

# R Notebook

## Allgemeines Setup der Daten

```
# read: https://cran.r-project.org/web/packages/mlogit/vignettes/mlogit.pdf

suppressMessages(library(AER))
suppressMessages(library(mlogit))
suppressMessages(library(stats))

# "Data on travel mode choice for travel between Sydney and Melbourne, Australia."
# used columns:
# choice      gibt die Wahl des Verkehrsmittels an (yes/no)
# wait        Wartezeit am Terminal (0 für Auto) *
# travel       Dauer der Reise mit dem gewählten Verkehrsmittel *
# vcost        Kosten des Transports *
# income       Einkommen des Haushalts **

# man erkennt: * ist abhängig von der Wahl, ** nur vom Individuum
data("TravelMode")
```

Korrektes Einsetzen der erklärenden Variable in das mlogit Modell

Alle Koeffizienten beziehen sich auf "car" ("Referenzlevel")

```
# mlogit:
# alternative specific variables xij with a generic coefficient  $\beta$  /
# individual specific variables zi with an alternative specific coefficients  $\gamma_j$  /
# alternative specific variables wij with an alternative specific coefficient  $\delta_j$ 

# Mixed-Logit Model
# zuerst mit Intercept, dann ohne
# choice wird regressiert auf {wait + travel + vcost} und {income}
# ersteres variiert über die Alternativen, zweiteres nicht
mlm.wI <- mlogit(choice ~ wait + travel + vcost | 1 + income | 0, data = TravelMode,
  shape = "long", alt.var = "mode", reflevel = "car")
mlm.woI <- mlogit(choice ~ wait + travel + vcost | 0 + income | 0, data = TravelMode,
  shape = "long", alt.var = "mode", reflevel = "car")

# man sieht in der summary die Frequenzen (Häufigkeiten) der auftretenden Entscheidungen
# und die geschätzten Parameter, kann eigentlich nur Aussagen über die Vorzeichen treffen
summary(mlm.wI)

##
## Call:
## mlogit(formula = choice ~ wait + travel + vcost | 1 + income |
##       0, data = TravelMode, reflevel = "car", shape = "long", alt.var = "mode",
##       method = "nr", print.level = 0)
##
## Frequencies of alternatives:
##      car      air    train      bus
## 0.28095 0.27619 0.30000 0.14286
##
## nr method
## 5 iterations, 0h:0m:0s
```

```
## g'(-H)^-1g = 0.000546
## successive function values within tolerance limits
##
## Coefficients :
##               Estimate Std. Error z-value Pr(>|z|)
## air:(intercept)  4.24742503  1.00650942  4.2200 2.444e-05 ***
## train:(intercept) 5.48954901  0.65069739  8.4364 < 2.2e-16 ***
## bus:(intercept)   4.06305942  0.68715749  5.9129 3.362e-09 ***
## wait              -0.09528341  0.01035524 -9.2015 < 2.2e-16 ***
## travel            -0.00366471  0.00086797 -4.2222 2.420e-05 ***
## vcost             -0.00449878  0.00721124 -0.6239  0.5327
## air:income        -0.00210282  0.01209542 -0.1739  0.8620
## train:income      -0.05799787  0.01438418 -4.0321 5.529e-05 ***
## bus:income        -0.02521351  0.01567725 -1.6083  0.1078
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -182.22
## McFadden R^2:  0.35784
## Likelihood ratio test : chisq = 203.08 (p.value = < 2.22e-16)
```

```
summary(mlm.wol)
```

```
##
## Call:
## mlogit(formula = choice ~ wait + travel + vcost | 0 + income |
##         0, data = TravelMode, reflevel = "car", shape = "long", alt.var = "mode",
##         method = "nr", print.level = 0)
##
## Frequencies of alternatives:
##      car      air    train      bus
## 0.28095 0.27619 0.30000 0.14286
##
## nr method
## 4 iterations, 0h:0m:0s
## g'(-H)^-1g = 4.37E-08
## gradient close to zero
##
## Coefficients :
##               Estimate Std. Error z-value Pr(>|z|)
## wait          -0.04675702  0.00618667 -7.5577 4.108e-14 ***
## travel        -0.00301083  0.00065043 -4.6290 3.674e-06 ***
## vcost          0.00083156  0.00564134  0.1474 0.8828128
## income:air     0.02787801  0.00845071  3.2989 0.0009707 ***
## income:train   0.02787408  0.00769462  3.6225 0.0002917 ***
## income:bus     0.02399300  0.00829853  2.8912 0.0038373 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -238.18
```

Nochmal, diesmal mit einem Nested Modell

```
# Nested-Logit Model
nlm <- mlogit(choice ~ wait + travel + vcost | 1 + income | 0, data = TravelMode,
             shape = "long", alt.var = "mode", reflevel = "car",
```

```

    nests = list(bt = c("bus", "train"), ca = c("car", "air"))

summary(nlm)

```

```

##
## Call:
## mlogit(formula = choice ~ wait + travel + vcost | 1 + income |
## 0, data = TravelMode, reflevel = "car", nests = list(bt = c("bus",
## "train"), ca = c("car", "air")), shape = "long", alt.var = "mode")
##
## Frequencies of alternatives:
##      car      air   train      bus
## 0.28095 0.27619 0.30000 0.14286
##
## bfgs method
## 16 iterations, 0h:0m:0s
## g'(-H)^-1g = 8.35E-07
## gradient close to zero
##
## Coefficients :
##              Estimate Std. Error z-value Pr(>|z|)
## air:(intercept)  2.7962212  1.4963957  1.8686 0.0616733 .
## train:(intercept) 6.9440379  1.3142291  5.2837 1.266e-07 ***
## bus:(intercept)   5.3127359  1.1042537  4.8112 1.501e-06 ***
## wait             -0.1144873  0.0227964 -5.0222 5.109e-07 ***
## travel            -0.0077779  0.0020172 -3.8559 0.0001153 ***
## vcost             -0.0078017  0.0101448 -0.7690 0.4418752
## air:income         0.0059258  0.0268487  0.2207 0.8253188
## train:income      -0.0573975  0.0174383 -3.2915 0.0009967 ***
## bus:income        -0.0182483  0.0244624 -0.7460 0.4556846
## iv:bt              0.9445671  0.3140772  3.0074 0.0026346 **
## iv:ca              2.3272270  0.7135084  3.2617 0.0011076 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Log-Likelihood: -176.69
## McFadden R^2:  0.37731
## Likelihood ratio test : chisq = 214.13 (p.value = < 2.22e-16)

```

```

# erneent wieder nur Vorzeichen aussagekräftig -> nun "besseres" Vorzeichen bei air

```

```

# zusätzlich aus dem helpfile

```

```

# iv:* sind die jeweiligen "inculsive values" (Formel im helpfile entspricht "unserer inclusive sum") d

```

Optional:

Wir wollen die IIA testen. Gilt diese, so ist das Ordnen der Alternativen in Nests nicht zielführend. Es läuft also auf einen Vergleich des MLM mit dem NLM hinaus. Einzig wichtig dabei ist, dass wir beide mit/ohne Intercept vergleichen (die Modelle müssen bsi auf das Nesting ident sein). Somit "entspricht" unser MLM der Parameterrestriktion der Nullhypothese ( $\rho_i = 1$ ), bzw. das NLM der Alternativhypothese

```

lrtest(mlm.wI, nlm)

```

```

## Likelihood ratio test
##
## Model 1: choice ~ wait + travel + vcost | 1 + income | 0

```

```
## Model 2: choice ~ wait + travel + vcost | 1 + income | 0
##   #Df LogLik Df   Chisq Pr(>Chisq)
## 1    9 -182.22
## 2   11 -176.69  2 11.048   0.003991 **
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

 $H_0 : \rho_1 = \rho_2 = 1$ , also die  $u_{ij1}, \dots, u_{ijm_j}$  (also innerhalb des Nests) sind unabhängig
 $H_1 : \rho_i \neq 1$  für ein  $i$ , also keine IIA Struktur

Wir erstellen also:
 $T_{LR} = \frac{L_n(H_0)}{L_n(H_1)}$ , dann ist
 $-2 \ln T_{LR}$  in Verteilung  $\chi_q^2$  mit  $q = 11 - 9 = 2$  Freiheitsgraden, also
 $\bar{T} = 2(\ln(L_n(H_1)) - \ln(L_n(H_0)))$ 
# as.double um den tag der Log.Lik mit df=11 zu entfernen (der stimmt ja jetzt nicht mehr)
T <- 2 * as.double(nlm$logLik - mlm.wI$logLik)
T

## [1] 11.0476
# 99% Quantil einer chi-sqr Verteilung mit 2 Freiheitsgraden
qchisq(.99, df=2)

## [1] 9.21034
# lower.tail erzwingt P[X > x] statt P[X <= x]
pchisq(T, df=2, lower.tail = FALSE)

## [1] 0.003990658
```