Recall the linear regression model with $k$ regressors:

$$y_i = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik} + \varepsilon_i$$

Assumptions:

1. The error term has zero mean, or equivalently,
   $$E(y_i) = \beta_0 + \beta_1 x_{i1} + \cdots + \beta_k x_{ik}.$$  
2. The error term has constant variance.
3. The errors are uncorrelated.
4. $\varepsilon_i \sim \text{Normal}$. 
We shall examine the model adequacy by analyzing residuals.

The residual is $e_i = y_i - \hat{y}_i$, or $e = (I - H)y$ in a matrix form.

Standardized residuals to remove the scale:

$$s_i = \frac{e_i}{\hat{\sigma}} = \frac{e_i}{\sqrt{MS_{Res}}}$$

The plain residual and its plots are useful for checking the model assumptions:

- QQ plot
- Residual vs. fitted values
- Residual vs. regressors
A quantile-quantile normal plot, or simply QQ plot, plots sample quantiles vs. theoretical quantiles of a standard normal.

In the ideal case where a sample is i.i.d. from a normal distribution, we expect to see a straight line in its QQ plot.

QQ plots may help diagnose heavy/light tailed or skewed error distributions.

Possible solutions to violated normal assumption:

- to cite George Box’s quote
- robust regression
- transformation of response

It is sometimes well behaved even if the errors are not normal.
Residuals vs. fitted values

- We expect to see a random scatter of points around the horizontal axis.
- This is because: under MLR, $e'\hat{y} = 0$; under the normal assumption, $e$ and $\hat{y}$ are independent.

(a) Satisfactory; (b) and (c) Heterogeneous variances; (d) Nonlinearity
Possible solutions: transformation of response/regressor, adding polynomial terms, etc.
It is satisfactory to have a horizontal band containing the residuals without any clear pattern.

For example, the plots below exhibit a pattern of autocorrelation.

(a) Positively correlated errors; (b) Negatively correlated errors.

Possible solutions: to build a time series model that specifically addresses the autocorrelation.