Below is a sample data set that we will be using for today's exercise. It lists the heights & weights for 10 men and 12 women.

Male	Height (in)	69	70	65	72	76	70	70	66	68	73	
	Weight (lb)	192	148	140	190	248	197	170	137	160	185	

Female	Height (in)	65	61	67	65	70	62	63	60	66	66	65	64
	Weight (lb)	110	105	136	135	187	125	147	118	128	175	147	120

Entering the data
If you haven't entered the data check the Minitab introductory tutorial on the proper method. When you are finished you should have 3 columns in your data view spreadsheet, a sample of which is shown below.

↓	C1 Height	C2 Weight	C3 Gender
1	69	192	1
2	70	148	1
3	65	140	1
4	72	190	1
5	76	248	1
6	70	197	1
7	70	170	1
8	66	137	1
9	68	160	1
10	73	185	1
11	65	110	2

Creating a scatterplot
Before any type of regression analysis is begun a simple scatterplot of the data should be created. The reasoning for this is twofold. The first and most important is to verify the quality of your data. Many times unusual points and outliers can be identified easily on the plots. Additionally, we can check to see if the assumptions for linear regression seem valid model for the data. If there appears to be curvature of the data or non-homogenous variance we might not be able to use simple linear regression.

By clicking on the "Graph" and then "Scatterplot…" button, the scatterplot window will be opened. Choose the "With Groups" to graph males and females separately, and click ok.
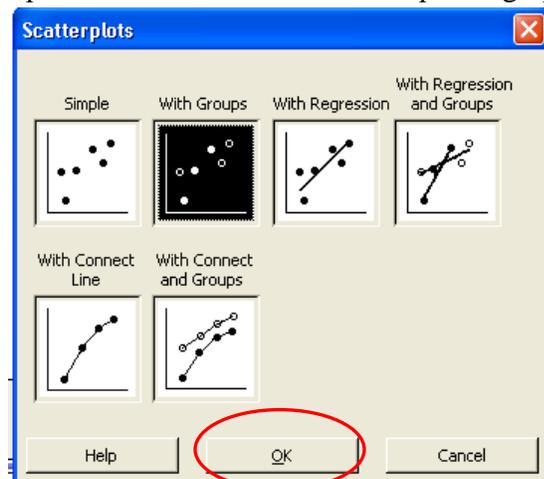
We then enter the columns into the appropriate areas and the graph will appear in a new window. It would also be beneficial to add a title to the graph. This can be done by clicking on "Labels…" in the Scatterplot-With Groups window.

As the default setting for Minitab is to use color and shapes to differentiate the groups [and most prints are in black and white]. You should immediately edit the graph if you have the same shape within the graph. To do this, double click on the graph to bring it into edit mode. Click <u>once</u> on a male symbol on the graph (all symbols will be highlighted). Click <u>once</u> on a male symbol again (male symbols will be highlighted).Double click a male symbol and the properties window should appear. Choose a symbol and hit apply. The male symbol should change. Now close the window and the graph is done. [Note if you click a third time on a symbol before double clicking you can change the symbol for a single point]
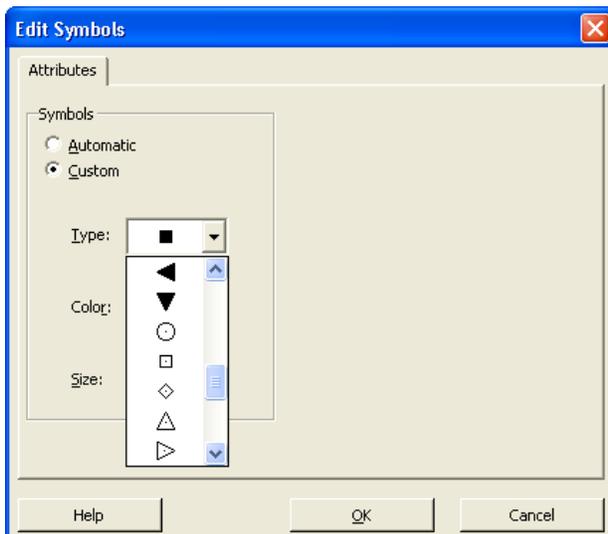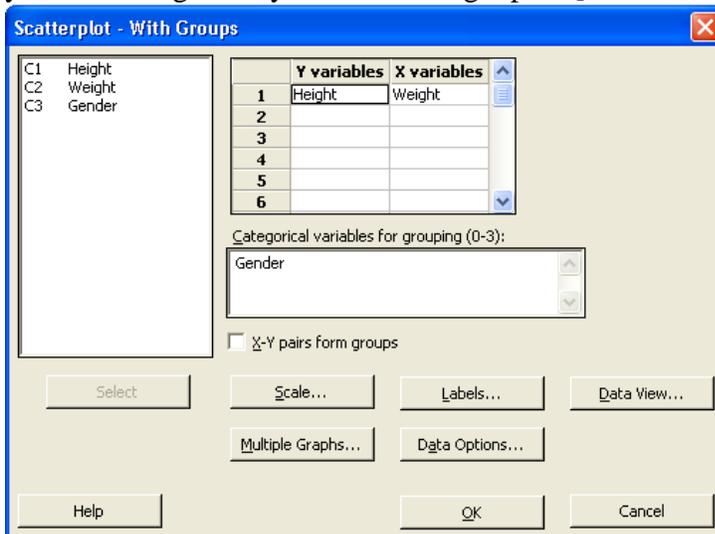
You will now be able to differentiate the groups if you do a plot in black and white.
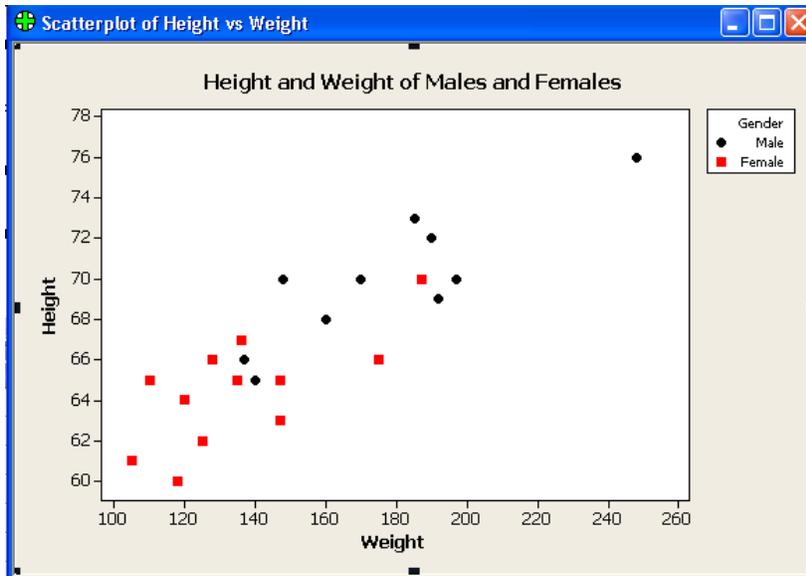
(Note: to change the title or labels just double click the label and an edit window will appear. The font, font size, color, etc. can also be changed.)

Seeing no problems with the data we can now run the regression for weight versus height. We select **Stat-Regression-Regression…** from the pull-down menu.
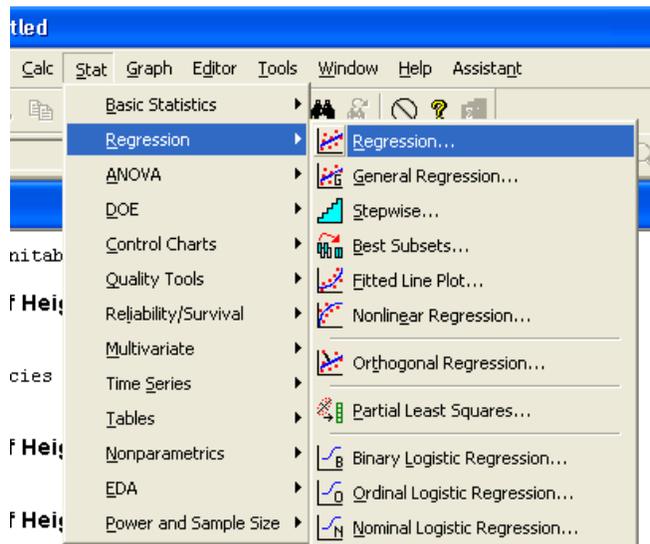
Placing the variable we would like to predict, weight, in the dependent variable and the variable we will use for prediction, height, in the independent variable, we hit OK.
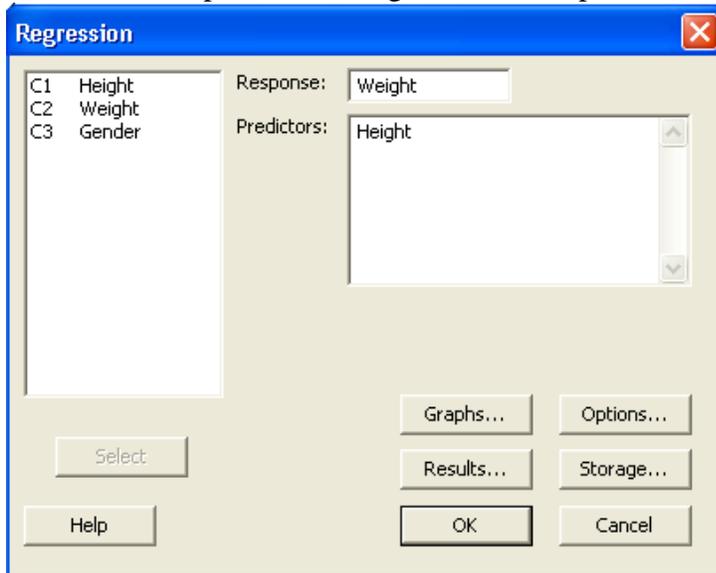
This part of the Minitab output gives R^2, which indicates how much of the variation in the response variable Y, is explained by the fitted regression line. We can see that there is a strong relationship between the 2 variables (75% of the variation in y is explained by the regression line), indicating if I know your height I should be able to make some prediction about your weight.

```
S = 17.7596    R-Sq = 75.6%    R-Sq(adj) = 74.3%
```

The next part of the output is the statistical analysis (ANOVA-analysis of variance) for the regression model. The ANOVA represents a hypothesis test with where the null hypothesis is

$H_o : \beta_i = 0$ for all i   (i = 1 to k)
$H_A : \beta_i \neq 0$ for at least 1 coefficient

The ANOVA table uses an F-statistic to check the hypothesis. It similar to the Z & T tests, in that large values of F indicate a rare test score (unusual data) under the null hypothesis and indicate that it is unlikely the null hypothesis is true. The significance level (or p-value) for the test is less than 0.05, so we would reject the null hypothesis and conclude that at least one coefficient is not zero (there is a significant linear relationship between weight and height).

```
Analysis of Variance

Source          DF     SS      MS      F      P
Regression       1  19503   19503  61.84  0.000
Residual Error  20   6308     315
Total           21  25811
```

The coefficient table contains the coefficients for the least square (fitted) line and other relative information about the coefficients. In the column B, the constant represents the y-intercept and the Height represents our slope. Minitab also generates the regression equation automatically.

```
The regression equation is
Weight = - 355 + 7.61 Height


Predictor      Coef   SE Coef       T        P
Constant    -354.84     64.89   -5.47    0.000
Height       7.6080    0.9675    7.86    0.000
```

A review of the table also indicates several other statistical tests that Minitab is performing. You'll note that Minitab tests both of the coefficients to see if they are equal to zero with t-tests. We can see that both of the coefficients are significantly different from zero.

$$H_o : \beta_1 = 0 \quad H_A : \beta_1 \neq 0$$

$$H_o : \beta_o = 0 \quad H_A : \beta_o \neq 0$$

Creating a fitted line plot
After you have created the scatterplot of the data, you can add a regression line. When the scatterplot is open go to the pull-down menu **Editor-Add-Regression Fit…**
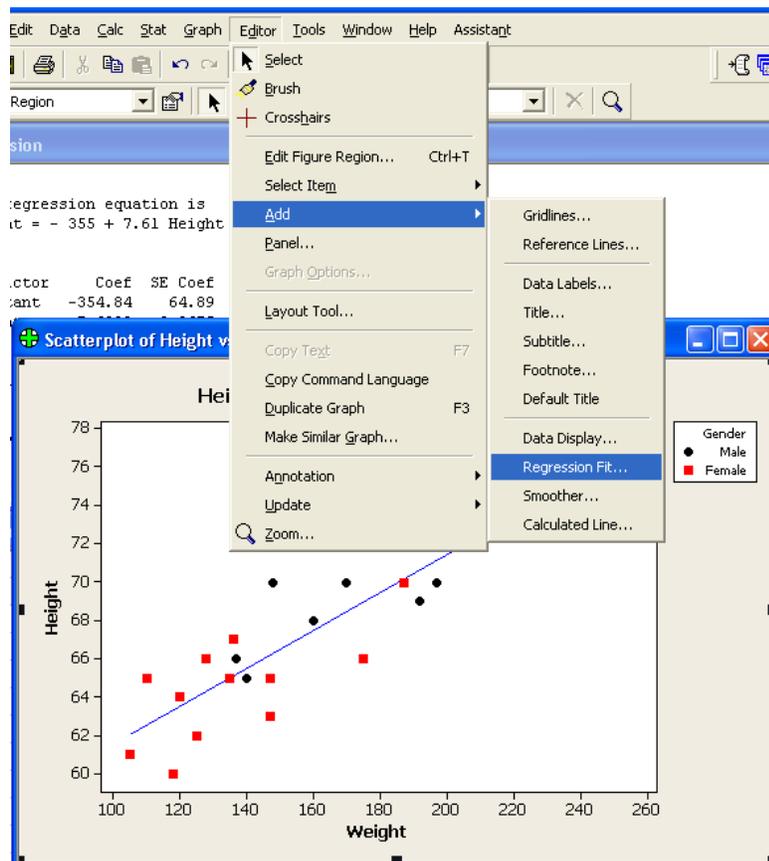
This brings up the **Add Regression Fit** window. Under Model Order choose the "Linear" option (you can leave the "Fit intercept" box checked). Click OK.
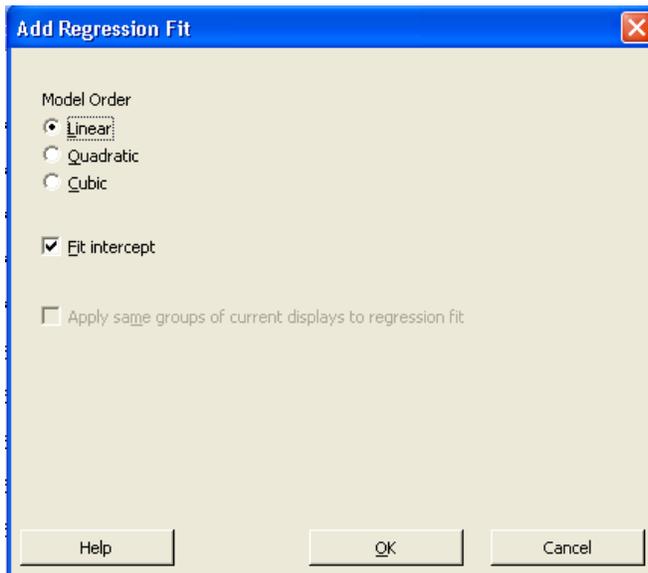
If you wish it is also possible to fit a line to each of the groups on a graph. The easiest way to do this is to make a new scatterplot by going to the pull-down menu **Graph-Scatterplot…** then choosing the "With Regression and Groups" option. Fill in the X-variable, Y-variable, and Categorical variables the same way you did before for making the first scatterplot.

Creating confidence intervals on the regression coefficients
Conduct the regression as before but select **Stat-Regression-General Regression…** from the pull-down menu. In the **General Regression** window select the "Results" button at the bottom. This will bring up the **General Regression-Results** window. Check the "Display confidence intervals" box under the Coefficient table section. Click OK. Then run the test.
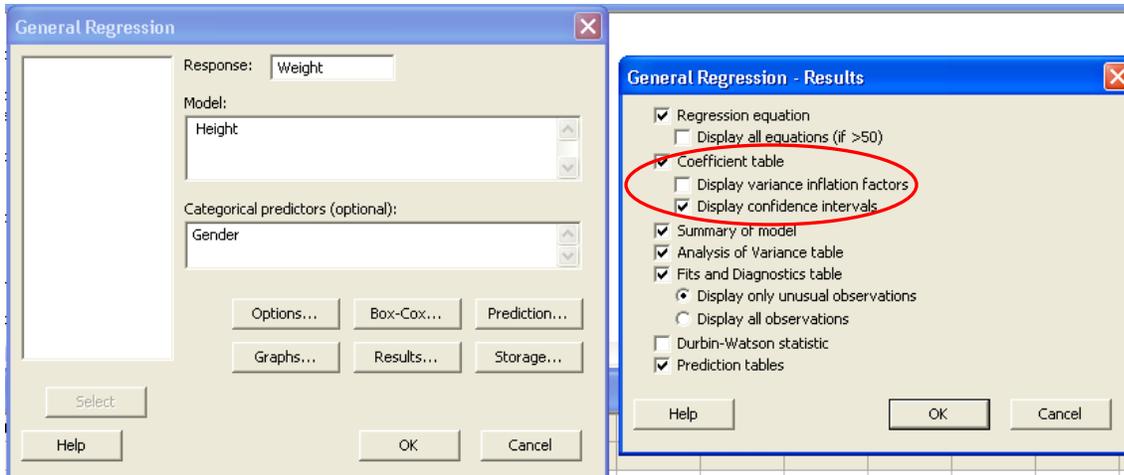
The following table will be created in your output:

```
Coefficients

Term          Coef   SE Coef         T      P          95% CI
Constant  -354.844   64.8884  -5.46852  0.000  (-490.199, -219.489)
Height       7.608    0.9675   7.86363  0.000  (   5.590,     9.626)
```

Creating fitted values with confidence & prediction intervals

Conduct the regression as before by selecting **Stat-Regression-Regression…** from the pull-down menu. In the **Regression** window select the "Options" button at the bottom. This will bring up the **Regression-Options** window. Under "Prediction intervals for new observations:" put the variable you are interested in predicting. Then check the "Confidence limits" box and click OK. Click OK to run the test, a new part of the output will look like the image below, with CI and PI for each observation given based on the variable you were interested in.
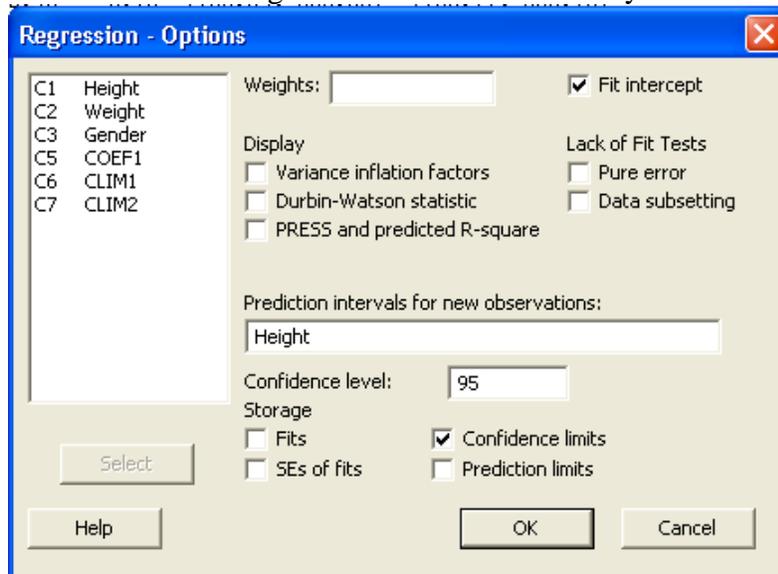

```
Predicted Values for New Observations

New Obs     Fit   SE Fit        95% CI              95% PI
      1  170.11     4.27  (161.20, 179.02)  (132.00, 208.21)
      2  177.72     4.80  (167.71, 187.72)  (139.34, 216.09)
      3  139.68     4.23  (130.85, 148.50)  (101.59, 177.76)
      4  192.93     6.18  (180.04, 205.82)  (153.71, 232.15)
      5  223.36     9.54  (203.47, 243.25)  (181.32, 265.41)X
      6  177.72     4.80  (167.71, 187.72)  (139.34, 216.09)
      7  177.72     4.80  (167.71, 187.72)  (139.34, 216.09)
      8  147.28     3.90  (139.15, 155.41)  (109.36, 185.21)
      9  162.50     3.92  (154.32, 170.67)  (124.56, 200.44)
     10  200.54     6.97  (186.01, 215.07)  (160.74, 240.33)
     11  139.68     4.23  (130.85, 148.50)  (101.59, 177.76)
```