Lab Assignment 7: Multiple Regression

Multiple Regression
Open the file Fitness.JMP from the subdirectory JMP Extra Data.

1. Look at each of the variables in the data set - histograms, means, medians, skewness, normality, correlations, outliers.

2. After this initial investigation into the data we are ready to fit a multiple regression model to predict "Oxy" based on the other variables found in the data set. To fit a multiple regression model, first go to Analyze → Fit Model and then chose "Oxy" as the dependent variable, Y. Next, choose the independent variables (X's), that will be used to predict "Oxy". We will begin with the full model, which includes all of the continuous variables found in this data set. To implement the full model, choose the variables "Age", "Weight", "Runtime", "RunPulse", "RstPulse", "MaxPulse" and then click on Add under the Construct Model Effects section. Once you have the model all set up, click on OK to run the multiple regression model. Today, we will only focus on the sections Summary of Fit, Parameter Estimates and Residual by Predicted Plot.

3. Write out the full multiple regression model based on the values found in the Parameter Estimates section. To write out the model, use the column of values called Estimates (these are the partial coefficients for each of the independent variables found in the model). What is the multiple regression equation for the full model?

4. Now we need to decide if the model we have just constructed is a good model for predicting "Oxy". This will be done by considering several of the values found in the previously mentioned sections.
   - What is the $R^2$ value and what does it tell us about our model?
   - What does the R.M.S. Error value mean in relation to the regression equation?

5. Look at the column of values called Prob > $|t|$ found in the Parameter Estimates section, the goal will be to have these values less than or equal to 0.05.
   - Which variables in the model which have Prob > $|t|$ values greater than 0.05? How large are their values?

6. We may decide to remove one of these variables later, which can be done by going back to Fit Model screen, Window → Fit Model, clicking on the variable we want to remove, "RstPulse" then Remove and finally we can rerun the model, Run Model. Look at the new output, how did $R^2$, R.M.S. Error, regression equation and Prob > $|t|$ values change?

7. Let’s go back to the first multiple regression model, and look at the residual plot.
   - How would you describe the plot - homoscedastic or heteroscedastic?
   - Is there a visible trend in the residual values? (You may want to define residual value so you can accurately answer these questions.)
• How does the residual plot for the newer model (without RstPulse) compare to the full model?

8. Once we have chosen a multiple regression model, we need to interpret the coefficients in terms of the dependent variable (Oxygen uptake). Let’s interpret several of the coefficients from the full model. For example, the coefficient for ”RunPulse” is -0.367 (rounded), this means that if we hold the every variable constant but ”RunPulse” increases by one unit, then Oxygen uptake will decrease by 0.367 units. Try interpreting each of the other values found in the full model.

Remember to Close all JMP IN windows.