Application of statistics in engineering

Snezhana G. Gocheva-Ilieva
University of Plovdiv Bulgaria, snow@uni-plovdiv.bg

"All models are wrong, but some are useful."

Annotation. **The treatment of experimental data is widely encountering in engineering practice. This course is oriented to application of different statistical techniques such as regression analysis, factor analysis or cluster analysis in modeling different problems in physics and engineering. The use of SPSS software to solve this type of problems is also demonstrated. In particular, new results about the statistical study in the field of metal vapor lasers are presented, with emphasize on planning the future experiment with improved output lasing characteristics.**

For more Materials, please visit
https://wiki.helsinki.fi/display/mathstatKurssit/Application+of+statistics+in+engineering%2C+spring+2009

1. Collecting data: retrospective study, observational study and designed experiment. Mechanistic and empirical models

A common case from the engineering practice is to treat data, obtained by measurements. For instance, physical laws (such as Ohm's law or pressure law) are often applied to help design products and processes.

The data in natural sciences and engineering are dominantly of quantitative (continuous) type, for instance the measurements of a pressure are rational or real numbers, belonging to some interval.

A good data collection procedure can greatly simplify the analysis and lead to improved understanding of the population or process that is being studied. There are three basic data collection methods:

- **A retrospective study using historical data**
- **An observational study**
- **A designed experiment**

Retrospective Study

A retrospective study use either all or a sample of the historical process data over some period of time. It is also assumed that the researcher could not or is very restricted in new measurements of the observed process. The study objective might be to discover some relationships among the available data.

Usual problems in retrospective study:

1. It must be difficult to find a relationship between some subsets of data (variables) if one of them did not be varied enough or, in the worst case, has been fixed, over the considered period. In that case it will be impossible to assess its real impact on other observed data.
2. The archived data on the two or more variables may be obtained by different people and by different methods without a clear recording.

As we can see, a retrospective study may involve a lot of data, but that data may contain relatively little useful information about the problem. Furthermore, some of the relevant data may be missing, there may be transcription errors, or data on other important factors may not have been collected and archived.

As a consequence, statistical analysis of historical data can not provide the desired results.

Observational Study

In an observational study, the engineer observes the process or population, disturbing it as little as possible, and records the quantities of interest. Because these studies are usually conducted for a relatively short time period, sometimes variables that are not routinely measured can be included. Generally, an observational study tends to solve problems 1 and 2 above and goes a long way toward obtaining accurate and reliable data.

Designed Experiments

Designed experiments play a very important role in engineering design and development and in the improvement of manufacturing processes. In a designed experiment the engineer makes deliberate or purposeful changes in the controllable variables of the system or process, observes the resulting system output data, and then makes an inference or decision about which variables are responsible for the observed changes in output performance.

Example 1. Designed experiment –two levels

Consider the problem with an acetone distillation column. This process has three factors, the two types of temperatures (a reboil temperature and a condensate temperature) and the reflux rate, and we want to investigate the effect of these three factors on produced output acetone concentration. The specified values of the three factors used in the experiment are called factor levels. Typically, we use a small number of levels for each factor, such as two or three.

For the distillation column problem, suppose we use a **“high,’’ and “low,’’ level (denoted +1 and -1, respectively)** for each of the factors. **We thus would use two levels for each of the three factors.** A very reasonable experiment design strategy **uses every possible combination of the factor levels to form a basic experiment** with eight different settings for the process. This type of experiment is called a factorial experiment. Table 1 presents this experimental design.

Table 1. The designed experiment for the distillation column

Reboil temperature	Condensate temperature	Reflux rate
-1	-1	-1
+1	-1	-1
-1	+1	-1
+1	+1	-1
-1	-1	+1
+1	-1	+1
-1	+1	+1
+1	+1	+1

With each setting of the process conditions, we allow the column to reach equilibrium, take a sample of the product stream, and determine the acetone concentration. We then can draw specific inferences about the effect of these factors. Such an approach allows us to proactively study a population or process.

1.2. Mechanistic and empirical models

Mechanistic models

Any physical, chemical or other law in science is actually some model of the real process or phenomena. For instance, consider the simple relationship between distance, time and speed:

$$t = \frac{S}{v},$$

where t is the time, S is the distance, and v is the speed. We call this type of model a **mechanistic model**. Ordinary, it is used as an approximate equality to calculate the time at measured distance and speed. For instance, if a sprint champion runs one and the same distance on ten consecutive days at almost the same conditions, the observed results will differ slightly because of

small changes or variations in factors that are not completely the same, such as changes in speed or slights detours in the direction. The more appropriate expression is

$$t = \frac{S}{v} + \varepsilon .$$

where ε is the term added to the model to account for the fact that the observed values of time do not perfectly much to the mechanistic model. The more common meaning of ε includes the effects of all of the unmodeled sources of variability that affect this measurements.

Empirical models

Sometimes, in an arbitrary study, we do not know the corresponding deterministic model.

Example 2. Let we have some measurements on thermal resistance of the wall as a function of its thickness, given in Table 2. We wish to establish some relationship between the thickness x and the resistance y. Drawing the graphics (Fig. 1) we observe, that the 8 points with coordinates (x_i, y_i) are similar to fit the straight line. For this reason we may tray to establish an approximate linear model in the form

$$y \approx b_0 + b_1 x,$$

where the coefficients have to be found. This type of equation is called **empirical model**, because we do not use an previously known deterministic model.

Table 1.

Thermal resistance of the wall as a function of its thickness								
x, mm	2	4	6	8	10	12	15	20
y, $mm^2.^oC$	0,83	1,34	1,63	2,29	2,44	2,93	4,06	4,48

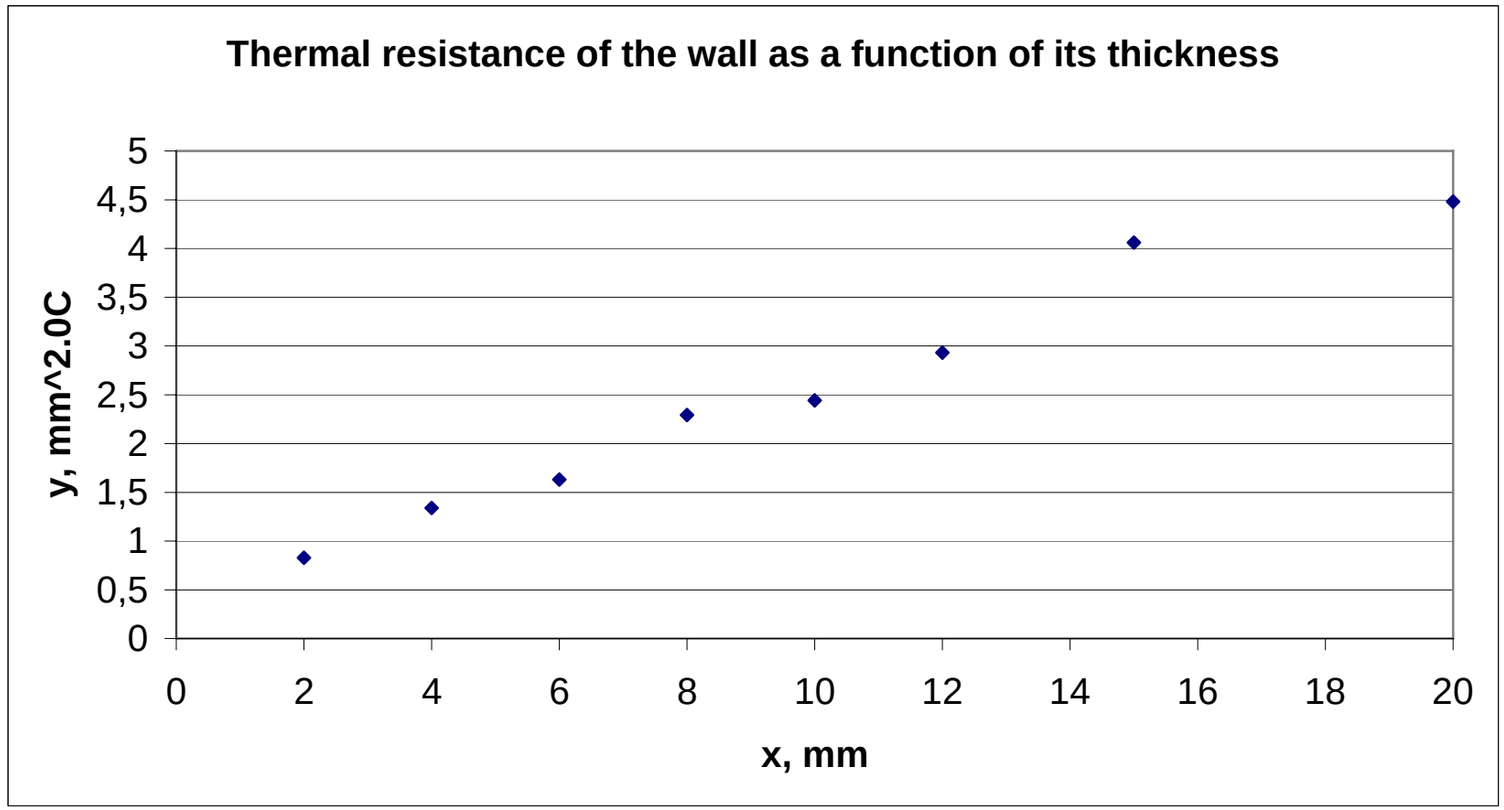


Fig. 1. The graphics with data from Table 1.

It will be found that the best fit of data from Table 1 to a straight line takes place for $b_0 = 0.4544,\ b_1 = 0.2125$. This way the empirical model is (see Fig. 2):

$$y \approx 0.4544 + 0.2125x\,. \quad (*)$$

The “exact” model will be written as

$$y = 0.4544 + 0.2125x + \varepsilon\,,$$

where ε is the model error and is accounted for the distinguish between the value, calculated from (*) and the measured value.

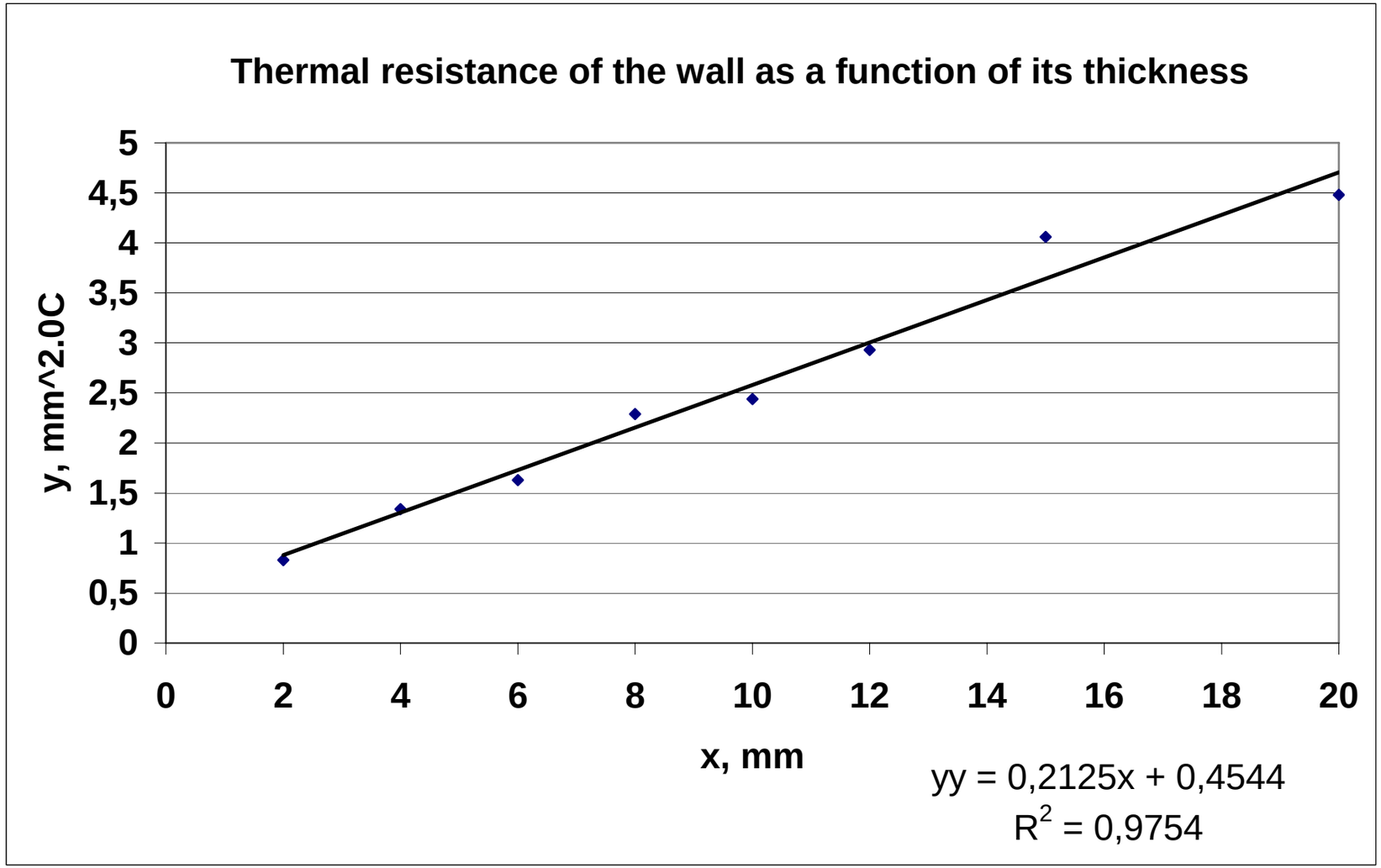


Fig. 2. Empirical linear model using data of Table 1.

More

The mechanistic and empirical models can be also classified as *deterministic classical models.* These approaches have the focus on the model-estimating parameters of the model and generating predicted values from the model. Deterministic models include, for example, regression models and analysis of variance (ANOVA) models.

A non-very formalised is the *exploratory model*. The exploratory data analysis (EDA) approach does not impose deterministic or probabilistic models on the data. On the contrary, the EDA approach allows the data to suggest admissible models that best fit the data. It focuses on the data -its structure, outliers, and models suggested by the data, being examined. A detailed description of the methods of EDA and many classical methods can be found in the reach internet “Engineering statistics handbook”: http://www.itl.nist.gov/div898/handbook/.

3. Introduction in regression empirical models: linear regression. Examples. Solving examples using SPSS.

Regression analysis is widely used statistical method that can establish the explicit relationship between one or more dependent (response) variables and one or more independent (predictor) variables. As a result any of dependent variables is expressed as a function of the predictor variables. This function is also called regression model. The models can be linear or nonlinear. They are usually used to predict the unknown values of the dependent variable at given selection of the predictors. The fit and the accuracy depend on the data used. Hence non-representative or improperly compiled data result in poor fits and conclusions.

An example of a multiple regression model is the linear regression model which is a linear relationship between dependent variable, y and the independent variables, $x_i, i = 1,2...,n$ of the form

$$y = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n + \varepsilon \quad (1)$$

where $b_0, b_1, ..., b_n$ are called regression coefficients (unknown model parameters), and ε is the error due to variability in the observed responses. The meaning of the regression coefficients is that they indicate how the change in one of the independent variables affects the values taken by the dependent variable. The values of $b_0, b_1, ..., b_n$ are easily interpreted by its sign. If $b_i > 0$, then any increase in x_i leads to increase in *y* and vice versa: if $b_i < 0$, then the decrease in x_i involves the decrease in *y*. The bigger the coefficient b_i, the bigger is the strength of x_i on the response *y*.

In the common case depending on the area of applications, the dependent variable is a quantitative measure of some property or behavior. When the dependent variable is qualitative or categorical, then other methods (such as logit or probit analysis) might be more appropriate.

To compare the strengths of the predictors on the dependent variable(s) it is more convenient to use the so called z-variables (with mean =0 and standard deviation = 1). We will obtain the equation

$$\frac{y-\overline{y}}{s_y}=\beta_1\frac{x_1-\overline{x}_1}{s_1}+\beta_2\frac{x_2-\overline{x}_2}{s_2}+\ldots+\beta_n\frac{x_n-\overline{x}_n}{s_n}\ , \qquad (2)$$

where $\beta_1,\ldots,\beta_n$ are called standardized regression coefficients.

Example 3.

In the transformation of raw or uncooked potato to cooked potato, heat is applied for some type of cooking. One might postulate that the amount of untransformed portion of the starch (y) in potato is a linear function of time (t) and temperature (θ) of cooking. This is represented as

$$y=b_0+b_1t+b_2\theta+\varepsilon \qquad (3)$$

Linear as used in linear regression refers formally to the form of occurrence of the unknown parameters, b_1 and b_2 as simple linear multipliers of the predictor variable. This means, that the two equations below can be also considered both linear, by introducing for example a new independent variable $\rho = t\theta$:

$$y = b_0 + b_1 t + b_2 t\theta + b_3\theta + \varepsilon \quad \text{or} \quad y = b_0 + b_1 t + b_2\rho + b_3\theta + \varepsilon; \tag{4}$$

$$y = b_0 + b_1 t\theta + b_2\theta + \varepsilon \quad \text{or} \quad y = b_0 + b_1\rho + b_2\theta + \varepsilon. \tag{5}$$

As an other example, a nonlinear model can be presented, depending on the behavior of data, in the form

$$y = b_0 + b_1 e^{\lambda_1 x} + b_2 e^{\lambda_2 x} + \varepsilon \tag{6}$$

3.2. Idea of the least squares method

We will briefly discuss the well-known least squares method (LSM) of estimation for the regression model. Let the predicted vector from the model (1) is $\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + ... + b_n x_n$,
or in the case of a sample size M:

$$\hat{y}_i = b_0 + b_1 x_{1i} + b_2 x_{2i} + ... + b_n x_{ni}, \quad i = 1,2,\ldots,M.$$

Denote the error ε_i, called the residual $\varepsilon_i = y_i - \hat{y}_i$.
The LSM finds the constants $b_0, b_1, ..., b_n$ to minimize the sum of squared residuals, or

$$\sum_{i=1}^{M} \varepsilon_i^2 = \min.$$

This problem leads to solve a linear system of order n+1.

3.3. Assumptions for regression analysis (RA)

Basic assumptions

Data. The dependent and independent variables should be quantitative. Categorical variables, such as religion, major field of study, or region of residence, need to be recoded to binary (dummy) variables or other types of contrast variables.

Assumptions. For each value of the independent variable, the distribution of the dependent variable must be normal. The variance of the distribution of the dependent variable should be constant for all values of the independent variable. The relationship between the dependent variable and each independent variable should be linear, and all observations should be independent.

The best regression models are those in which the predictor variables each correlate highly with the dependent (outcome) variable but correlate at most only minimally with each other. Such a model is often called "low noise" and will be statistically robust (that is, it will predict reliably across numerous samples drawn from the same statistical population).

Some aspects and more detailed assumptions

The least squares fitting procedure is usually used for data analysis as a descriptive technique. However, the procedure has strong theoretical justification if a few assumptions are made about how the data are obtained. The starting point is the regression equation presented above which describes some causal or behavioral process. The independent variables play the role of experimental or treatment variables. The error term ε gives the effects of all omitted variables. In an experiment, randomization of the treatments (independent variables) ensures that the omitted factors (the disturbances) are uncorrelated with the treatments. This greatly simplifies the further inference. Although, the regression analysis is valid if some special assumptions are satisfied. Here are the assumptions, known as the Gauss-Markov assumptions, that are sufficient to guarantee that ordinary regression estimates will have good properties.

1. The errors ε_i have an expected value of zero: E(ε_i) = 0. This means that on average the errors are random and balance out.
2. The independent variables can be non-random. In an experiment, the values of the independent variable would be fixed by the experimenter and repeated samples could be drawn with the independent variables fixed at the same values in each sample. As a consequence of this assumption, the indenpendent variables will in fact be

independent of the disturbance. For non-experimental work, this will need to be assumed directly along with the assumption that the independent variables have finite variances.

3. The independent variables are linearly independent. That is, no independent variable can be expressed as a (non-zero) linear combination of the remaining independent variables. The failure of this assumption, known as *multicollinearity*, clearly makes it infeasible to discover the effects of the observed independent variables.

4. The variance of the errors ε_i is the same for each observation:
   $E(\varepsilon_i^2) = \sigma^2, \quad i = 1, 2, ..., M$.

5. The residuals are not autocorrelated: $E(\varepsilon_i \varepsilon_j) = 0$ for $i \neq j$.

3.4. How to carry out regression analysis by SPSS?

Example 4 (RA-2).

The tensile strength of a certain synthetic fiber is thought to be related to x1, the percentage of cotton in the fiber, and x2, the drying time of the fiber. A test of 10 pieces of fiber produced under different conditions yielded the following results.

(a) Fit a multiple regression equation.

(b) Determine a 90 percent confidence interval for the mean tensile strength of a synthetic fiber having 21 percent cotton whose drying time is 3,6.

The data are given in Table 2.

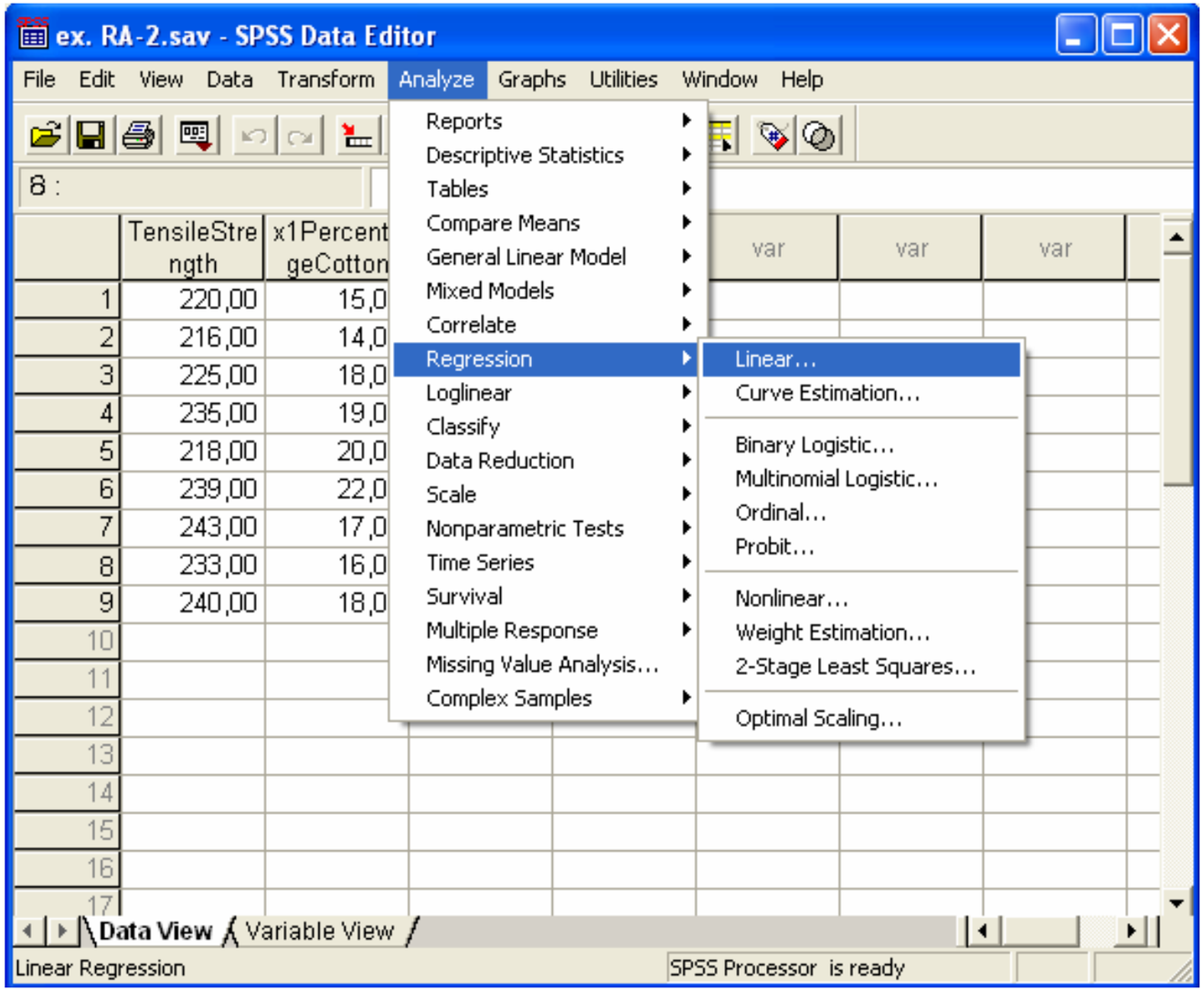
Table 2. The data of measurements of a tensile strength as a function of x1 and x2.

Tensile Strength	x1 Percentage of Cotton	x2 Drying Time
220	15	2,3
216	14	2,2
225	18	2,5
235	19	3,2
218	20	2,4
239	22	3,4
243	17	4,1
233	16	4,0
240	18	4,3

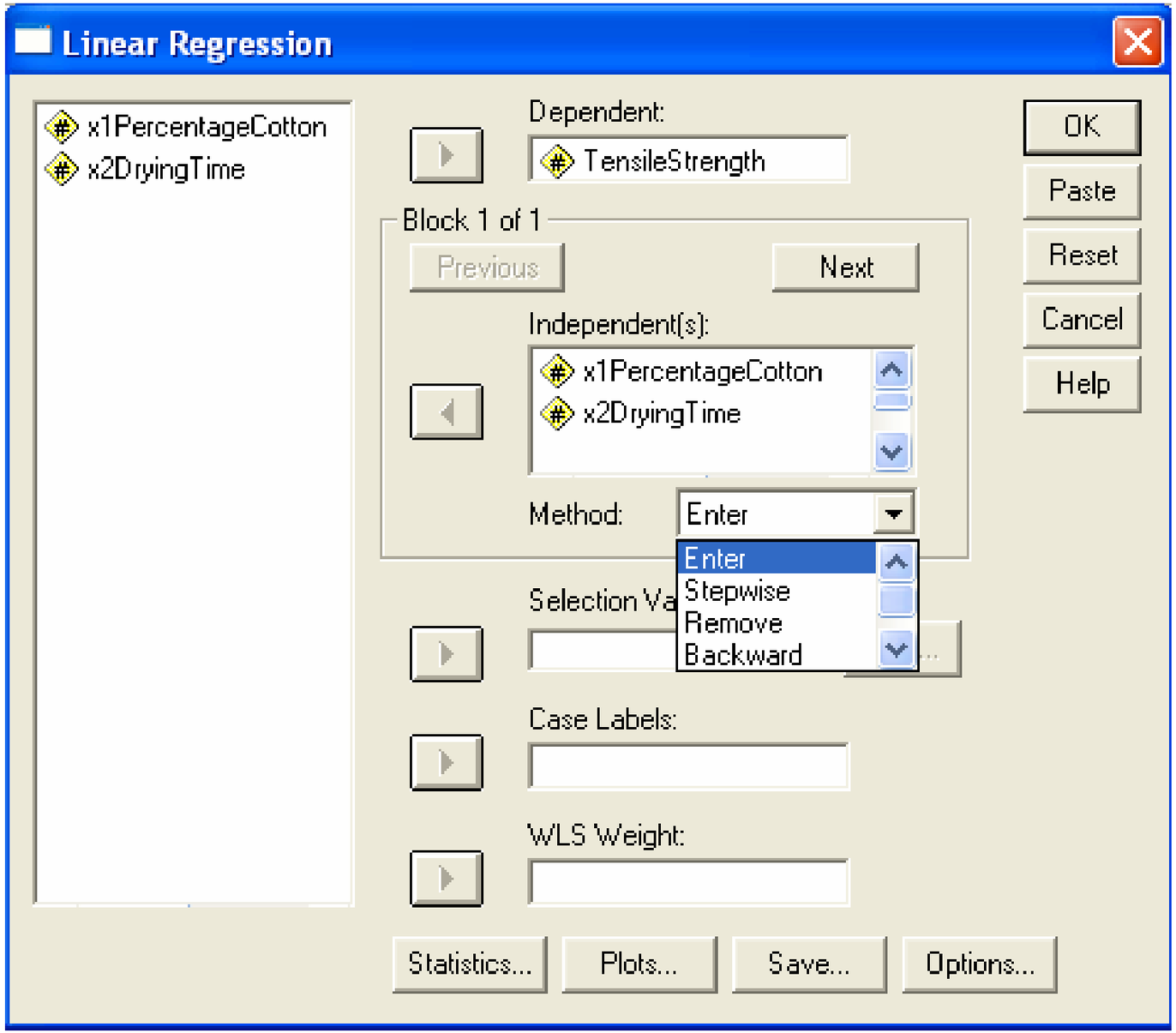
This example is solving near by SPSS.

1. Enter your data
2. From the main menu of SPSS choose:

Analyze/Regression/Linear



It will appear the Linear regression window

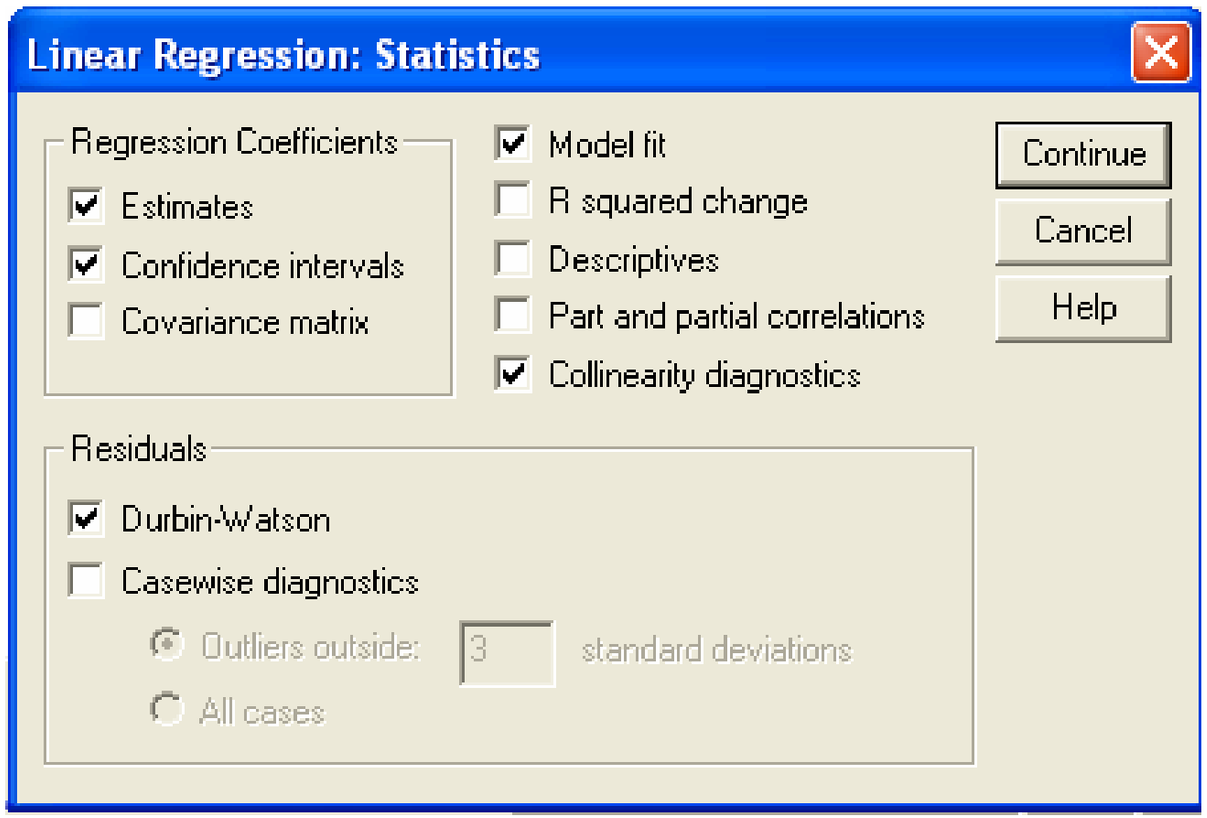

Linear Regression
x1PercentageCotton
x2DryingTime
Dependent:
TensileStrength
Block 1 of 1
Previous
Next
Independent(s):
x1PercentageCotton
x2DryingTime
Method:
Enter
Enter
Stepwise
Remove
Backward
Selection Va
Case Labels:
WLS Weight:
OK
Paste
Reset
Cancel
Help
Statistics...
Plots...
Save...
Options...

There exist many methods to explore the data with automatic selection of variables

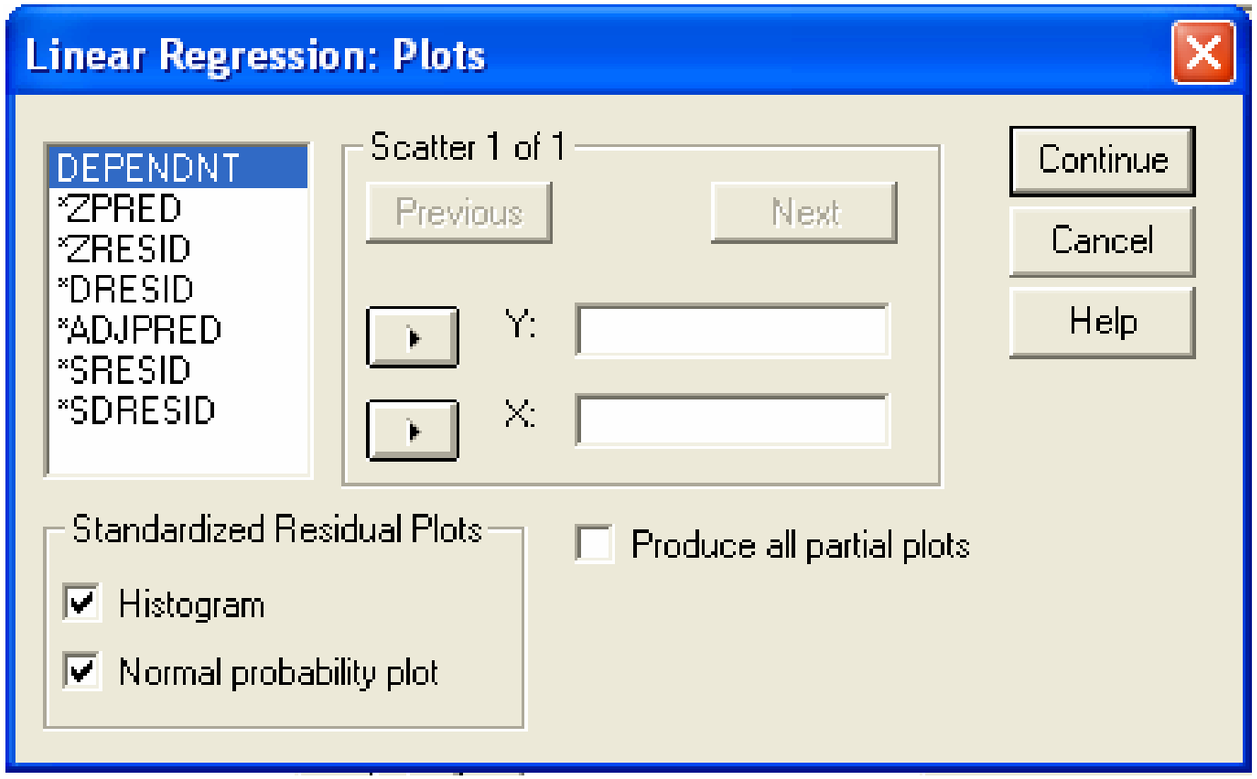
- Enter (linear)
- Stepwise - add one variable per step, remove another one if necessary
- Backward - remove one variable per step
- Forward - add one variable per step

The more usual are Linear and Stepwise methods. In the stepwise method the model begins with all selected independent variables with consequent removing of the non-significant of them. This provides a natural way to determine the predictor variables which have strength on the dependent variable.

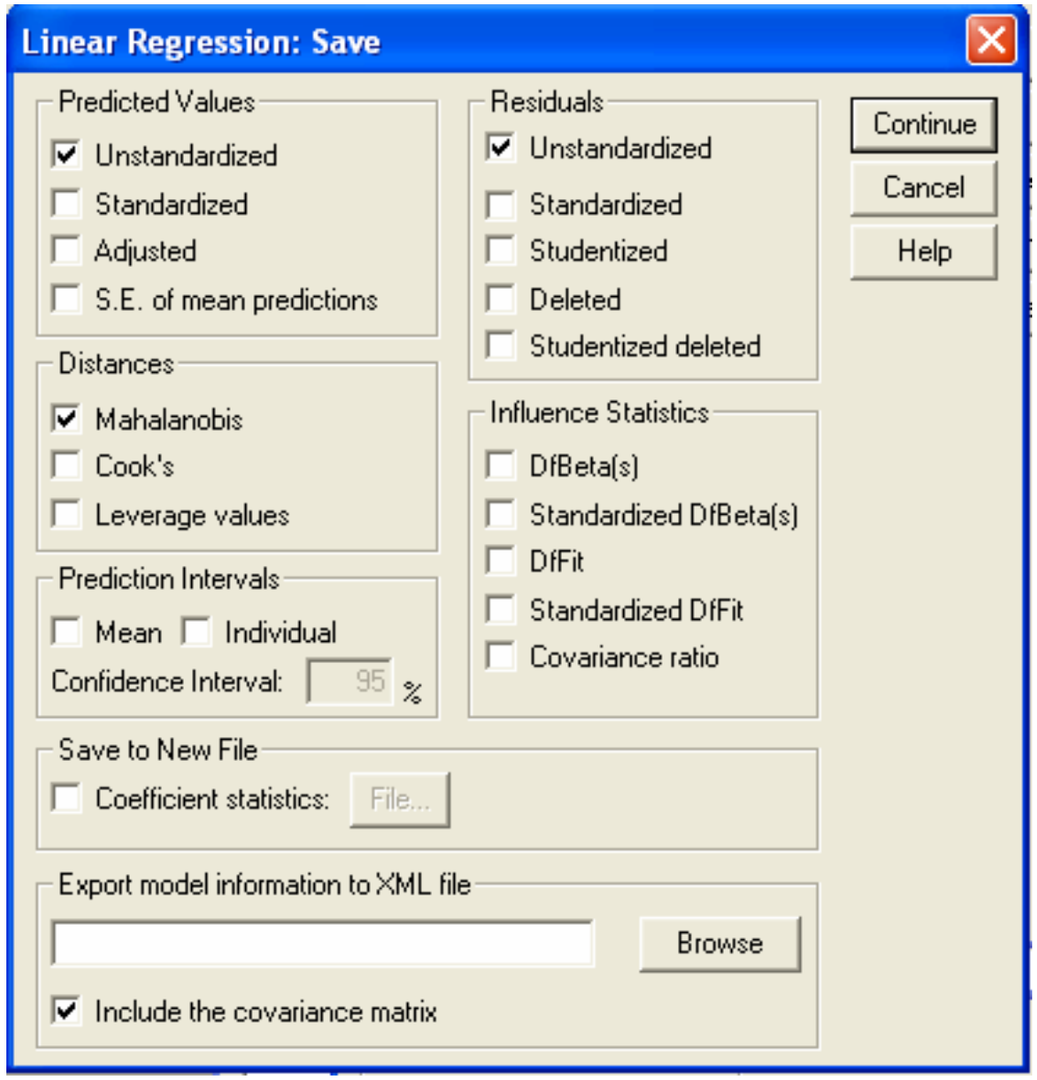
3. Now from the buttons below open **Statistics** and check the desired boxes + Continue:



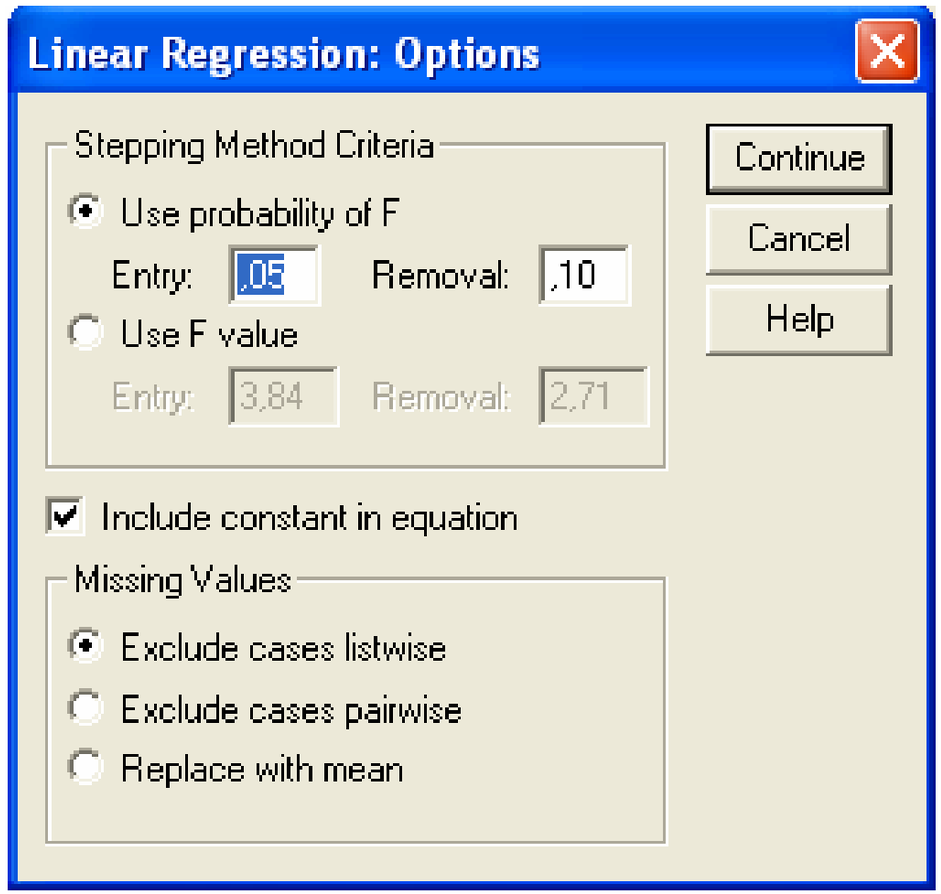
4. Choose Plots + Continue:



5. Save possibilities are:



6. Adjust Options if necessary + Continue:



7. Click OK in the Linear regression window to start the regression calculations.

3.5. Basic linear regression statistics

♦ Model Summary:

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,941[a]	,885	,846	4,03817	2,447

a. Predictors: (Constant), x2DryingTime, x1PercentageCotton

b. Dependent Variable: TensileStrength

R Square, the coefficient of determination, is the squared value of the (multiple) correlation coefficient R. It shows the percent of overall variation in data sample which is explained by the model. The higher the R-square the better the fit.

R squared change. The change in the R^2 statistic that is produced by adding or deleting an independent variable. If the R^2 change associated with a variable is large, that means that the variable is a good predictor of the dependent variable. Residuals. Displays the Durbin-Watson test for serial correlation of the

residuals and casewise diagnostics for the cases meeting the selection criterion (outliers above n standard deviations). The Durbin-Watson statistic tests the null hypothesis that the residuals from an ordinary least-squares regression are not autocorrelated. The Durbin-Watson statistic ranges in value from 0 to 4. A value near 2 indicates non-autocorrelation, a value toward 0 indicates positive autocorrelation, and a value toward 4 indicates negative autocorrelation. For more details see SPSS Help/Durbin-Watson tables.

- **Analysis of variance table**

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	751,048	2	375,524	23,029	,002[a]
	Residual	97,841	6	16,307		
	Total	848,889	8			

a. Predictors: (Constant), x2DryingTime, x1PercentageCotton

b. Dependent Variable: TensileStrength

Empirical F-value. The null hypothesis (H_0) to verify is that there is no effect on

dependent variable, i.e. $b_i = 0, i = 0,1,...,n$. The alternative hypothesis (H_A) is that this is not the case, i.e. at least one of the coefficients is not zero. For instance, in the upper example in ANOVA table Sig. =0.002 < 0.05, thus we can reject H_0 in favor of H_A. This means that the model that has been estimated is not only a theoretical construct, but exists and is statistically significant.

♦ Regression Coefficients

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B		Collinearity Statistics	
		B	Std. Error	Beta			Lower Bound	Upper Bound	Tolerance	VIF
1	(Constant)	179,055	10,755		16,648	,000	152,737	205,372		
	x1PercentageCotton	,979	,582	,238	1,681	,144	-,446	2,404	,962	1,039
	x2DryingTime	10,629	1,736	,865	6,121	,001	6,380	14,878	,962	1,039

a. Dependent Variable: TensileStrength

Estimates displays Regression coefficients B, standard errors of B, standardized coefficients Beta, t value for B, and two-tailed significance level of t. If the Sig. >0.05 for one variable, this means that this variable is non-significant for the model and can be omitted in further considerations. The relative absolute magnitudes of the standardized coefficients Beta reflect their relative importance in predicting the dependent variable.

Confidence intervals displays 95%-confidence intervals for each regression coefficient, or a covariance matrix.

Collinearity diagnostics can be observed by the values of variance inflation factors (VIF) and tolerances for individual variables. Collinearity (or multicollinearity) is the undesirable situation when one independent variable is a linear function of other independent variables. The tolerance for a variable is (1 - R-squared) for the regression of that variable on all the other independents, ignoring the dependent. When tolerance is close to 0 there is high multicollinearity of that variable with other independents and the

coefficients will be unstable. VIF is the variance inflation factor, which is simply the reciprocal of tolerance. Therefore, when VIF is high there is high multicollinearity and instability of the coefficients. As a rule, if tolerance is less than .20, a problem with multicollinearity is indicated. If VIF is about 1 the multicollinearuty is not the case.

More statistics

Descriptives. Provides the number of valid cases, the mean, and the standard deviation for each variable in the analysis. A correlation matrix with a one-tailed significance level and the number of cases for each correlation are also displayed.

Partial Correlation. The correlation that remains between two variables after removing the correlation that is due to their mutual association with the other variables. The correlation between the dependent variable and an independent variable when the linear effects of the other independent variables in the model have been removed from both.

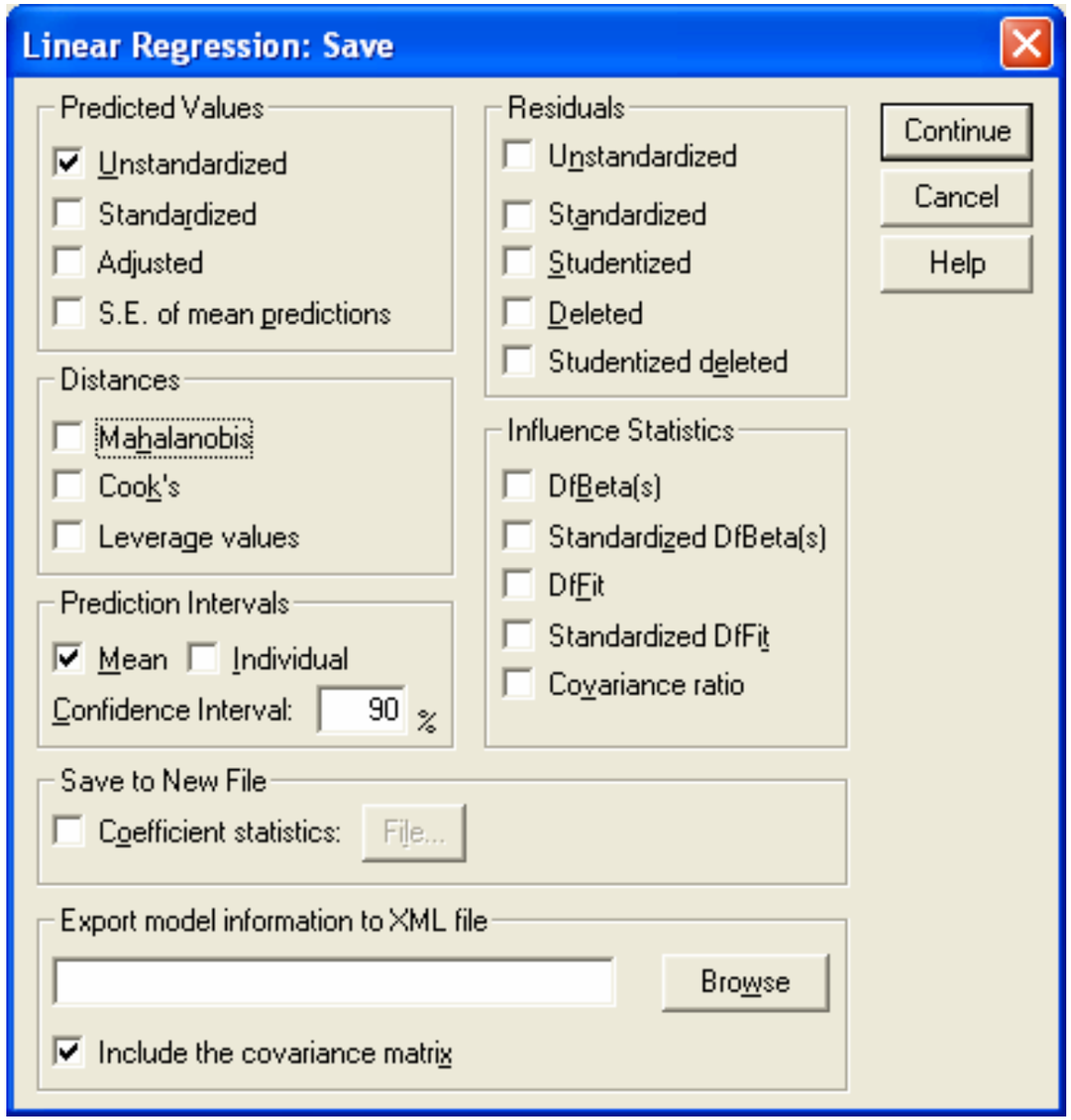
Part Correlation. The correlation between the dependent variable and an independent variable when the linear effects of the other independent variables in the model have been removed from the independent variable. It is related to the change in R squared when a variable is added to an equation. Sometimes called the semipartial correlation.

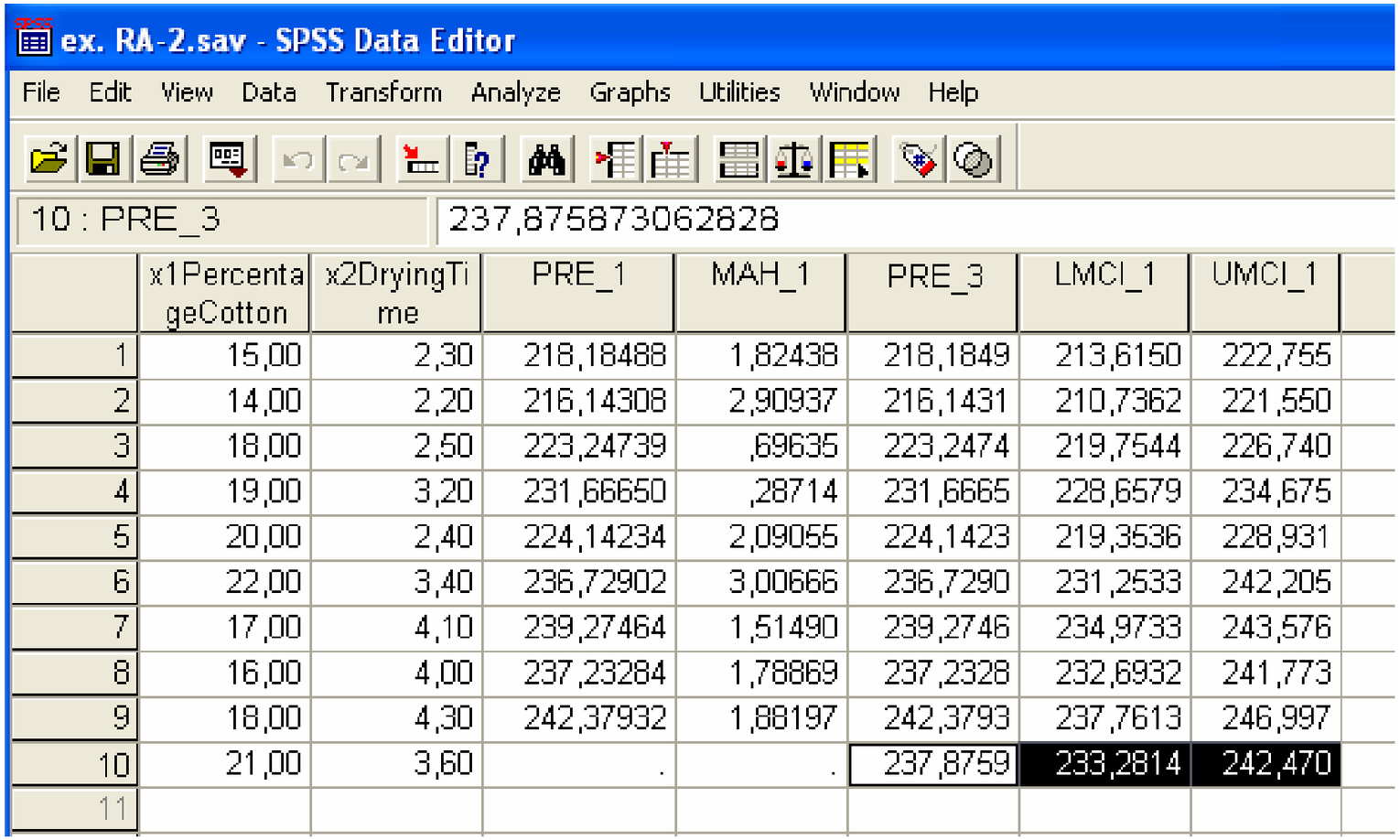
With our example RA-2 we obtained:

(a) Very good coefficient of determination Rsquared=0,885, so the model is accounted for 88,5% of the data sample. Durbin-Watson is = 2,447, approximately closed to 2 – the autocorrelation in residuals can be considered as missing. However that exact values of the possible values can be seen in the SPSS D-W tables. From ANOVA the Sig.=0,002<0,005, thus the model in total is valid. The coefficients table show, that the first independent variable x1PercentageCotton has a Sig. level = 0,144, or it is non-significant for the data used. This variable could be omitted in the model. This way we finally obtain the linear model:

$$\text{TensileStrenght} = 179,055 + 10,629 * \text{x2DryingTime} \qquad (7)$$

(b) To obtain the prediction value for x1=21 and x2=3,6 we add this data to the sample. Now we might change the Save box in Linear regression windows as it is shown below (for 90% percent confidence interval) and repeat the regression analysis. We find the approximate predicted value Tensile=237,88 with prediction interval [233,28; 242,47].


Linear Regression: Save
Predicted Values
Unstandardized
Standardized
Adjusted
S.E. of mean predictions
Residuals
Unstandardized
Standardized
Studentized
Deleted
Studentized deleted
Continue
Cancel
Help
Distances
Mahalanobis
Cook's
Leverage values
Influence Statistics
DfBeta(s)
Standardized DfBeta(s)
DfFit
Standardized DfFit
Covariance ratio
Prediction Intervals
Mean
Individual
Confidence Interval:
90
%
Save to New File
Coefficient statistics:
File...
Export model information to XML file
Browse
Include the covariance matrix



	x1PercentageCotton	x2DryingTime	PRE_1	MAH_1	PRE_3	LMCI_1	UMCI_1
1	15,00	2,30	218,18488	1,82438	218,1849	213,6150	222,755
2	14,00	2,20	216,14308	2,90937	216,1431	210,7362	221,550
3	18,00	2,50	223,24739	,69635	223,2474	219,7544	226,740
4	19,00	3,20	231,66650	,28714	231,6665	228,6579	234,675
5	20,00	2,40	224,14234	2,09055	224,1423	219,3536	228,931
6	22,00	3,40	236,72902	3,00666	236,7290	231,2533	242,205
7	17,00	4,10	239,27464	1,51490	239,2746	234,9733	243,576
8	16,00	4,00	237,23284	1,78869	237,2328	232,6932	241,773
9	18,00	4,30	242,37932	1,88197	242,3793	237,7613	246,997
10	21,00	3,60	.	.	237,8759	233,2814	242,470
11							

3.6. Remarks and further solutions of Example 4 (RA-2)

In 3.5. we obtained the Linear regression model of the TensileStrenght variable by using x1 and x2 parameters in the form

$$\text{TensileStrenght} = 179,055 + 10,629 * \text{x2DryingTime}\,, \qquad (8)$$

as it was established that the variable x1 is not statistically significant.

Solution 2. Having in mind the previous remark we may repeat the regression analysis without the variable x1. We will obtain

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	,911[a]	,830	,806	4,53457	2,170

a. Predictors: (Constant), x2DryingTime

b. Dependent Variable: TensileStrength

ANOVA[b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	704,953	1	704,953	34,284	,001[a]
	Residual	143,936	7	20,562		
	Total	848,889	8			

a. Predictors: (Constant), x2DryingTime

b. Dependent Variable: TensileStrength

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.	95% Confidence Interval for B	
		B	Std. Error	Beta			Lower Bound	Upper Bound
1	(Constant)	194,554	6,221		31,273	,000	179,843	209,265
	x2DryingTime	11,198	1,912	,911	5,855	,001	6,675	15,720

a. Dependent Variable: TensileStrength

This way the more precise regression model is written in the form

$$\text{TensileStrenght} = 194,554 + 11,198 * \text{x2DryingTime}\,. \qquad (9)$$

The prediction value for (b) is 234,866 with the 90% confidence interval [225,67;

244,06]. A relative error within 1% in the predicted values occurs with respect to the previous value 237,88.

Solution 3. In many cases it is preferable to apply the regression analysis with the Stepwise method, instead of the usual Linear. This will start with all selected independent variables and will automatically exclude the variables that are not statistically significant. By choosing “Stepwise” in the Linear regression window we will obtain the same Model summary, ANOVA and Coefficients tables with additional table of the removed variables. In our example this is

Excluded Variables[b]

Model		Beta In	t	Sig.	Partial Correlation	Collinearity Statistics: Tolerance
1	x1PercentageCotton	,238[a]	1,681	,144	,566	,962

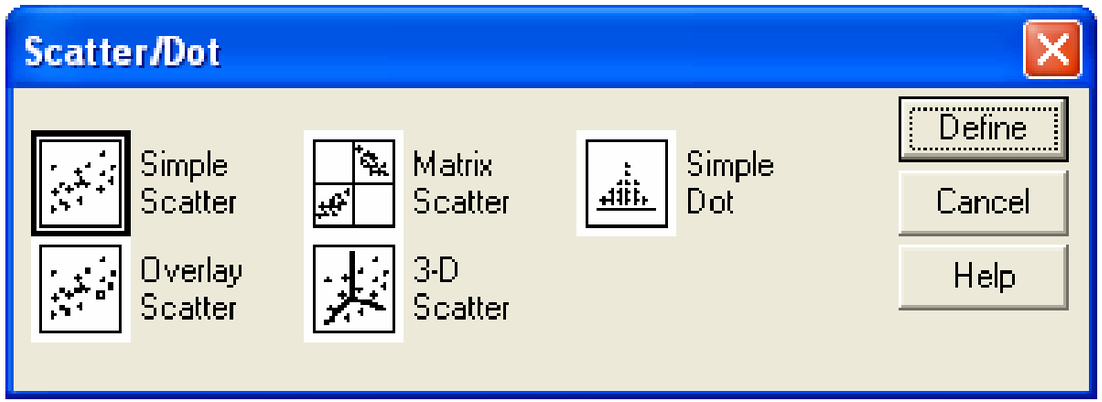
a. Predictors in the Model: (Constant), x2DryingTime

b. Dependent Variable: TensileStrength

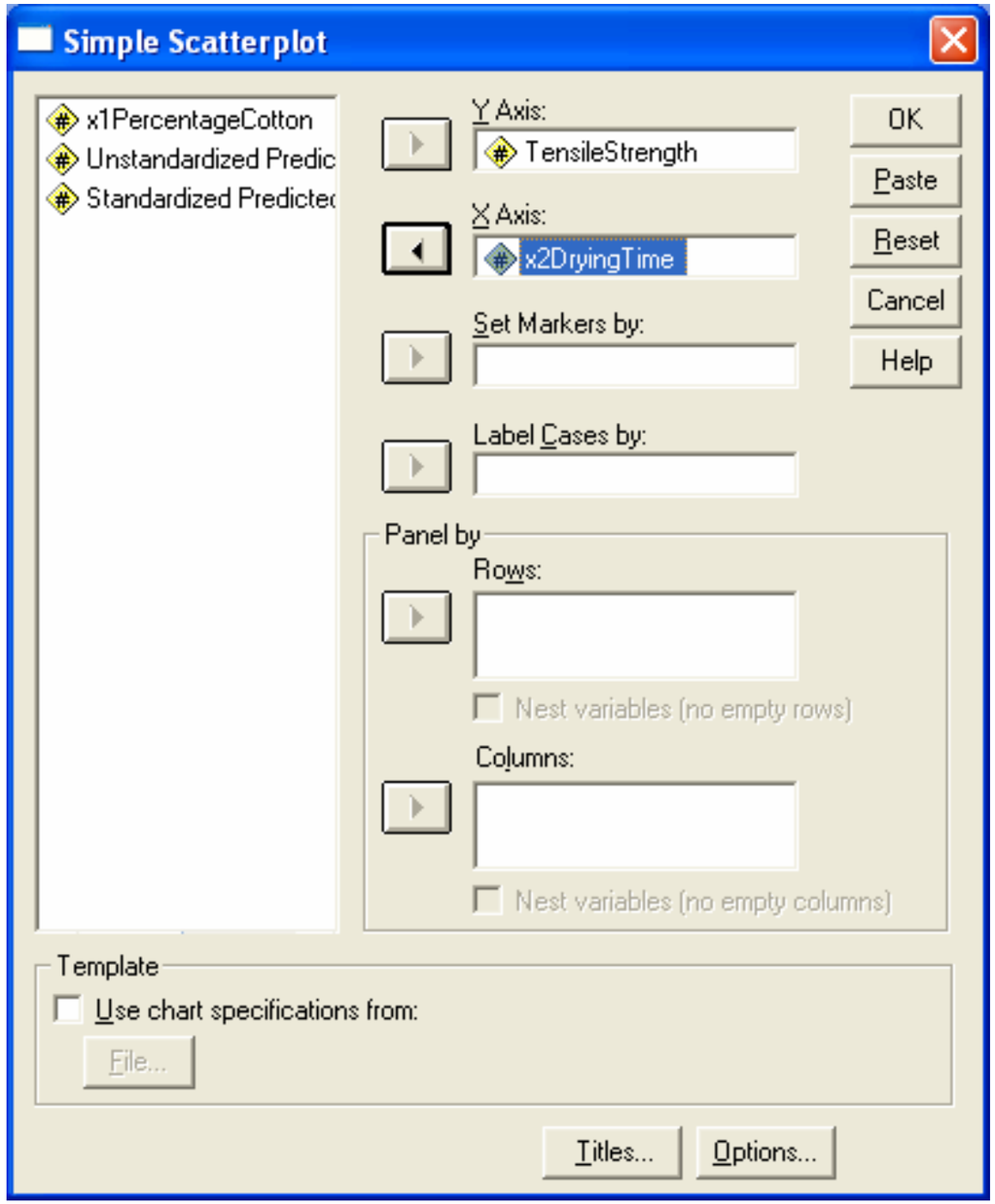
3.7. Plot of the regression line

To draw plot with the model line follow the next steps:

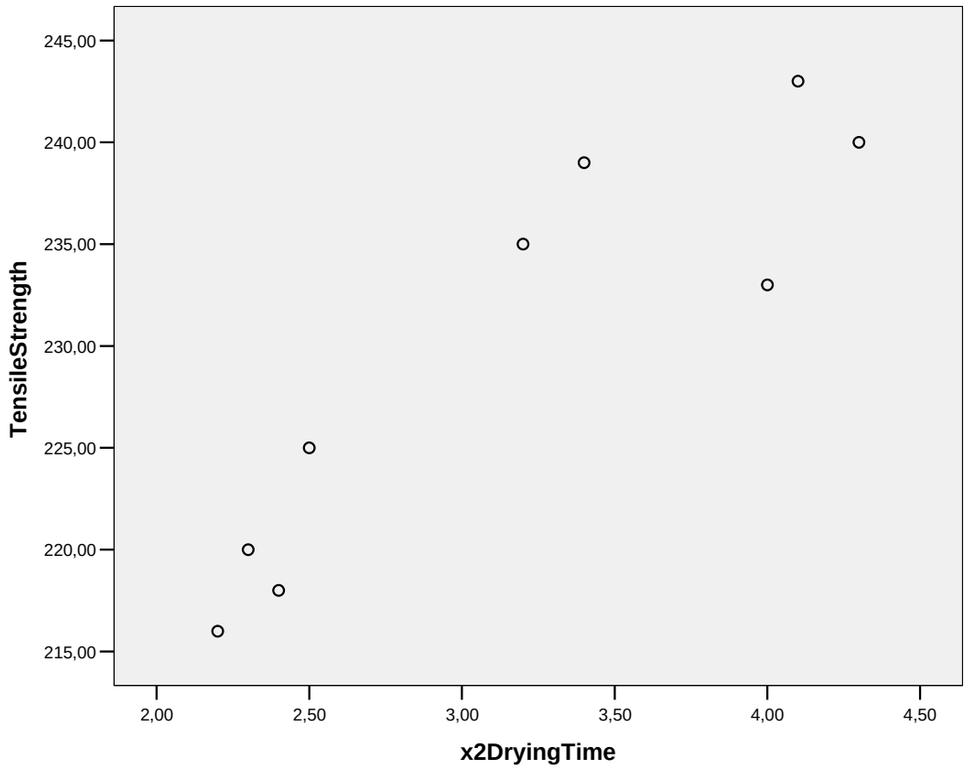
- From the main SPSS menu choose: Graphs/Scatter/Dot…



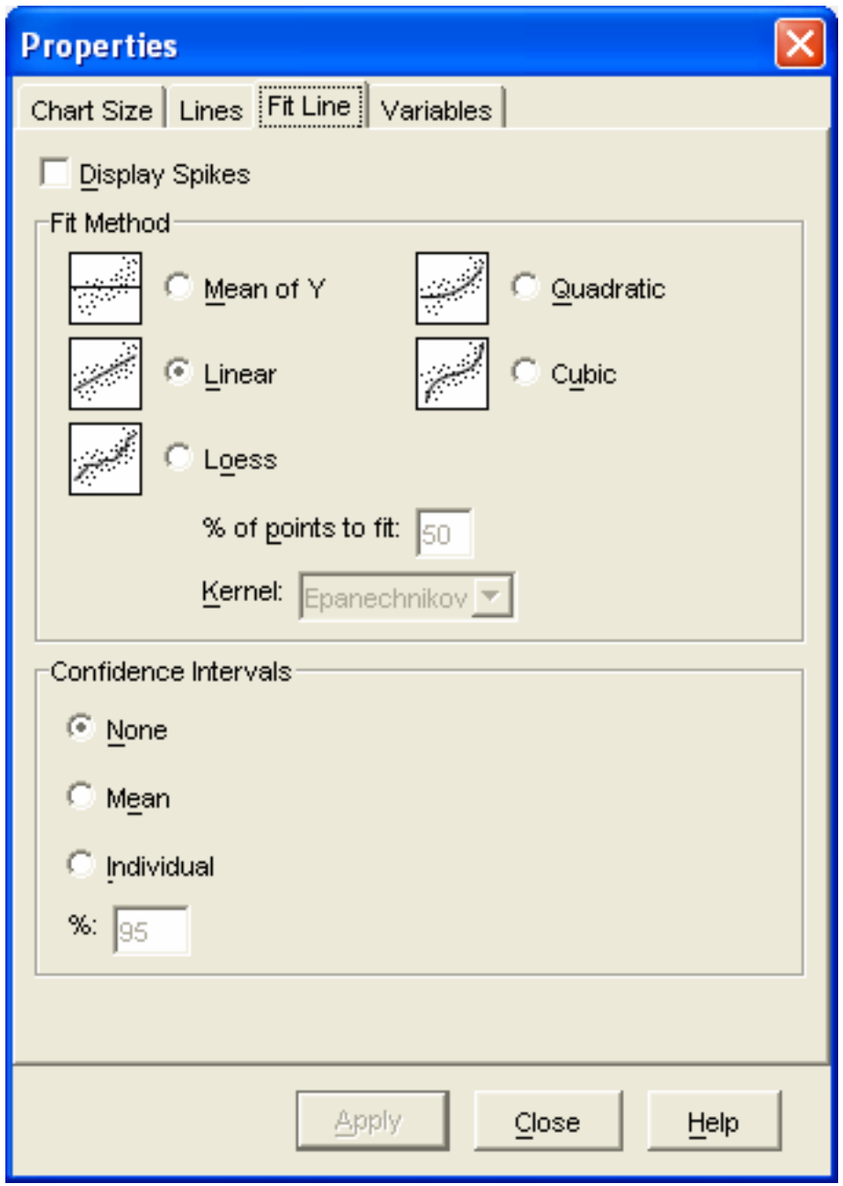
- In the Scatter/Dot window click on the Simple Scatter and validate by Define. It will appear the corresponding window, where you can move the desired dependent and independent variables to the right + OK:


Simple Scatterplot
x1PercentageCotton
Unstandardized Predic
Standardized Predicte
Y Axis:
TensileStrength
X Axis:
x2DryingTime
Set Markers by:
Label Cases by:
OK
Paste
Reset
Cancel
Help
Panel by
Rows:
Nest variables (no empty rows)
Columns:
Nest variables (no empty columns)
Template
Use chart specifications from:
File...
Titles...
Options...

♦ The scatter plot of the selected data will appear (see below).



♦ A double click on this window will open the Chart Editor. Now choose Chart/Elements/Fit Line at Total/. The window Properties appears:


Properties
Chart Size
Lines
Fit Line
Variables
Display Spikes
Fit Method
Mean of Y
Quadratic
Linear
Cubic
Loess
% of points to fit:
50
Kernel:
Epanechnikov
Confidence Intervals
None
Mean
Individual
%:
95
Apply
Close
Help

- Check the radio button Linear. The scatter plot with the model line is drawn. There is an option to show the confidence intervals.
- Click in arbitrary place outside of the Chart Editor.