

Assumptions of Logistic Regression

[Logistic regression](#) does not make many of the key assumptions of [linear regression](#) and [general linear models](#) that are based on ordinary least squares algorithms – particularly regarding linearity, normality, homoscedasticity, and measurement level.

Firstly, it does not need a linear relationship between the dependent and independent variables. [Logistic regression](#) can handle all sorts of relationships, because it applies a non-linear log transformation to the predicted odds ratio. Secondly, the independent variables do not need to be multivariate normal – although multivariate normality yields a more stable solution. Also the error terms (the residuals) do not need to be multivariate normally distributed. Thirdly, homoscedasticity is not needed. Logistic regression does not need variances to be heteroscedastic for each level of the independent variables. Lastly, it can handle ordinal and nominal data as independent variables. The independent variables do not need to be metric (interval or ratio scaled).

However some other assumptions still apply.

Binary logistic regression requires the dependent variable to be binary and ordinal logistic regression requires the dependent variable to be ordinal. Reducing an ordinal or even metric variable to dichotomous level loses a lot of information, which makes this test inferior compared to ordinal logistic regression in these cases.

Secondly, since logistic regression assumes that $P(Y=1)$ is the probability of the event occurring, it is necessary that the dependent variable is coded accordingly. That is, for a binary regression, the factor level 1 of the dependent variable should represent the desired outcome.

Thirdly, the model should be fitted correctly. Neither over fitting nor under fitting should occur. That is only the meaningful variables should be included, but also all meaningful variables should be included. A good approach to ensure this is to use a stepwise method to estimate the logistic regression.

Fourthly, the error terms need to be independent. Logistic regression requires each observation to be independent. That is that the data-points should not be from any dependent samples design, e.g., before-after measurements, or matched pairings. Also the model should have little or no multicollinearity. That is that the independent variables should be independent from each other. However, there is the option to include interaction effects of categorical variables in the analysis and the model. If multicollinearity is present centering the variables might resolve the issue, i.e. deducting the mean of each variable. If this does not lower the multicollinearity, a factor analysis with orthogonally rotated factors should be done before the logistic regression is estimated.

Fifthly, logistic regression assumes linearity of independent variables and log odds. Whilst it does not require the dependent and independent variables to be related linearly, it requires that the

independent variables are linearly related to the log odds. Otherwise the test underestimates the strength of the relationship and rejects the relationship too easily, that is being not significant (not rejecting the null hypothesis) where it should be significant. A solution to this problem is the categorization of the independent variables. That is transforming metric variables to ordinal level and then including them in the model. Another approach would be to use discriminant analysis, if the assumptions of homoscedasticity, multivariate normality, and absence of multicollinearity are met.

Lastly, it requires quite large sample sizes. Because maximum likelihood estimates are less powerful than ordinary least squares (e.g., simple linear regression, multiple linear regression); whilst OLS needs 5 cases per independent variable in the analysis, ML needs at least 10 cases per independent variable, some statisticians recommend at least 30 cases for each parameter to be estimated.

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

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

Related Pages:

[Conduct and Interpret a Logistic Regression](#)

[What is Logistic Regression?](#)

[The Logistic Regression Analysis in SPSS](#)