

Package ‘mlogit’

April 17, 2009

Version 0.1-2

Date 2008-12-17

Title multinomial logit model

Author Yves Croissant <yves.croissant@let.ish-lyon.cnrs.fr>

Maintainer Yves Croissant <yves.croissant@let.ish-lyon.cnrs.fr>

Depends R (>= 1.8.0), maxLik

Suggests Ecdat, lmtest, car

Description Estimation of the multinomial logit model with choice specific variables

License GPL (>= 2)

URL <http://www.r-project.org>

Repository CRAN

Date/Publication 2008-12-19 19:55:07

R topics documented:

Car	2
Catsup	3
Cracker	4
Fishing	5
HC	6
Heating	6
hmftest	7
Ketchup	9
logitform	10
mlogit	11
mlogit.data	14
Mode	16
Train	17
Tuna	18
Yogurt	19

Car

*Stated Preferences for Car Choice***Description**

a cross-section

number of observations : 4654*observation* : individuals*country* : United States**Usage**

data (Car)

Format

A dataframe containing :

choice choice of a vehicle among 6 propositions**college** college education ?**hsg2** size of household greater than 2 ?**coml5** commute lower than 5 miles a day ?**typez** body type, one of regcar (regular car), sportuv (sport utility vehicle), sportcar, stwagon (station wagon), truck, van, for each proposition z from 1 to 6**fuelz** fuel for proposition z, one of gasoline, methanol, cng (compressed natural gas), electric.**pricez** price of vehicle divided by the logarithme of income**rangez** hundreds of miles vehicle can travel between refuelings/rechargings**accz** acceleration, tens of seconds required to reach 30 mph from stop**speedz** highest attainable speed in hundreds of mph**pollutionz** tailpipe emissions as fraction of those for new gas vehicle**sizez** 0 for a mini, 1 for a subcompact, 2 for a compact and 3 for a mid-size or large vehicle**spacez** fraction of luggage space in comparable new gas vehicle**costz** cost per mile of travel (tens of cents) : home recharging for electric vehicle, station refueling otherwise**stationz** fraction of stations that can refuel/recharge vehicle**Source**McFadden, Daniel and Kenneth Train (2000) "Mixed MNL models for discrete response", *Journal of Applied Econometrics*, **15(5)**, 447–470.**References**Journal of Applied Econometrics data archive : <http://jae.wiley.com/jae/>.

Catsup

Choice of Brand for Catsup

Description

a cross-section

number of observations : 2798

observation : individuals

country : United States

Usage

data(Catsup)

Format

A dataframe containing :

id individuals identifiers

choice one of heinz41, heinz32, heinz28, hunts32

disp.z is there a display for brand z ?

feat.z is there a newspaper feature advertisement for brand z ?

price.z price of brand z

Source

Jain, Dipak C., Naufel J. Vilcassim and Pradeep K. Chintagunta (1994) "A random-coefficients logit brand-choice model applied to panel data", *Journal of Business and Economics Statistics*, **12(3)**, 317.

References

Journal of Business Economics and Statistics web site : <http://www.amstat.org/publications/jbes/>.

Cracker

Choice of Brand for Crakers

Description

a cross-section

number of observations : 3292

observation : individuals

country : United States

Usage

```
data(Cracker)
```

Format

A dataframe containing :

id individuals identifiers

choice one of sunshine, kleebler, nabisco, private

disp.z is there a display for brand z ?

feat.z is there a newspaper feature advertisement for brand z ?

price.z price of brand z

Source

Jain, Dipak C., Naufel J. Vilcassim and Pradeep K. Chintagunta (1994) "A random-coefficients logit brand-choice model applied to panel data", *Journal of Business and Economics Statistics*, **12(3)**, 317.

Paap, R. and Philip Hans Frances (2000) "A dynamic multinomial probit model for brand choices with different short-run effects of marketing mix variables", *Journal of Applied Econometrics*, **15(6)**, 717-744.

References

Journal of Business Economics and Statistics web site : <http://www.amstat.org/publications/jbes/>.

Fishing

Choice of Fishing Mode

Description

a cross-section

number of observations : 1182

observation : individuals

country : United States

Usage

```
data(Fishing)
```

Format

A dataframe containing :

mode recreation mode choice, on of : beach, pier, boat and charter

price price for chosen alternative

catch catch rate for chosen alternative

pbeach price for beach mode

ppier price for pier mode

pboat price for private boat mode

pcharter price for charter boat mode

cbeach catch rate for beach mode

cpier catch rate for pier mode

cboat catch rate for private boat mode

ccharter catch rate for charter boat mode

income monthly income

Source

Herriges, J. A. and C. L. Kling (1999) "Nonlinear Income Effects in Random Utility Models", *Review of Economics and Statistics*, **81**, 62-72.

References

Cameron, A.C. and P.K. Trivedi (2005) *Microeconometrics : methods and applications*, Cambridge, pp. 463–466, 486 and 491–495.

HC *Heating and Cooling System Choice in Newly Built Houses in California*

Description

a cross-section
number of observations : 250
observation : households
country : California

Usage

data (HC)

Format

A dataframe containing :

depvar heating system, one of gcc (gas central heat with cooling), ecc (electric central resistance heat with cooling), erc (electric room resistance heat with cooling), hpc (electric heat pump which provides cooling also), gc (gas central heat without cooling), ec (electric central resistance heat without cooling), er (electric room resistance heat without cooling)

ich.z installation cost of the heating portion of the system

icca installation cost for cooling

och.z operating cost for the heating portion of the system

occa operating cost for cooling

income annual income of the household

References

Kenneth Train's home page : <http://elsa.berkeley.edu/~train/>.

Heating *Heating System Choice in California Houses*

Description

a cross-section
number of observations : 900
observation : households
country : California

Usage

```
data(Heating)
```

Format

A dataframe containing :

idcase id

depvar heating system, one of gc (gas central), gr (gas room), ec (electric central), er (electric room), hp (heat pump)

ic.z installation cost for heating system z (defined for the 5 heating systems)

oc.z annual operating cost for heating system z (defined for the 5 heating systems)

pb.z ratio oc.z/ic.z

income annual income of the household

agehed age of the household head

rooms numbers of rooms in the house

References

Kenneth Train's home page : <http://elsa.berkeley.edu/~train/>.

hmftest

Hausman-McFadden Test

Description

Test the IIA hypothesis (independence of irrelevant alternatives) for a multinomial logit model.

Usage

```
hmftest(x, ...)
## S3 method for class 'mlogit':
hmftest(x, z, ...)
## S3 method for class 'formula':
hmftest(x, alt.subset, ...)
```

Arguments

x an object of class `mlogit` or a formula,

z an object of class `mlogit` or a subset of alternatives for the `mlogit` method. This should be the same model as `x` estimated on a subset of alternatives,

alt.subset a subset of alternatives,

... further arguments passed to `mlogit` for the `formula` method.

Details

This is an implementation of the Hausman's consistency test for multinomial logit models. If the independence of irrelevant alternatives applies, the probability ratio of every two alternatives depends only on the characteristics of these alternatives. Consequently, the results obtained on the estimation with all the alternatives or only on a subset of them are consistent, but more efficient in the first case. On the contrary, only the results obtained from the estimation on a relevant subset are consistent. To compute this test, one needs a model estimated with all the alternatives and one model estimated on a subset of alternatives. This can be done by providing two objects of class `mlogit`, one object of class `mlogit` and a character vector indicating the subset of alternatives, or a formula and a subset of alternatives.

Value

an object of class "htest".

Author(s)

Yves Croissant

References

Hausman, J.A. and D. McFadden (1984), A Specification Test for the Multinomial Logit Model, *Econometrica*, **52**, pp.1219–1240.

Examples

```
## from Greene's Econometric Analysis p. 731

data("TravelMode", package="AER")
TravelMode <- mlogit.data(TravelMode, choice="choice", shape="long",
                          alt.var="mode", id.var="individual")

## Create a variable of income only for the air mode

TravelMode$avinc <- with(TravelMode, (mode=='air')*income)

## Estimate the model on all alternatives, with car as the base level
## like in Greene's book.

x <- mlogit(choice~wait+gcost+avinc, TravelMode, reflevel="car")

## Estimate the same model for ground modes only (the variable avinc
## must be dropped because it is 0 for every observation

g <- mlogit(choice~wait+gcost, TravelMode, reflevel="car",
            alt.subset=c("car", "bus", "train"))

## Compute the test

hmftest(x, g)
```

Ketchup

Choice of Brand for Ketchup

Description

a cross-section

number of observations : 4956

observation : individuals

country : United States

Usage

data(Ketchup)

Format

A dataframe containing :

hid individuals identifiers

id purchase identifiers

choice one of heinz, hunts, delmonte, stb (store brand)

price.z price of brand z

Source

Kim, Byong–Do, Robert C. Blattberg and Peter E. Rossi (1995) “Modeling the distribution of price sensitivity and implications for optimal retail pricing”, *Journal of Business Economics and Statistics*, **13(3)**, 291.

References

Journal of Business Economics and Statistics web site : <http://www.amstat.org/publications/jbes/>.

logitform

Model formula for logit models

Description

Two kinds of variables are used in logit models: alternative specific and individual specific variables. `logitform` provides a relevant class to deal with this specificity and suitable methods to extract the elements of the model.

Usage

```
logitform(object)
## S3 method for class 'formula':
logitform(object)
## S3 method for class 'list':
logitform(object)
is.logitform(object)
## S3 method for class 'logitform':
model.matrix(object, data, ...)
## S3 method for class 'logitform':
model.frame(formula, data, ...)
## S3 method for class 'logitform':
terms(x, ...)
## S3 method for class 'logitform':
update(object, new, ...)
```

Arguments

<code>object, x</code>	for the <code>logitform</code> function, a formula or a list, for the <code>update</code> and <code>model.matrix</code> methods, a <code>logitform</code> object,
<code>new</code>	the new formula for the <code>update</code> method,
<code>formula</code>	a <code>logitform</code> object,
<code>data</code>	a <code>data.frame</code> ,
<code>...</code>	further arguments.

Details

a `logitform` is a formula for which the right hand side may contain two parts: the first one contains the alternative specific variables and the second one the individual specific variables. The two parts are separated by a “|” sign. If a standard formula is written, it is assumed that there are no individual specific variables. Specific methods are provided to build correctly the model matrix and to update the formula. The `logitform` function is not intended to be use directly. While using the `mlogit` function, the first argument is automatically coerced to a `logitform` object.

Value

an object of class `logitform`.

Author(s)

Yves Croissant

Examples

```
v <-logitform(y~x1+x2|z1+z2)
update(v, .~.+x3-x1|. +z3-z1)
update(v, .~-x1-x2|. )
update(v, .~.|.-z1-z2)
v <-logitform(y~x1+x2)
update(v, .~-x1+x3)
update(v, .~.|z1+z2)

data("Fishing", package="Ecdat")
colnames(Fishing)[4:11] <- c("pr.beach", "pr.pier", "pr.boat", "pr.charter",
                           "ca.beach", "ca.pier", "ca.boat", "ca.charter")
Fish <- mlogit.data(Fishing, varying=c(4:11), shape="wide", choice="mode")
f <- logitform(mode~pr+ca|income)
X <- model.matrix(f, data=Fish)
X[1:6,]
```

mlogit

Multinomial logit model

Description

Estimation by maximum likelihood of the multinomial logit model, with alternative-specific and/or individual specific variables.

Usage

```
mlogit(formula, data, subset, weights, na.action,
        alt.subset = NULL, relevel = NULL, ...)
## S3 method for class 'mlogit':
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'mlogit':
summary(object, ...)
## S3 method for class 'summary.mlogit':
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'mlogit':
print(x, digits = max(3, getOption("digits") - 2),
      width = getOption("width"), ...)
## S3 method for class 'mlogit':
logLik(object, ...)
## S3 method for class 'mlogit':
```

```
vcov(object, ...)
## S3 method for class 'mlogit':
residuals(object, outcome = TRUE, ...)
## S3 method for class 'mlogit':
fitted(object, outcome = TRUE, ...)
## S3 method for class 'mlogit':
df.residual(object, ...)
## S3 method for class 'mlogit':
terms(x, ...)
## S3 method for class 'mlogit':
model.matrix(object, ...)
## S3 method for class 'mlogit':
update(object, new, ...)
```

Arguments

<code>x</code> , <code>object</code>	an object of class <code>mlogit</code>
<code>formula</code>	a symbolic description of the model to be estimated,
<code>new</code>	an updated formula for the <code>update</code> method,
<code>data</code>	the data,
<code>subset</code>	an optional vector specifying a subset of observations,
<code>weights</code>	an optional vector of weights,
<code>na.action</code>	a function which indicates what should happen when the data contains 'NA's,
<code>alt.subset</code>	a vector of character strings containing the subset of alternative on which the model should be estimated,
<code>reflevel</code>	the base alternative (the one for which the coefficients of individual-specific variables are normalized to 0),
<code>digits</code>	the number of digits,
<code>width</code>	the width of the printing,
<code>outcome</code>	a boolean which indicates, for the <code>fitted</code> and the <code>residuals</code> methods whether a matrix (for each choice, one value for each alternative) or a vector (for each choice, only a value for the alternative chosen) should be returned,
<code>...</code>	further arguments.

Details

Let J being the number of alternatives. The formula may include alternative-specific and individual specific variables. For the latter, $J-1$ coefficients are estimated for each variable. Alternative and individual specific variables are separated by a `|`. For example, if x_1 and x_2 are alternative specific and z_1 and z_2 are individual specific, the formula $y \sim x_1 + x_2 | z_1 + z_2$ describe a model with one coefficient for x_1 and x_2 and $J-1$ coefficients for z_1 and z_2 . $J-1$ intercepts are also estimated. A model without intercepts is defined by the formula: $y \sim x_1 + x_2 - 1 | z_1 + z_2$. To obtain alternative specific coefficients for the alternative-specific variable x_2 , use: $y \sim x_1 + x_2 + x_2 : alt | z_1 + z_2$ (replace `alt` by the relevant variable name if the alternative index is provided). Models with only alternative-specific or individual-specific variables are respectively estimated by $y \sim x_1 + x_2$ and $y \sim 1 | z_1 + z_2$.

The model is estimated with the `maxLik` package and the Newton-Raphson method, using analytic gradient and hessian.

Value

An object of class "mlogit", a list with elements:

<code>coefficients</code>	the named vector of coefficients,
<code>logLik</code>	the value of the log-likelihood,
<code>hessian</code>	the hessian of the log-likelihood at convergence,
<code>gradient</code>	the gradient of the log-likelihood at convergence,
<code>call</code>	the matched call,
<code>est.stat</code>	some information about the estimation (time used, optimisation method),
<code>freq</code>	the frequency of choice,
<code>residuals</code>	the residuals,
<code>fitted.values</code>	the fitted values,
<code>formula</code>	the formula (a <code>logitform</code> object),
<code>expanded.formula</code>	the formula (a <code>formula</code> object),
<code>model</code>	the model frame used,
<code>index</code>	the index of the choice and of the alternatives.

Author(s)

Yves Croissant

References

McFadden, D. (1973) Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka ed., *Frontiers in Econometrics*, New-York: Academic Press.

McFadden, D. (1974) "The Measurement of Urban Travel Demand", *Journal of Public Economics*, 3, pp. 303-328.

Train, K. (2004) *Discrete Choice Modelling, with Simulations*, Cambridge University Press.

See Also

`mlogit.data` to shape the data. `multinom` from package `nnet` performs the estimation of the multinomial logit model with individual specific variables

Examples

```
## Cameron and Trivedi's Microeconometrics p.493
## There are two alternative specific variables : pr (price) and ca (catch)
## and four fishing mode : beach, pier, boat, charter

data("Fishing",package="mlogit")
Fish <- mlogit.data(Fishing,varying=c(4:11),shape="wide",choice="mode")

## a pure "conditional" model without intercepts

summary(mlogit(mode~pr+ca-1,data=Fish))

## a pure "multinomial model"

summary(mlogit(mode~1|income,data=Fish))

## which can also be estimated using multinom (package nnet)

library(nnet)
summary(multinom(mode~income,data=Fishing))

## a "mixed" model

m <- mlogit(mode~pr+ca|income,data=Fish)
summary(m)
```

mlogit.data

data.frame for logit model

Description

shape a `data.frame` in a suitable form for the use of the `mlogit` function.

Usage

```
mlogit.data(data, choice, shape = c("wide", "long"), varying = NULL,
            sep=".", alt.var = NULL, id.var = NULL,
            alt.levels = NULL, opposite = NULL, ...)
```

Arguments

<code>data</code>	a <code>data.frame</code> ,
<code>choice</code>	the variable indicating the choice made: it can be either a logical vector, a numerical vector with 0 where the alternative is not chosen, a factor with level 'yes' when the alternative is chosen
<code>shape</code>	the shape of the <code>data.frame</code> : whether <code>long</code> if each row is an alternative or <code>wide</code> if each row is an observation,

varying	the indexes of the variables that are alternative specific,
sep	the separator of the variable name and the alternative name (only relevant for a wide data.frame),
alt.var	the name of the variable that contains the alternative index (for a long data.frame only) or the name under which the alternative index will be stored (the default name is alt),
id.var	the name of the variable that contains the choice index or the name under which the choice index will be stored,
alt.levels	the name of the alternatives: if null, for a wide data.frame, they are guessed from the variable names and the choice variable (both should be the same), for a long data.frame, they are guessed from the alt.var argument,
opposite	returns the opposite of the specified variables,
...	further arguments passed to reshape.

Value

A data.frame in long format, *i.e.* one line for each alternative. The first two columns contain the choice index and the alternative index. The choice variable is a boolean which indicates the choice made. This function use reshape if the data.frame is in wide format.

Author(s)

Yves Croissant

See Also

reshape

Examples

```
# ModeChoice is a long data.frame

data("TravelMode", package="AER")
TM <- mlogit.data(TravelMode, choice="choice", shape="long",
                 alt.levels=c("air", "train", "bus", "car"))

# Same but the alt variable called mode is provided

TM <- mlogit.data(TravelMode, choice="choice", shape="long",
                 alt.var="mode")

# Same but the chid variable called individual is provided

TM <- mlogit.data(TravelMode, choice="choice", shape="long",
                 id.var="individual", alt.levels=c("air", "train", "bus", "car"))

# Same but with two own provided variables
```

```

TM <- mlogit.data(TravelMode,choice="choice",shape="long",
                 id.var="individual",alt.var="mode")

# Train is a wide data.frame with columns 'choiceid' is the choice
# index, the alternatives are named "ch1" and "ch2", the opposite
# of the variables is returned

data("Train",package="Ecdat")
Train <- mlogit.data(Train,choice="choice",shape="wide",
                   varying=4:11,alt.levels=c("ch1","ch2"),
                   opposite=c("price","time","change","comfort"), sep="")

# Car is a wide data.frame, indexes names are provided

data("Car",package="Ecdat")
Car <-
mlogit.data(Car,varying=5:70,shape="wide",sep="",choice="choice",
            id.var="Mr",alt.var="Car",alt.levels=1:6)

```

Mode

Mode Choice

Description

a cross-section

number of observations : 453

observation : individuals

Usage

data (Mode)

Format

A dataframe containing :

choice one of car, carpool, bus or rail

cost.z cost of mode z

time.z time of mode z

References

Kenneth Train's home page : <http://elsa.berkeley.edu/~train/>.

Train

Stated Preferences for Train Traveling

Description

a cross-section from 1987

number of observations : 2929

observation : individuals

country : Netherland

Usage

```
data(Train)
```

Format

A dataframe containing :

id individual identifiant

choiceid choice identifiant

choice one of choice1, choice2

pricez price of proposition z (z=1,2) in cents of guilders

timez travel time of proposition z (z=1,2) in minutes

comfortz comfort of proposition z (z=1,2), 0, 1 or 2 in decreasing comfort order

changez number of changes for proposition z (z=1,2)

Source

Meijer, Erik and Jan Rouwendal (2005) "Measuring welfare effects in models with random coefficients", *Journal of Applied Econometrics*, **forthcoming**.

Ben-Akiva, M., D. Bolduc and M. Bradley (1993) "Estimation of travel choice models with randomly distributed values of time", *Transportation Research Record*, **1413**, 88–97.

Carson, R.T., L. Wilks and D. Imber (1994) "Valuing the preservation of Australia's Kakadu conservation zone", *Oxford Economic Papers*, **46**, 727–749.

References

Journal of Applied Econometrics data archive : <http://jae.wiley.com/jae/>.

Tuna

Choice of Brand for Tuna

Description

a cross-section

number of observations : 13705

observation : individuals

country : United States

Usage

data(Tuna)

Format

A dataframe containing :

hid individuals identifiers

id purchase identifiers

choice one of skw (Starkist water), cosw (Chicken of the sea water), pw (store-specific private label water), sko (Starkist oil), coso (Chicken of the sea oil)

price.z price of brand z

Source

Kim, Byong-Do, Robert C. Blattberg and Peter E. Rossi (1995) "Modeling the distribution of price sensitivity and implications for optimal retail pricing", *Journal of Business Economics and Statistics*, **13**(3), 291.

References

Journal of Business Economics and Statistics web site : <http://www.amstat.org/publications/jbes/>.

Yogurt

Choice of Brand for Yogurts

Description

a cross-section

number of observations : 2412

observation : individuals

country : United States

Usage

`data(Yogurt)`

Format

A dataframe containing :

id individuals identifiers

choice one of yoplait, dannon, hiland, weight (weight watcher)

feat.z is there a newspaper feature advertisement for brand z ?

price.z price of brand z

Source

Jain, Dipak C., Naufel J. Vilcassim and Pradeep K. Chintagunta (1994) "A random-coefficients logit brand-choice model applied to panel data", *Journal of Business and Economics Statistics*, **12(3)**, 317.

References

Journal of Business Economics and Statistics web site : <http://www.amstat.org/publications/jbes/>.

Index

- *Topic **attribute**
 - mlogit.data, 14
- *Topic **datasets**
 - Car, 1
 - Catsup, 3
 - Cracker, 4
 - Fishing, 5
 - HC, 6
 - Heating, 6
 - Ketchup, 9
 - Mode, 16
 - Train, 17
 - Tuna, 18
 - Yogurt, 19
- *Topic **htest**
 - hmfctest, 7
- *Topic **models**
 - logitform, 10
- *Topic **regression**
 - mlogit, 11

Car, 1

Catsup, 3

Cracker, 4

df.residual.mlogit (*mlogit*), 11

Fishing, 5

fitted.mlogit (*mlogit*), 11

HC, 6

Heating, 6

hmfctest, 7

is.logitform (*logitform*), 10

Ketchup, 9

logitform, 10

logLik.mlogit (*mlogit*), 11

mlogit, 11

mlogit.data, 14

Mode, 16

model.frame.logitform
(*logitform*), 10

model.matrix.logitform
(*logitform*), 10

model.matrix.mlogit (*mlogit*), 11

print.mlogit (*mlogit*), 11

print.summary.mlogit (*mlogit*), 11

residuals.mlogit (*mlogit*), 11

summary.mlogit (*mlogit*), 11

terms.logitform (*logitform*), 10

terms.mlogit (*mlogit*), 11

Train, 17

Tuna, 18

update.logitform (*logitform*), 10

update.mlogit (*mlogit*), 11

vcov.mlogit (*mlogit*), 11

Yogurt, 19