**10.0 - Introduction**

In this section we will consider multiple linear regression where all of the predictors are numeric (continuous/discrete) and focus on “interpretation” of the estimated parameters. We will see it is not as straight forward is it in SLR where we only have a single term/predictor.

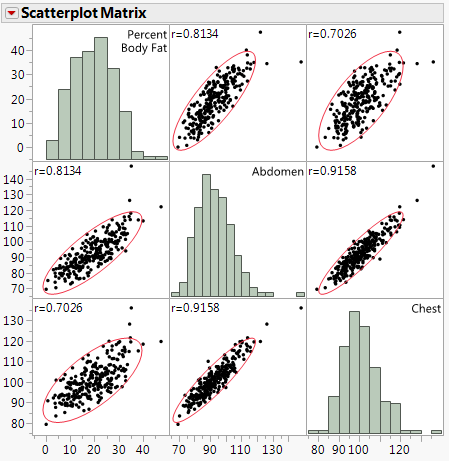
**10.1 – Regression with All Numeric Predictors**

As we saw in Section 8 when adding another numeric predictor to a SLR regression model we need to consider the relationship between the two predictors in understanding what their combined effect will be in the multiple regression model.

We will begin by revisiting Example 8.2 where we considered the regression of percent body fat () on two predictors where and .

**Example 10.1 – Percent Body Fat and Abdominal & Chest Circumferences**

As a first step in performing a multiple regression analysis it is a good idea examine a scatterplot matrix of the response (Y) and any numeric predictors. To obtain a scatterplot matrix in JMP select **Analyze > Multivariate Methods > Multivariate** and put the response all numeric predictors of interest in the ***Y,Columns*** box, response first.



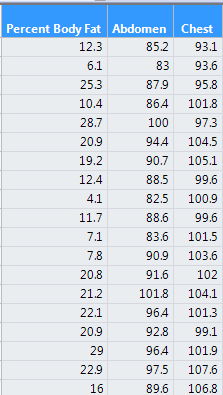
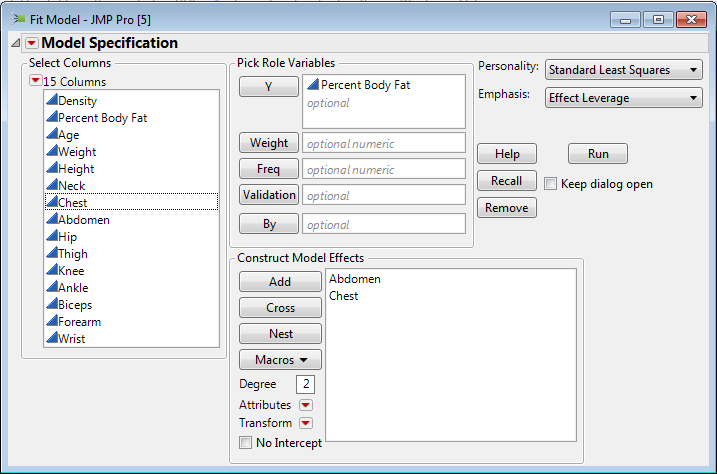


**Comments:**

We begin by fitting the MLR model

and

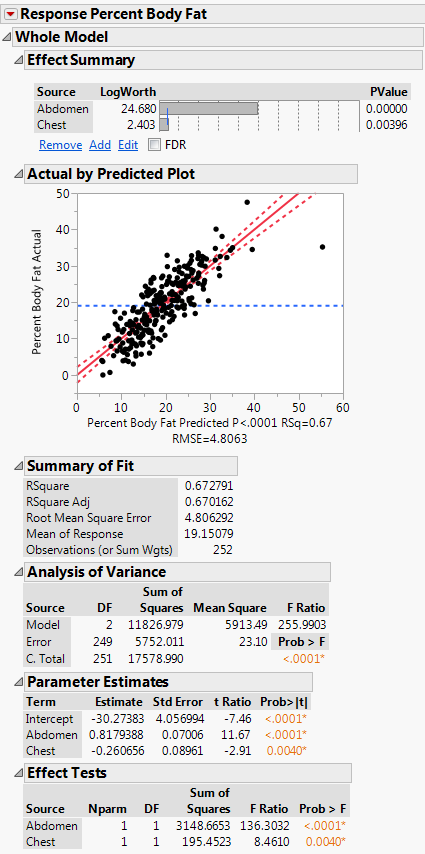
in JMP. Note that the both terms are just the predictors themselves therefore we have not used the term () notation as the terms are simply . To fit a multiple linear regression (MLR) model in JMP we use **Analyze > Fit Model** as shown below.  
  
**Data Table The Fit Model dialog box**

   
Model in Matrix Notation:

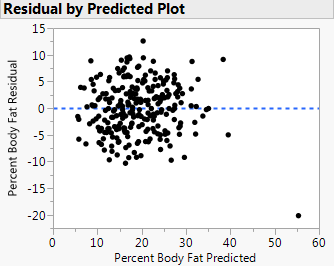
We could write out the model in more descriptive terms as:  
  
The complete summary of the fitted model is given on the following page.



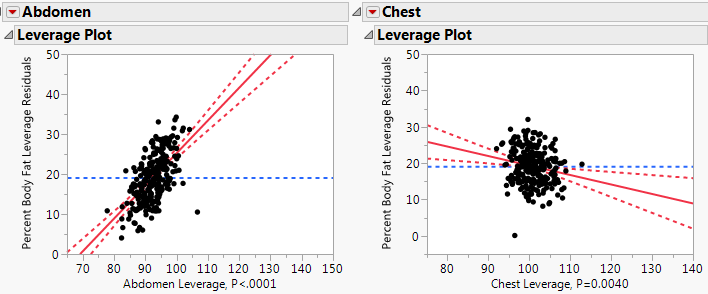
Summary of





**Residuals vs. Fitted Values** ()  


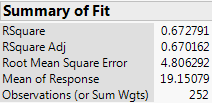
**Effect Leverage Plots (Added Variable Plots)**



**General Discussion:**

**Interpretation of Regression Output**



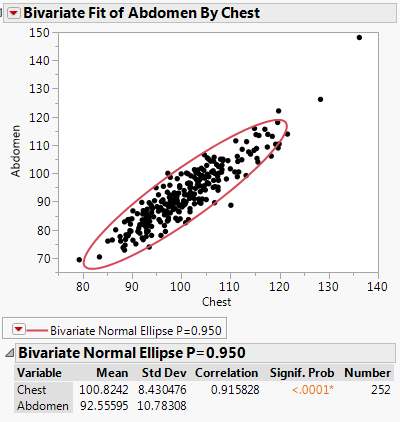


|  |  |  |
| --- | --- | --- |
| **Quantity** | **Value** | **Interpretation** |
|  |  |  |
|  |  |  |
| **Estimated Mean Function:** | | |
| : |  |  |
| : |  | |  |  |  |  | | --- | --- | --- | --- | | Case | Abdominal Circum. (cm) | Chest Circum. (cm) | Prediction | | 1 |  |  |  | | 2 |  |  |  | |
| : |  | |  |  |  |  | | --- | --- | --- | --- | | Case | Abdominal Circum. (cm) | Chest Circum. (cm) | Prediction | | 1 |  |  |  | | 2 |  |  |  | |

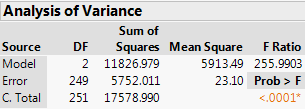
Is it reasonable to hold one predictor constant while changing the other? For example can we examine the estimated effect on mean percent body fat associated with a 5 cm unit increase chest circumference while holding abdominal circumference fixed?



In answer the previous question we need to consider how the predictors/terms are related to each other.

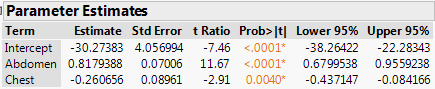


**Tests of Significance –** for the regression of % body fat on abdominal and chest circumference

|  |  |
| --- | --- |
| Tests of Significance | |
| Overall Test |  |
| Test for  Y-Intercept |  |
| Test for Abdominal Slope |  |
| Test for Chest Slope |  |

**Interpretation of Parameter Estimates**



What are the units on the parameter estimates?

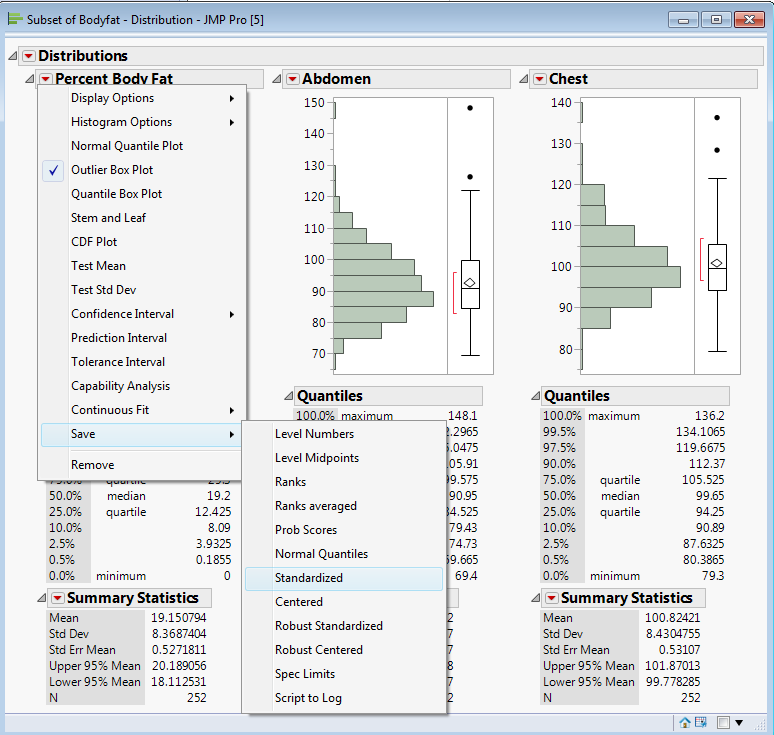
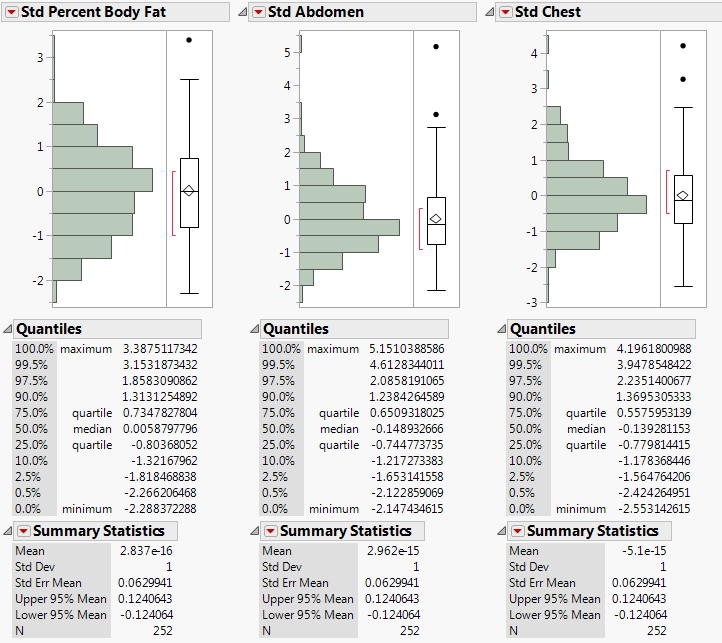
What is the interpretation of the parameter estimates?

Can we conclude that because that abdominal circumference is more important than chest circumference in terms of explaining percent body fat?

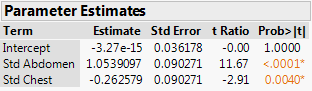
Here both variables are in the same units and in roughly the same scale, but in general this will NOT be true. For this reason, some people consider using **standardized coefficients** or **standardized betas**. Standardized coefficients are obtained from the regression of the standardized response on the standardized predictors/terms. To standardize a variable we subtract the mean and divide by the standard deviation, i.e. convert the variable to z-scores.

For example to standardize the response, % body fat ()

**Select Save > Standardized for each variable Distribution of the standardized response and predictors**

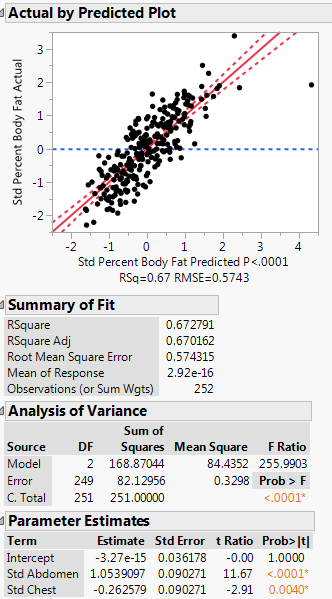
 

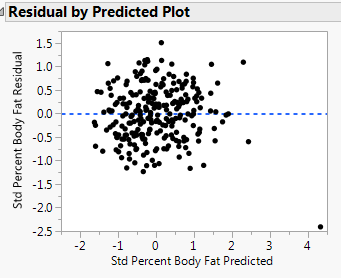
Below are the parameter estimates from the regression of the standardized response on the standardized predictors/terms.

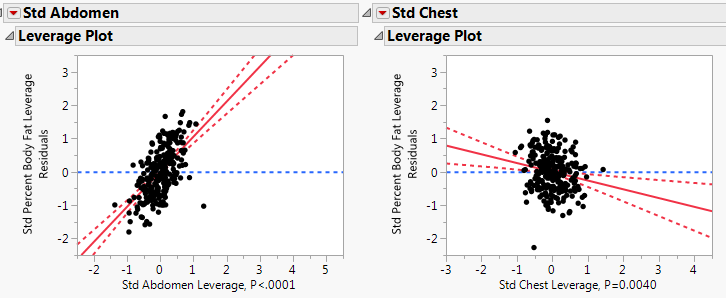


Interpretation of standardized betas:

Here is the full model from the regression using the standardized variables.



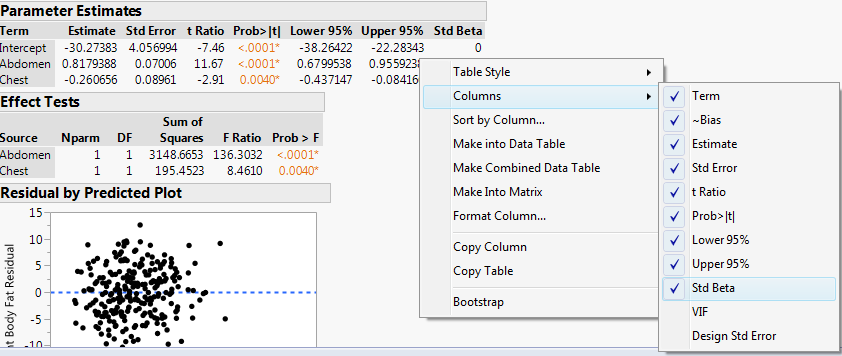
**Residual Plot:** 

**Effect Leverage (Added Variable) Plots**

Notice aside from the parameter estimates and their standard errors, all other output is exactly the same. Also note that the estimated intercept is numerically 0, this will always be the case when using standardized variables.

As before changing one variable while holding the other constant is a major issue, especially we are considering a one standard deviation increase (or decrease) in a standardized predictor/term.

NOTE: To obtain standardized betas we do NOT need to actually standardize the variables and re-run the regression. You can add the standardized betas for any model by adding that column to the Parameter Estimates section of the model summary.



Note: To add columns to the Parameter Estimates report, right-click on the table and select the additional columns you want from the **Columns** pull-out menu. This the way you add CI’s for parameters as well.

**Example 10.2 – Percent Body Fat and All Body Measurements & Age**

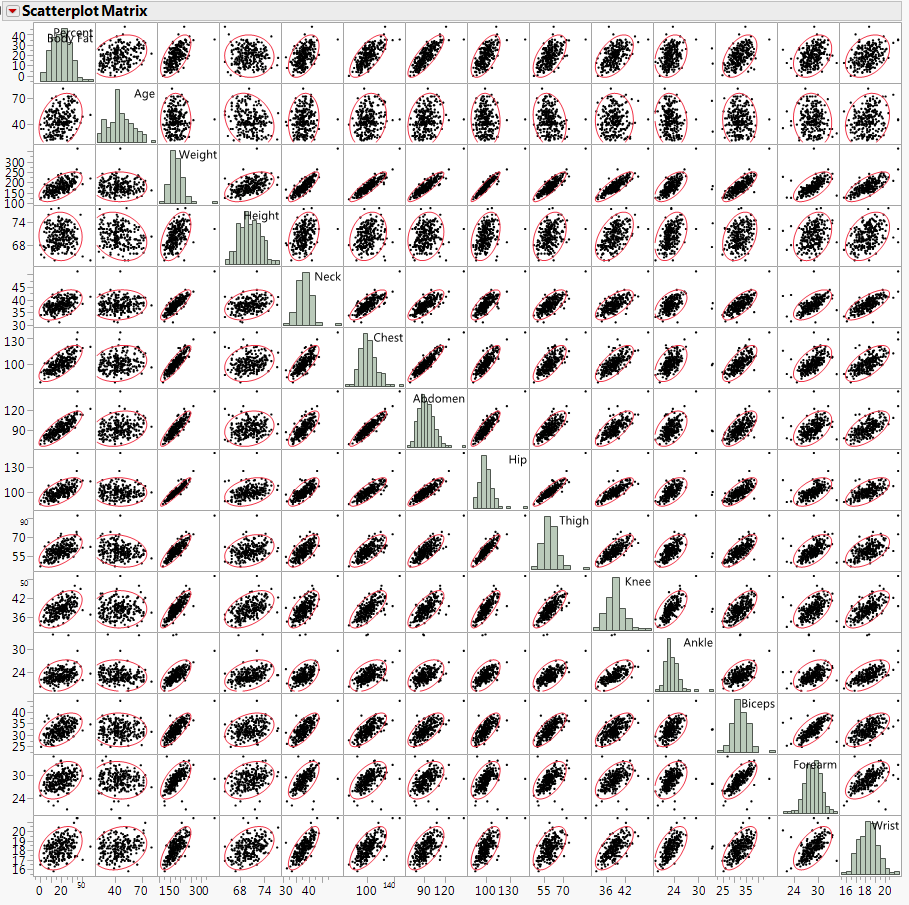
We now consider the multiple regression of percent body fat on all available body dimension measurements plus age (yrs.). We again begin by examining a scatterplot matrix and pairwise correlations between all of these variables (response & the potential predictors).

Correlation Matrix:



Which variable(s) are most correlated with the response?

Which variables have the weakest relationship with the response?



Would we characterize all of these variables as normally distributed?

Would we characterize the each of bivariate distributions as bivariate normal (BVN)?

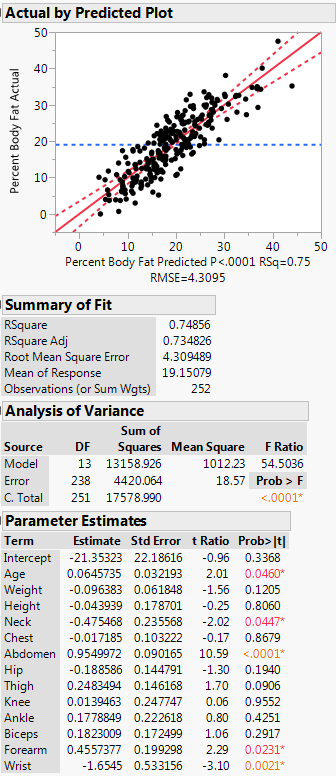
Do any of the relationships exhibited in the scatterplot matrix appear to be nonlinear?

Do the conditional variances appear to be constant?

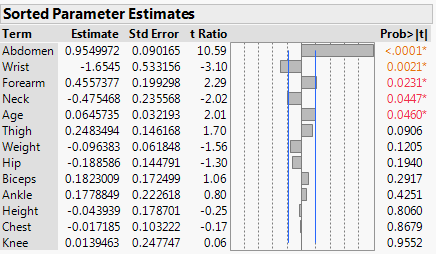
Do there appear to be any points/cases that could be problematic?

Let’s fit the full regression model including as the terms all the predictors themselves, i.e. using  as terms in our model.

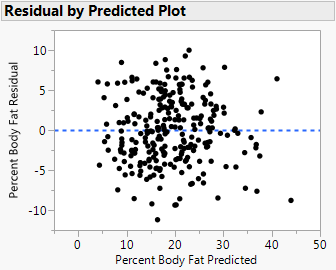
The summary of the full regression model is given below. Note this is summary from using **Effect Screening** as the ***Model Emphasis*** which JMP will default to when *p* is big.



Parameter Estimates sorted by t-statistic/p-value for testing NH: vs. AH:



Residual Plot:



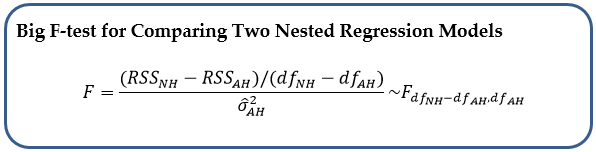
The Overall Regression test which testing if at least one the terms in our regression is useful, i.e.

is clearly significant (, thus something in our model is worthwhile, which is hardly surprising.

Clearly though, not all terms/predictors in our model are significant as several have p-values > .05 for testing:

A guiding principle in multiple regression is to fit the simplest model possible that adequately explains the variation in the response, i.e. we seek a ***parsimonious*** model. With that in mind we should consider dropping some insignificant terms from our model. The first term we would probably consider dropping from our model is \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_. We might try dropping several predictors/terms from our model at one time, but we need to cautious in doing this. When a single term is dropped from the multiple regression model the adjusted relationships for the remaining terms all change! Think of it this way, if we remove a predictor/term that is competing against another in the current model, it is possible the predictor left in the model could become highly significant. It can also be the case that when we drop a term from the model, another term that was significant may become insignificant in the absence of the dropped term.

To compare a “full” model to a “reduced” model with terms dropped from the “full” model we use the “Big F-test” which is shown below.



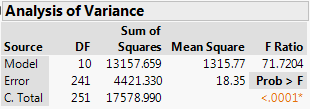
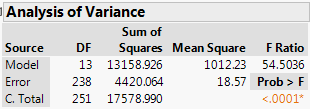
For example, from the parameter estimates above we can see that Knee & Chest circumferences and Height all have large p-values (p > 0.80 for all), thus we might consider dropping them all from model. Thus we would be testing the NH and AH hypotheses below:

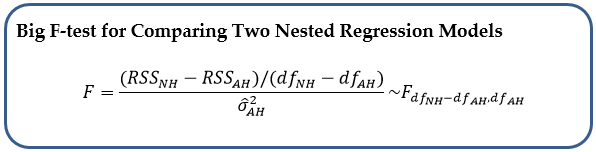
🡨 Model will all potential terms/predictors included.

More generically,

Below are the Analysis of Variance portion for Full and Reduced Models fit to these data.

Reduced Model ANOVA (NH model) Full Model ANOVA (AH model)

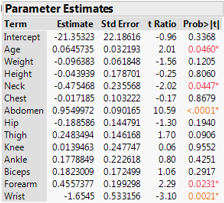
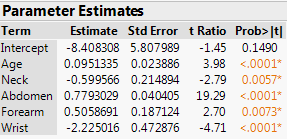


Thus the “Big F-test” statistic is

> pf(.023,3,238,lower.tail=F)

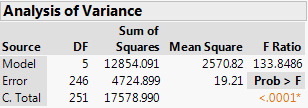
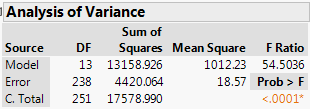
[1] 0.9952644

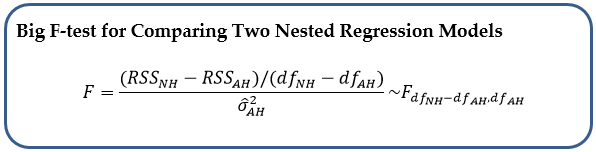
Clearly we fail to reject the NH and conclude the “full” model is not necessary. We can be more aggressive and consider dropping all terms/predictors whose p-value > .05. Thus we would consider dropping: Weight, Height, Chest, Hip, Thigh, Knee, Ankle, and Biceps. We conduct the test comparing the full model to this reduced model below.

Parameter Estimates Parameter Estimates  
with p-values from Full Model with p-values from Reduced Model  
 

Below are the Analysis of Variance portion for Full and Reduced Models fit to these data.

Reduced Model ANOVA (NH model) Full Model ANOVA (AH model)



Thus the “Big F-test” statistic is

> pf(2.052,8,238,lower.tail=F)

[1] 0.04134839

Here we reject NH and conclude the reduced model is NOT sufficient (p < .05). Thus at this point we might consider dropping a smaller subset or use some other logical scheme for reducing the number of terms in our regression model.

A cleaner representation of the test for comparing a “full”/larger model to a reduced model obtained by dropping terms from the larger model is as follows:

Let

Note: the NH model is obtained by setting all the parameters associated with the collection of terms in to zero, i.e. **.**

In the last example,

Age,Neck,Abdomen,Forearm,Wrist)

(Weight, Height, Chest, Hip, Thigh, Knee, Ankle, Biceps)

and in that example we found we could not drop all of the terms in without significantly degrading the model.

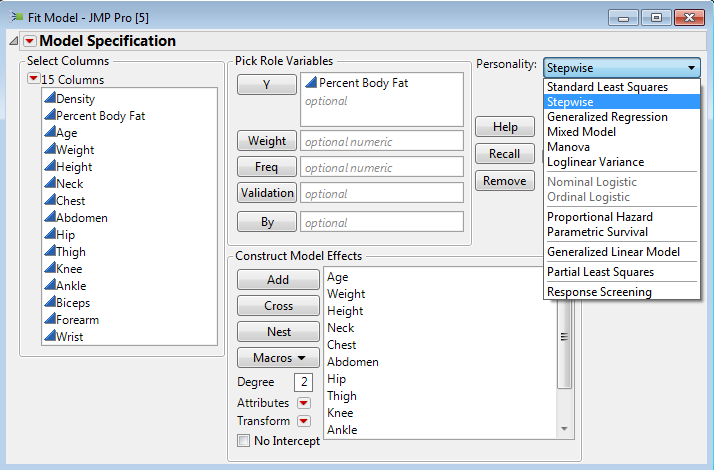
**Backward Elimination**

One method that is particularly easy to implement when trying to reduce the number of terms a regression model is called ***“Backward Elimination”***. In backward elimination we start with a “full” model and drop one term at a time until no further terms can be dropped without significantly degrading the model according to some criteria. We will discuss model selection criteria in much more detail in Section 17 of the notes.

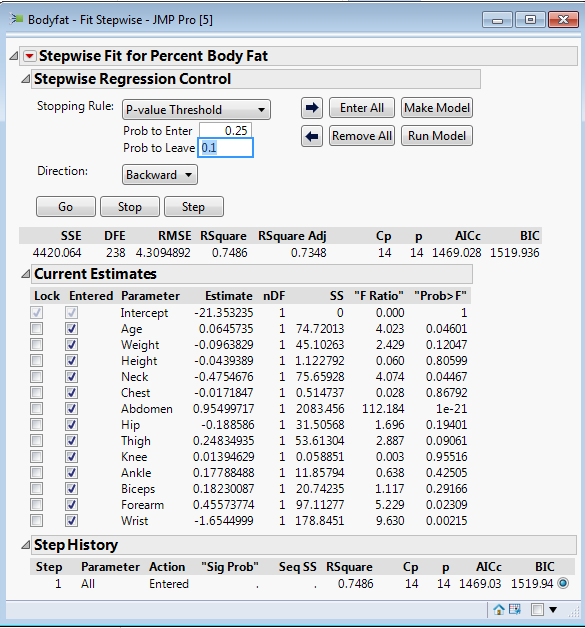
For now we will use Backward Elimination as follows:

1. Fit a “full” model.
2. Drop the term from the model with the largest p-value.
3. Fit the model with term dropped, this becomes the new “full model”.
4. Repeat Steps 1 -3 until no further terms can be dropped. We generally are more generous with the p-value for significance in model building, so we choose to only a drop term at a given step if the associated p-value > *p-value threshold.*Typically *p-value threshold = 0.10* (or possibly higher) rather than the usual .05.

Performing Backward Elimination (or other stepwise selection method) is easy in JMP. First select **Analyze > Fit Model** specifying the response Y with all potential terms in the **Construct Model Effects** box and changing **Personality** to **Stepwise** as shown below.



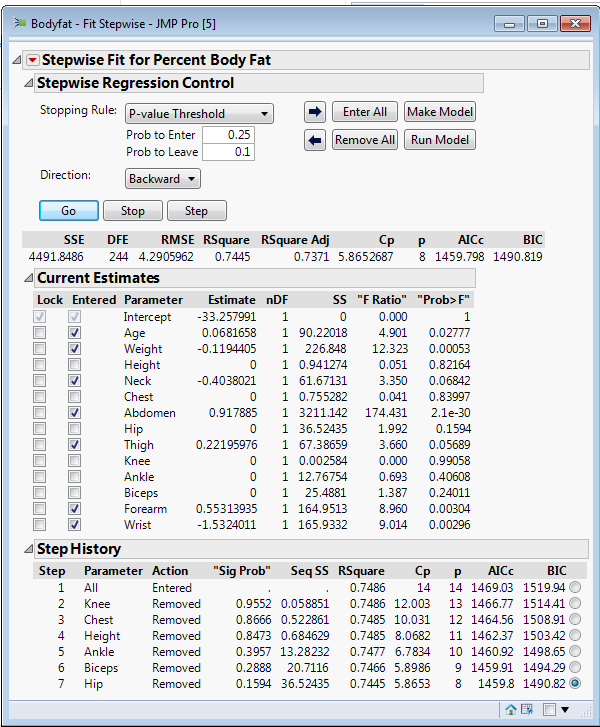
We then set-up the **Stepwise** fitting platform for Backward Elimination using P-value Threshold as follows:



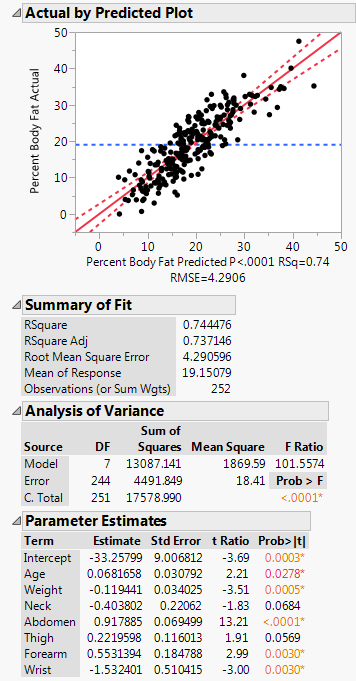
**Comments:**

The result of performing Backward Elimination starting with the full model including all available predictors as terms is shown on the next page.

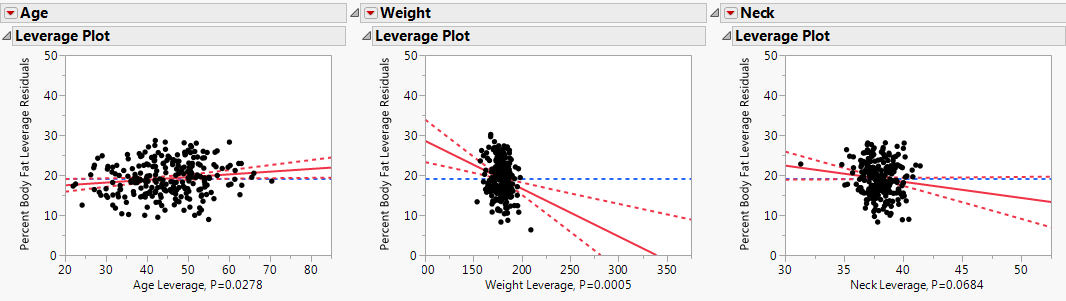
Model chosen via Backward Elimination using p-value threshold = .10.

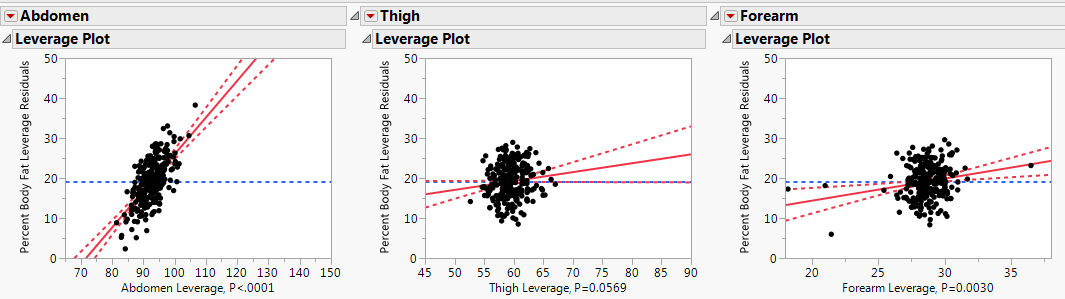


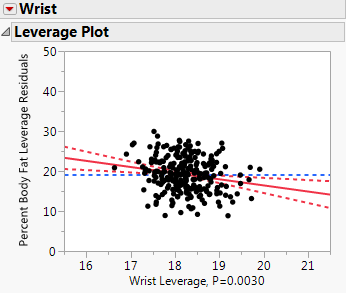
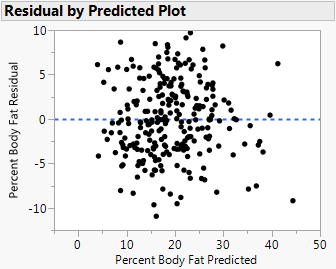
We can then select **Make** or **Run Model** to fit the backward selected model. **Make Model** sets up the dialog box for **Fit Model** with the selected terms and will give the complete multiple regression output once you run it. The regression summary for the reduced model is shown below. Which variable has the strongest adjusted relationship with % body fat?



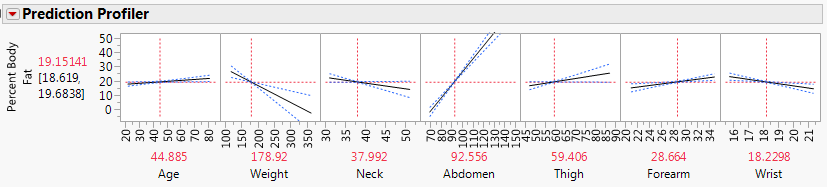
Do the residuals suggest any model violations?





The prediction profiler for this model is shown below with each term/predictor set equal to its sample mean.

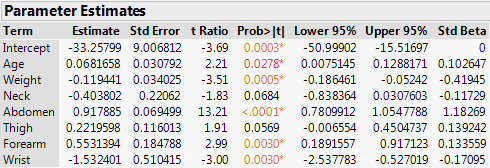


The negative coefficients for weight, neck circumference, and wrist circumference seem counterintuitive as these variables are individually positively associated with the response.

To interpret any one term/predictor in our model we need to be careful! Independently changing one predictors value is not in theory possible as the predictors themselves are almost always related. For example we could not necessarily consider increasing a person’s weight by say 10 lbs., without also expecting a change in their other body measurements as well to go along with the increased weight.

Technically the parameter estimates tell us the change in the mean response associated with a 1-unit increase in the associated predictor while HOLDING ALL PREDICTORS FIXED.

Parameter Estimates from the Backward Elimination Model



**Interpretation of**

Holding all other body measurements constant/fixed a 1 cm increase in the abdominal circumference corresponds to a .918 unit increase in the percent body fat or a percent body fat increase between .781 and 1.055 percentage points with 95% confidence.

**Interpretation of**

Holding all other body measurements constant, a 1 cm increase in wrist circumference is associated with a 1.53 decrease in the percent body fat or a decrease between 2.538 and .527 percentage points with 95% confidence.

If we use the standardized beta instead the interpretation would be; if we increase wrist circumference by 1 standard deviation the percent body fat deceases by .171 standard deviations.

Note: A one standard deviation increase in a term value while holding terms constant may not be even remotely realistic!

Other coefficients could be interpreted in a similar fashion.

**Example 10.3 – Auction Sales of Grandfather Clocks**  
Consider the following data regarding the price of Grandfather clocks bought at auction. The age of the clock will certainly affect the price and the number of bidder bidding on the clock is known to influence the price as well.

|  |  |
| --- | --- |
| Data Table in JMP | A grandfather clock  http://www.1-800-4clocks.com/miva/graphics/full/R2232.jpg |

Multiple Linear Regression Setup

* Response Variable : Sale Price
* Predictors: Age and Number of Bidders
* Assumed Model using the predictorsas the model terms

* Matrix Representation

|  |  |
| --- | --- |
| Response Vector: | Model Terms:  matrix |

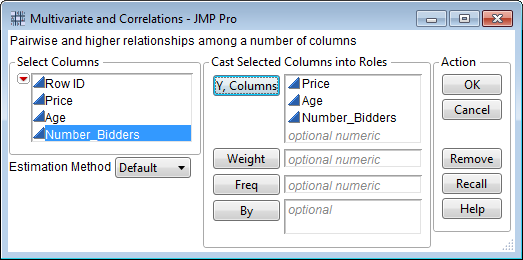
Considering each Predictor Individually

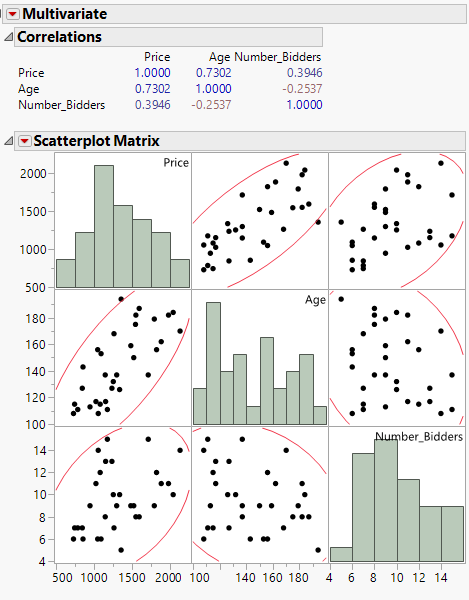
|  |  |
| --- | --- |
|  |  |

Questions

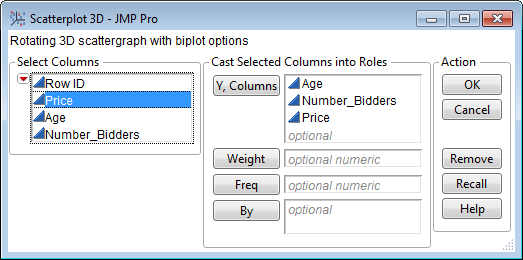
1. What is the impact of Age on Price? Is age a useful predictor of price? Discuss.
2. What is the impact of number of bidders on Price? Is the number of bidders a useful predictor of price? Discuss.

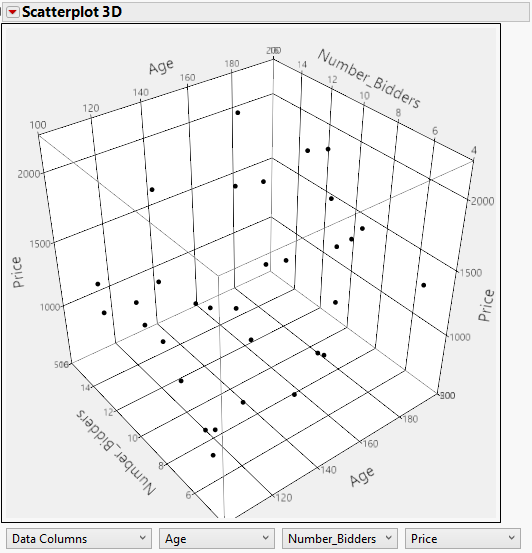
Scatterplot Matrix: Again and important first step in conducting a multiple regression analysis is to construct a scatterplot matrix. In JMP this is done by selecting **Analyze > Multivariate Methods** and putting the response followed by the predictors/terms in the **Y, Columns** box.





3-D Scatterplot:   
To construct a three dimensional scatterplot in JMP, select **Graph > Scatterplot 3D**





|  |  |
| --- | --- |
| Profile of Age | Profile of Number of Bidders |

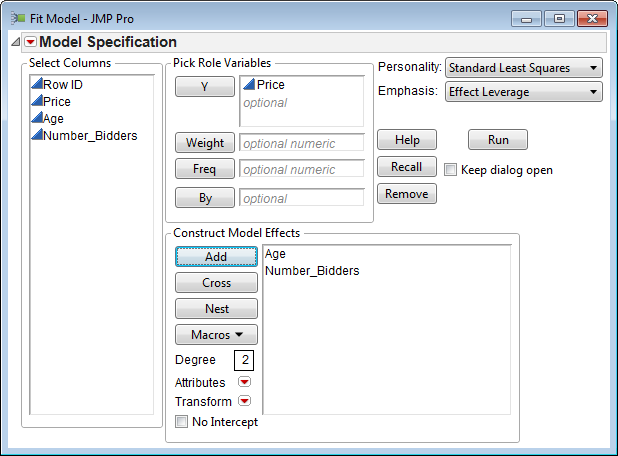
Fitted Plane:



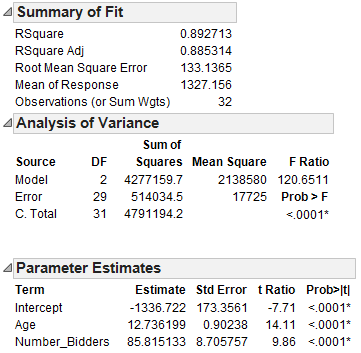
Recall, Multiple Linear Regression Setup:

* Response Variable : Price
* Predictor Variable: Age and Number of Bidders
* *Model:*

To fit this model in JMP, select **Analyze > Fit Model**.



Regression Output from JMP:



Comparing the Models:

|  |  |  |
| --- | --- | --- |
| **Model** |  |  |
|  |  |  |
|  |  |  |
|  |  |  |

Understanding the Regression Output:

|  |  |  |
| --- | --- | --- |
| Quantity | Interpretation | |
|  |  | |
|  |  | |
| Estimated Mean Function: | | |
| : |  | |
| : | |  |  |  |  | | --- | --- | --- | --- | | Case | Age | Number of Bidders | Prediction | | 1 |  |  |  | | 2 |  |  |  | | |
| : | |  |  |  |  | | --- | --- | --- | --- | | Case | Age | Number of Bidders | Prediction | | 1 |  |  |  | | 2 |  |  |  | | |
| Is it reasonable to hold one predictor constant while changing the other? | |

|  |  |
| --- | --- |
| Tests of Significance | |
| Overall Test |  |
| Test for Y-Intercept |  |
| Test for Slope for Age |  |
| Test for Slope for Num\_Bidders |  |

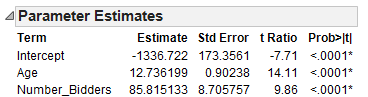
**Comments**:

Caution: Caution should be used when investigation the individual importance of a predictor in a multiple linear regression model.

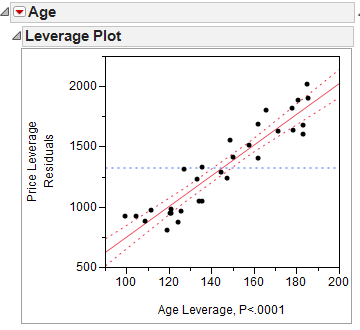
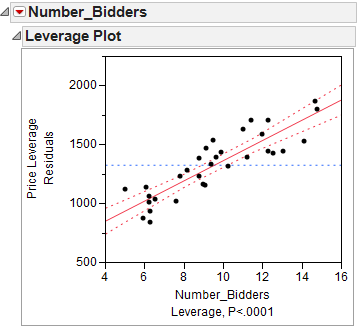
Rationale: If a predictor variable is not significant, it doesn’t necessarily mean that it is not related to (i.e. impact) the response—it could simply mean that the information has already been explained by other predictor variables. To overcome this problem, ***partial correlations*** can be used instead of simple correlations. These partial correlations are displayed graphically in an added variable plot.

**Partial Correlations and Added Variable Plots (AVPs):**

A review of the added variable plots for the Grandfather Clock example implies that both Age and Number of Bidders has something to contribute to our understanding of Price.



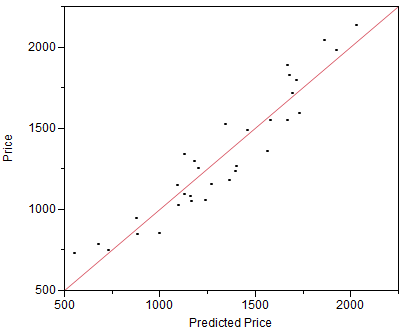
Added variable plot for Age Added variable plot for Number of Bidders

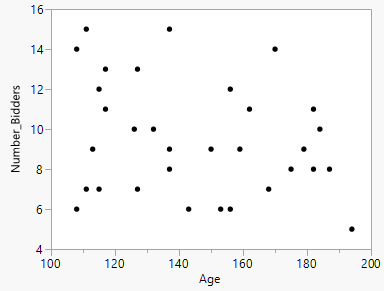
**Concept of Partial Correlations**: As previously discussed there exists a simple relationship between Pearson’s Correlation and R2 for simple linear regression, i.e. . The in the added variable plot visually communicates the true effect of a particular predictor variable on the response void of any potential effect of the other predictor variables. A **partial correlation coefficient** can accomplish the same thorough a correlation type measurement.

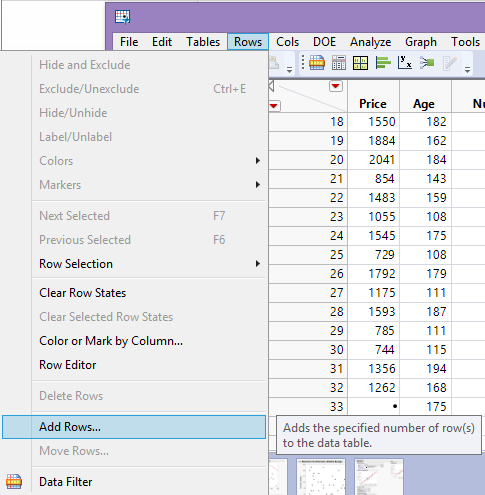
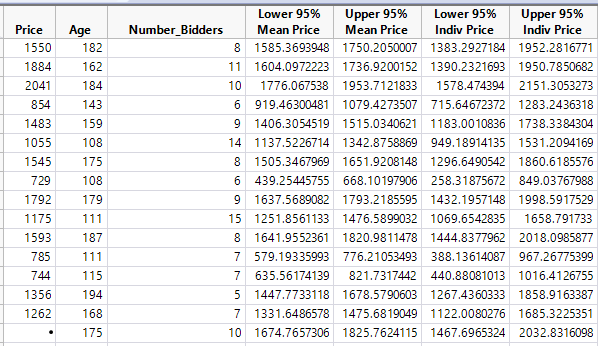
|  |  |
| --- | --- |
| In JMP, **Analyze > Multivariate Methods > Multivariate** enter the response and predictor variables, then select **Partial Correlations** from red drop down arrow in the upper-left corner. | The added variable plot for Age from the Grandfather clock example. |

How well is our model doing?



Making a Prediction for a New Observation: **Rows > Add Rows**, after saving **Mean & Indiv Confidence Formulae**Obtain an estimated for the average price of a grandfather clock that is Age = 175 and Number of Bidders = 10.

  
Use this plot to determine which combinations of Age and Number of Bidders are valid.

Prediction by hand: