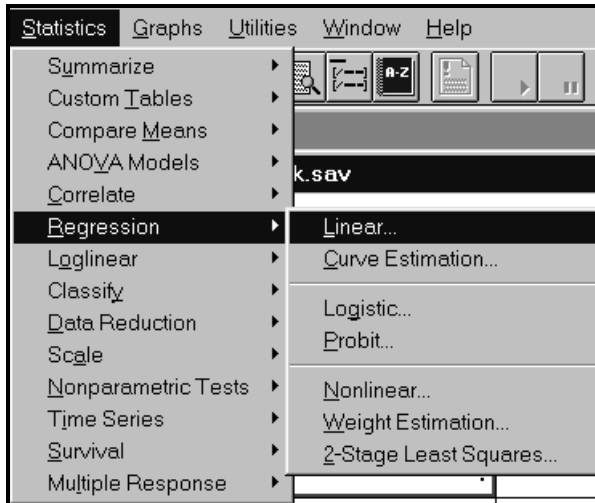


WINE CONSUMPTION AND HEART DISEASE

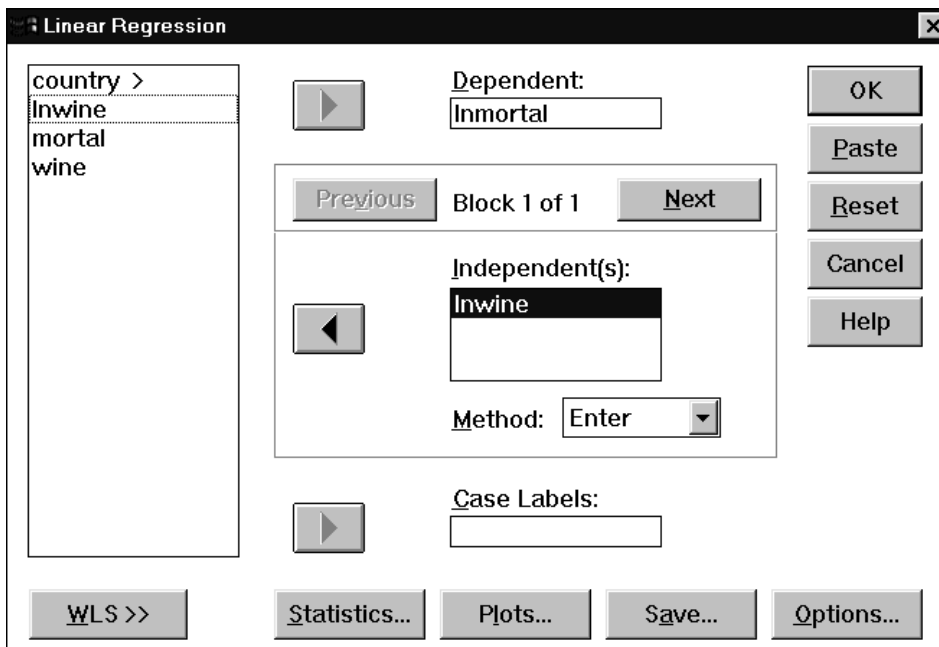
8. Linear Regression Analysis with SPSS

In this section we will demonstrate how to apply the linear regression procedure in SPSS to the wine consumption and heart disease data. The numerical outputs and graphical displays for the data were discussed in Sections 5 and 6.

In order to access the simple linear regression procedure in SPSS, choose *Statistics* and then *Regression*. The following menu will be displayed:



Click on *Linear* to open the *Linear Regression* dialog box. The five variable names *country*, *lnmortal*, *lnwine*, *mortal*, *wine* will appear in the left-hand box. In this case the dependent variable is *lnmortal* and the independent variable is *lnwine*. Transfer these variables names into appropriate boxes in the dialog box by clicking on the variable name and then on >.



Click on *Statistics...* tab to obtain statistics for the linear regression. *Estimates* of the regression coefficients and *Model fit* are default options. In order to obtain the mean and standard deviation for each variable, and a correlation matrix, check *Descriptives* box. It is also very useful to obtain the 95% confidence intervals for the regression coefficients.

Linear Regression: Statistics

Regression Coefficients

- Estimates
- Confidence intervals
- Covariance matrix

Descriptives

- Descriptives
- Model fit
- Block summary
- Durbin-Watson
- Collinearity diagnostics

Continue
Cancel
Help

Click on *Continue* to return to the *Linear Regression* dialog box. Information about residuals is obtained by clicking on the *Plots* tab. *Linear Regression: Plots* dialog box opens.

Linear Regression: Plots

DEPENDNT

- *ZPRED
- *ZRESID
- *DRESID
- *ADJPRED
- *SRESID
- *SDRESID

Previous Scatter 1 of 1 Next

Y:

X:

Produce all partial plots

Standardized Residual Plots

- Histogram
- Casewise plot
 - Outliers outside std. deviations
 - All cases
- Normal probability plot

Continue
Cancel
Help

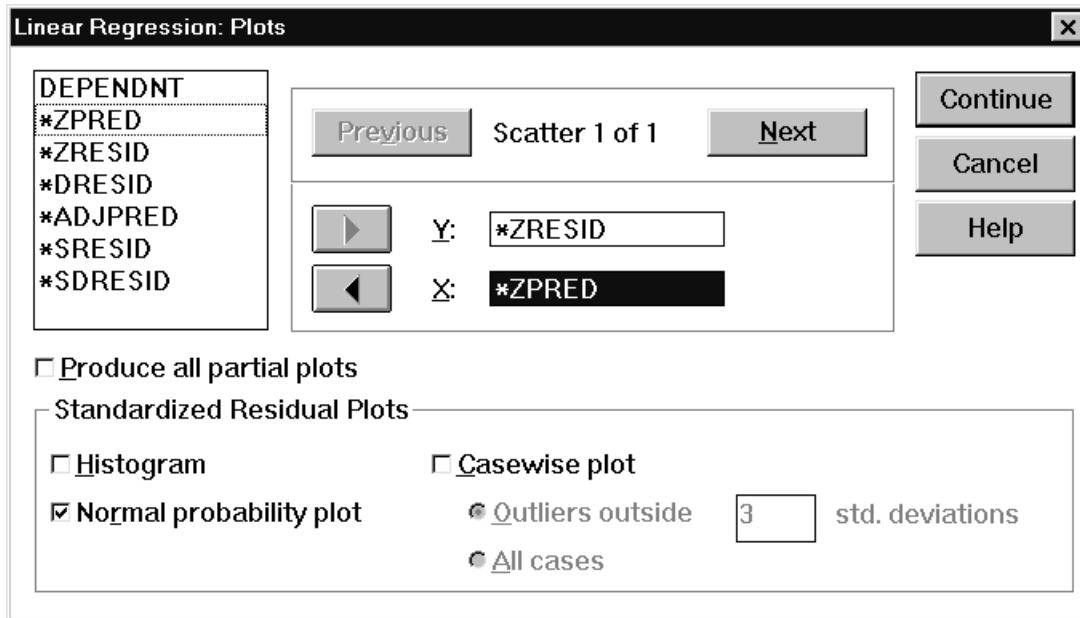
Plots can aid in the validation of the assumptions of normality, linearity, and equality of variances. Plots also allow you to check whether there are any cases, which might be considered as outliers and so dropped from the analysis. Click on the *Casewise plot* check box to obtain a listing of any exceptionally large residuals. We recommend that this is done for an initial run of the procedure. Click on *Continue* and the on *OK* to run the regression for the first time.

No outliers have been found in the wine consumption and heart disease data. The listing states: *No outliers found. No casewise plot produced.*

Now we can request a plot of residuals against the predicted values. The plot allows you to check whether the assumption of constant variance is not violated. We then look for a change in the spread or dispersion of the plotted points.

SPSS creates several temporary variables (prefaced with *) during execution of a regression analysis. *PRED comprises the unstandardized predicted values, *RESID is the set of unstandardized residuals, *ZPRED contains the standardized predicted values (i.e. *PRED has been transformed to a scale with mean 0 and standard deviation of 1), and *ZRESID comprises the standardized residuals (i.e. *RESID standardized to a scale with mean 0 and standard deviation of 1).

In order to obtain a plot of residuals against the predicted values, click on *ZRESID and then on > to transfer the variable to the Y: box, and on *ZPRED and then on > to transfer the variable to the X: box. The completed *Linear Regression: Plots* dialog box is shown below.



The normality assumption can be verified by looking at the plot of residuals. In order to assess whether the normality assumption is not violated with SPSS, the normal P-P plot of regression standardized residuals is obtained. The plot plots the cumulative proportions of standardized residuals against the cumulative proportions of the normal distribution. If the normality assumption is not violated, points will cluster around a straight line. In order to obtain the plot, check the *Normal probability plot* box.

Click on *Continue* to return to the *Linear Regression* dialog box. Click on *Save...* to save predicted values, residuals, and other statistics useful for diagnostics. Each selection adds one or more new variables to your active data file. In particular, the predicted values (mean estimates) are obtained as the variables *pre_1* and *sep_1*, respectively. The *Linear Regression: Save New Variables* dialog box is displayed below.

Linear Regression: Save New Variables

Unstandardized
 Standardized
 Adjusted
 S.E. of mean predictions

Mahalanobis
 Cook's
 Leverage values

Mean Individual
Confidence Interval: %

Unstandardized
 Standardized
 Studentized
 Deleted
 Studentized deleted

DfBeta(s)
 Standardized DfBeta(s)
 DfFit
 Standardized DfFit
 Covariance ratio

Continue
Cancel
Help

Click on *Continue* to return to the *Linear Regression* dialog box. Click on *OK* to obtain the linear regression output. The numerical output is discussed in **Section 5**. The plots obtained to validate the assumptions of the linear regression model are discussed in **Section 6**.