**17.1 – Introduction to Weighted Least Squares (WLS)**

As we have seen several examples the assumption that the conditional variance of the response given the predictors is constant is often times violated, i.e. . One way to address this problem is to use a response transformation as discussed in Section 14 – Response Transformations. Another approach is to allow for the non-constant conditional variance in modelling process. One way to do this is to use *Weighted Least Squares (WLS)*. The weighted least squares model for the case is given below:

where are *known positive numbers*. To estimate the variance function we still only need to estimate the parameter and the weights are known. If the weight is large then the conditional variance is “small” and if the weight is small then the conditional variance is “big”.

To estimate parameters in the mean function we use *Weighted Least Squares (WLS)* instead of *Ordinary Least Squares (OLS)*. WLS chooses the parameters to minimize the weighted least squares criterion shown below

OLS is a special case of WLS where for all . Notice that in choosing the obtaining the parameter estimates the weighted least squares criteria gives more weight to observations with smaller variance and less weight to observations with larger variance. This should make sense intuitively.

The WLS estimates of the in the mean function in matrix terms are given by:

which is an ***diagonal*** matrix with the weights on the diagonal and zeroes in all the off-diagonal elements.

**17.2 – Weights and Weighted Least Squares**

Where do the weights come from? We will consider some common scenarios in the context of simple linear regression (i.e. one predictor ).

1. Suppose that the response values ( are actually sample means (based on samples of size from the population defined by , then the weights because .

Note: We are still assuming , the weights reflect the sample fact that variability of the sample mean is proportional to the sample size.

1. Suppose the response represents a **total** of observations () made when then

.

1. Suppose we assume that the variance is proportional to , i.e.

1. Suppose we have visual evidence that the variance is not constant, generally this will come from a plot of the residuals vs. fitted values or from a NCV plot . We can use our data to estimate weights following the procedure outlined below:
2. Store the residuals and fitted values from an OLS regression.
3. Calculate the absolute value of the residuals ().
4. Regression the absolute residuals () vs. the fitted values () and store the fitted values from this regression. These are an estimate of the error standard deviations, i.e. .
5. Calculate the weights .

Notes on (4):  
This process can be repeated if the WLS regression residuals still show evidence of   
 heteroscedasticity (i.e. nonconstant variation). If we repeat steps (a)-(d) in order to stabilize   
 the variance then we are actually performing a version of *iteratively reweighted least squares (IRLS)* regression. This procedure could be employed in a multiple regression setting as well.

1. Robust Regression

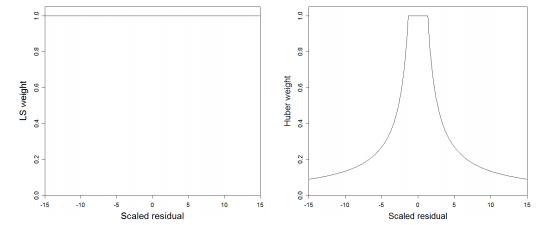
Robust statistical methods are designed to work well when distributional assumptions are violated, such as normality. If we have outliers, which are indicative of heavy tail distributions, robust procedures protect us from their effect on estimated parameters such a coefficients in multiple regression model. One form of robust estimation is called ***M-estimation***, where in the process of fitting these models Iteratively Reweighted Least Squares (IRLS) is used to down weight “outliers” minimizing their effect on the parameter estimates.

At each stage we essentially choose our parameter estimates to minimize

where the weights are determined iteratively at each stage, down weighting

observations with large *robustly* standardized residuals. Below are plots of some   
 of the commonly employed the weight functions.

OLS Huber

  
 Hampel Bisquare

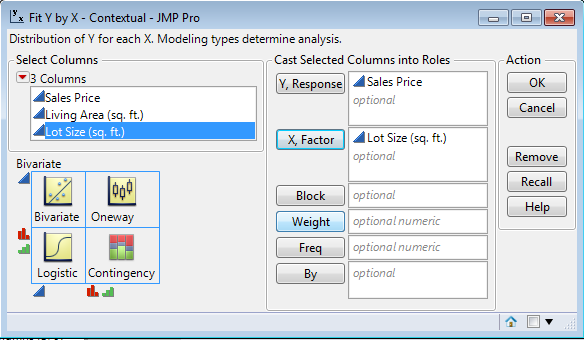
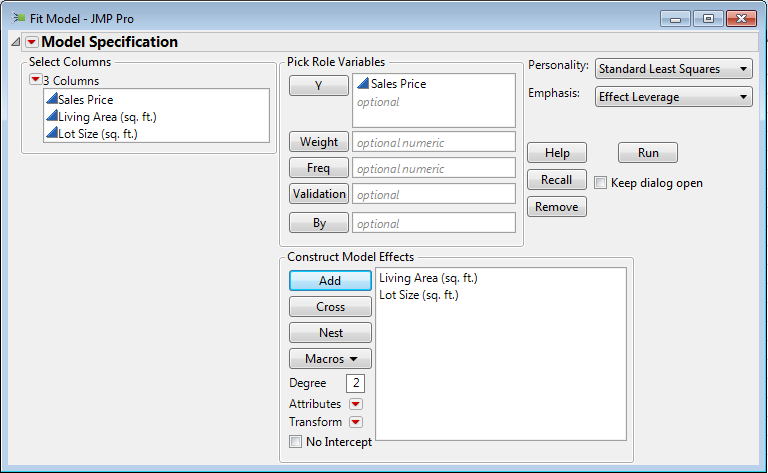


**17.3 – Examples of WLS Regression**

In this section we consider some example of weighted least squares where the weights used demonstrate some of the common situations where WLS is performed.

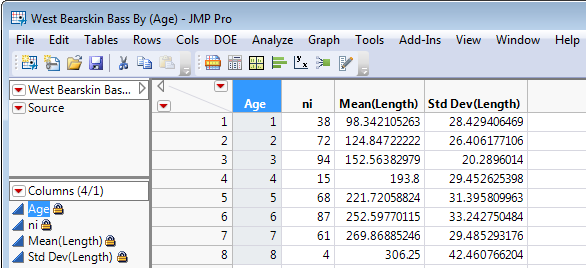
To perform WLS in JMP we simply add the weights we wish to use to the **Weights** portion of the dialog box in the **Fit Y by X** (SLR) or **Fit Model** (MLR) platforms (see below).

**Fit Y by X (Simple Linear Regression)** **Fit Model (Multiple Regression Analysis)**

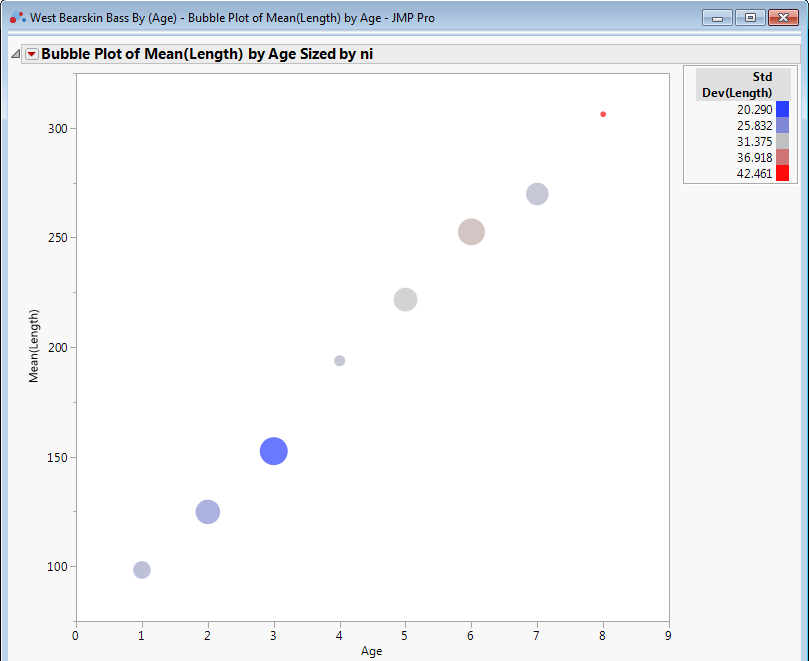
**Example 17.1 – Length and Age of West Bearskin Smallmouth Bass**

Suppose the data from the smallmouth bass study on West Bearskin Lake came to use in the following format instead of the raw data which had a row for each fish sampled. Here the data has been aggregated by age in years. For example, for Age = 1 yr. we have a total sample size of fish with an observed sample mean length and .

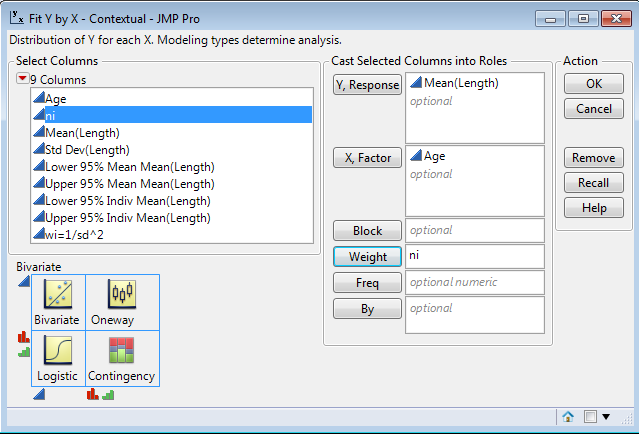


We know , thus the variance is not-constant and the weights .

Below is a scatterplot of mean length vs. with bubbles proportional to the sample size from the age class and color determined by the sample SD for that age class.



Here we will use the sample size for each age class () as a weight



OLS Parameter Estimates WLS Estimates

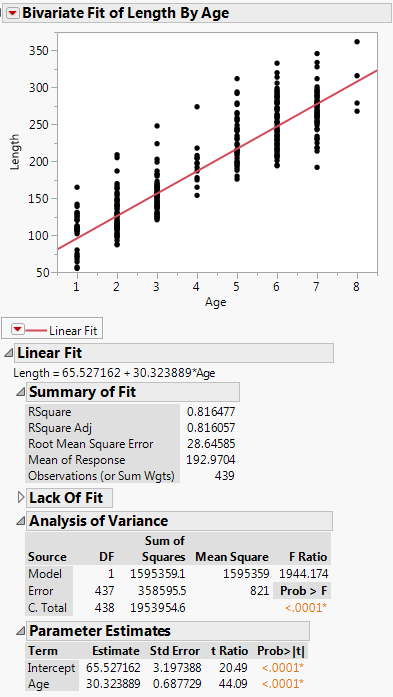
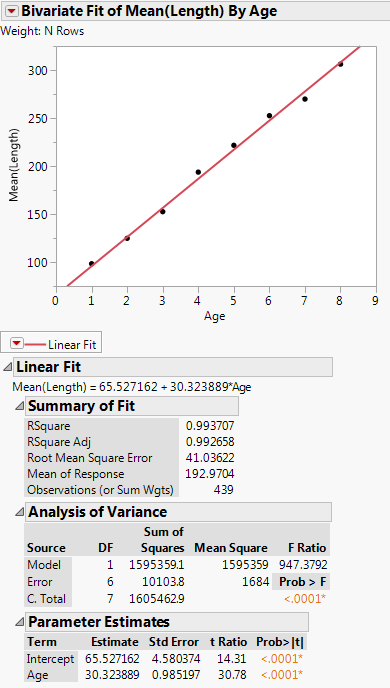
As we also have the we could use as weights.

WLS Estimates



All other quantities (e.g. ) in the regression summary are interpreted in the same manner. We can compare the WLS regression estimates to the same model fit to the full dataset (i.e. without aggregating lengths for each age class).

Using All Fish - not averaged Using Sample Means – average length at each age

CI and PI for Age = 5 yrs. - using individual fish (OLS)

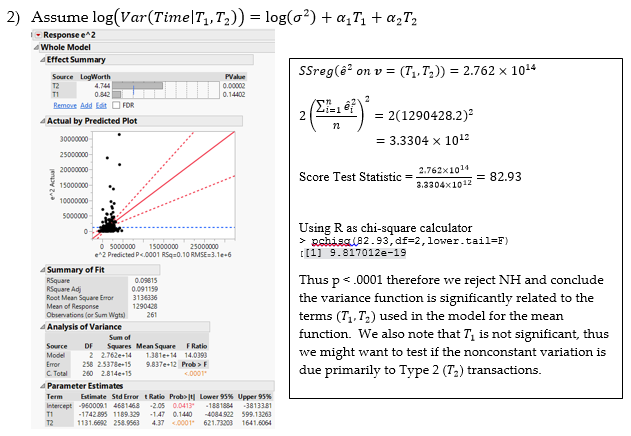


CI and PI for Age = 5 yrs. – using mean length for fish at each age class (WLS)

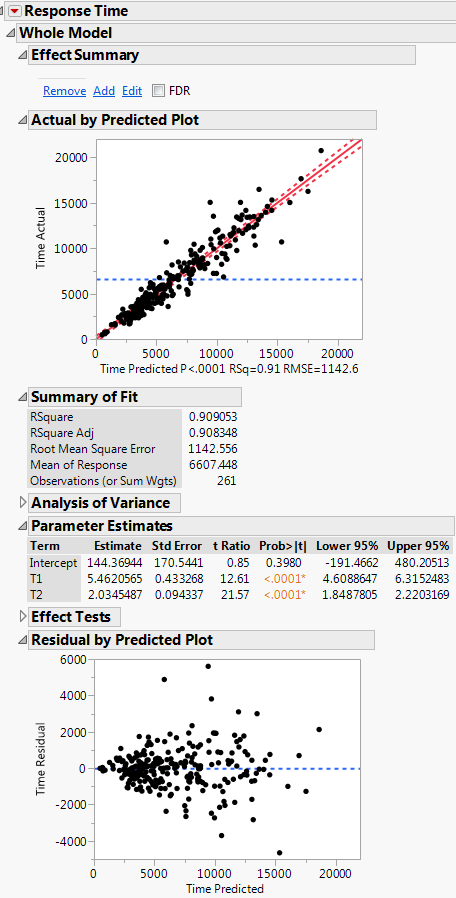


**Example 17.2 – Bank Transactions**

We examined these data in Section 12 when we covered the Score Test for testing if there was evidence on nonconstant variation. In our investigation it was clear that the variation in total transaction time increased as the mean increased or as the total number of transactions of each type of type (either individually or collectively). Below is the regression summary from the regression of total transaction time on both the number of Type 1 (T1) and Type 2 (T2) transactions.

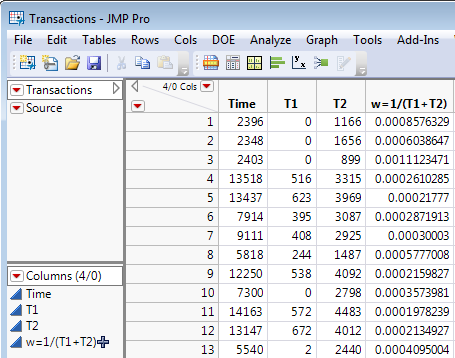


This is the results of the Score Test for testing if the was a function of the number of Type 1 () and Type 2 () transactions (Section 12 – pg. 12).

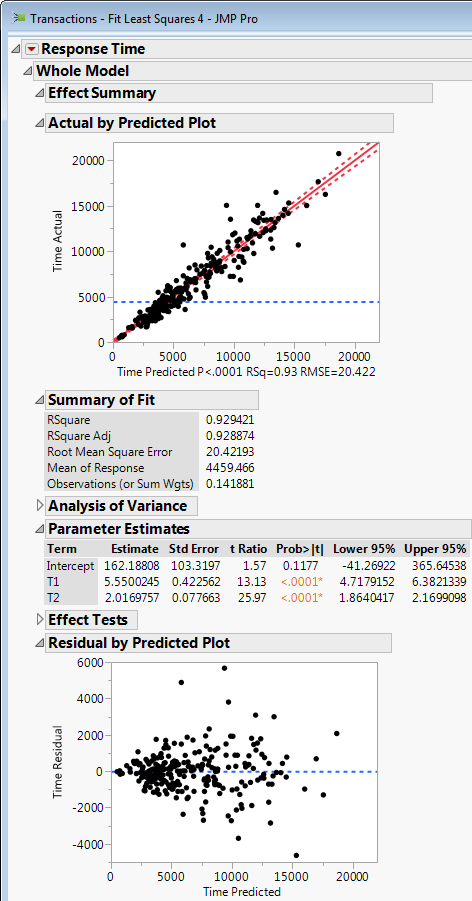


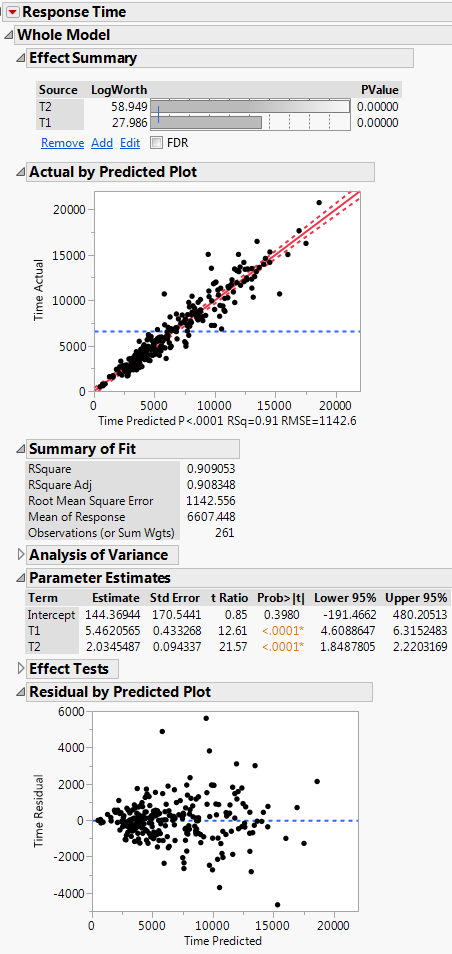
Given the results of the Score Test above we might consider using weights proportional to   
 the total number of transactions of the two types combined, i.e. we will assume .

We can form the weights and use WLS to estimate the model parameters.

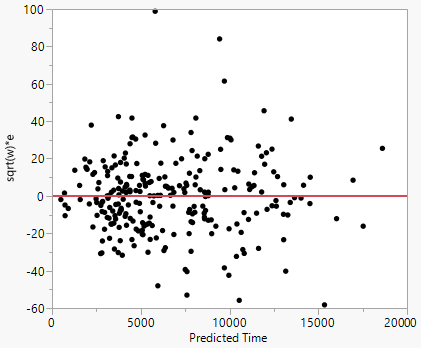


OLS Estimates WLS Estimates ()

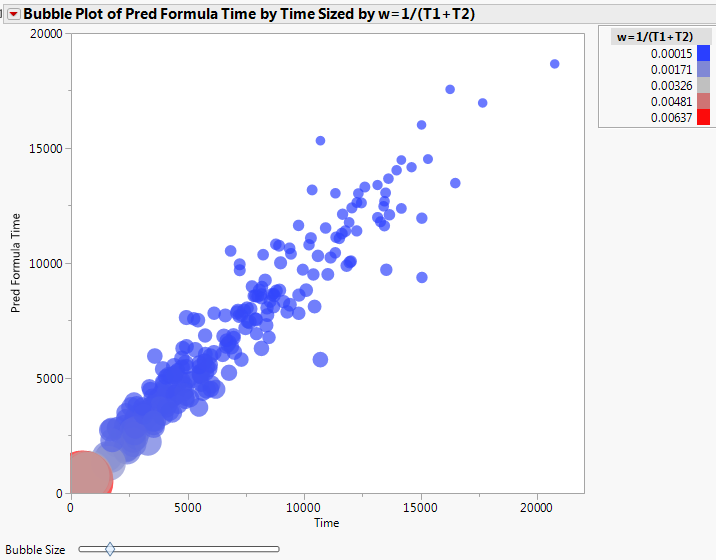




To assess the model fit when using weight least squares it is better to plot vs. ) rather than the usual ().

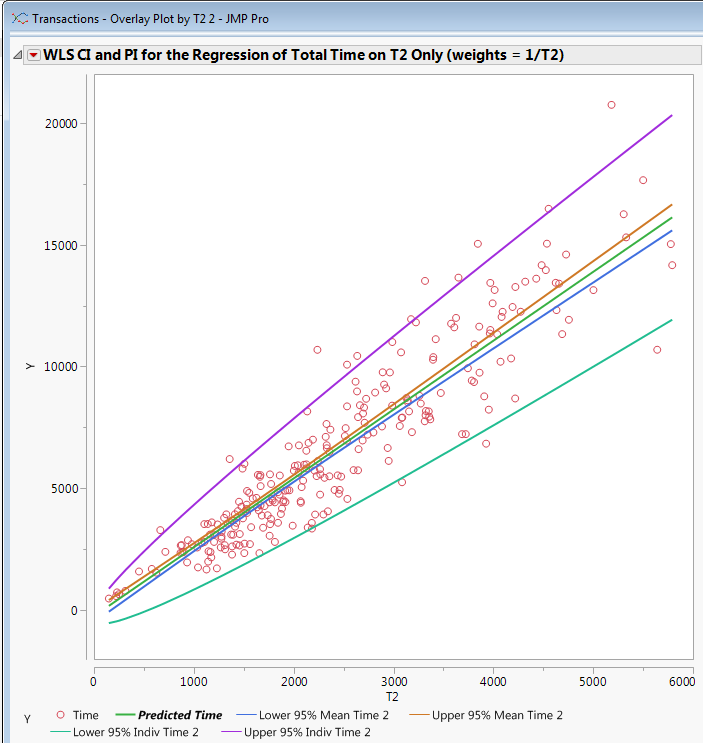
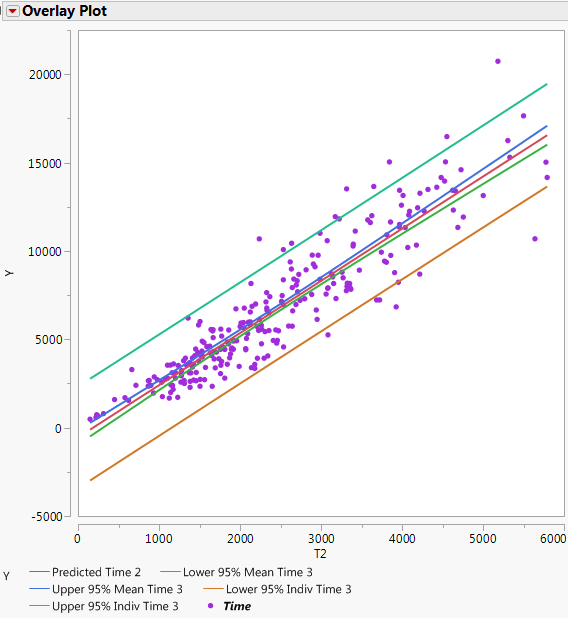


To get a feel for the effect of weighted least squares consider a plot of the fitted values vs. the actual response values () with bubbles proportional to the weights ).



Clearly the weights with a small number of transactions and hence smaller time and predicted time receive far more weight in estimating the model parameters.

The effect of weighting is also seen in the CI for the and PI for **.** To demonstrate this we consider the regression of total transaction time on the number of Type 2 transactions only using weights equal to the reciprocal of the number type 2 transactions (), i.e. we are assuming .

WLS CI and PI for Time|T2 only OLS CI and PI for Time|T2 only

**Example 17.3 -**