

**Home-Assignment 6**      **Wednesday, April 19, 2006**  
**Due Friday April 21, 2006**

*Telemarketing.JMP*

A Company that sells transportation services uses a telemarketing division to help sell its services. The division manager is interested in the time spend on the phone by the telemarketers in the division. Data on the number of *MONTHS* of employment and the number of *CALLS* placed per day is recorded for 20 employees. The division of the manager, think that there may be a relationship between time on the job and the number of calls.

1. Give the scatterplot of calls (y) versus Months (x). Are *CALLS* and *MONTHS* collinear?
2. We want to add a new variable *XQRS*, which is a variable created by squaring each value of the *MONTHS* variable.
3. Test whether the coefficient of the two variables (*MONTHS* and *XQRS*) is significantly different from zero.
4. Test the adequacy of this model.
5. We want to check if *MONTHS* and *XQRS* are collinear, calculate  $VIF_{XQRS}$ . What is your conclusion?
6. Predict number of *CALLS* for one year and for 3 years and provide their confidence interval.

**Problem 2:**

*C6\_EPS.JMP*

Many investors rely on predictions of a company's performance compiled by professional analysis. Here we examine a company's predicted earning per share (*PRE\_EPS*) and attempt to explain this in terms of number of firm and industry variables which are *PRE\_IND*, *CL5Y*, *IL5Y*, *CR*, *TL/TA* and *NI/NS*. *PRE\_EPS* is the predicted earnings per share, *PRE\_IND* is the expected industry change in eps for the next five years and *CL5* is the company change in EPS over the last five years etc...

1. Now consider first a regression with all independent variables. You are concerned that several variables are all measures of different aspects of company's performance.
2. Describe whether or not there is a serious collinearity between *CL5Y* and the other explanatory variables and between *PRE\_IND* and the other explanatory variables.
3. To search for a better regression model, you decide to use stepwise regression to explore the data. Run a stepwise regression with "Forward" option. Fit the best model chosen by this option
4. Run stepwise regression with "Backward" option. Fit the best model chosen by this option
5. The variables *NI/NS* and *CL5Y* are initially considered important by the stepwise regression algorithm but are dropped in the latter stages. Provide some intuition as

to how this could happen?

6. At the end, a regression model was run with two independent variables,  $IL5Y$  and  $TL/TA$ . It turns out that the standardized residuals plot has a particular shape. Describe what kind of shape it is and how we should deal with this problem.

### Problem 3

*Fitness.JMP*

$Y$ : oxygene

$X_1$  : Runtime

$X_2$ : Runpulse

$X_3$ =Rst Pulse

$X_4$ : Maxpulse

Using different combination of  $X_1, X_2, X_3$ , and  $X_4$ , we thought that the following models are important to investigate

1.  $Oxy = \beta_0 + \beta_1 Runtime + e$
2.  $Oxy = \beta_0 + \beta_3 RstPulse + e$
3.  $Oxy = \beta_0 + \beta_1 Runtime + \beta_2 Runpulse + e$
4.  $Oxy = \beta_0 + \beta_1 Runtime + \beta_4 Maxpulse + e$
5.  $Oxy = \beta_0 + \beta_2 Runpulse + \beta_3 RstPulse + e$
6.  $Oxy = \beta_0 + \beta_3 RstPulse + \beta_4 Maxpulse + e$
7.  $Oxy = \beta_0 + \beta_1 Runtime + \beta_2 Runpulse + \beta_4 Maxpulse + e$
8.  $Oxy = \beta_0 + \beta_1 Runtime + \beta_3 RstPulse + \beta_4 Maxpulse + e$

Calculate  $C_p$  and  $PRESS$  for each one of those models

Which one is the best according to  $C_p$  and to  $PRESS$ . Does both model selections agree on choosing the best model?