1. Find researching problem i.e. collect statistical data. Choose topics that matter to you and others. The best research is targeted at issues that are relevant both scientifically (for example your own experimental data) and socially which can be found in published articles, journals, and magazines or on Internet (for example from http://www.pordata.pt/en/Home or from http://stat.wto.org/Home/WSDBHome.aspx?Language or other statistical databases). The way of collecting data is also via questionnaires. Decide on the method of sampling (random, systematic, convenience or stratified). Your data should have at least 30 cases and 2 variables which seem to you be correlated. You need program STATISTICA to analyze your statistical data.

2. Once you have selected a topic, focus on research design. Determine what type of data you collected (qualitative: ordinal, nominal or dichotomous or quantitative: discrete, quasi-continuous or continuous). To prepare this project your data should be quantitative. Determine what type of statistics you will use.

3. Build frequency tables with adequate number of classes:
   - for quasi-continuous and continuous data take ($k \approx \sqrt{n}$): [Statistics] – [Basic Statistics/Tables] – [Frequency tables] - [Advanced] – "No. of exact intervals" or "'Neat' intervals; approximate no." or "step size and starting at";
   - for discrete data: [Statistics] – [Basic Statistics/Tables] – [Frequency tables] - [Advanced] – "All distinct values"

4. Draw:
   - frequency histograms (experimental distributions):
     [Graphs] – [Histograms] – [Advanced] – [Regular] –
     "Categories" (for continuous or quasi-continuous data selecting the same number of categories as in the previous section) or
     "Integer mode" (for discrete data)

   - experimental cumulative distribution functions by changing showing type from "Standard" to "Cumulative"

5. Describe the resulting graphs due to the type of symmetry and shape. Decide what compliance hypotheses should you verify: compliance with the Poisson distribution (discrete data) or normal distribution (continuous or quasi-continuous data);

6. Calculate and comment statistical parameters: mean μ, variance σ^2, standard deviation σ, median m_0, skewness g_1, kurtosis K for both variables 1 and 2.
   [Statistics] – [Basic Statistics/Tables] – [Descriptive Statistics] - [Advanced]

7. Draw box and whisker plots.
   [Statistics] – [Basic Statistics/Tables] – [Descriptive Statistics] - [Options] – "Mean/Quartiles/Range" – [Quick] – "Box&whisker plot"

8. Check normality

9. Verify the hypothesis that your data:
   a. are normally distributed if your data are continuous or quasi-continuous type
      - using the chi^2, Kolmogorov tests;

[Statistics] – [Distribution Fitting] – [Continuous Distributions] – [Normal] – [Ok] – [Options] – [Parameters] – [Quick];

- using the Shapiro-Wilk and Lilliefors tests
[Statistics] – [Basic Statistics/Tables] – [Descriptive Statistics] – [Normality] – "Kolmogorov-Smirnov & Lilliefors" – "Shapiro-Wilk's test" – [Frequency Tables] or [Histograms];

- construct normality graphs
[Statistics] – [Basic Statistics/Tables] – [Descriptive Statistics] – [Prob.&Scatterplots] – [Normal Probability Plot].

b. follow Poisson distribution if your data are discrete type
- using the chi^2, Kolmogorov tests;
[Statistics] – [Distribution Fitting] – [Discrete Distributions] – [Poisson] – [Ok] – [Options] – [Parameters] – [Quick];

10. Comment above result. Should you reject the null hypothesis or not? Compare test's results. Decide to reject the null hypothesis (if $p < 0.05$) or accept the null hypothesis (if $p > 0.05$).

11. Investigate a dependence between your two variables:
a) scatterplots: Verify the hypothesis about the significance of the correlation coefficient. In case of significance of this factor ($p<0.05$) decide which variable is the dependent variable (described), and which is independent variable (describing variable). On the basis of this decision construct an appropriate regression line. Explain the importance of the coefficients *a* and *b* in this equation and the value of the correlation coefficient *r* and determination coefficient R^2.
[Graphs] – [Scatterplots] – [Advanced] – "R square", "Corr. And p", "Regression (fit) equation"

b) Correlation matrix: [Statistics] – [Basic Statistics/Tables] – [Correlation matrices] – [Two lists (rect. matrix)]
In the obtained correlation matrix a correlation coefficient *r* is calculated. If it is marked in red it will mean that tested variables are dependent (the null hypothesis of independence of these variables i.e that correlation coefficient *r* is equal zero should be rejected ($p <0.05$)).

c) Bivariate histogram: [Graphs] – [3D Sequential Graphs] – [Bivariate Histograms]
If the highest columns lie along or near the main diagonal, it will mean that the dependence is directly proportional. If they lie along or near the anti-diagonal it will mean that the relationship is inversely proportional.