Nested Logit regression model Harvard Case Solution & Analysis

TheCaseSolutions.com

Nested Logit regression model Harvard Case Solution & Analysis

TheCaseSolutions.com

Logistic Regression

In statistics, logistic regression or logit regression is a type of regression analysis used for predicting the outcome of a categorical dependent variable based on one or more predictor variables. It is used in estimating empirical values of the parameters in a qualitative response model.

Logistic regression measures the relationship between a categorical dependent variable and one or more independent variables, which are usually, by using probability scores as the predicted values of the dependent variable.

Logistic Regression

Logistic regression can be binomial or multinomial. Binomial or binary logistic regression deals with situations in which the observed outcome for a dependent variable can have only two possible types.

In binary logistic regression, the outcome is usually coded as "0" or "1", as this leads to the most straightforward interpretation.

The logit of success is then fit to the predictors using linear regression analysis. The predicted value of the logit is converted back into predicted odds via the inverse of the natural logarithm, namely the exponential function.

Assumption

- Lack of Multicollinearity

Logistic Function

Logistic function always takes on values between zero and one

$$F(t) = \frac{e^t}{e^t+1} = \frac{1}{1+e^{-t}}$$

and viewing t as a linear function of an explanatory variable x, the logistic function can be written as:

$$\pi(x) = \frac{e^{(\beta_0+\beta_1 x)}}{e^{(\beta_0+\beta_1 x)}+1} = \frac{1}{1+e^{-(\beta_0+\beta_1 x)}}$$

This will be interpreted as the probability of the dependent variable equaling a "success" or "case" rather than a failure or non-case. We also define the inverse of the logistic function, the logit:

$$g(x) = \ln \frac{\pi(x)}{1-\pi(x)} = \beta_0 + \beta_1 x$$

and equivalently:

$$\frac{\pi(x)}{1-\pi(x)} = e^{(\beta_0+\beta_1 x)}$$

In the above equations, g(x) refers to the logit function of some given linear combination x of the predictors, ln denotes the natural logarithm.

Logistic Function (cont.)

The formula for $\pi(x)$ illustrates that the probability of the dependent variable equaling a case is equal to the value of the logistic function of the linear regression expression.

The equation for g(x) illustrates that the logit is equivalent to the linear regression expression.

Logistic Regression

Logistic regression can be binomial or multinomial. Binomial or binary logistic regression deals with situations in which the observed outcome for a dependent variable can have only two possible types.

In binary logistic regression, the outcome is usually coded as "0" or "1", as this leads to the most straightforward interpretation.

The logit of success is then fit to the predictors using linear regression analysis. The predicted value of the logit is converted back into predicted odds via the inverse of the natural logarithm, namely the exponential function.

Assumption

- Lack of Multicollinearity

Logistic Function

Logistic function always takes on values between zero and one

$$F(t) = \frac{e^t}{e^t + 1} = \frac{1}{1 + e^{-t}}$$

and viewing t as a linear function of an explanatory variable x), the logistic function can be written as

$$\pi(x) = \frac{e^{(\beta_0 + \beta_1 x)}}{e^{(\beta_0 + \beta_1 x)} + 1} = \frac{1}{1 + e^{-(\beta_0 + \beta_1 x)}}$$

This will be interpreted as the probability of the dependent variable equalling a "success" or "case" rather than a failure or non-case. We also define the inverse of the logistic function, the logit $g(x) = \ln \frac{\pi(x)}{1 - \pi(x)} = \beta_0 + \beta_1 x$, and equivalently $\frac{\pi(x)}{1 - \pi(x)} = e^{(\beta_0 + \beta_1 x)}$

In the above equations, g(x) refers to the logit function of some given linear combination x of the predictors, 'ln' denotes the natural logarithm

Logistic Function (cont.)

The formula for $\pi(x)$ illustrates that the probability of the dependent variable equaling a case is equal to the value of the logistic function of the linear regression expression.

The equation for g(x) illustrates that the logit is equivalent to the linear regression expression.

Logistic Regression Model (cont.)

To Fit Logistic Regression Model

- Deviance of Likelihood Ratio test: Deviance is analogous to the sum of squares calculations in linear regression and is a measure of the lack of fit to the data in a logistic regression model. Deviance is calculated by comparing a given model with the saturated model – a model with a theoretically perfect fit. This computation is called the likelihood-ratio test: $D = -2\ln\frac{\text{likelihood of the fitted model}}{\text{likelihood of the saturated model}}$

Smaller values indicate better fit as the fitted model deviates less from the saturated model. When assessed upon a chi-square distribution, nonsignificant chi-square values indicate very little unexplained variance and thus, good model fit, while

[illegible]

[illegible]

To Fit Logistic Regression Model (cont.) & Inter[illegible]

then $D_{fitted} - D_{null} = -2\ln\frac{\text{likelihood of null model}}{\text{likelihood of the fitted model}}$

- We use Maximum-likelihood estimation to estimate coefficeince

We interpret Logistic Regression Model by odds ratio

Power of Predictive

Cox & Snell Pseudo R^2
Nagelkerke R^2