **brand preference** data described in KNNL Problem 6.5. Run the linear regression with moisture and sweetness of the product as the explanatory variables and degree of liking as the response variable.

   a. Summarize the regression results by giving the fitted regression equation, the value of R^2.

   b. State the results of the significance test for the null hypothesis that the two regression coefficients for the explanatory variables are *all* zero (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and a brief conclusion in words).

   c. Describe the results of the hypothesis tests for the individual regression coefficients (give null and alternative hypotheses, test statistic with degrees of freedom, p-value, and a brief conclusion in words).

   d. Give separate 95% confidence intervals for the regression coefficients of sweetness and moisture. What is the relationship between these confidence intervals and the above hypothesis results?