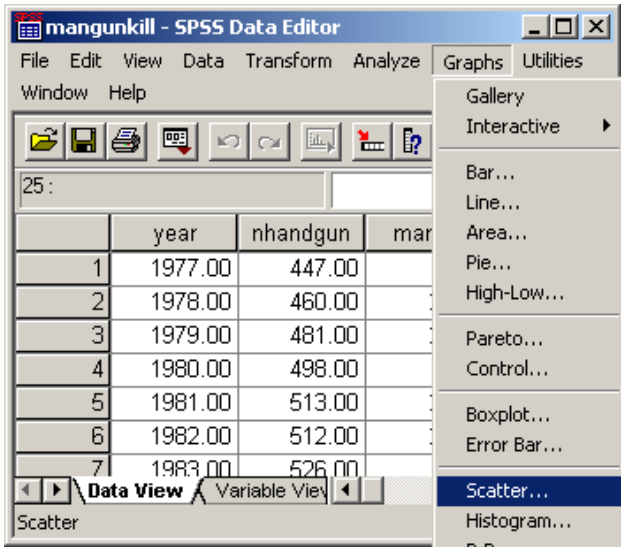
Linear Regression in SPSS

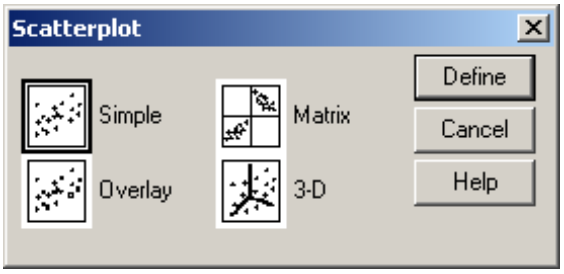
Data: mangunkill.sav
Goals:

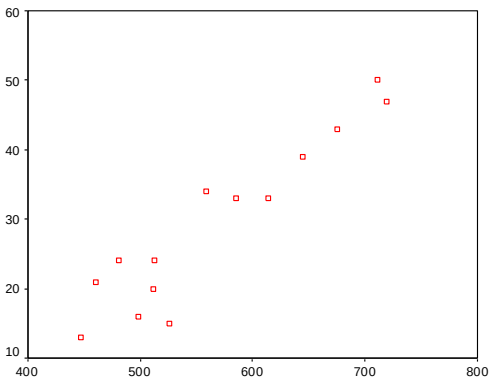
- Examine relation between number of handguns registered (**nhandgun**) and number of man killed (**mankill**)
- Model checking
- Predict number of man killed using number of handguns registered

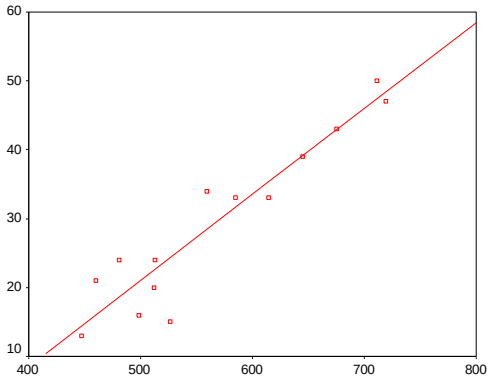
I. View the Data with a Scatter Plot

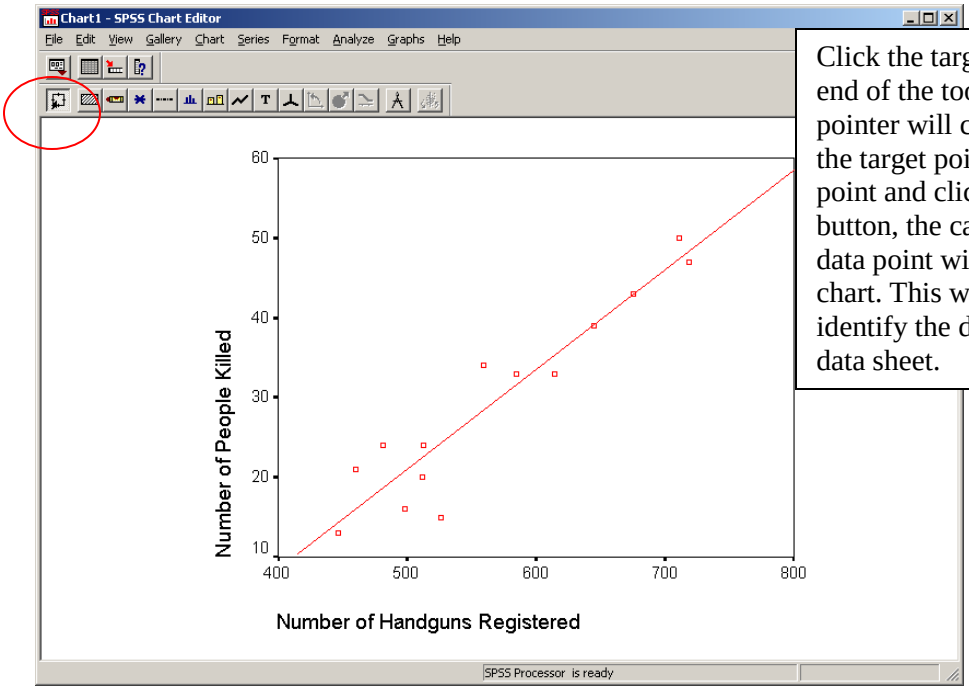
To create a scatter plot, click through **Graphs\Scatter\Simple\Define**. A Scatterplot dialog box will appear. Click **Simple** option and then click **Define** button. The Simple Scatterplot dialog box will appear. Click **mankill** to the Y-axis box and click **nhandgun** to the **x-axis** box. Then hit **OK**. To see fit line, double click on the scatter plot, click on **Chart\Options**, then check the **Total box** in the box labeled **Fit Line** in the upper right corner.











Click the target button on the left end of the tool bar, the mouse pointer will change shape. Move the target pointer to the data point and click the left mouse button, the case number of the data point will appear on the chart. This will help you to identify the data point in your data sheet.

II. Regression Analysis

To perform the regression, click on **Analyze\Regression\Linear**. Place **nhandgun** in the **Dependent** box and place **mankill** in the **Independent** box. To obtain the 95% confidence interval for the slope, click on the **Statistics** button at the bottom and then put a check in the box for **Confidence Intervals**. Hit **Continue** and then hit **OK**.

The independent variable (**nhandgun**) is said to be useful in predicting the dependent variable (**mankill**) when the level of significance (P-value labeled with **Sig.** on the Output) is below 0.05.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.941[a]	.886	.877	4.2764

a. Predictors: (Constant), Number of Handguns Registered

b. Dependent Variable: Number of People Killed

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	1711.979	1	1711.979	93.615	.000[a]
	Residual	219.450	12	18.287		
	Total	1931.429	13			

a. Predictors: (Constant), Number of Handguns Registered

b. Dependent Variable: Number of People Killed

Coefficients[a]

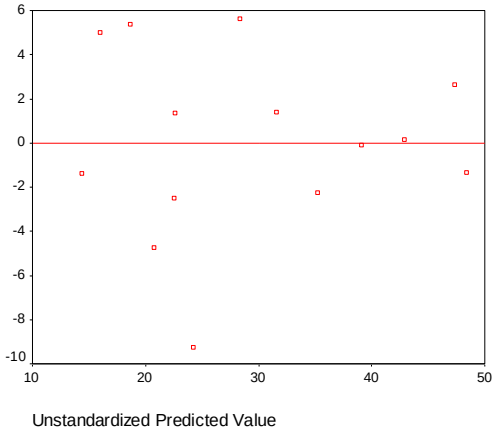
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	-41.430	7.412		-5.589	.000	-57.580	-25.281
	Number of Handguns Registered	.125	.013	.941	9.675	.000	.097	.153

a. Dependent Variable: Number of People Killed

III. Residual Analysis

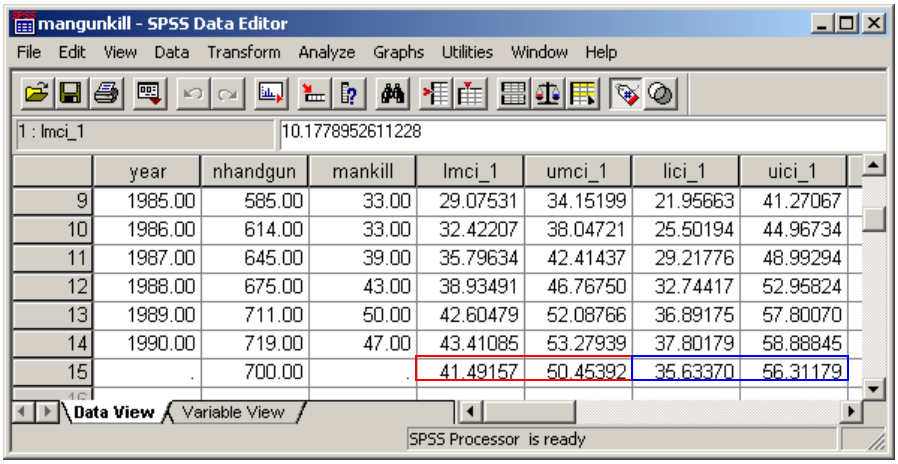
To assess the fit of the model, when performing the regression, also click on the **Save** button at the bottom of the dialogue box. In the boxes labeled **Predicted Values** and **Residuals**, click **Unstandardized** in both boxes. In the Hit **Continue** and then hit **OK**. There will now be a column labeled **res_1** and a column labeled **pre_1** in the data window. To make a residual plot, click **Graphs\Scatter\Simple\Define** and put **res_1** in the **Y-axis** box and again put **pre_1** in the **X-axis** box. Hit **OK**.

The goal of a residual plot is to see a random scatter of residuals. The normality test in the Explore… option can be used to check for normality.



IV. Prediction Intervals

To calculate the mean prediction intervals and the individual prediction intervals, use the **Save** button that appears after clicking **Analyze\Regression\Linear**. Now in the box labeled **Prediction Values**, click on **Unstandardized.** This will give the predicted Y-values from the model. The data window will have a column labeled **pre_1**. For the prediction intervals, in the boxes near the bottom labeled **Prediction Intervals,** put check marks in front of **Mean** and **Individual**. In the data window, will now be columns, labeled **lmci_1, umci_1, lici_1,** and **uici_1**. The first two columns are for the lower and upper bounds for the 95% mean prediction interval. (We are 95% sure that the average y for the given x values is within that interval.) The second two columns are the 95% prediction intervals. To do a prediction, simply enter the value of the predictor variable at the last row of the data sheet under the predictor variable and go through the model building. For instance, to predict the average number of people killed if the number of handguns registered is 700. The confidence interval for mean response is (41.49, 50.45), and for an individual response is (35.63, 56.31).



	year	nhandgun	mankill	lmci_1	umci_1	lici_1	uici_1
9	1985.00	585.00	33.00	29.07531	34.15199	21.95663	41.27067
10	1986.00	614.00	33.00	32.42207	38.04721	25.50194	44.96734
11	1987.00	645.00	39.00	35.79634	42.41437	29.21776	48.99294
12	1988.00	675.00	43.00	38.93491	46.76750	32.74417	52.95824
13	1989.00	711.00	50.00	42.60479	52.08766	36.89175	57.80070
14	1990.00	719.00	47.00	43.41085	53.27939	37.80179	58.88845
15	.	700.00	.	41.49157	50.45392	35.63370	56.31179

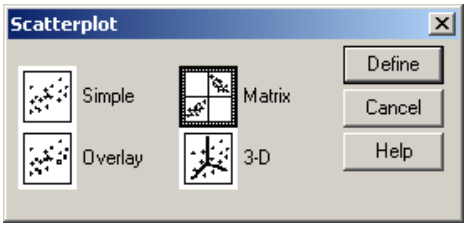
Multiple Regression in SPSS

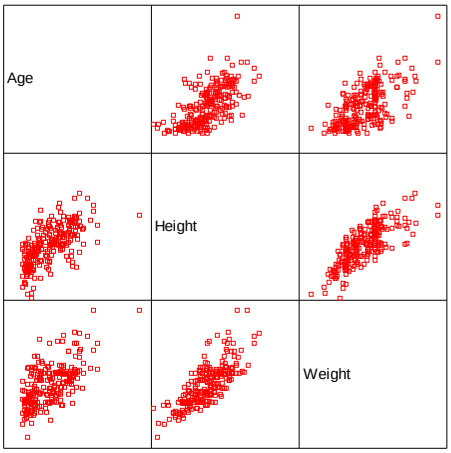
Data: WeightbyAgeHeight.sav
Goals:

- Examine relation between weight (response) and age and height (explanatory)
- Model checking
- Predict weight

I. View the Data with a Scatter Plot

To create a scatter plot, click through **Graphs\Scatter\Simple\Define**. A Scatterplot dialog box will appear. Click **Matrix** option and then click **Define** button. The Simple Scatterplot dialog box will appear. Click **age**, **weight** and **height** variables into matrix variables box.





II. Regression Analysis

To perform the regression, click on **Analyze\Regression\Linear**. Place **weight** in the **Dependent** box and place **age** and **height** in the **Independent** box. Click the Statistics button to select the **collinearity diagnostics** and click **Continue**. To obtain the 95% confidence interval for the slope, click on the **Statistics** button at the bottom and then put a check in the box for **Confidence Intervals**. Hit **Continue** and then hit **OK**.

The independent variables (**age, height**) are useful in predicting the dependent variable (**weight**) when the level of significance (P-value labeled with **Sig.** on the Output) is below 0.05.

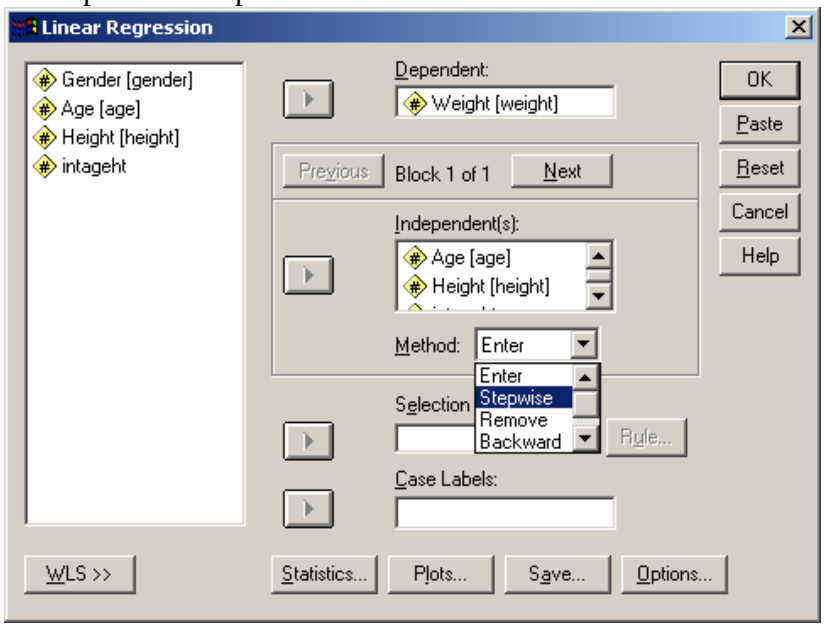
Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	-127.820	12.099		-10.565	.000		
	Age	.240	.055	.228	4.360	.000	.579	1.727
	Height	3.090	.257	.627	12.008	.000	.579	1.727

a. Dependent Variable: Weight

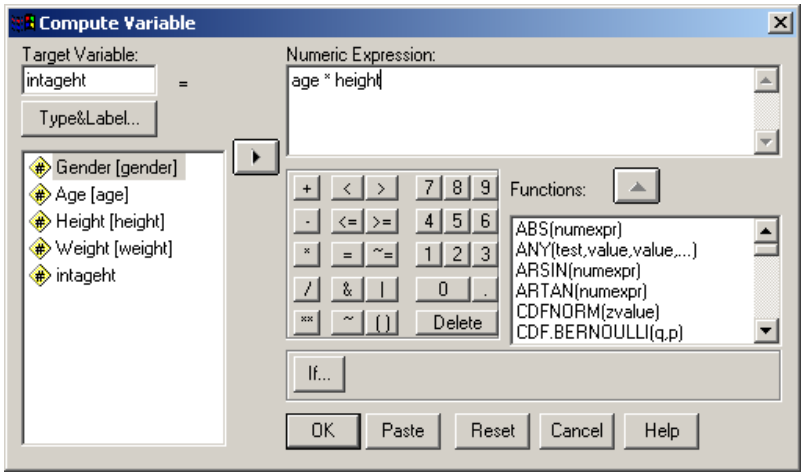
Stepwise Regression

To perform stepwise regression for automatically selecting significant variables, check the **Method** drop down list and choose the desired one and click OK. SPSS will produce an output table to present the final model with a coefficients table.



Interaction Term

To examine the interaction between age and height variables, first create the interaction variable (**intageht**). Click **Transform\Compute**, and in the compute variable dialog box, enter a name for the interaction term, **intageht**. Click and enter **age** variable and click "*" for multiplying and then click and enter **height** variable in the Numeric Expression box and then click **OK**.



Collinearity Diagnostics

To perform the regression, click on **Analyze\Regression\Linear**. Place **weight** in the **Dependent** box and place **age**, **height** and **intageht** in the **Independent** box. Click the Statistics button to select the **collinearity diagnostics** and click **Continue**, and then hit **OK**. In the coefficients table, VIF are all greater than 10 which implies collinearity. In this example, centralizing the variables

may reduce the possible collinearity from the interaction term that is generated by the product of the two variables.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
		B	Std. Error	Beta			Tolerance	VIF
1	(Constant)	66.996	106.189		.631	.529		
	Age	-.973	.660	-.923	-1.476	.141	.004	250.009
	Height	-3.13E-02	1.710	-.006	-.018	.985	.013	77.016
	INTAGEHT	1.936E-02	.010	1.636	1.847	.066	.002	501.996

a. Dependent Variable: Weight

If categorical variables are to be included in the model, the indicator variables will need to be created. **Recode** under transform option is one of the methods to generate the indicator variables.

III. Residual Analysis

To assess the fit of the model, when performing the regression, also click on the **Save** button at the bottom of the dialogue box. In the boxes labeled **Predicted Values** and **Residuals**, click **Unstandardized** in both boxes. In the Hit **Continue** and then hit **OK**. There will now be a column labeled **res_1** and a column labeled **pre_1** in the data window. To make a residual plot, click **Graphs\Scatter\Simple\Define** and put **res_1** in the **Y-axis** box and again put **pre_1** in the **X-axis** box. Hit **OK**. The goal of a residual plot is to see a random scatter of residuals. The normality test in the Explore… option can be used to check for normality.

IV. Prediction Intervals

To calculate the mean prediction intervals and the individual prediction intervals, use the **Save** button that appears after clicking **Analyze\Regression\Linear**. Now in the box labeled **Prediction Values**, click on **Unstandardized.** This will give the predicted Y-values from the model. The data window will have a column labeled **pre_1**. For the prediction intervals, in the boxes near the bottom labeled **Prediction Intervals,** put check marks in front of **Mean** and **Individual**. In the data window, will now be columns, labeled **lmci_1, umci_1, lici_1,** and **uici_1**. The first two columns are for the lower and upper bounds for the 95% mean prediction interval. (We are 95% sure that the average y for the given x values is within that interval.) The second two columns are the 95% prediction intervals. To do a prediction, simply enter the value of the predictor variable at the last row of the data sheet under the predictor variable and go through the model building. The procedure is similar to that of simple linear regression.

Logistic Regression in SPSS

Data: logdisea.sav
Goals:

- Examine relation between disease (binary response) and other explanatory variables such as age, socioeconomic status, sector, and savings account.
- Model checking
- Predict probability of getting disease and estimating the odds ratio

To perform the regression, click on **Analyze\Regression\Binary Logistic**. Place **disease** in the **Dependent** box and place **age**, **sciostat**, **sector** and **savings** in the **covariates** box. Click the **Categorical** button for creating indicator variables for the categorical variables. To obtain the the predicting probability of getting disease, click on the **Save** button at the bottom and then put a check in the **Probabilities** box of **Predicted Values** and click **Continue.** Also, click the **Option** button and check the **goodness of fit** box to see how well the model fit the data, and check the **CI for exp(B)** to obtain confidence interval for odds ratio. Hit **Continue** and then hit **OK**.

The independent variables (**age, sector**) are significant in predicting the dependent variable (**disease**) when the level of significance (P-value labeled with **Sig.** on the Output) is below 0.05.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95.0% C.I.for EXP(B)	
								Lower	Upper
Step 1[a]	AGE	.027	.009	8.646	1	.003	1.027	1.009	1.045
	SCIOSTAT			.439	2	.803			
	SCIOSTAT(1)	-.278	.434	.409	1	.522	.757	.323	1.775
	SCIOSTAT(2)	-.219	.459	.227	1	.634	.803	.327	1.976
	SECTOR(1)	-1.234	.357	11.970	1	.001	.291	.145	.586
	SAVINGS	.061	.386	.025	1	.874	1.063	.499	2.264
	Constant	-.814	.452	3.246	1	.072	.443		

a. Variable(s) entered on step 1: AGE, SCIOSTAT, SECTOR, SAVINGS.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	10.853	8	.210

Stepwise Regression

To perform stepwise regression for automatically selecting significant variables, check the **Method** drop down list and choose the desired one and click OK. SPSS will produce an output table to present the final model with a coefficients table.