

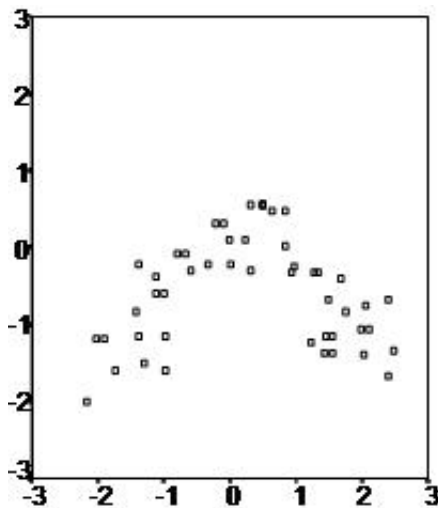
Assumptions of Multiple Linear Regression

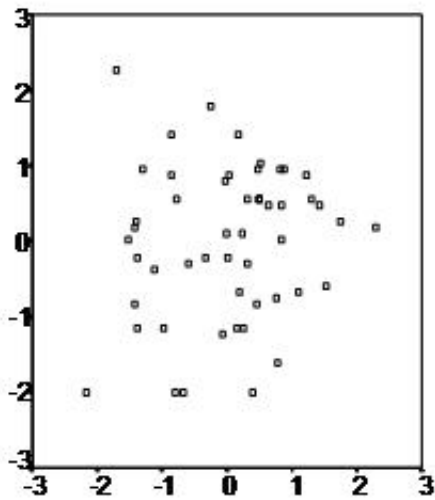
Multiple linear regression analysis makes several key assumptions:

- Linear relationship
- Multivariate normality
- No or little multicollinearity
- No auto-correlation
- Homoscedasticity

[Multiple linear regression](#) needs at least 3 variables of metric (ratio or interval) scale. A rule of thumb for the sample size is that regression analysis requires at least 20 cases per independent variable in the analysis, in the simplest case of having just two independent variables that requires $n > 40$. G*Power can also be used to calculate a more exact, appropriate sample size.

Firstly, multiple linear regression needs the relationship between the independent and dependent variables to be linear. It is also important to check for outliers since multiple linear regression is sensitive to outlier effects. The linearity assumption can best be tested with scatter plots, the following two examples depict two cases, where no and little linearity is present.





Secondly, the multiple linear regression analysis requires all variables to be normal. This assumption can best be checked with a histogram and a fitted normal curve or a Q-Q-Plot. Normality can be checked with a goodness of fit test, e.g., the Kolmogorov-Smirnov test. When the data is not normally distributed a non-linear transformation, e.g., log-transformation might fix this issue. However it can introduce effects of multicollinearity.

Thirdly, multiple linear regression assumes that there is little or no multicollinearity in the data. Multicollinearity occurs when the independent variables are not independent from each other. A second important independence assumption is that the error of the mean is uncorrelated; that is that the standard mean error of the dependent variable is independent from the independent variables.

Multicollinearity is checked against 4 key criteria:

- 1) Correlation matrix – when computing the matrix of Pearson's Bivariate Correlation among all independent variables the correlation coefficients need to be smaller than .08.
- 2) Tolerance – the tolerance measures the influence of one independent variable on all other independent variables; the tolerance is calculated with an initial linear regression analysis. Tolerance is defined as $T = 1 - R^2$ for these first step regression analysis. With T
- 3) Variance Inflation Factor (VIF) – the variance inflation factor of the linear regression is defined as $VIF = 1/T$. Similarly with $VIF > 10$ there is an indication for multicollinearity to be present.
- 4) Condition Index – the condition index is calculated using a factor analysis on the independent variables. Values of 10-30 indicate a mediocre multicollinearity in the regression variables, values > 30 indicate strong multicollinearity.

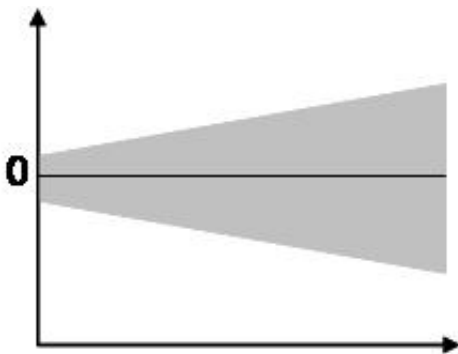
If multicollinearity is found in the data one remedy might be centering the data. To center the data you would simply deduct the mean score. This typically helps in cases where multicollinearity sneaked into the model when applying non-linear transformations to correct missing multivariate normality.

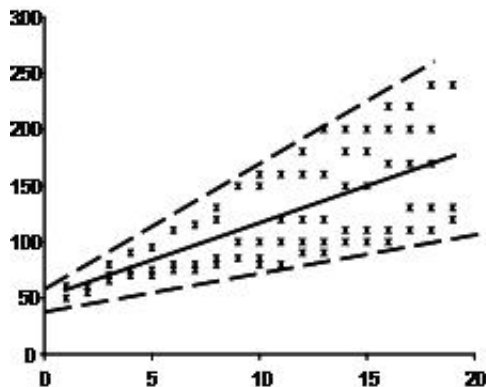
Other alternatives to tackle the problem of multicollinearity in multiple linear regression is to conduct a factor analysis before the regression analysis and to rotate the factors to insure independence of the factors in the linear regression analysis.

Fourthly, multiple linear regression analysis requires that there is little or no autocorrelation in the data. Autocorrelation occurs when the residuals are not independent from each other. In other words when the value of $y(x+1)$ is not independent from the value of $y(x)$. This for instance typically occurs in stock prices, where today's price is not independent from yesterday's price.

While a scatter plot let's you check for autocorrelations, you can test the multiple linear regression model for autocorrelation with the Durbin-Watson test. Durbin-Watson's d tests the null hypothesis that the residuals are not linearly auto-correlated. While d can assume values between 0 and 4, values around 2 indicate no autocorrelation. As a rule of thumb values of 1.5

The last assumption the multiple linear regression analysis makes is homoscedasticity. The scatter plot is good way to check whether homoscedasticity (that is the error terms along the regression line are equal) is given. If the data is heteroscedastic the scatter plots look like the following examples:





The Goldfeld-Quandt Test can test for heteroscedasticity. The test splits the multiple linear regression data in high and low value to see if the samples are significantly different. If homoscedasticity is present in our multiple linear regression model, a non-linear correction might fix the problem, but might sneak multicollinearity into the model.

Statistics Solutions can assist with your quantitative or qualitative analysis by assisting you to develop your methodology and results chapters. The services that we offer include:

Data Analysis Plan

- Edit your research questions and null/alternative hypotheses
- Write your data analysis plan; specify specific statistics to address the research questions, the assumptions of the statistics, and justify why they are the appropriate statistics; provide references
- Justify your sample size/power analysis, provide references
- Explain your data analysis plan to you so you are comfortable and confident
- Two hours of additional support with your statistician

Quantitative Results Section (Descriptive Statistics, Bivariate and Multivariate Analyses, Structural Equation Modeling, Path analysis, HLM, Cluster Analysis)

- Clean and code dataset
- Conduct descriptive statistics (i.e., mean, standard deviation, frequency and percent, as appropriate)
- Conduct analyses to examine each of your research questions
- Write-up results
- Provide APA 6th edition tables and figures
- Explain chapter 4 findings
- Ongoing support for entire results chapter statistics

****Please call 877-437-8622 to request a quote based on the specifics of your research, or***

email [**Info@StatisticsSolutions.com**](mailto:Info@StatisticsSolutions.com)

Related Pages:

Multicollinearity

[What is Multiple Linear Regression](#)

[Multiple Linear Regression Video Tutorial](#)

[Conduct and Interpret a Multiple Linear Regression](#)