

Logistic Regression Extras - Estimating Model Parameters, Comparing Models and Assessing Model Fit

1. Estimating Model Parameters (Coefficients)

Ordinary Least Squares regression uses the Minimum least squares method to estimate parameters so that the values of α and β are chosen so that they minimize the sum of squared deviations of the observed values of Y from the predicted values based upon the model. However, this method does not work equally well for a categorical response variable.

Logistic regression therefore uses the **Maximum Likelihood Estimation** method to estimate the model coefficients. This method yields values of α and β which maximize the probability of obtaining the observed set of data.

Conceptually, it works like this:

First construct a likelihood function which expresses the probability of the observed data as a function of the unknown parameters α and β .

In the univariate case, the contribution to the likelihood function for a given value of the predictor X , is

$$P(Y = 1 | x)^y * P(Y = 0 | x)^{1-y}$$

Thus when $Y = 1$, the contribution is: $P(Y = 1 | x)$

when $Y = 0$, the contribution is: $P(Y = 0 | x)$

Since the sample observations are assumed to be independent, the likelihood function for the dataset is just the product of the individual contributions:

$$L = \prod P(Y = 1 | x)^y * P(Y = 0 | x)^{1-y}$$

A more tractable version of this function is obtained by taking the natural logarithm of the likelihood function, called the Log Likelihood function:

$$LL = \sum y \text{Log}[P(Y = 1 | x)] + (1 - y) \text{Log}[P(Y = 0 | x)]$$

To find the values of the parameters that maximize the above function, we differentiate this function with respect to α and β and set the two resulting expressions to zero. An iterative method is used to solve the equations and the resulting values of α and β are called the maximum likelihood estimates of those parameters. The same approach is used in the multiple predictor case where we would have $(p+1)$ equations corresponding to the p predictors and the constant α .

2. Comparing Alternative Models and Assessing Model Fit

Once we've fit a particular logistic regression model, the first step is to assess the significance of the overall model with p coefficients for the predictors included. This is done via a Likelihood Ratio Test which tests the null hypothesis $H_0: \beta_1 = \beta_2 \dots = \beta_p = 0$

The test statistic is given by:

$$G = -2 [LL(\text{constant only model}) - LL(\text{chosen model with } p \text{ predictors})]$$

G has a Chi-square distribution with (p) degrees of freedom. A small p -value leads to rejecting H_0 and the conclusion that at least one (or more) of the p coefficients are different from zero.

We first test the individual predictors in the model using z tests or univariate Wald tests (very similar to the z test, or t test). Again, small p -values indicate the corresponding predictor is significant. A more robust alternative is the likelihood ratio test which computes the G -statistic with a slight difference – at each step it tests the model with a predictor to one without that predictor – a small p -value indicates the significance of that predictor variable in explaining Y .

In assessing model fit, we examine how well the model describes (fits) the observed data. One way to assess goodness of fit (GOF) is to examine how likely the sample results are, given the parameter estimates.

Many different statistics are used to assess the fit of the chosen model. Among these:

1. **The pseudo R^2 :** This statistic is slightly analogous to the coefficient of determination statistic from linear regression analysis, in that it takes values in the range 0 to 1 and larger values indicate a better fit. However, it cannot be interpreted as the amount of variance in Y explained by the predictor variable X .

$$\text{Pseudo } R^2 = 1 - \frac{LL_{full}}{LL_{constant\ only}}$$

Where LL_{full} = Log likelihood of the full (chosen) model, $LL_{constant\ only}$ = Log likelihood of the full model

2. **Pearson Goodness Of Fit Chi-Square test:** The null hypothesis is that the chosen model fits the data. The test statistic computes the overall difference between the observed probabilities and those estimated from the fitted model. Unlike most tests of significance, we want this test to be non-significant (large p-value desired) to indicate that our model is a good fit. This means is that the difference between the expected values using this model and the actual values is non-significant.

3. **Deviance Test:** The deviance statistic is calculated as the sum of the differences between the log likelihoods of the saturated model (which has as many coefficients as observations in the dataset) and the chosen model, for all the observations in the sample. It follows a Chi-square distribution with df = difference in the number of parameters in the two models. The null hypothesis sets the coefficients that are in the saturated model but not in the fitted model, to zero. A large p-value indicates that none of the excluded variables is significant; that the fitted model is as good as the saturated model.

$$D = -2\sum\{LL(\text{saturated}) - LL(\text{model})\}$$

4. **Hosmer-Lemeshow Test:** Considered one of the best tests of fit, this approach divides the range of probability values into groups based on covariate patterns and compares the observed and expected counts within these groups using a Chi-square statistic. The smaller the differences in expected and observed counts, the smaller the overall variance and thus the test statistic value. Therefore, a large p-value indicates a good fit.

5. **Information Criteria:** These are relatively new tests and are comparative in nature. There are two main types of information criteria: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). All things being equal, smaller values of the criteria indicate a better fitting model.

- **Akaike Information Criterion (AIC)**

$$AIC = \frac{-2 (LL - k)}{n} \quad k = \# \text{ predictors, } n = \text{sample size}$$

- **Bayesian Information Criterion (BIC)**

$$BIC = -2*LL + \log(n)*k \quad k = \# \text{ predictors, } n = \text{sample size}$$