

STAT 302 - Chapter 1 : Simple Linear Regression - Part 2

Harsha Perera

Assessing Conditions

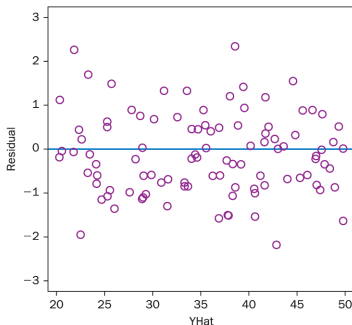
- ▶ Linearity
- ▶ Constant Variance
- ▶ Normality

Linearity

- ▶ A scatterplot with the fitted regression line provides one visual method of checking linearity. Points will be randomly scattered above and below the line when the linear model is appropriate.
- ▶ Clusters of points above and below the line in a systematic fashion, indicate that the simple linear model is not appropriate.
- ▶ A more informative plot is a scatterplot of the residuals versus the fitted values.
- ▶ In this scatterplot the regression line is represented as a horizontal line through zero. This plot allows us to focus on the estimated errors and look for any clear patterns.

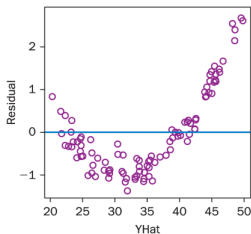
Constant Variance

- ▶ Use a scatterplot of the residuals versus the fitted values as before.
- ▶ The ideal pattern will be a random variation above and below zero in a band of relatively constant width.
- ▶ The figure below shows a typical residual versus fitted values scatterplot when both **Linearity** and **Constant Variance** conditions are satisfied.

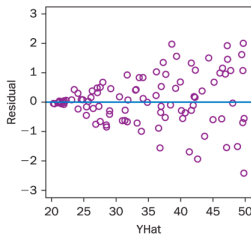


Residuals versus Fitted Values Plots

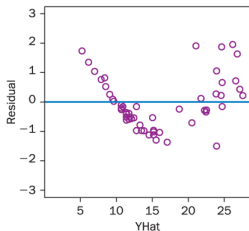
The figure below shows some examples of residuals versus fitted values plots that exhibit some typical patterns indicating a problem with **Linearity**, **Constant Variance**, or **both conditions**.



(a) Nonlinear



(b) Nonconstant variance

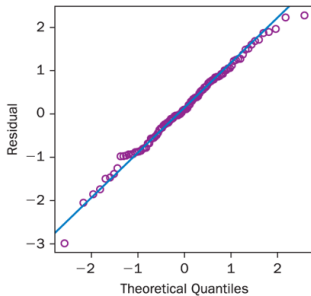


(c) Both nonlinear and nonconstant variance

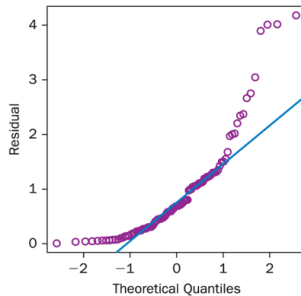
Normality

- ▶ **Normal Plots** - Different view of the data where an ideal pattern for a normal sample is a straight line. There are generally two common methods for constructing a normal plot. They are **Normal Quantile Plot** and **Normal Probability Plot**.
- ▶ **Normal Quantile Plot** - Scatterplot of the ordered observed data versus values (the theoretical quantiles) that we would expect to see from a "perfect" normal sample of the same size. If the ordered data are coming from a normal sample the resulting scatterplot is a **straight line**. If the distribution of the data is skewed or has long tails due to some extreme outliers, then the **normal quantile plot** will bend away from a straight line.

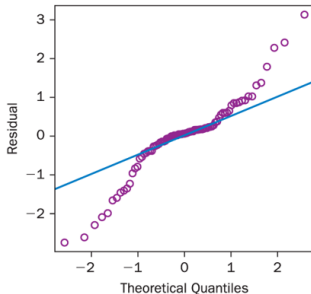
Normal Quantile Plot - Examples



(a) Normal residuals



(b) Skewed right residuals

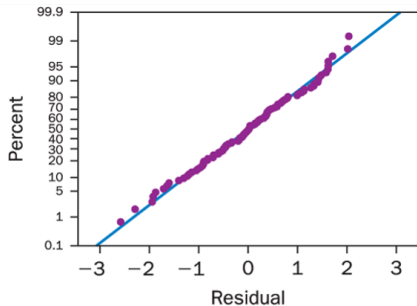


(c) Long-tailed residuals

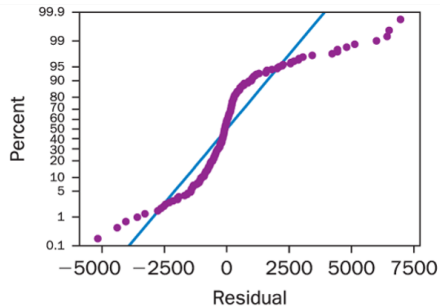
Normality

- ▶ **Normal Probability Plot** - Residuals are plotted on the horizontal axis (x-axis) while the vertical axis (y-axis) is the rate that normal probability grow.
- ▶ The interpretation is same as the normal quantile plot. A linear pattern indicates a good agreement with normality, and curvature or bending away from a straight line shows a departures from normality.

Normal Probability Plot - Examples



(a) Normal residuals



(b) Nonnormal residuals