

Multinomial Logistic Regression

- ▶ In statistics, multinomial logistic regression is a classification method that generalizes logistic regression to multiclass problems, i.e. with more than two possible discrete outcomes.
- ▶ That is, it is a model that is used to predict the probabilities of the different possible outcomes of a categorically distributed dependent variable, given a set of independent variables (which may be real-valued, binary-valued, categorical-valued, etc.).

Multinomial Logistic Regression

- ▶ Multinomial logistic regression is used to model nominal outcome variables, in which the log odds of the outcomes are modeled as a linear combination of the predictor variables.

Multinomial Regression with R

The main package we will use is the **nnet** package. We will also use the **ggplot2** and **reshape2** package.

```
install.packages("nnet")  
library(nnet)  
  
library(ggplot2)  
library(reshape2)
```

nnet

- ▶ Title : Feed-forward Neural Networks and Multinomial Log-Linear Models
- ▶ Description : Software for feed-forward neural networks with a single hidden layer, and for multinomial log-linear models.
- ▶ Authors : Brian Ripley ,William Venables
- ▶ URL : <http://www.stats.ox.ac.uk/pub/MASS4/>

Multinomial Logistic regression R

Examples of multinomial logistic regression

- ▶ Which major will a college student choose, given their grades, stated likes and dislikes, etc.?
- ▶ Which blood type does a person have, given the results of various diagnostic tests?
- ▶ In a hands-free mobile phone dialing application, which person's name was spoken, given various properties of the speech signal?
- ▶ Which candidate will a person vote for, given particular demographic characteristics?
- ▶ Which country will a firm locate an office in, given the characteristics of the firm and of the various candidate countries?

Multinomial Regression with R

Examples of multinomial logistic regression

Example 1. Entering high school students make program choices among general program, vocational program and academic program. Their choice might be modeled using their writing score and their social economic status.

Multinomial Regression with R

Examples of multinomial logistic regression

Example 2. People's occupational choices might be influenced by their parents' occupations and their own education level. We can study the relationship of one's occupation choice with education level and father's occupation. The occupational choices will be the outcome variable which consists of categories of occupations.

Multinomial Regression with R

Examples of multinomial logistic regression

Example 3. A biologist may be interested in food choices that alligators make. Adult alligators might have different preferences from young ones. The outcome variable here will be the types of food, and the predictor variables might be size of the alligators and other environmental variables.

Description of the Data

We will use the third example using the *multilog* data set (or *ml*).

```
ml <- read.csv("multilog.csv",header=T)
```

Description of the Data

- ▶ The outcome variable is **prog**, program type.
- ▶ The two predictor variables are
 1. social economic status, **ses**, (a three-level categorical variable)
 2. writing score, **write**, (a continuous variable).
- ▶ The data set contains variables on 200 students.

Multinomial Regression with R

```
table(ml$ses, ml$prog)
```

	general	academic	vocation
low	16	19	12
middle	20	44	31
high	9	42	7

	M	SD
general	51.33	9.398
academic	56.26	7.943
vocation	46.76	9.319