

Unusual Values of Independent Values

Leverage is the distance between the values of the independent variables for a case and the average for all cases. <<Scatter diagram: **Sequence * Leverage** (X*Y)>>

Mahalanobis distances is the leverage multiplied by one less than the number of cases ($n - 1$). <<Scatter diagram: **Sequence * Mahalanobis Distances** (X*Y)>>

Leverage and Mahalanobis distances are very similar.

Influential Points

Cook's distance is a summary measure that is based on how much regression coefficients for all variables change when a case is removed from the regression.

<<Changes in all coefficients>> <<Scatter diagram: **Sequence * Cook's distance** (X*Y)>>

DfBeta and Standardized DfBeta measure the effect on each coefficient of deleting each case from the regression equation. <<Changes in individual coefficients>>

<<Scatter diagram: **Sequence * Independent Variable DfBeta** (X*Y)>>

DfFit and Standardized DfFit examine the change in the predicted value when a case is deleted from the computation of the model. <<Changes in predicted values>>
<<Scatter diagram: **Sequence * DfFit** (X*Y)>>

Deleted residual and Studentized deleted residual are used to find out unusual points. <<Changes in residuals>> <<Scatter diagram: **Deleted residual * Studentized deleted residual** (X*Y). Expect a straight line>>

Partial Regression Plots

Purpose: To **check linearity** and **identify influential observations**.

Scatter diagram: The residuals when the dependent variable is predicted from all of the variables except that variable, and the residuals when that independent variable is predicted from all of the other independent variables.

Interpretation:

If the plot shows a linear relationship, and it does. If the points fall along a curve, the relationship between the two variables may not be linear. Transformation, such as variable squared, may be used.

Points that are far removed from the rest are influential points for determining the regression coefficient for the variable.