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 SPSS 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.

	height	weight	gender
1	69	192	1
2	70	148	1
3	65	140	1
4	72	190	1
5	78	248	1
6	70	197	1
7	70	170	1
8	66	137	1
9	68	160	1
10	75	185	1
11	66	110	2
12	61	105	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 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 scatterplot button, the scatterplot window will be opened. Choose the simple scatterplot and click on define.
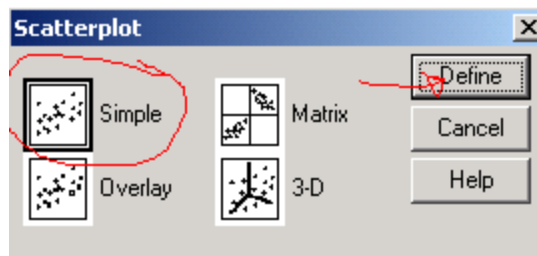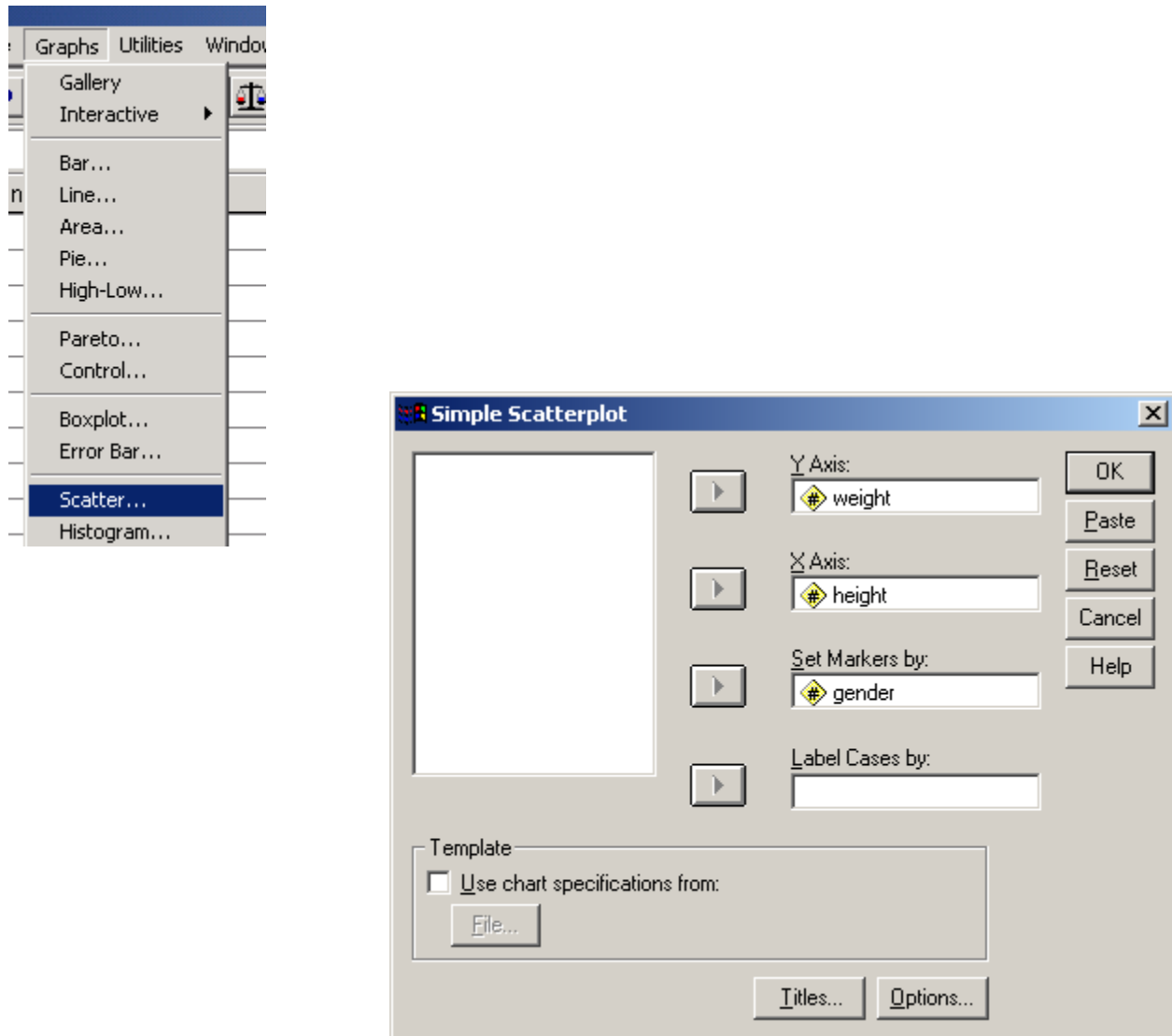
We then enter the columns into the appropriate areas and the graph will appear in the output window. It would also be beneficial to add a title to the graph. This can be done by clicking on **Chart -Title** in the pull-down menu.

As the default setting for SPSS is to use color to differentiate the groups [and most prints are in black and white] you should immediately edit the graph and change the symbols for the groups. To do this, double click on the graph to bring it into edit mode. Click once on a male symbol on the graph (all symbols will be highlighted). Click once 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]
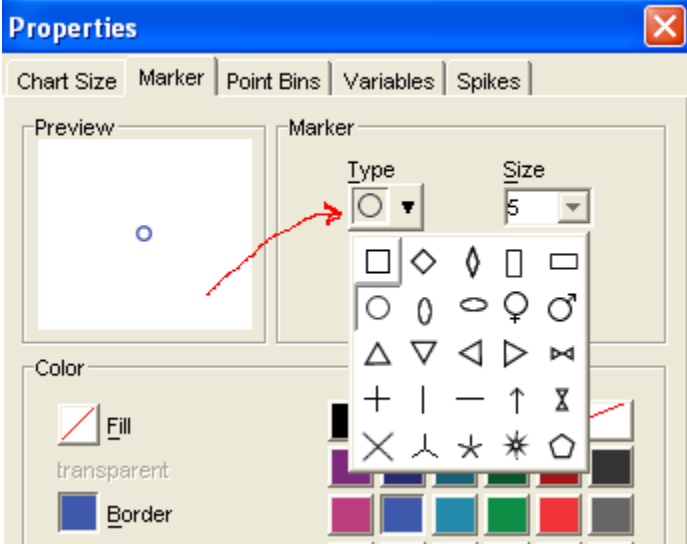
You will now be able to differentiate the groups if you do a plot in black and white.
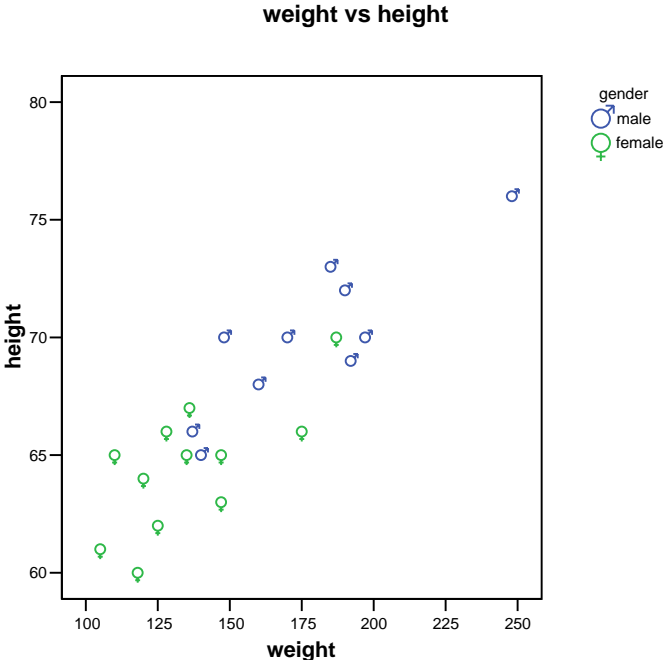
weight vs height

Seeing no problems with the data we can now run the regression for weight versus height. We select **Analyze-Regression-Linear** from the pull-down menu.
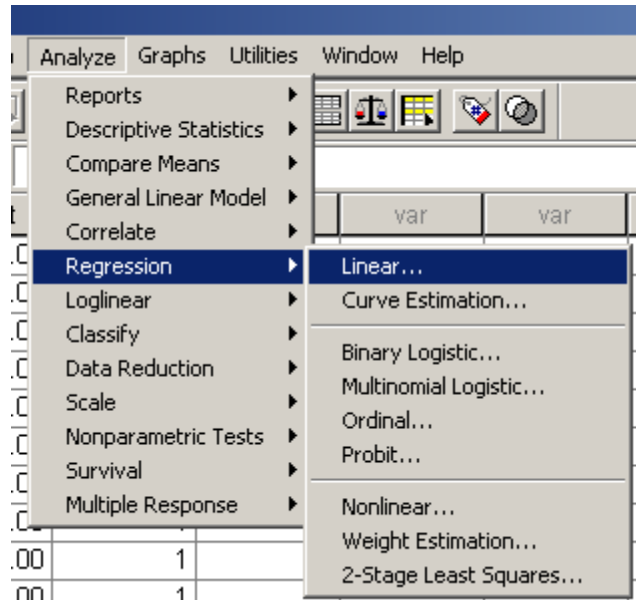
Placing the variable we would like to predict, weight, in the dependent variable and the variable we will use for prediction, height, in the independent variable, we hit OK.
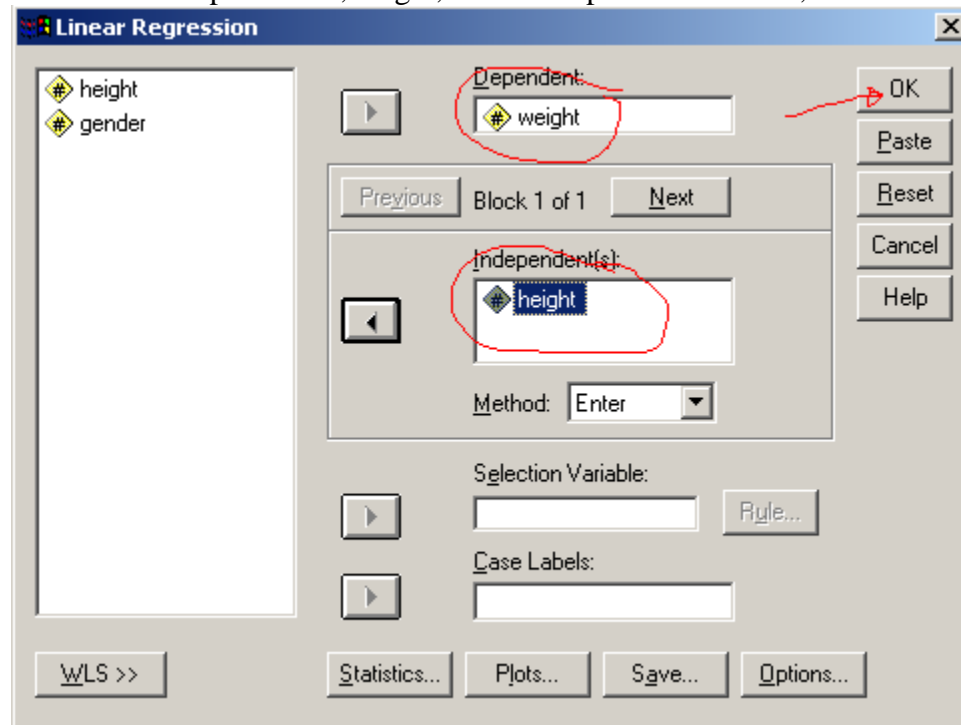
This part of the SPSS output gives the correlation, r, for the regression. This represents the strength of the linear relationship between weight and height. It also gives R^2, which indicates how much of the variation in the response variable Y, is explained by the fitted regression line. We can see 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.

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.869[a]	.756	.743	17.7596

a. Predictors: (Constant), HEIGHT

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 that we have done in the past, 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 none zero (there is a significant linear relationship between weight and height).

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	19503.418	1	19503.418	61.837	.000[a]
	Residual	6308.037	20	315.402		
	Total	25811.455	21			

a. Predictors: (Constant), HEIGHT

b. Dependent Variable: WEIGHT

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. The equation of the line found from the output is

$$\hat{w} = -354.844 + 7.608(height)$$

Coefficients[a]

| Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	-354.844	64.888		-5.469	.000
	HEIGHT	7.608	.967	.869	7.864	.000

a. Dependent Variable: WEIGHT

A review of the table also indicates several other statistical tests that SPSS is performing. You'll note that SPSS test 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.

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

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

Creating a fitted line plot
After you have created the scatterplot of the data, double click on the graph as if you were editing it. When the edit window comes up, double click once on the data to highlight them. Then click on the fitted line symbol in the tool box.
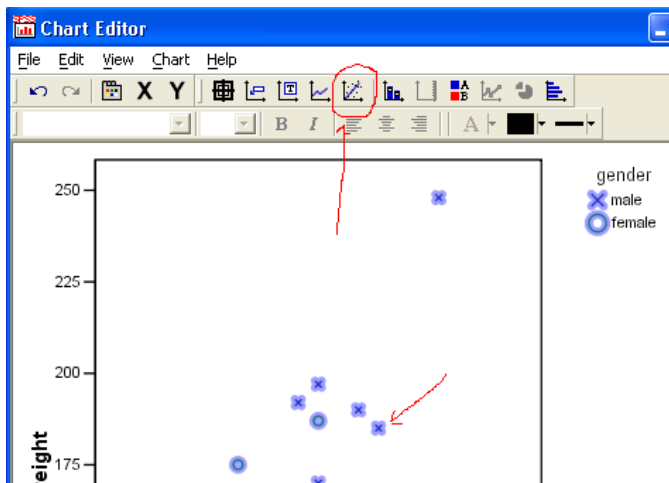

This brings up the **Properties** window. Click on the "Fit Line" tab and choose the Linear option. A line will now appear on the plot.
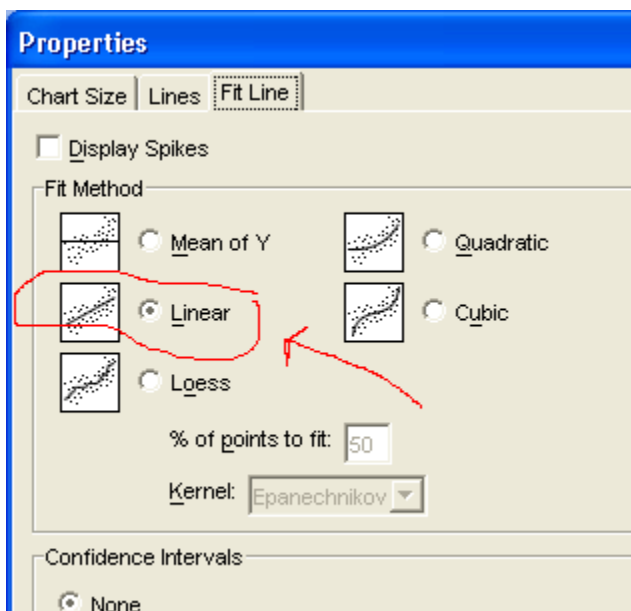
If you wish it is also possible to fit a line to each of the groups on a graph. To do this simple click on the **sub-groups** box within the **scatterplot options** screen shown above.

Creating Confidence Intervals and Prediction Intervals on your graph
The predicted value from a regression equation is a point estimate for the mean value of Y for that particular X [in other words its our best guess of the average value of Y for that value of X]. Interval estimate provides an idea of its accuracy. We commonly add confidence intervals [which indicate reasonable values for the average Y at an X] and prediction intervals [which indicate reasonable values for an individual Y at an X] to our estimates. To get these on the graph simply follow the instructions above for adding a fitted line. When the **Properties** window appears click on the appropriate checkboxes found on the bottom of the window. If you wish to put both the CI and PI on the same plot you must apply, then close the window, and then start again as if adding a new line. Note you can change the 95% interval if you so desire.
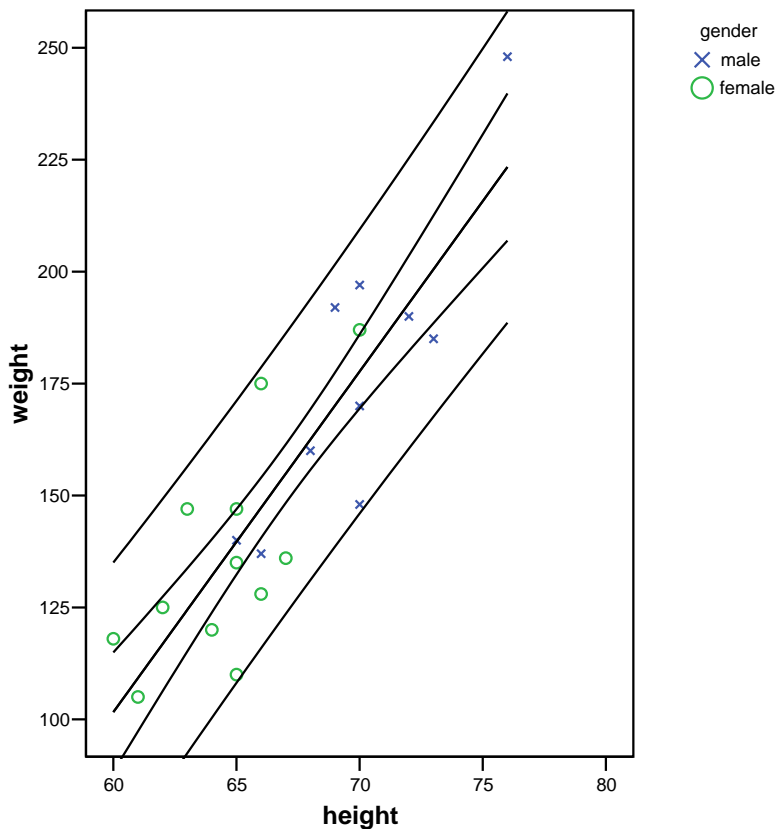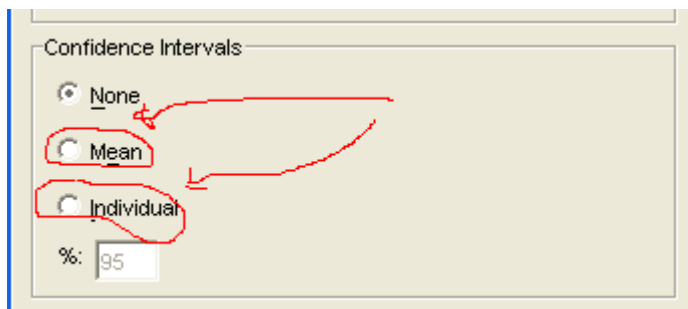
Creating confidence intervals on the regression coefficients

Conduct the regression as before by selecting **Analyze-Regression-Linear** from the pull-down menu. In the **Linear Regression** window select the statistics button at the bottom. This will bring up the **Linear Regression: Statistics** window. Check the box for confidence intervals and hit continue. This will create a new output in the output screen which contains the CI along with the hypothesis tests.
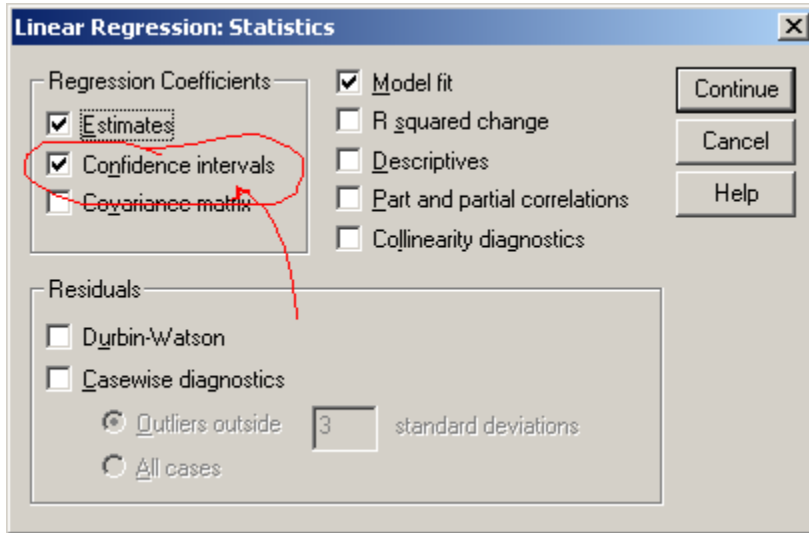
Coefficients^a

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-354.844	64.888		-5.469	.000	-490.199	-219.489
	HEIGHT	7.608	.967	.869	7.864	.000	5.590	9.626

a. Dependent Variable: WEIGHT

Creating fitted values with confidence & prediction intervals

Conduct the regression as before by selecting **Analyze-Regression-Linear** from the pull-down menu. In the **Linear Regression** window select the "SAVE" button at the bottom. This will bring up the **Linear Regression: Save** window. Note SPSS offers you a prediction interval on a mean (what we call a confidence interval) and a prediction interval on an individual (what we call a prediction interval). The predicted values along with the respective CI & PI's can be found on the data view spreadsheet.