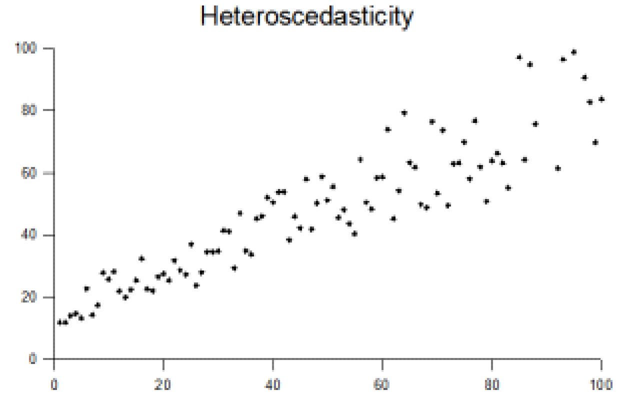
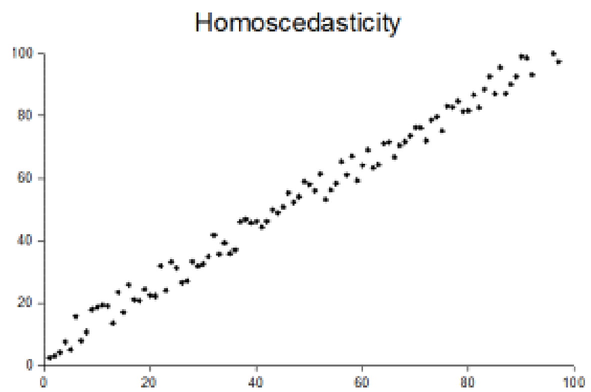
Homoscedasticity

A simple bivariate example can help to illustrate heteroscedasticity:

Imagine we have data on family income and spending on luxury items. Using bivariate regression, we use family income to predict luxury spending. As expected, there is a strong, positive association between income and spending. Upon examining the residuals we detect a problem – the residuals are very small for low values of family income (almost all families with low incomes don't spend much on luxury items) while there is great variation in the size of the residuals for wealthier families (some families spend a great deal on luxury items while some are more moderate in their luxury spending). This situation represents heteroscedasticity because the size of the error varies across values of the independent variable. Examining a scatterplot of the residuals against the predicted values of the dependent variable would show a classic cone-shaped pattern of heteroscedasticity.





Simply put, homoscedasticity means "having the same scatter." For it to exist in a set of data, the points must be about the same distance from the line, as shown in the picture above. The opposite is heteroscedasticity ("different scatter"), where points are at widely varying distances from the regression line.

Testing for Homogeneity of Variance

Tests that you can run to check your data meets this assumption include:

- Bartlett's Test
- Box's M Test
- Brown-Forsythe Test
- Hartley's Fmax test
- Levene's Test