



IBM® SPSS® Statistics
Version 22

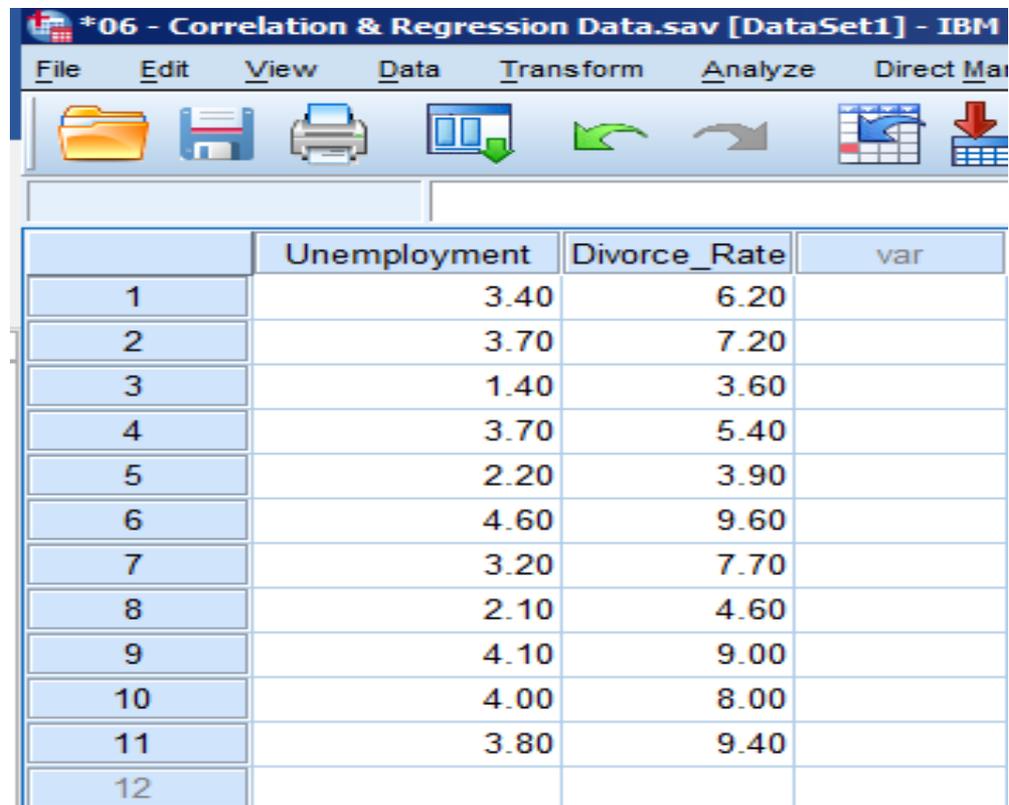
Licensed Materials - Property of IBM Corp. © Copyright IBM Corporation and other(s) 1989, 2013. IBM, the IBM logo, ibm.com, Cognos and SPSS are trademarks or registered trademarks of International Business Machines Corp., registered in many jurisdictions worldwide. A current list of IBM trademarks is available on the Web at www.ibm.com/legal/copytrade.shtml. Java and all Java-based trademarks and logos are trademarks or registered trademarks of Oracle and/or its affiliates. Other product and service names might be trademarks of IBM or other companies. This Program is licensed under the terms of the license agreement accompanying the Program. This license agreement may be either located in a Program directory folder or library identified as "License" or "Non_IBM_License", if applicable, or provided as a printed license agreement. Please read the agreement carefully before using the Program. By using the Program you agree to these terms.



Psyc 381
Correlation & Regression
with two variables
A brief how-to guide

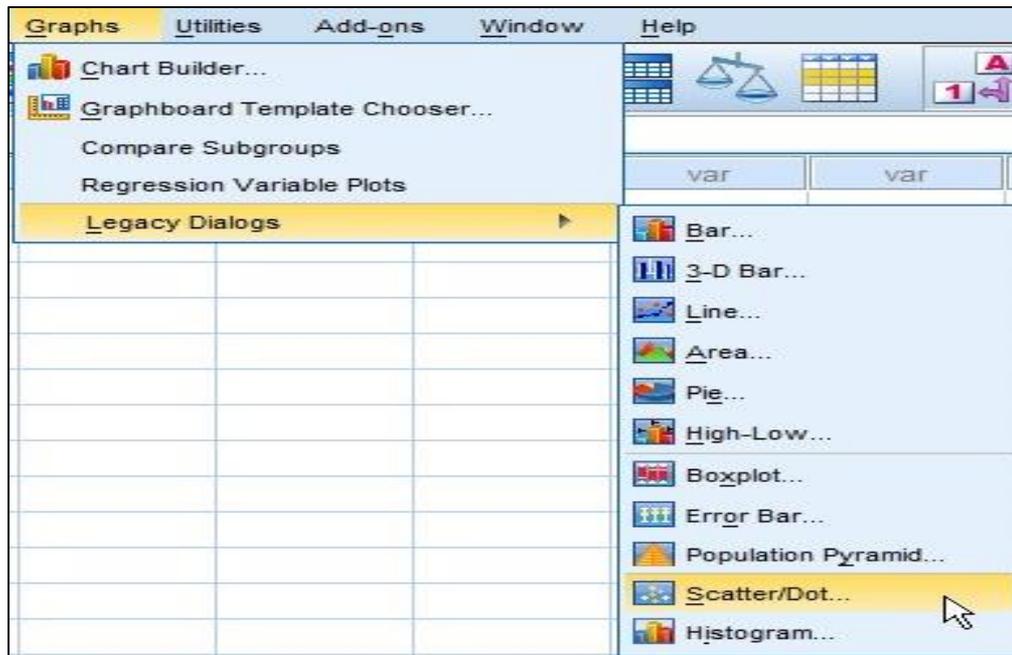
Data Entry

- Correlation and (simple) regression analyses are used to identify the statistical relationship between two (or more) variables. As a result, entering data into SPSS will require the use of (at least) two columns of data points. Each measured variable will need data entered into its own column. See the example below:

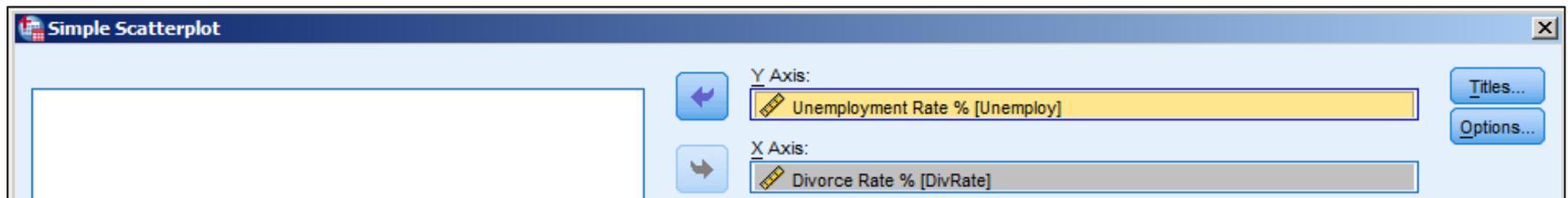
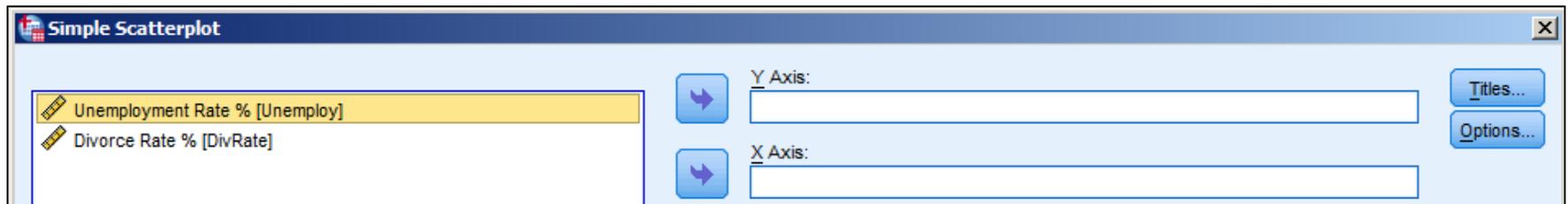
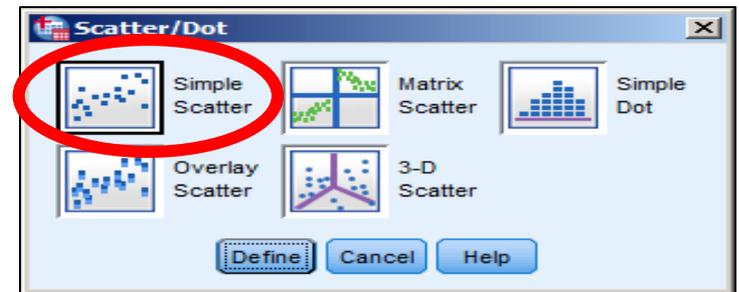


	Unemployment	Divorce_Rate	var
1	3.40	6.20	
2	3.70	7.20	
3	1.40	3.60	
4	3.70	5.40	
5	2.20	3.90	
6	4.60	9.60	
7	3.20	7.70	
8	2.10	4.60	
9	4.10	9.00	
10	4.00	8.00	
11	3.80	9.40	
12			

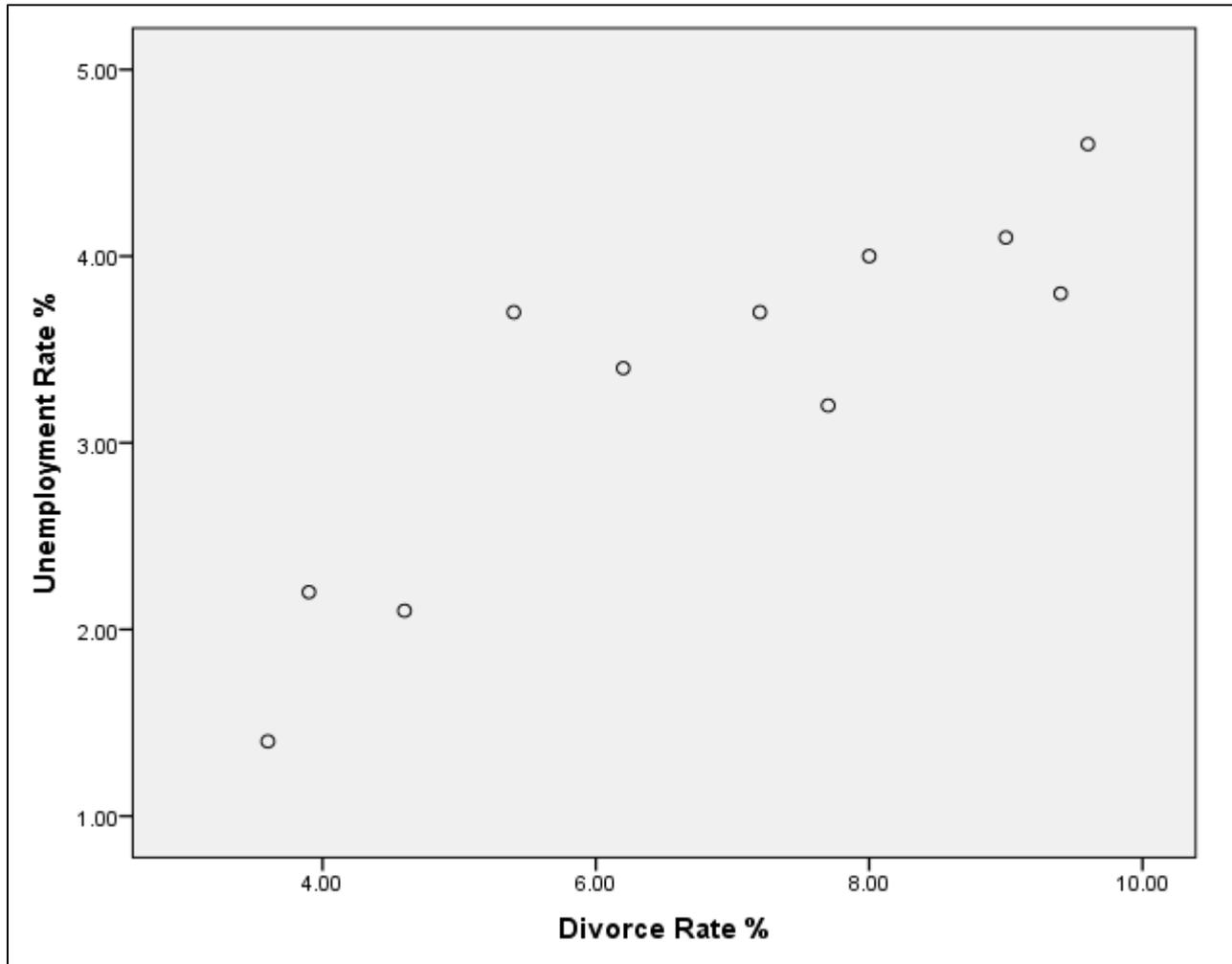
Obtaining a scatterplot



- Using the legacy dialogue under “Graphs” in the menu, navigate to the “Scatter/Dot” menu and press the “Simple Scatter” option
- Drag the variables to the X-axis and Y-axis as needed and then confirm your selection (continue) to produce your scatterplot

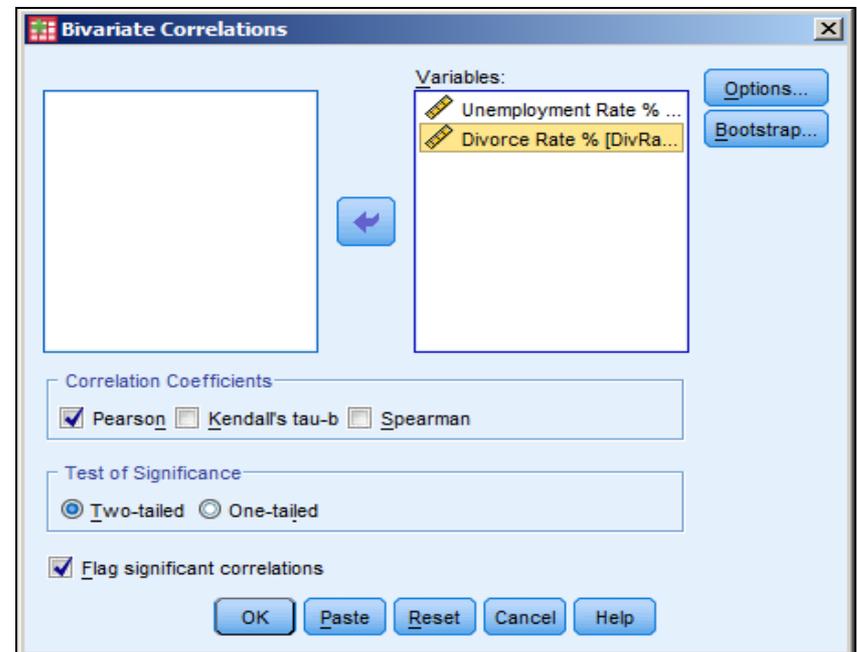
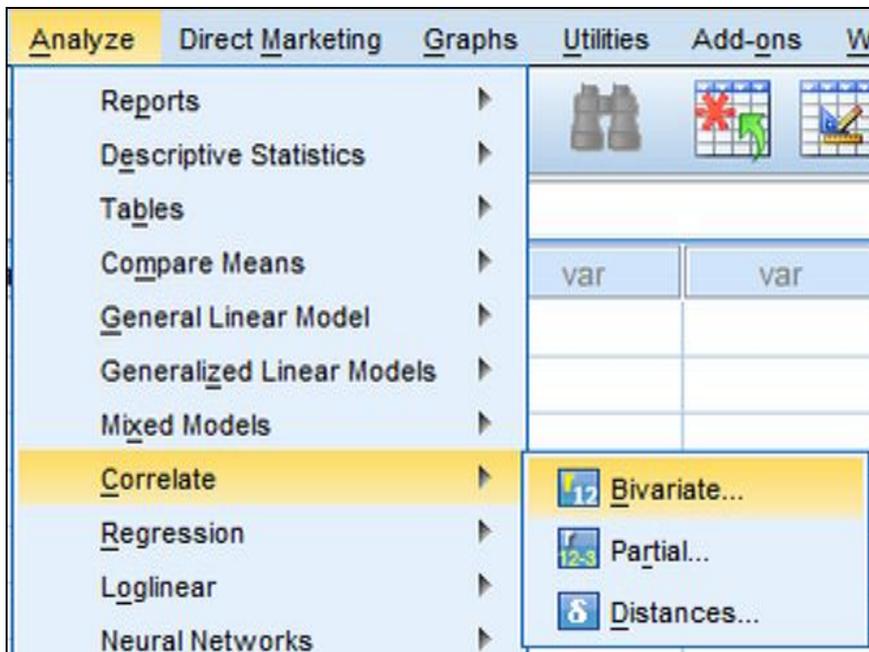


Obtaining a scatterplot



Obtaining Pearson's r

- Navigate to the following menu option to obtain Pearson's r between two variables. In the menu, move the variables of interest from the left to the right window.
- Keep in mind that all variables in the rightmost window will be analyzed and produce all possible pairwise r values so to avoid extra output make sure you only have the variables you want to analyze in the right window at once.
- If desired, descriptive statistics (means, SDs, SEs) and other information can be obtained for the analyzed variables using the "Options" menu.



Obtaining Pearson's r

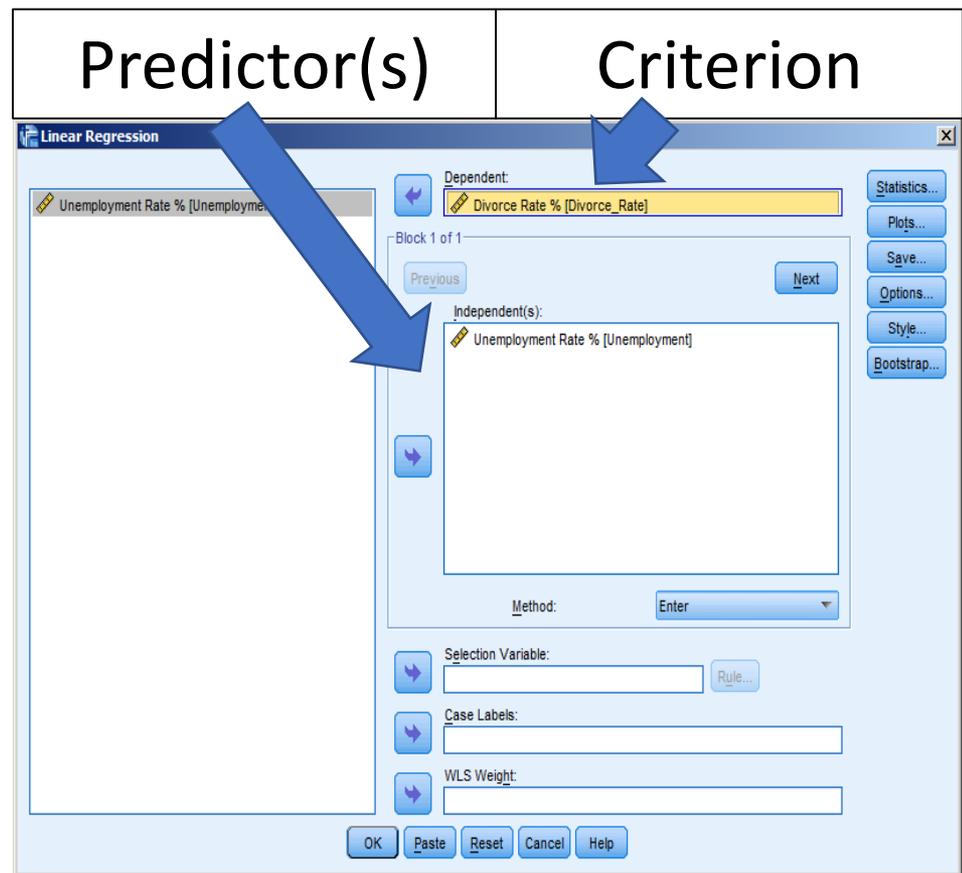
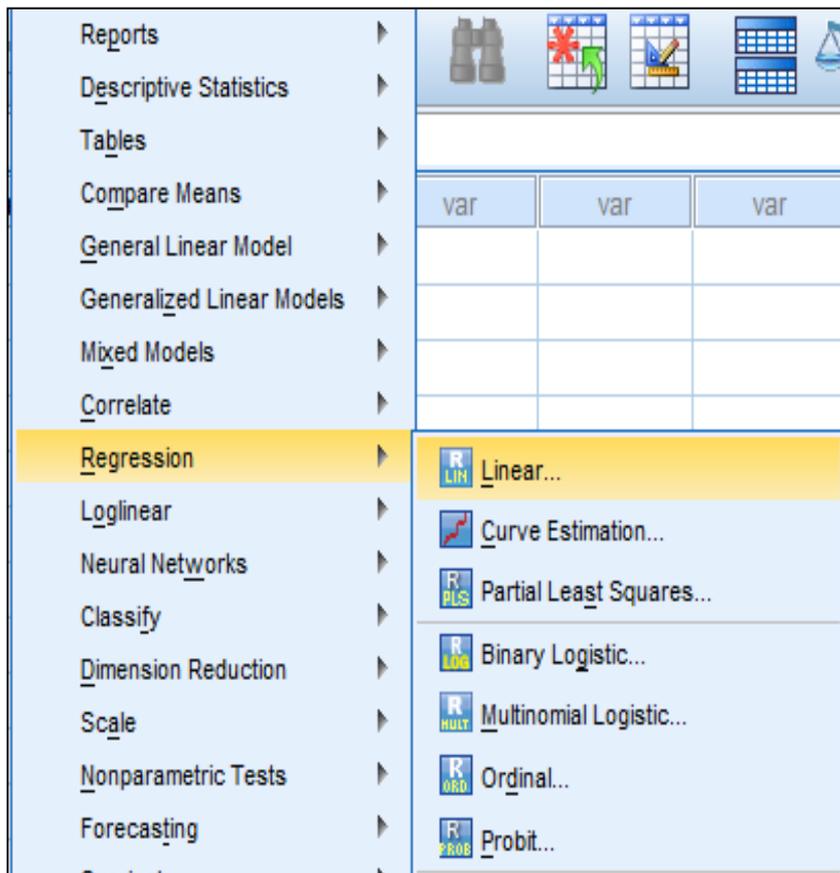
- SPSS will produce a correlation matrix (see below) in which all pairwise combinations of the selected variables are analyzed. Each box can be matched up based on the column and row with each variable name and the respective r -value, its p -value (Sig 2-tailed in the box), and the sample size (N). Keep in mind you will have to determine the degrees of freedom value ($N - 2$) on your own for each r value.
- Note: SPSS will generate repeated r values because it computes the r value for each pairwise combination of variables possible (e.g., A & B, B & A). This is normal SPSS behavior and you can simply focus your attention on one of the repeated boxes for the appropriate information.

		Unemployment Rate %	Divorce Rate %
Unemployment Rate %	Pearson Correlation	1	.880 ^{**}
	Sig. (2-tailed)		.000
	N	11	11
Divorce Rate %	Pearson Correlation	.880 ^{**}	1
	Sig. (2-tailed)	.000	
	N	11	11

** . Correlation is significant at the 0.01 level (2-tailed).

Conducting a simple regression analysis

- Navigate to the following menu and select “Linear” to run a linear regression analysis. You must first determine which variable you want to use as the predictor and which you want to use as the criterion. SPSS cannot make this determination for you – use your best judgement.
- Drag the “predictor” variable into the “Independent” box and drag the “criterion” variable into the “Dependent” box and then hit “OK”.
- Additional information concerning the analysis (e.g., descriptive statistics) and other forms of regression can be produced using the various menu options but they are not required to obtain the results of a simple regression analysis.



Interpreting a simple regression analysis

- With the default options, SPSS will generate four pieces of output. The first two are described below and the remaining two are discussed on the next page.
- The first piece of output describes the requested regression model. As seen below, the output is telling you that the “Unemployment rate” is being used as the predictor for the criterion (or dependent) “Divorce Rate”.
- The second piece of output describes two important things
 - The “R Square” is the overall proportion (or percentage if x100) of the variability in the criterion associated with variability in the predictor
 - The Std Error of the Estimate is the measure of precision of the prediction; on average, the predicted values will be off 1.09 from the actual (or observed) values using the regression line as a prediction tool

Regression			
Variables Entered/Removed^a			
Model	Variables Entered	Variables Removed	Method
1	Unemployment Rate % ^b	.	Enter

a. Dependent Variable: Divorce Rate %
b. All requested variables entered.

Model Summary				
Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.880 ^a	.775	.750	1.09138

a. Predictors: (Constant), Unemployment Rate %

Interpreting a simple regression analysis

- With the default options, SPSS will generate four pieces of output. The last two pieces of output are described below.
- The third piece of output describes the results of an analysis of variance (ANOVA). In brief, this box describes the math that underlies the regression approach. Although valuable information on its own, this box will typically not be reported as part of a result summary.
- The fourth piece of output contains the information to create the regression equation to predict (in this example) “Divorce Rate” using “Unemployment Rate).
 - For an unstandardized regression equation, the regression constant can be found in the “unstandardized coefficients” box under the B column (0.350) and the regression coefficient (or weight) of the predictor (Unemployment Rate) can be found under the same B column (1.954).
 - These two values can be combined to create the following regression equation. This equation can then be used to determine the predicted value of “Divorce Rate” based on any value plugged into X (Unemployment Rate).

$$Y' = .350 + .1.954(x)$$

ANOVA ^a						
Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	36.936	1	36.936	31.010	.000 ^b
	Residual	10.720	9	1.191		
	Total	47.656	10			

a. Dependent Variable: Divorce Rate %
b. Predictors: (Constant), Unemployment Rate %

Coefficients ^a						
Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	.350	1.201		.291	.777
	Unemployment Rate %	1.954	.351	.880	5.569	.000

a. Dependent Variable: Divorce Rate %