

How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar?

Authored by
stats writer

June 30, 2024

RECOMMENDED CITATION

stats writer (2024). *How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar?*. PSYCHOLOGICAL SCALES. Retrieved from <https://scales.arabpsychology.com/?p=161048>

The likelihood ratio, Wald, and Lagrange multiplier (score) tests are statistical methods used to test the significance of parameters in a regression model.

The likelihood ratio test compares the likelihood of the model with all parameters included to the likelihood of a reduced model with some parameters omitted. It measures the overall fit of the model and determines if the included parameters significantly improve the model's prediction.

The Wald test, on the other hand, tests the significance of individual parameters by comparing their estimated values to their expected values under the null hypothesis. It is based on the assumption that the parameter estimates follow a normal distribution.

Similarly, the Lagrange multiplier (score) test also evaluates the significance of individual parameters, but it uses a different approach by calculating the score function, which measures the sensitivity of the model to changes in the parameters.

While all three tests serve the same purpose of testing the significance of model parameters, they differ in their underlying assumptions and methodologies. The likelihood ratio test is more powerful and preferred for comparing nested models, while the Wald and Lagrange multiplier tests are suitable for testing individual parameters.

FAQ: How are the likelihood ratio, Wald, and Lagrange multiplier (score) tests different and/or similar?

Purpose: This page introduces the concepts of the a) likelihood ratio test, b) Wald test, and c) score test. To see how the likelihood ratio test and Wald test are implemented in Stata refer to How can I perform the likelihood ratio and Wald test in Stata?

A researcher estimated the following model, which predicts high versus low writing scores on a standardized test (hiwrite), using students' gender (female), and scores on standardized test scores in

reading (read), math (math), and science (science). The output for the model looks like this:

Logistic regression Number of obs = 200

LR chi2(4) = 105.99

Prob > chi2 = 0.0000

Log likelihood = -84.419842 Pseudo R2 = 0.3857

hiwrite | Coef. Std. Err. z P>|z|
 -----+-----

female | 1.805528 .4358101 4.14 0.000 .9513555 2.6597

read | .0529536 .0275925 1.92 0.055 -.0011268 .107034

math | .1319787 .0318836 4.14 0.000 .069488 .1944694

science | .0577623 .027586 2.09 0.036 .0036947 .1118299

_cons | -13.26097 1.893801 -7.00 0.000 -16.97275

-9.549188

The researcher would like to know whether this model (with four predictor variables) fits significantly better than a model with just female and read as predictors. How can the researcher accomplish this? There are three common tests that can be used to test this type of

question, they are the likelihood ratio (LR) test, the Wald test, and the Lagrange multiplier test (sometimes called a score test). These tests are sometimes described as tests for differences among nested models, because one of the models can be said to be nested within the other. The null hypothesis for all three tests is that the smaller model is the "true" model, a large test statistics indicate that the null hypothesis is false. While all three tests address the same basic question, they are slightly different. In this page we will describe how to perform these tests and discuss the similarities and differences among them. (Note: these tests are very general and are used to test other types of hypotheses that involve testing whether fixing a parameter significantly harms model fit.)

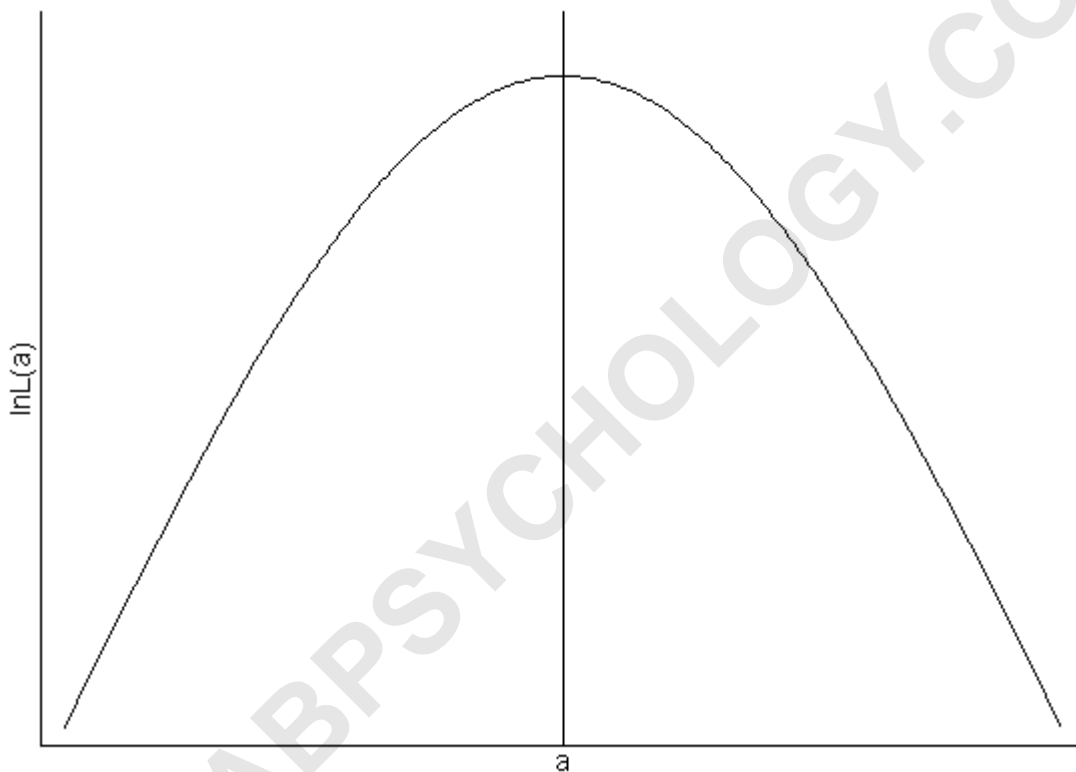
The likelihood

All three tests use the likelihood of the models being compared to assess their fit. The likelihood is the probability the data given the parameter estimates. The goal of a model is to find values for the parameters (coefficients) that maximize value of the likelihood function, that is, to find the set of parameter estimates that make the data most likely. Many procedures use

the log of the likelihood, rather than the likelihood itself, because it is easier to work with. The log likelihood (i.e., the log of the likelihood) will always be negative, with higher values (closer to zero) indicating a better fitting model. The above example involves a logistic regression model, however, these tests are very general, and can be applied to any model with a likelihood function. Note that even models for which a likelihood or a log likelihood is not typically displayed by statistical software (e.g., ordinary least squares regression) have likelihood functions.

As mentioned above, the likelihood is a function of the coefficient estimates and the data. The data are fixed, that is, you cannot change them, so one changes the estimates of the coefficients in such a way as to maximize the probability (likelihood). Different parameter estimates, or sets of estimates give different values of the likelihood. In the figure below, the arch or curve shows the changes in the value of the likelihood for changes in one parameter (a). On the x-axis are values of a , while the y-axis is the value of the likelihood at the

appropriate value of a . Most models have more than one parameter, but, if the values of all the other coefficients in the model are fixed, changes in a will show a similar picture. The vertical line marks the value of a that maximizes the likelihood.



The likelihood ratio test

The LR test is performed by estimating two models and comparing the fit of one model to the fit of the other. Removing predictor variables from a model will almost always make the model fit less well (i.e., a model will

have a lower log likelihood), but it is necessary to test whether the observed difference in model fit is statistically significant. The LR test does this by comparing the log likelihoods of the two models, if this difference is statistically significant, then the less restrictive model (the one with more variables) is said to fit the data significantly better than the more restrictive model. If one has the log likelihoods from the models, the LR test is fairly easy to calculate. The formula for the LR test statistic is:

$$LR = -2 \ln \left(\frac{L(m_1)}{L(m_2)} \right) = 2(\loglik(m_2) - \loglik(m_1))$$

Where $L(m_*)$ denotes the likelihood of the respective model (either Model 1 or Model 2), and $\loglik(m_*)$ the natural log of the model's final likelihood (i.e., the log likelihood). Where m_1 is the more restrictive model, and m_2 is the less restrictive model.

The resulting test statistic is distributed chi-squared, with degrees of freedom equal to the number of parameters that are constrained (in the current example, the number of variables removed from the

model, i.e., 2).

Using the same example as above, we will run both the full and the restricted model, and assess the difference in fit using the LR test. Model one is the model using female and read as predictors (by not including math and science in the model, we restrict their coefficients to zero). Below is the output for model 1. We will skip the interpretation of the results because that is not the focus of our discussion, but we will make note of the final log likelihood printed just above the table of coefficients ($\$loglik(m_1) = -102.45$).

Logistic regression Number of obs = 200

LR chi2(2) = 69.94

Prob > chi2 = 0.0000

Log likelihood = -102.44518 Pseudo R2 = 0.2545

hiwrite | Coef. Std. Err. z P>|z|

-----+-----
female | 1.403022 .3671964 3.82 0.000 .6833301 2.122713
read | .1411402 .0224042 6.30 0.000 .0972287 .1850517
_cons | -7.798179 1.235685 -6.31 0.000 -10.22008

-5.376281

Now we can run Model 2, in which coefficients for science and math are freely estimated, that is, a model with the full set of predictor variables. Below is output for Model 2. Again, we will skip the interpretation, and just make note of the log likelihood ($\$loglik(m_2) = -84.42$).

Logistic regression Number of obs = 200

LR chi2(4) = 105.99

Prob > chi2 = 0.0000

Log likelihood = -84.419842 Pseudo R2 = 0.3857

hiwrite | Coef. Std. Err. z P>|z|

female | 1.805528 .4358101 4.14 0.000 .9513555 2.6597

read | .0529536 .0275925 1.92 0.055 -.0011268 .107034

math | .1319787 .0318836 4.14 0.000 .069488 .1944694

science | .0577623 .027586 2.09 0.036 .0036947 .1118299

_cons | -13.26097 1.893801 -7.00 0.000 -16.97275

-9.549188

Now that we have both log likelihoods, calculating the test statistic is simple:

$$LR = 2 * (-84.419842 - (-102.44518)) = 2 * (-84.419842 + 102.44518) = 36.050676$$

So our likelihood ratio test statistic is 36.05 (distributed chi-squared), with two degrees of freedom. We can now use a table or some other method to find the associated p -value, which is $p < 0.001$, indicating that the model with all four predictors fits significantly better than the model with only two predictors. Note that many statistical packages will perform an LR test comparing two models, we have done the test by hand because it is easy to calculate, and because doing so makes it clear how the LR test works.

The Wald test

The Wald test approximates the LR test, but with the advantage that it only requires estimating one model. The Wald test works by testing the null hypothesis that

a set of parameters is equal to some value. In the model being tested here, the null hypothesis is that the two coefficients of interest are simultaneously equal to zero. If the test fails to reject the null hypothesis, this suggests that removing the variables from the model will not substantially harm the fit of that model, since a predictor with a coefficient that is very small relative to its standard error is generally not doing much to help predict the dependent variable. The formula for a Wald test is a bit more daunting than the formula for the LR test, so we won't write it out here (see Fox, 1997, p. 569, or other regression texts if you are interested). To give you an intuition about how the test works, it tests how far the estimated parameters are from zero (or any other value under the null hypothesis) in standard errors, similar to the hypothesis tests typically printed in regression output. The difference is that the Wald test can be used to test multiple parameters simultaneously, while the tests typically printed in regression output only test one parameter at a time.

Returning to our example, we will use a statistical

package to run our model and then to perform the Wald test. Below we see output for the model with all four predictors (the same output as model 2 above).

Logistic regression Number of obs = 200

LR chi2(4) = 105.99

Prob > chi2 = 0.0000

Log likelihood = -84.419842 Pseudo R2 = 0.3857

 hiwrite | Coef. Std. Err. z P>|z|

-----+-----
 female | 1.805528 .4358101 4.14 0.000 .9513555 2.6597

read | .0529536 .0275925 1.92 0.055 -.0011268 .107034

math | .1319787 .0318836 4.14 0.000 .069488 .1944694

science | .0577623 .027586 2.09 0.036 .0036947 .1118299

_cons | -13.26097 1.893801 -7.00 0.000 -16.97275
 -9.549188

After running the logistic regression model, the Wald test can be used. The output below shows the results of the Wald test. The first thing listed in this particular output (the method of obtaining the Wald test and the

output may vary by package) are the specific parameter constraints being tested (i.e., the null hypothesis), which is that the coefficients for math and science are simultaneously equal to zero. Below the list of constraints we see the chi-squared value generated by the Wald test, as well as the p-value associated with a chi-squared of \$27.53\$ with two degrees of freedom. The p-value is less than the generally used criterion of \$0.05\$, so we are able to reject the null hypothesis, indicating that the coefficients are not simultaneously equal to zero. Because including statistically significant predictors should lead to better prediction (i.e., better model fit) we can conclude that including math and science results in a statistically significant improvement in the fit of the model.

(1) math = 0

(2) science = 0

chi2(2) = 27.53

Prob > chi2 = 0.0000

The Lagrange multiplier or score test

As with the Wald test, the Lagrange multiplier test

requires estimating only a single model. The difference is that with the Lagrange multiplier test, the model estimated does not include the parameter(s) of interest. This means, in our example, we can use the Lagrange multiplier test to test whether adding science and math to the model will result in a significant improvement in model fit, after running a model with just female and read as predictor variables. The test statistic is calculated based on the slope of the likelihood function at the observed values of the variables in the model (female and read). This estimated slope, or "score" is the reason the Lagrange multiplier test is sometimes called the score test. The scores are then used to estimate the improvement in model fit if additional variables were included in the model. The test statistic is the expected change in the chi-squared statistic for the model if a variable or set of variables is added to the model. Because it tests for improvement of model fit if variables that are currently omitted are added to the model, the Lagrange multiplier test is sometimes also referred to as a test for omitted variables. They are also sometimes referred to as modification indices, particularly in the structural equation modeling

literature.

Below is output for the logistic regression model using the variables female and read as predictors of hiwrite (this is the same as Model 1 from the LR test).

Logistic regression Number of obs = 200

LR chi2(2) = 69.94

Prob > chi2 = 0.0000

Log likelihood = -102.44518 Pseudo R2 = 0.2545

```

-----
hiwrite | Coef. Std. Err. z P>|z|
-----+-----
female | 1.403022 .3671964 3.82 0.000 .6833301 2.122713
read | .1411402 .0224042 6.30 0.000 .0972287 .1850517
_cons | -7.798179 1.235685 -6.31 0.000 -10.22008
-5.376281
-----

```

After running the above model, we can look at the results of the Lagrange multiplier test. Unlike the previous two tests, which are primarily used to assess the change in model fit when more than one variable is added to the model, the Lagrange multiplier test can be

used to test the expected change in model fit if one or more parameters which are currently constrained are allowed to be estimated freely. In our example, this means testing whether adding math and science to the model would significantly improve model fit. Below is the output for the score test. The first two rows in the table give the test statistics (or scores) for adding either variable alone to the model. To carry on with our example, we will focus on the results in the third row labeled "simultaneous test," which shows the test statistic for adding both math and science to our model. The test statistic for adding both math and science to the model is 35.51, it is distributed chi-squared, with degrees of freedom equal to the number of variables being added to the model, so in our example, 2. The p-value is below the typical cutoff of 0.05, suggesting that including the variables math and science in the model would create a statistically significant improvement in model fit. This conclusion is consistent with the results of both the LR and Wald tests.

logit: score tests for omitted variables

Term | score df p

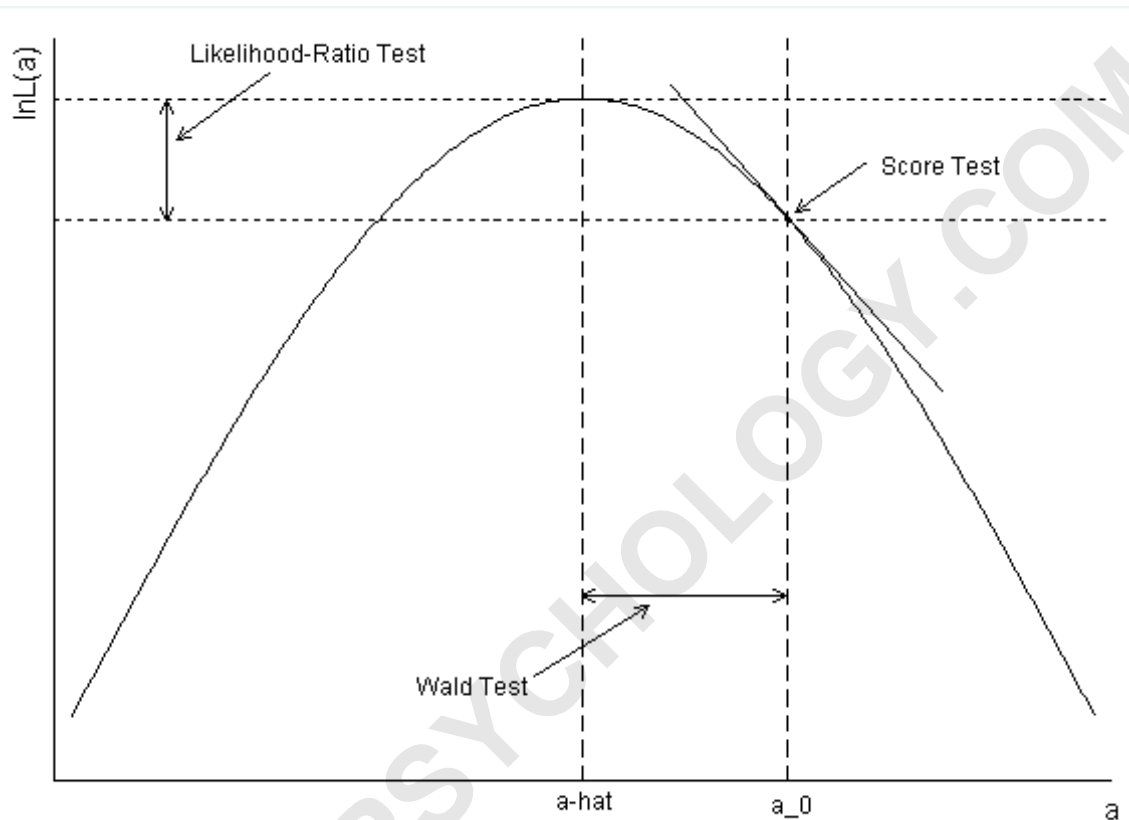
math		28.94	1	0.0000
science		15.39	1	0.0001
simultaneous test		35.51	2	0.0000

A comparison of the three tests

As discussed above, all three tests address the same basic question, which is, does constraining parameters to zero (i.e., leaving out these predictor variables) reduce the fit of the model? The difference between the tests is how they go about answering that question. As you have seen, in order to perform a likelihood ratio test, one must estimate both of the models one wishes to compare. The advantage of the Wald and Lagrange multiplier (or score) tests is that they approximate the LR test, but require that only one model be estimated. Both the Wald and the Lagrange multiplier tests are asymptotically equivalent to the LR test, that is, as the sample size becomes infinitely large, the values of the Wald and Lagrange multiplier test statistics will become increasingly close to the test statistic from the LR test.

In finite samples, the three will tend to generate somewhat different test statistics, but will generally come to the same conclusion. An interesting relationship between the three tests is that, when the model is linear the three test statistics have the following relationship $Wald \geq LR \geq \text{score}$ (Johnston and DiNardo 1997 *p.* 150). That is, the Wald test statistic will always be greater than the LR test statistic, which will, in turn, always be greater than the test statistic from the score test. When computing power was much more limited, and many models took a long time to run, being able to approximate the LR test using a single model was a fairly major advantage. Today, for most of the models researchers are likely to want to compare, computational time is not an issue, and we generally recommend running the likelihood ratio test in most situations. This is not to say that one should never use the Wald or score tests. For example, the Wald test is commonly used to perform multiple degree of freedom tests on sets of dummy variables used to model categorical predictor variables in regression (for more information see our webbooks on Regression with Stata, SPSS, and SAS, specifically Chapter 3 - Regression with Categorical Predictors.) The advantage

of the score test is that it can be used to search for omitted variables when the number of candidate variables is large.



Figure

based on a figure in Fox (1997, p. 570); used with author's permission.

One way to better understand how the three tests are related, and how they are different, is to look at a graphical representation of what they are testing. The figure above illustrates what each of the three tests does. Along the x-axis (labeled "a") are possible values

of the parameter a (in our example, this would be the regression coefficient for either math or science). Along the y -axis are the values of the log likelihood corresponding to those values of a . The LR test compares the log likelihoods of a model with values of the parameter a

constrained to some value (in our example zero) to a model where a is freely estimated. It does this by comparing the height of the likelihoods for the two models to see if the difference is statistically significant (remember, higher values of the likelihood indicate better fit). In the figure above, this corresponds to the vertical distance between the two dotted lines. In contrast, the Wald test compares the parameter estimate \hat{a} to a_0 ; a_0 is the value of a under the null hypothesis, which generally states that $a = 0$. If \hat{a} is significantly different from a_0 , this suggests that freely estimating a (using \hat{a}) significantly improves model fit. In the figure, this

is shown as the distance between a_0 and \hat{a} on the x -axis (highlighted by the solid lines). Finally, the score test looks at the slope of the log likelihood when a is constrained (in our example to zero). That is, it looks at how quickly the likelihood is changing at the (null)

hypothesized value of a . In the figure above this is shown as the tangent line at a_0 .

References

Fox, J. (1997) Applied regression analysis, linear models, and related methods. Thousand Oaks, CA: Sage Publications.

Johnston, J. and DiNardo, J. (1997) Econometric Methods Fourth Edition. New York, NY: The McGraw-Hill Companies, Inc.

ARABPSYCHOLOGY.COM