LogisticRegressionAnalysis.com
*Fast, easy guide to understanding,
running, and interpreting multivariate
logistic regression*

What are Z-Values in Logistic Regression?

Posted on May 31, 2014 by StatsProf

Very Short Answer

The z-value is the regression coefficient divided by its standard error. It
is also sometimes called the z-statistic. It is usually given in the third
column of the logistic regression regression coefficient table output.
Thus, in the example below, the z-value for the regression coefficient
for ResidenceLength is $0.024680/0.013800 = 1.79$.

If the z-value is too big in magnitude (i.e., either too positive or too
negative), it indicates that the corresponding true regression coefficient
is not 0 and the corresponding X-variable matters. A good rule of
thumb is to use a cut-off value of 2 which approximately corresponds to
a two-sided hypothesis test with a significance level of $\alpha = 0.05$. So, for
the ResidenceLength variable, the z-value is 1.79 which is not large
enough to provide strong evidence that ResidenceLength matters.

Note: The relationship between the regression coefficient, its standard
error, the z-value, and the p-value is virtually identical both logistic
regression and regular least-squares regression. So if you understand
this is regular regression, you also understand it in logistic regression.

Detailed Explanation

In statistics, the letter "Z" is often used to refer to a random variable
that has a standard normal distribution. A standard normal distribution
is a normal distribution with expectation 0 and standard deviation 1.
This is the normal distribution that is generally tabulated in the back of
any basic statistics book.

Because of this, the term "z-value" is often used to refer to the value of
a statistic that has a standard normal distribution. Sometimes it is also
used to refer to percentile points from the standard normal distribution

that are used to compare to the value of statistic. For example, one might refer to "the z-value corresponding to a 95% confidence interval" (which would be 1.96).

In basic univariate statistics, z-statistics and z-values usually come about as a result of standardizing a statistic such as the sample mean $\bar{X}$ or sample proportion $\hat{p}$. Standardizing a statistic means subtracting its expected value μ_{stat} and then dividing by its standard error σ_{stat} (the standard error of a statistic is its standard deviation). The leading example of this from basic statistics would be a z-statistic derived from from the sample mean:

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}}$$

Here $\bar{X}$ is the statistic to be standardized (the sample mean), μ is its expectation (which, for the sample mean, is the same as the population mean), and the standard deviation of the statistic σ_{stat} is $\sigma_{\bar{X}} = \sigma/\sqrt{n}$. Here σ is the population standard deviation, and the formula $\sigma_{\bar{X}} = \sigma/\sqrt{n}$ comes about because of the relationship between the standard deviation of a sample mean and the population mean. Finally, the statistic has a normal distribution as the sample size gets large because of the central limit theorem.

Hopefully, the above reminds you about enough of your basic statistics that using these ideas in the context of logistic regression will make sense.

So where do z-values come about in logistic regression? They primarily come about as a result of standardizing the logistic regression coefficients when testing whether or not the individual X-variables are related to the Y-variables. For example, consider the coefficient table output from the logistic regression in the "Kid Creative" example I discussed in the post Understanding Logistic Regression Output: Part 2 — Which Variables Matter.

Coefficients:

	Estimate	Std. Error	z value	Pr(>|z|)	Odds Ratio
(Intercept)	-17.910000	2.223000	-8.06	0.0000	
Income	0.000202	0.000024	8.55	0.0000	1.0002
IsFemale	1.646000	0.465100	3.54	0.0004	5.1862
IsMarried	0.566200	0.586400	0.97	0.3343	1.7616
HasCollege	-0.279400	0.443700	-0.63	0.5290	0.7562
IsProfessional	0.225300	0.465000	0.49	0.6280	1.2527
IsRetired	-1.159000	0.932300	-1.24	0.2140	0.3138
Unemployed	0.988600	4.690000	0.21	0.8330	2.6875
ResidenceLength	0.024680	0.013800	1.79	0.0738	1.0250
DualIncome	0.451800	0.521500	0.87	0.3863	1.5711
Minors	1.133000	0.463500	2.44	0.0145	3.1050
Own	1.056000	0.559400	1.89	0.0590	2.8748
House	-0.926500	0.621800	-1.49	0.1362	0.3959
White	1.864000	0.545400	3.42	0.0006	6.4495
English	1.530000	0.840700	1.82	0.0687	4.6182
PrevChildMag	1.557000	0.711900	2.19	0.0287	4.7446
PrevParentMag	0.477700	0.624000	0.77	0.4439	1.6124

You will see that the z-values are given in the third column of numbers in the table. These z-values are computed as the test statistic for the hypothesis test that the true corresponding regression coefficient β is 0. (Note: The p-values computed from the z-values are given in the 4th column of numbers in the regression coefficient output table. I generally do not look at the z-values, but rather use the p-values.

More specifically, suppose we want to determine if an X-variable matters (that is, has a significant relationship to the Y variable). We determine this by testing the null hypothesis that the corresponding regression coefficient β is 0. In hypothesis testing, we assume the null hypothesis is true, and then see if the data provide evidence against it. So in this case, we assume β is 0. That is, we assume the expectation of the fitted regression coefficient $\hat{\beta}$ is 0. So we standardize the regression coefficient as follows:

$$Z = \frac{\hat{\beta} - 0}{\hat{\sigma}_{\hat{\beta}}} = \hat{\beta}/\hat{\sigma}_{\hat{\beta}}$$

Note that there is no closed-form formula for $\hat{\sigma}_{\hat{\beta}}$. It is computed as the solutions to a non-linear system of equations.

So, for example, consider the ResidenceLength regression coefficient in the coefficient output table above. For this variable $\hat{\beta} = 0.024680$ and $\hat{\sigma}_{\hat{\beta}} = 0.013800$, so the Z-value is $\hat{\beta}/\hat{\sigma}_{\hat{\beta}} = 0.024680/0.013800 = 1.79$. So the value in the third column of numbers is $Z = 1.79$.

How do we interpret the Z-values? As a rough rule of thumb, if the absolute value of the Z-value is bigger that 2.0, the variable is significant (which means that there is statistical evidence that it is related to the Y variable). This gives a rough hypothesis test with a significance level of about $\alpha = 0.05$.

More precisely, in the hypothesis test, select a significance level such as $\alpha = 0.05$. Determine the corresponding critical value for the test. This will depend on whether or not the hypothesis test is one-sided or two-sided. If it is one-sided, the critical value will be the upper α percentage point of the standard normal distribution (generally referred to as z_α). If it is a two sided test (most common), then the critical value is the upper $\alpha/2$ percentage point (generally referred to as $z_{\alpha/2}$). The absolute value of the Z-value is then compared to the appropriate critical value to determine if the test is significant. That is, the regression coefficient is significantly different from 0 if:<\p>

$$|\text{Z-value}| \geq z_\alpha \quad (\text{one-sided})$$

or

$$|\text{Z-value}| \geq z_{\alpha/2} \quad (\text{two-sided})$$

This entry was posted in FAQ. Bookmark the permalink.

3 Responses to *What are Z-Values in Logistic Regression?*

Sai Potturi *says:*

April 14, 2017 at 3:03 pm

this was really useful. Thank you.

Reply

AEM *says:*

February 2, 2017 at 7:37 am

This is really useful, thanks. Maybe a more general/basic question. If z-scores are so correlated to p-values and we actually attribute a cut-off of 2 for hypothesis testing, what does it brings in addition to the p-value information?

Cheers!

Reply

gangan *says:*

I think your address is very useful to us.I also appreciate for you had written your post. long time no update. Wish you can back to this blog.

Reply

LogisticRegressionAnalysis.com

Proudly powered by WordPress.