**7 Classical Assumptions of Ordinary Least Squares (OLS) Linear Regression**

June 1, 2018 By [Jim Frost](http://statisticsbyjim.com/author/statis11_wp/) [7 Comments](http://statisticsbyjim.com/regression/ols-linear-regression-assumptions/#comments)

[Ordinary Least Squares](http://statisticsbyjim.com/glossary/ordinary-least-squares/) (OLS) is the most common estimation method for linear models—and that’s true for a good reason. As long as your model satisfies the OLS assumptions for linear [regression](http://statisticsbyjim.com/glossary/regression-analysis/), you can rest easy knowing that you’re getting the best possible [estimates](http://statisticsbyjim.com/glossary/estimator/).

Regression is a powerful analysis that can analyze multiple variables simultaneously to answer complex research questions. However, if you don’t satisfy the OLS assumptions, you might not be able to trust the results.

In this post, I cover the OLS linear regression assumptions, why they’re essential, and help you determine whether your model satisfies the assumptions.

**What Does OLS Estimate and What are Good Estimates?**

First, a bit of context.

Regression analysis is like other [inferential methodologies](http://statisticsbyjim.com/basics/descriptive-inferential-statistics/). Our goal is to draw a [random sample](http://statisticsbyjim.com/glossary/sample/) from a [population](http://statisticsbyjim.com/glossary/population/) and use it to estimate the properties of that population.

In regression analysis, the [coefficients](http://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/) in the regression equation are estimates of the actual population [parameters](http://statisticsbyjim.com/glossary/parameter/). We want these [coefficient](http://statisticsbyjim.com/glossary/regression-coefficient/) estimates to be the best possible estimates!

Suppose you request an estimate—say for the cost of a service that you are considering. How would you define a reasonable estimate?

1. The estimates should tend to be right on target. They should not be systematically too high or too low. In other words, they should be unbiased or correct on [average](http://statisticsbyjim.com/glossary/mean/).
2. Recognizing that estimates are almost never exactly correct, you want to minimize the discrepancy between the estimated value and actual value. Large differences are bad!

These two properties are exactly what we need for our coefficient estimates!

When your linear regression model satisfies the OLS assumptions, the procedure generates unbiased coefficient estimates that tend to be relatively close to the true population values (minimum variance). In fact, the Gauss-Markov theorem states that OLS produces estimates that are better than estimates from all other linear model estimation methods when the assumptions hold true.

For more information about the implications of this theorem on OLS estimates, read my post: [The Gauss-Markov Theorem and BLUE OLS Coefficient Estimates](http://statisticsbyjim.com/regression/gauss-markov-theorem-ols-blue/).

The Seven Classical OLS Assumptions

Like many statistical analyses, ordinary least squares (OLS) regression has underlying assumptions. When these classical assumptions for linear regression are true, ordinary least squares produces the best estimates. However, if some of these assumptions are not true, you might need to employ remedial measures or use other estimation methods to improve the results.

Many of these assumptions describe properties of the error term. Unfortunately, the error term is a population value that we’ll never know. Instead, we’ll use the next best thing that is available—the [residuals](http://statisticsbyjim.com/glossary/residuals/). Residuals are the sample estimate of the error for each observation.

Residuals = Observed value – the [fitted value](http://statisticsbyjim.com/glossary/fitted-values/)

When it comes to checking OLS assumptions, assessing the residuals is crucial!

There are seven classical OLS assumptions for linear regression. The first six are mandatory to produce the best estimates. While the quality of the estimates does not depend on the seventh assumption, analysts often evaluate it for other important reasons that I’ll cover.

**OLS Assumption 1: The regression model is linear in the coefficients and the error term**

This assumption addresses the functional form of the model. In [statistics](http://statisticsbyjim.com/glossary/statistics/), a regression model is linear when all terms in the model are either the constant or a parameter multiplied by an [independent variable](http://statisticsbyjim.com/glossary/predictor-variables/). You build the model equation only by adding the terms together. These rules constrain the model to one type:

Y =\beta _{0} + \beta _{1}X_{1} + \beta _{2}X_{2} + \cdots + \beta _{k}X_{k} + \epsilon

In the equation, the betas (βs) are the parameters that OLS estimates. Epsilon (ε) is the random error.

In fact, the defining characteristic of linear regression is this functional form of the *parameters* rather than the ability to model curvature. Linear models can model curvature by including nonlinear *variables* such as polynomials and transforming exponential functions.

To satisfy this assumption, the correctly specified model must fit the linear pattern.

**Related posts**: [The Difference Between Linear and Nonlinear Regression](http://statisticsbyjim.com/regression/difference-between-linear-nonlinear-regression-models/) and [How to Specify a Regression Model](http://statisticsbyjim.com/regression/model-specification-variable-selection/)

**OLS Assumption 2: The error term has a population mean of zero**

The error term accounts for the variation in the [dependent variable](http://statisticsbyjim.com/glossary/response-variables/) that the independent variables do not explain. Random chance should determine the values of the error term. For your model to be unbiased, the average value of the error term must equal zero.

After all, if the average error is +7, this non-zero error indicates that our model systematically underpredicts the observed values. Statisticians refer to systematic error like this as bias, and it signifies that our model is inadequate because it is not correct on average.

Stated another way, we want the expected value of the error to equal zero. If the expected value is +7 rather than zero, part of the error term is predictable, and we should add that information to the regression model itself. We want only random error left for the error term.

You don’t need to worry about this assumption when you include the constant in your regression model because it forces the mean of the residuals to equal zero. For more information about this assumption, read my post about the [regression constant](http://statisticsbyjim.com/regression/interpret-constant-y-intercept-regression/).

**OLS Assumption 3: All independent variables are uncorrelated with the error term**

If an independent variable is correlated with the error term, we can use the independent variable to predict the error term, which violates the notion that the error term represents unpredictable random error. We need to find a way to incorporate that information into the regression model itself.

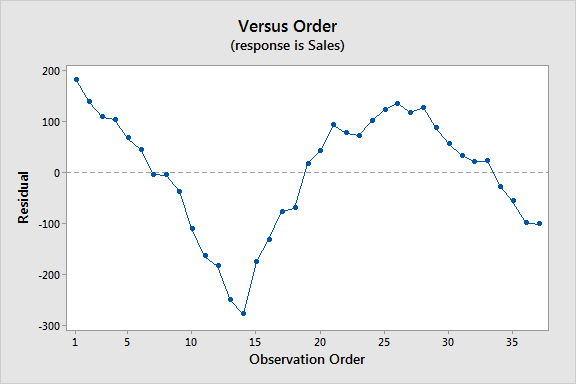
This assumption is also referred to as exogeneity. When this type of [correlation](http://statisticsbyjim.com/glossary/correlation/) exists, there is endogeneity. Violations of this assumption can occur because there is simultaneity between the independent and dependent variables, omitted variable bias, or measurement error in the independent variables.

Violating this assumption biases the coefficient estimate by making it larger (in absolute terms) than it should be. To understand why this bias occurs, keep in mind that the error term always explains some of the variability in the dependent variable. However, when an independent variable is correlated with the error term, OLS incorrectly attributes some of the variance that the error term actually explains to the independent variable instead.

**OLS Assumption 4: Observations of the error term are uncorrelated with each other**

One observation of the error term should not predict the next observation. For instance, if the error for one observation is positive and that systematically increases the probability that the following error is positive, that is a positive correlation. If the subsequent error is more likely to have the opposite sign, that is a negative correlation. This problem is known both as serial correlation and autocorrelation.

Assess this assumption by graphing the residuals in the order that the data were collected. You want to see a randomness in the plot. In the graph for a sales model, there appears to be a cyclical pattern with a positive correlation.



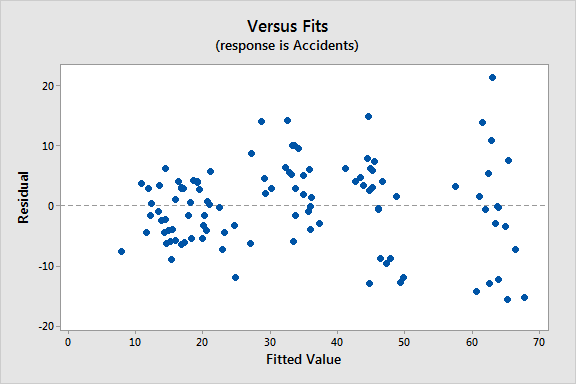
As I’ve explained, if you have information that allows you to predict the error term for an observation, you need to incorporate that information into the model itself. Serial correlation is most likely to occur in time series models. To resolve this issue, you might need to add an independent variable to the model that captures this information. For the sales model above, we probably need to add variables that explains the cyclical pattern.

Serial correlation reduces the precision of OLS estimates.

**OLS Assumption 5: The error term has a constant variance (no heteroscedasticity)**

The variance of the errors should be consistent for all observations. In other words, the variance does not change for each observation or for a range of observations. This preferred condition is known as homoscedasticity (same scatter). If the variance changes, we refer to that as heteroscedasticity (different scatter).

The easiest way to check this assumption is to create a residuals versus fitted value plot. On this type of graph, heteroscedasticity appears as a cone shape where the spread of the residuals increases in one direction. In the graph below, the spread of the residuals increases as the fitted value increases.



Heteroscedasticity reduces the precision of the estimates in OLS linear regression.

**Related post**: [Heteroscedasticity in Regression Analysis](http://statisticsbyjim.com/regression/heteroscedasticity-regression/" \t "_blank)

Note: When assumption 5 (no autocorrelation) and 6 (homoscedasticity) are both true, statisticians say that the error term is independent and identically distributed (IID) and refer to them as spherical errors.

**OLS Assumption 6: No independent variable is a perfect linear function of other explanatory variables**

Perfect correlation occurs when two variables have a [Pearson’s correlation coefficient](http://statisticsbyjim.com/basics/correlations/) of +1 or -1. When one of the variables changes, the other variable also changes by a completely fixed proportion. The two variables move in unison.

Perfect correlation suggests that two variables are different forms of the same variable. For example, games won and games lost have a perfect negative correlation (-1). The temperature in Fahrenheit and Celsius have a perfect positive correlation (+1).

Ordinary least squares cannot distinguish one variable from the other when they are perfectly correlated. If you specify a model that contains independent variables with perfect correlation, your statistical software can’t fit the model, and it will display an error message. You must remove one of the variables from the model to proceed.

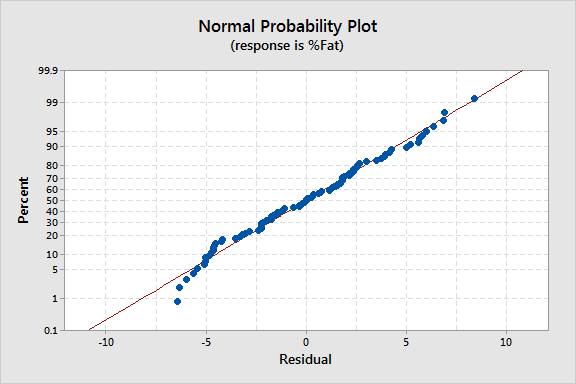
Perfect correlation is a show stopper. However, your statistical software can fit OLS regression models with imperfect but strong relationships between the independent variables. If these correlations are high enough, they can cause problems. Statisticians refer to this condition as multicollinearity, and it reduces the precision of the estimates in OLS linear regression.

**Related post**: [Multicollinearity in Regression Analysis: Problems, Detection, and Solutions](http://statisticsbyjim.com/regression/multicollinearity-in-regression-analysis/" \t "_blank)

**OLS Assumption 7: The error term is normally distributed (optional)**

OLS does not require that the error term follows a [normal distribution](http://statisticsbyjim.com/basics/normal-distribution/) to produce unbiased estimates with the minimum variance. However, satisfying this assumption allows you to perform statistical hypothesis testing and generate reliable [confidence intervals](http://statisticsbyjim.com/glossary/confidence-interval/) and [prediction intervals](http://statisticsbyjim.com/glossary/prediction-intervals/).

The easiest way to determine whether the residuals follow a normal distribution is to assess a normal probability plot. If the residuals follow the straight line on this type of graph, they are normally distributed. They look good on the plot below!



If you need to obtain [p-values for the coefficient estimates](http://statisticsbyjim.com/regression/interpret-coefficients-p-values-regression/) and the [overall test of significance](http://statisticsbyjim.com/regression/interpret-f-test-overall-significance-regression/), check this assumption!

**Why You Should Care About the Classical OLS Assumptions**

In a nutshell, your linear model should produce residuals that have a mean of zero, have a constant variance, and are not correlated with themselves or other variables.

If these assumptions hold true, the OLS procedure creates the best possible estimates. In statistics, estimators that produce unbiased estimates that have the smallest variance are referred to as being “efficient.” Efficiency is a statistical concept that compares the quality of the estimates calculated by different procedures while holding the sample size constant. OLS is the most efficient linear regression estimator when the assumptions hold true.

Another benefit of satisfying these assumptions is that as the sample size increases to infinity, the coefficient estimates converge on the actual population parameters.

If your error term also follows the normal distribution, you can safely use hypothesis testing to determine whether the independent variables and the entire model are statistically significant. You can also produce reliable confidence intervals and prediction intervals.

Knowing that you’re maximizing the value of your data by using the most efficient methodology to obtain the best possible estimates should set your mind at ease. It’s worthwhile checking these OLS assumptions! The best way to assess them is by using residual plots. To learn how to do this, read my post about [using residual plots](http://statisticsbyjim.com/regression/check-residual-plots-regression-analysis/)!