Discrete Choice

AECN 396/896-002

Discrete Choice Analysis

- Focus on understanding choices that are discrete (not continuous)
  - Whether you own a car or not (binary choice)
  - Whether you use an iPhone, Android, or other types of cell phones (Multinomial choice)
  - Which recreation sites you visit this winter (multinomial)
- Linear models we have seen are often not appropriate

Binary Response Model

Binary response

$y = 0 \;$ (if you do not own a car)

$y = 1 \;$ (if you own at least one car)

Question we would like to answer

How do variables $x_1, \ldots, x_k$ affect the status of y (the choice of whether to own at least one car or not)?

Binary response

We try to model the probability of $y = 1$ (own at least one car)

$$Pr(y = 1|x_1, \ldots, x_k) = f(x_1, \ldots, x_k)$$

as a function of independent variables.

Linear Probability Model

$$Pr(y = 1|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Linear Probability Model

$$Pr(y = 1|x_1, \ldots, x_k) = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$$

Drawback

There is no guarantee that the predicted probability is bounded within [0, 1].

How about this?

$$Pr(y = 1|x_1, \dots, x_k) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

where $0 < G(z) < 1$ for all real numbers z

Different choices of $G()$ lead to different models.

Logit model

$G(z) = exp(z)/[1 + exp(z)] = \frac{e^z}{1+e^z}$

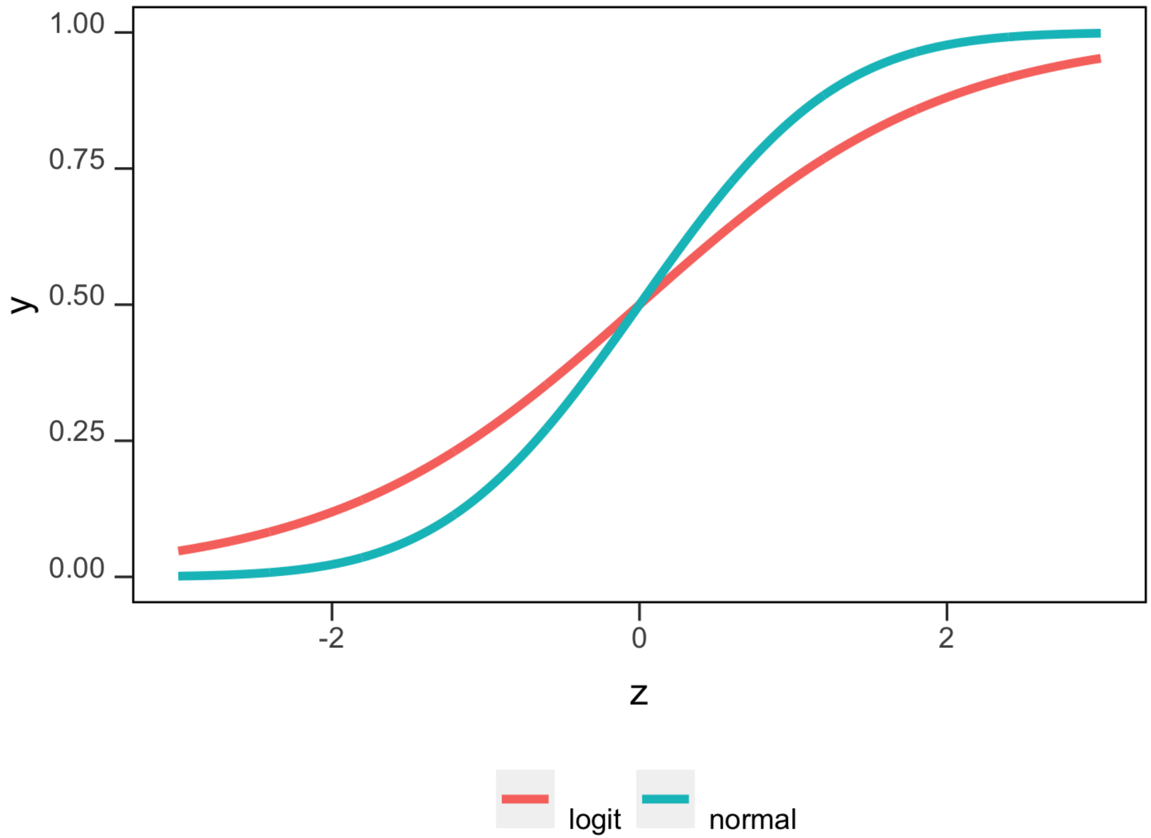
where $z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

Probit model

$G(z) = \Phi(z)$

where $\Phi(z)$ is the standard normal cumulative distribution function

This what $G()$ looks like for logit and probit.

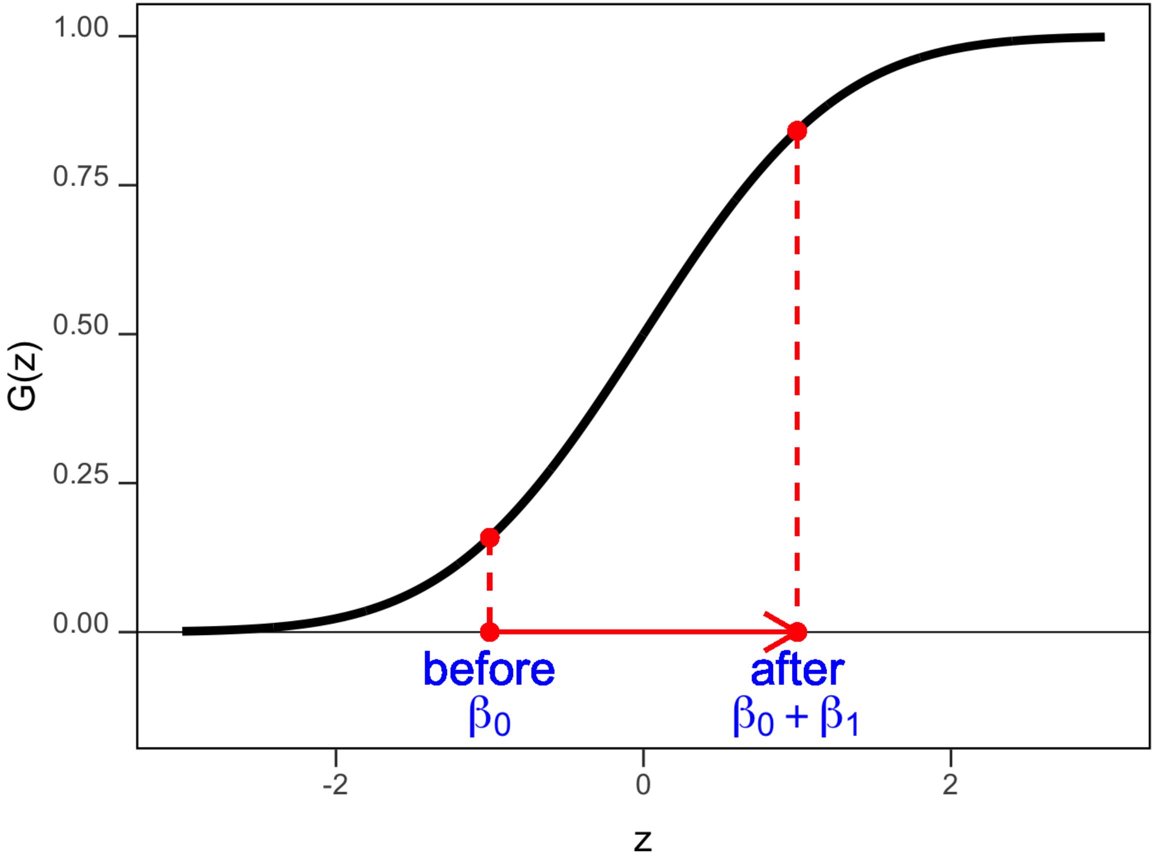


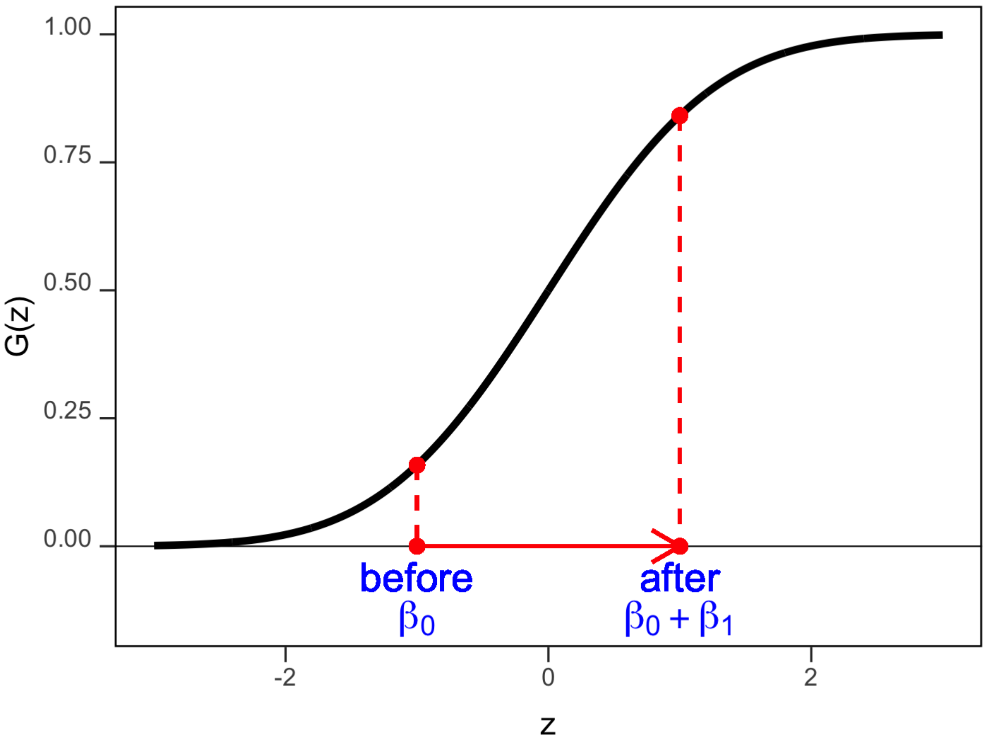
$$Pr(y = 1|x_1, \ldots, x_k) = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- What do βs measure?
- How do we interpret them?

Before: $x_1 = 0, \ldots, x_k = 0 \Rightarrow z = \beta_0$

After: $x_1 = 1$ and $x_2 = 0, \ldots, x_k = 0 \Rightarrow z = \beta_0 + \beta_1$





- βs measure how far you move along the x-axis
- βs does not directly measure how independent variables influence the probability of $y = 1$

To understand the marginal impact of x_k on $Prob(y = 1)$ (how a change in x_k affects the likelihood of owning a car), you need to do a bit of math.

Model

$Pr(y = 1|x_1, \ldots, x_k) = G(z)$

$z = \beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k$

marginal impact

Differentiating both sides with respect to x_k,

$$\begin{aligned}\frac{\partial Pr(y=1|x_1,\ldots,x_k)}{\partial x_k} &= G'(z) \times \frac{\partial z}{\partial x_k} \\ &= G'(z) \times \beta_k\end{aligned}$$

marginal impact

Differentiating both sides with respect to x_k,

$$\frac{\partial Pr(y=1|x_1,...,x_k)}{\partial x_k} = G'(z) \times \frac{\partial z}{\partial x_k}$$
$$= G'(z) \times \beta_k$$

Notes

- The marginal impact of an independent variable depends on the values of all the independent variables: $G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$
- Since $G'()$ is always positive, the sign of the marginal impact of an independent variable on $Prob(y = 1)$ is always the same as the sign of its coefficient

Estimation of Binary Choice Models

- Linear models: OLS
- Binary choice models: Maximum Likelihood Estimation (MLE)

OLS

Find parameters that makes the sum of residuals squared the smallest

MLE (very loosely put)

Find parameters (βs) that makes what we observed (collection of binary decisions made by different individuals) most likely (Maximum Likelihood)

Observed decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Observed decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Probability of individual decisions

- $\text{Individual 1} : Prob(y_1 = 1|\mathbf{x_1}) = G(z_1)$
- $\text{Individual 2} : Prob(y_2 = 0|\mathbf{x_2}) = 1 - G(z_2)$

where

- $\mathbf{x_i}$ is a collection of independent variables for individual i $(x_{1,i}, \ldots, x_{k,i})$.
- $z_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}$

Observed decisions made by two individuals

- Individual 1: $y = 1$ (own at least one car)
- Individual 2: $y = 0$ (does not own a car)

Probability of individual decisions

- $\text{Individual 1}: Prob(y_1 = 1|\mathbf{x_1}) = G(z_1)$
- $\text{Individual 2}: Prob(y_2 = 0|\mathbf{x_2}) = 1 - G(z_2)$

where

- $\mathbf{x_i}$ is a collection of independent variables for individual i $(x_{1,i}, \ldots, x_{k,i})$.
- $z_i = \beta_0 + \beta_1 x_{1,i} + \cdots + \beta_k x_{k,i}$

Probability of a collection of decisions

The probability that we observe a collection of choices made by them (if their decisions are independent)

$$Prob(y_1 = 1|\mathbf{x_1}) \times Prob(y_2 = 0|\mathbf{x_2}) = G(z_1) \times [1 - G(z_2)]$$

which we call likelihood function.

Probability of a collection of decisions

The probability that we observe a collection of choices made by them (if their decisions are independent)

$$Prob(y_1 = 1|\mathbf{x_1}) \times Prob(y_2 = 0|\mathbf{x_2}) = G(z_1) \times [1 - G(z_2)]$$

which we call likelihood function.

Probability of a collection of decisions

The probability that we observe a collection of choices made by them (if their decisions are independent)

$$Prob(y_1 = 1|\mathbf{x_1}) \times Prob(y_2 = 0|\mathbf{x_2}) = G(z_1) \times [1 - G(z_2)]$$

which we call likelihood function.

MLE

$$Max_{\beta_1,...,\beta_k} \; G(z_1) \times [1 - G(z_2)]$$

MLE of Binary Choice Model in General

Maximize the likelihood function:

$$Max_{\beta_1,...,\beta_k} \;\; L$$

where $L = \Pi_{i=1}^{n} \Big[y_i \times G(z_i) + (1 - y_i) \times (1 - G(z_i)) \Big]$ is the likelihood function.

Log-likelihood function

$$\begin{aligned} LL &= log\Big(\Pi_{i=1}^{n} \Big[y_i \times G(z_i) + (1 - y_i) \times (1 - G(z_i)) \Big] \Big) \\ &= \sum_{i=1}^{n} log\Big(y_i \times G(z_i) + (1 - y_i) \times (1 - G(z_i)) \Big) \end{aligned}$$

MLE with（LL）

$$argmax_{\beta_1,...,\beta_k} \;\; L \equiv argmax_{\beta_1,...,\beta_k} \;\; LL$$

Implementation in R with an example

Participation of females in labor force:

$$Pr(inlf = 1|\mathbf{x}) = G(z)$$

where

$$z = \beta_0 + \beta_1 nwifeinc + \beta_2 educ + \beta_3 exper$$

$$+\beta_4 exper^2 + \beta_5 age + \beta_6 kidslt6 + \beta_7 kidsge6$$

- $inlf$: 1 if in labor force in 1975, 0 otherwise
- $nwifeinc$: earning as a family if she does not work
- $kidslt6$: # of kids less than 6 years old
- $kidsge6$: # of kids who are 6-18 year old

```r
#--- import the data ---#
data <- read.dta13("MROZ.dta") %>%
  mutate(exper2 = exper^2)

#--- take a look ---#
dplyr::select(data, inlf, nwifeinc, kidslt6, kidsge6, educ) %>%
  head()
```

```
##   inlf  nwifeinc kidslt6 kidsge6 educ
## 1    1 10.910060       1       0   12
## 2    1 19.499981       0       2   12
## 3    1 12.039910       1       3   12
## 4    1  6.799996       0       3   12
## 5    1 20.100058       1       2   14
## 6    1  9.859054       0       0   12
```

```
##   inlf  nwifeinc kidslt6 kidsge6 educ
## 1    1 10.910060       1       0   12
## 2    1 19.499981       0       2   12
## 3    1 12.039910       1       3   12
## 4    1  6.799996       0       3   12
## 5    1 20.100058       1       2   14
## 6    1  9.859054       0       0   12
```

For individual 1 (row 1 of the data),

$$z_1 = \beta_0 + \beta_1 10.91 + \beta_2 12 + \beta_3 14 + \beta_4 196 + \beta_5 32 + \beta_6 1 + \beta_7 0$$

The probability that individual 1 would make the decision he/she made given βs is:

$G(z_1)$ (a function of βs)

```
##     inlf nwifeinc kidslt6 kidsge6 educ
## 748    0    5.330       0       2   12
## 749    0   28.200       0       2   13
## 750    0   10.000       2       3   12
## 751    0    9.952       0       0   12
## 752    0   24.984       0       0   12
## 753    0   28.363       0       3    9
```

For individual 753 (row 753 of the data),

$$z_{753} = \beta_0 + \beta_1 28.36 + \beta_2 9 + \beta_3 12 + \beta_4 144 + \beta_5 39 + \beta_6 0 + \beta_7 3$$

The probability that individual 753 would make the decision he/she made given βs is:

$1 - G(z_{753})$ (a function of βs)

Multiply all the probabilities of observed choices given βs,

$$L = G(z_1) \times G(z_2) \times \ldots [1 - G(z_753)]$$

Multiply all the probabilities of observed choices given βs,

$$L = G(z_1) \times G(z_2) \times \ldots [1 - G(z_753)]$$

$$LL = log\Big(G(z_1) \times G(z_2) \times \ldots [1 - G(z_753)]\Big)$$

Multiply all the probabilities of observed choices given βs,

$$L = G(z_1) \times G(z_2) \times \ldots [1 - G(z_753)]$$

$$LL = log\Big(G(z_1) \times G(z_2) \times \ldots [1 - G(z_753)]\Big)$$

Solve the following problems to estimate βs:

$$Max_{\beta_1, \ldots, \beta_7} \quad LL$$

Estimating binary choice model using `(R)`

You can use the `glm()` function (no new packages installation necessary) when using cross-sectional data

- `glm` refers to Generalized Linear Model, which encompass linear models we have been using
- you specify the `family` option to tell what kind of model you are estimating

Probit model estimation

```r
probit_lf <- glm(
  #--- formula ---#
  inlf ~ nwifeinc + educ + exper + exper2 + age + kidslt6 + kidsge6,
  #--- data ---#
  data = data,
  #--- models ---#
  family = binomial(link = "probit")
)
```

family option

- `binomial()`: tells R that your dependent variable is binary
- `link = "probit"`: tells R that you want to use the cumulative distribution function of the standard normal distribution as $G()$ in $Prob(y = 1|\mathbf{x}) = G(z)$

```r
msummary(
  probit_lf,
  stars = TRUE,
  gof_omit = "IC|F",
  output = "flextable"
) %>%
  fontsize(
    size = 9,
    part = "all"
  ) %>%
  autofit()
```

	Model 1
(Intercept)	0.270
	(0.508)
nwifeinc	-0.012
	(0.005)
educ	0.131**
	(0.025)
exper	0.123*
	(0.019)
exper2	-0.002
	(0.001)
age	-0.053
	(0.008)
kidslt6	-0.868
	(0.118)
kidsge6	0.036
	(0.044)
Num.Obs.	753
Log.Lik.	-401.302

+ p < 0.1, *p < 0.05,* ***p < 0.01,*** p < 0.001

Logit model estimation

```
logit_lf <- glm(
  #--- formula ---#
  inlf ~ nwifeinc + educ + exper + exper2 + age + kidslt6 + kidsge6,
  #--- data ---#
  data = data,
  #--- models ---#
  family = binomial(link = "logit")
)
```

family option

- `binomial()`: tells R that your dependent variable is binary
- `link = "logit"`: tells R that you want to use $G(z) = \frac{e^z}{1+e^z}$ in $Prob(y = 1|\mathbf{x}) = G(z)$

```r
msummary(
  logit_lf,
  stars = TRUE,
  gof_omit = "IC|F",
  output = "flextable"
) %>%
  fontsize(
    size = 9,
    part = "all"
  ) %>%
  autofit()
```

	Model 1
(Intercept)	0.425
	(0.860)
nwifeinc	-0.021
	(0.008)
educ	0.221**
	(0.043)
exper	0.206*
	(0.032)
exper2	-0.003
	(0.001)
age	-0.088
	(0.015)
kidslt6	-1.443
	(0.204)
kidsge6	0.060
	(0.075)
Num.Obs.	753
Log.Lik.	-401.765

+ p < 0.1, *p < 0.05,* ***p < 0.01,*** p < 0.001

Important

- You cannot directly compare the coefficient on the same variable from probit and logit! The fact that the coefficient on `educ` is higher from the logit model does not mean the logit model is suggesting `educ` is more influential than the probit model suggests. They are on different scales.

Post-estimation operations and diagnostics

Log-likelihood (fitted)

$LL = \sum_{i=1}^{n} log\Big(y_i \times G(\hat{z}_i) + (1 - y_i) \times (1 - G(\hat{z}_i))\Big)$

- $\hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
- $G(\hat{z}_i)$ is the fitted value of $Prob(y = 1|\mathbf{x})$

Example

- $G(\hat{z}_i) = 0.9$: predicted that individual i is very likely to own a car
- $y_i = 0$: in reality, individual i does not own a car

$\Rightarrow$

$log\Big(0 \times 0.9 + (1 - 0) \times (1 - 0.9)\Big) = log(0.1) = -2.3$

Log-likelihood (fitted)

$$LL = \sum_{i=1}^{n} log\Big(y_i \times G(\hat{z}_i) + (1 - y_i) \times (1 - G(\hat{z}_i))\Big)$$

- $\hat{z}_i = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
- $G(\hat{z}_i)$ is the fitted value of $Prob(y = 1|\mathbf{x})$

Example

- $G(\hat{z}_i) = 0.9$: predicted that individual i is very likely to own a car
- $y_i = 1$: in reality, individual i indeed owns a car

$\Rightarrow$

$$log\Big(1 \times 0.9 + (1 - 1) \times (1 - 0.9)\Big) = log(0.9) = -0.11$$

Log-likelihood (fitted)

So, the better your prediction (model fit) is, the the greater (less negative) LL is.

McFadden's pseudo- R^2

A measure of how much better your model is compared to the model with only the intercept.

$$pseudo - R^2 = 1 - LL/LL_0$$

where LL_0 is the log-likelihood when you include only the intercept.

R code

```r
logit_lf_0 <- glm(
  inlf ~ 1,
  data = data,
  family = binomial(link = "logit")
)

#--- extract LL using the logLik() function ---#
(LL0 <- logLik(logit_lf_0))
```

```
## 'log Lik.' -514.8732 (df=1)
```

```r
#--- extract LL using the logLik() function from your preferred model ---#
(LL <- logLik(logit_lf))
```

```
## 'log Lik.' -401.7652 (df=8)
```

```r
#--- pseudo R2 ---#
1 - LL / LL0
```

```
## 'log Lik.' 0.2196814 (df=8)
```

Alternatively

```
#--- or more easily ---#
1 - logit_lf$deviance / logit_lf$null.deviance
```

```
## [1] 0.2196814
```

```
#--- what are deviances? ---#
logit_lf$null.deviance # = -2*LL0
```

```
## [1] 1029.746
```

```
logit_lf$deviance # = -2*LL
```

```
## [1] 803.5303
```

- `null.deviance` $= -2 \times LL_0$
- `deviance` $= -2 \times LL$

Testing joint significance

You can do Likelihood Ratio (LR) test:

$$LR = 2(LL_{unrestricted} - LL_{restricted}) \sim \chi^2_{df\_restrictions}$$

where $df\_restrictions$ is the number of restrictions.

Note

LR test is very similar conceptually to F-test.

Example

- H_0 : the coefficients on $exper$, $exper2$, and age are 0
- H_1 : H_0 is false

```r
#--- unrestricted ---#
logit_ur <- glm(
  inlf ~ nwifeinc + educ + exper + exper2 + age + kidslt6 + kidsge6,
  data = data, family = binomial(link = "logit")
)

#--- restricted ---#
logit_r <- glm(
  inlf ~ nwifeinc + educ + kidslt6 + kidsge6,
  data = data, family = binomial(link = "logit")
)

#--- LR test using lrtest() from the lmtest package ---#
library(lmtest)
lrtest(logit_r, logit_ur)
```

```
## Likelihood ratio test
##
## Model 1: inlf ~ nwifeinc + educ + kidslt6 + kidsge6
## Model 2: inlf ~ nwifeinc + educ + exper + exper2 + age + kidslt6 + kidsge6
##   #Df  LogLik Df  Chisq Pr(>Chisq)
## 1   5 -464.92
## 2   8 -401.77  3 126.32  < 2.2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

Prediction

After estimating a binary choice model, you can easily predict the following two

- $\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$
- $Prob\widehat{(y = 1}|\mathbf{x}) = G(\hat{z}) = G(\hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k)$

R code

```r
#--- z hat ---#
z <- predict(probit_lf, type = "link")
head(z)
```

```
##         1         2         3         4         5         6
## 0.5071349 0.6624576 0.5116317 0.7423429 0.1972781 0.8837878
```

```r
#--- G(z) hat ---#
Gz <- predict(probit_lf, type = "response")
head(Gz)
```

```
##         1         2         3         4         5         6
## 0.6939699 0.7461610 0.6955456 0.7710602 0.5781950 0.8115946
```

Marginal effect of an independent variable

- Coefficient estimates across different models (probit and logit) are not meaningful because the same value of a coefficient estimate means different things
- They are the estimates of βs, not the direct impact of the independent variables on the $Prob(y = 1)$

Marginal effect of an independent variable

$$\frac{\partial Pr(y=1|x_1,\ldots,x_k)}{\partial x_k} = G'(z) \times \beta_k$$

- the marginal impact depends on the current levels of all the independent variables
- we typically report one of the two types of marginal impacts
    - (becoming obsolete) the marginal impact at the mean (average person): when all the independent variables take on their respective means
    - the average of the marginal impacts calculated for each of all the individuals observed

Marginal impact at the mean

$$\frac{\partial Pr(y=1|x_1,...,\bar{x_k})}{\partial x_k} = G'(\beta_0 + \beta_1 \bar{x_1} + \cdots + \beta_k \bar{x_k}) \times \beta_k$$

Mean marginal impact (MME)

$$\sum_{i=1}^{n} \frac{\partial Pr(y_i=1|x_{i,1},...,x_{i,k})}{\partial x_k} = \sum_{i=1}^{n} G'(z_i) \times \beta_k$$

R codes to get MME

$$\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

```
#--- get z for all the individuals ---#
z <- predict(probit_lf, type = "link")
```

R codes to get MME

$$\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

```
#--- get z for all the individuals ---#
z <- predict(probit_lf, type = "link")
```

$$G'(\beta_0 + \beta_1 \bar{x}_1 + \cdots + \beta_k \bar{x}_k)$$

where $G(z)$ is the cumulative distribution function for the standard normal distribution.

```
#--- get G'(z) ---#
Gz_indiv <- dnorm(z)
```

R codes to get MME

$$\hat{z} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \cdots + \hat{\beta}_k x_k$$

```
#--- get z for all the individuals ---#
z <- predict(probit_lf, type = "link")
```

$$G'(\beta_0 + \beta_1 \bar{x_1} + \cdots + \beta_k \bar{x_k})$$

where $G(z)$ is the cumulative distribution function for the standard normal distribution.

```
#--- get G'(z) ---#
Gz_indiv <- dnorm(z)
```

$$G'(\beta_0 + \beta_1 \bar{x_1} + \cdots + \beta_k \bar{x_k}) \times \beta_k$$

```
#--- mean marignal impact of eduction ---#
mean(Gz_indiv) * probit_lf$coef["educ"]
```

```
##       educ
## 0.03937009
```

Fortunately, the `margins` package provides you with a more convenient way of calculating MMEs.

```r
library(margins)

#--- calculate MME based on the probit estimation ---#
mme_lf <- margins(probit_lf, type = "response")

#--- get the summary ---#
summary(mme_lf)
```

```
##    factor     AME     SE       z      p   lower   upper
##       age -0.0159 0.0024 -6.7392 0.0000 -0.0205 -0.0113
##      educ  0.0394 0.0073  5.4186 0.0000  0.0251  0.0536
##     exper  0.0371 0.0052  7.1779 0.0000  0.0270  0.0472
##    exper2 -0.0006 0.0002 -3.2050 0.0014 -0.0009 -0.0002
##   kidsge6  0.0108 0.0132  0.8189 0.4129 -0.0151  0.0367
##   kidslt6 -0.2612 0.0319 -8.1860 0.0000 -0.3237 -0.1986
##  nwifeinc -0.0036 0.0015 -2.4604 0.0139 -0.0065 -0.0007
```

Multinomial Choice Model

Multinomial Choice

Instead of two options, you are picking one option out of more than two options

- which carrier?
  - Verizon
  - Sprint
  - AT\&T
  - T-mobile
- which transportation means to commute?
  - drive
  - Uber
  - bus
  - train
  - bike

Multinomial logit model

The most popular model to analyze multinomial choice

- environmental evaluation
- tranposrtation
- marketing

Understanding multinomial logit model

Choice of train route options

- 10 euros, 30 minutes travel time, one change
- 20 euros, 20 minutes travel time, one change
- 22 euros, 22 minutes travel time, no change

Understanding multinomial logit model

Choice of train route options

- 10 euros, 30 minutes travel time, one change
- 20 euros, 20 minutes travel time, one change
- 22 euros, 22 minutes travel time, no change

Associated utility

- $V_1 = \alpha_1 + \beta 10 + \gamma 30 + \rho 1 + v_1$
- $V_2 = \alpha_2 + \beta 20 + \gamma 20 + \rho 1 + v_2$
- $V_3 = \alpha_3 + \beta 22 + \gamma 22 + \rho 0 + v_3$

Choice probability

Logit model assumes that the probability of choosing an alternative is the following:

- $P_1 = \frac{e^{V_1}}{e^{V_1}+e^{V_2}+e^{V_3}}$
- $P_2 = \frac{e^{V_2}}{e^{V_1}+e^{V_2}+e^{V_3}}$
- $P_3 = \frac{e^{V_3}}{e^{V_1}+e^{V_2}+e^{V_3}}$

Choice probability

Logit model assumes that the probability of choosing an alternative is the following:

- $P_1 = \frac{e^{V_1}}{e^{V_1}+e^{V_2}+e^{V_3}}$
- $P_2 = \frac{e^{V_2}}{e^{V_1}+e^{V_2}+e^{V_3}}$
- $P_3 = \frac{e^{V_3}}{e^{V_1}+e^{V_2}+e^{V_3}}$

Notes

- $0 < P_j < 1, {}^\forall j = 1, 2, 3$
- $\sum_{j=1}^{3} = 1$

Modeled probability of choices

Modeled probability of observing individual i choosing the option i chose

$$P_i = \Pi_{j=1}^{3} y_{i,j} \times P_j$$

where $y_{i,j} = 1$ if i chose j, 0 otherwise.

Example

$$y_{i,1} = 0, \;\; y_{i,2} = 1, \;\; y_{i,3} = 0$$

$$P_i = \Pi_{j=1}^{3} y_{i,j} \times P_j = 0 \times P_1 + 1 \times P_2 + 0 \times P_3$$

The probability of observing a series of chocies made by all the subjects is

$$LL = \Pi_{i=1}^{n} P_i = \Pi_{i=1}^{n} \Pi_{j=1}^{3} y_{i,j} \times P_j$$

if choices made by the subjects are independent with each other.

The probability of observing a series of chocies made by all the subjects is

$$LL = \Pi_{i=1}^{n} P_i = \Pi_{i=1}^{n} \Pi_{j=1}^{3} y_{i,j} \times P_j$$

if choices made by the subjects are independent with each other.

MLE

$$Max_{\beta,\gamma,\rho} \;\; log(LL)$$

Interpretation of the coefficients

Model in general

$V_{i,j} = \alpha_j + \beta_1 x_{1,i,j} + \cdots + \beta_k x_{k,i,j}$

$P_{i,j} = \frac{e^{V_{i,j}}}{\sum_{k=1}^{J} e^{V_{i,k}}}$

Interpretation of the coefficients

Model in general

$$V_{i,j} = \alpha_j + \beta_1 x_{1,i,j} + \cdots + \beta_k x_{k,i,j}$$

$$P_{i,j} = \frac{e^{V_{i,j}}}{\sum_{k=1}^{J} e^{V_{i,k}}}$$

Interpretation of the coefficients

$$\frac{\partial P_{i,j}}{\partial x_{k,i,j}} = \beta_k P_{i,j}(1 - P_{i,j})$$

- A marginal change in kth variable for alternative j would change the probability of choosing alternative j by $\beta_k P_{i,j}(1 - P_{i,j})$
- the sign of the impact is the same as the sign of the coefficient

Implementation in `(R)`

You can use $mlogit$ package to estimate multinomial logit models:

- format your data in a specific manner
- convert your data using $mlogit.data()$
- estimate the model using $mlogit()$

```r
#--- library ---#
library(mlogit)

#--- get the travel mode data from the mlogit package ---#
data("TravelMode", package = "AER")

#--- take a look at the data ---#
# first 10 rows
head(TravelMode, 10)
```

```
##    individual  mode choice wait vcost travel gcost income size
## 1           1   air     no   69    59    100    70     35    1
## 2           1 train     no   34    31    372    71     35    1
## 3           1   bus     no   35    25    417    70     35    1
## 4           1   car    yes    0    10    180    30     35    1
## 5           2   air     no   64    58     68    68     30    2
## 6           2 train     no   44    31    354    84     30    2
## 7           2   bus     no   53    25    399    85     30    2
## 8           2   car    yes    0    11    255    50     30    2
## 9           3   air     no   69   115    125   129     40    1
## 10          3 train     no   34    98    892   195     40    1
```

R code: data preparation

```
#--- convert the data ---#
TM <- mlogit.data(TravelMode,
  shape = "long", # what format is the data in?
  choice = "choice", # name of the variable that indicates choice made
  chid.var = "individual", # name of the variable that indicates who made choices
  alt.var = "mode" # the name of the variable that indicates options
)
```

```r
#--- take a look at the data ---#
# first 10 rows
head(TM, 10)
```

```
## ~~~~~~~
##  first 10 observations out of 840
## ~~~~~~~
##    individual  mode choice wait vcost travel gcost income size    idx
## 1           1   air  FALSE   69    59    100    70     35    1  1:air
## 2           1 train  FALSE   34    31    372    71     35    1 1:rain
## 3           1   bus  FALSE   35    25    417    70     35    1  1:bus
## 4           1   car   TRUE    0    10    180    30     35    1  1:car
## 5           2   air  FALSE   64    58     68    68     30    2  2:air
## 6           2 train  FALSE   44    31    354    84     30    2 2:rain
## 7           2   bus  FALSE   53    25    399    85     30    2  2:bus
## 8           2   car   TRUE    0    11    255    50     30    2  2:car
## 9           3   air  FALSE   69   115    125   129     40    1  3:air
## 10          3 train  FALSE   34    98    892   195     40    1 3:rain
```

```r
#--- estimate ---#
ml_reg <- mlogit(choice ~ wait + vcost + travel, data = TM)
```

```r
summary(ml_reg)
```

```
##
## Call:
## mlogit(formula = choice ~ wait + vcost + travel, data = TM, method = "nr")
##
## Frequencies of alternatives:choice
##     air   train     bus     car
## 0.27619 0.30000 0.14286 0.28095
##
## nr method
## 5 iterations, 0h:0m:0s
## g'(-H)^-1g = 0.000192
## successive function values within tolerance limits
##
## Coefficients :
##                      Estimate  Std. Error z-value  Pr(>|z|)
## (Intercept):train -0.78666667  0.60260733 -1.3054   0.19174
## (Intercept):bus   -1.43363372  0.68071345 -2.1061   0.03520 *
## (Intercept):car   -4.73985647  0.86753178 -5.4636 4.665e-08 ***
## wait              -0.09688675  0.01034202 -9.3683 < 2.2e-16 ***
```

Understanding the results

```
summary(ml_reg)$coef
```

```
## (Intercept):train   (Intercept):bus   (Intercept):car              wait            \
##      -0.786666672      -1.433633718      -4.739856473      -0.096886747      -0.0139
## attr(,"names.sup.coef")
## character(0)
## attr(,"fixed")
## (Intercept):train   (Intercept):bus   (Intercept):car              wait            \
##             FALSE             FALSE             FALSE             FALSE            F
## attr(,"sup")
## character(0)
```

- intercept for air is dropped $($$air$ is the base)
  - train:(intercept) is -0.786 means that train is less likely to be chosen if all the other included variables are the same
- the greater the travel time, the less likely the option is chosen

Count data (Poisson)

Count data (Poisson)

Count variables take non-negative discrete integers $(0, 1, \ldots,)$

- the number of times individuals get arrested in a year
- the number of cars owned by a family
- the number of kids in a family

Poisson regression

By far the most popular choice to analyze count variables is Poisson regression

- The outcome (count) variable is assumed to be Poisson distributed
- The mean of the Poisson distribution is assumed to be a function of some variables you believe matter

Poisson regression

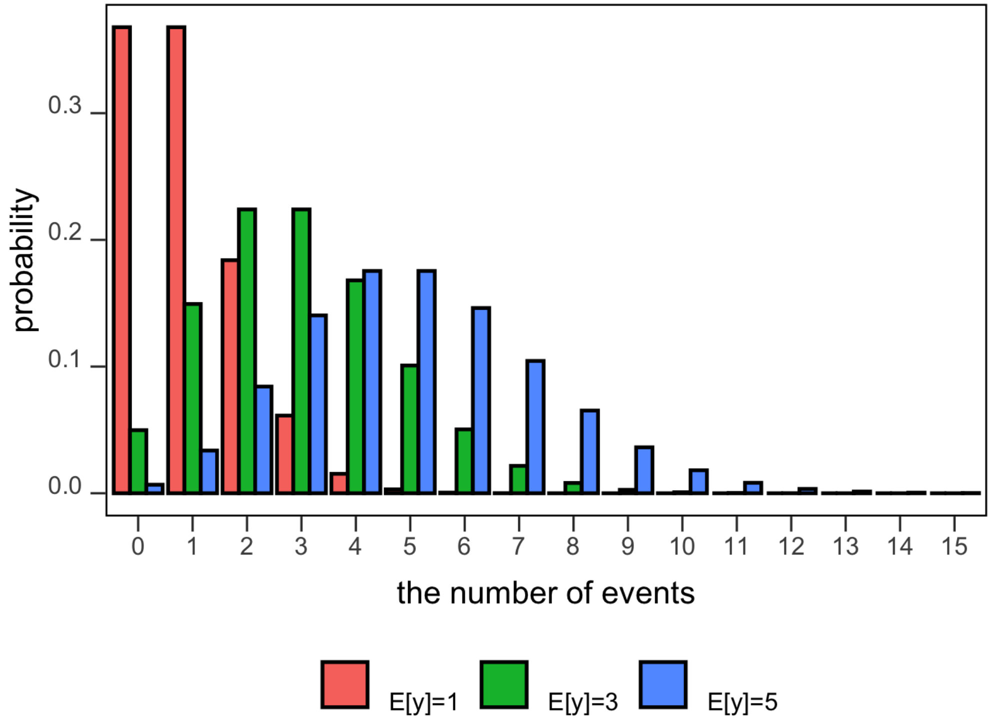
By far the most popular choice to analyze count variables is Poisson regression

- The outcome (count) variable is assumed to be Poisson distributed
- The mean of the Poisson distribution is assumed to be a function of some variables you believe matter

Poisson distribution

Poisson distribution is a discrete probability distribution that describes the probability of the number of events that occur in a fixed interval of time and/or space

$$Prob(y|\lambda) = \frac{\lambda^y e^{-\lambda}}{y!}, \quad \text{where} \quad \lambda = E[y]$$



Poisson regression

We try to learn what and how variables affect the expected value (the expected number of events conditional on independent variables).

Expected number of events conditional on independent variables

$$E[y|\mathbf{x}] = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- This is exactly the same modeling framework we used
  - Linear model: $G(z) = z$
  - Probit model: $G(z) = \Phi(z)$

Expected number of events conditional on independent variables

$$E[y|\mathbf{x}] = G(\beta_0 + \beta_1 x_1 + \cdots + \beta_k x_k)$$

- This is exactly the same modeling framework we used
  - Linear model: $G(z) = z$
  - Probit model: $G(z) = \Phi(z)$

A popular choice of $G()$

$$G(z) = exp(z)$$

This ensures that the expected value conditional on $\mathbf{x}$ is always positive

The number of events for two individuals

- Individual 1: $y = 3$ (own three cars)
- Individual 2: $y = 1$ (own one car)

The number of events for two individuals

- Individual 1: $y = 3$ (own three cars)
- Individual 2: $y = 1$ (own one car)

Expected number of events observed

- Individual 1: $\lambda_1 = exp(z_1)$
- Individual 2: $\lambda_2 = exp(z_2)$

Probability of observing the number of events we observed

$$\text{Individual 1:} Prob(y = 3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \text{ Individual 2 :} \qquad Prob(y = 1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

Probability of observing the number of events we observed

$$\text{Individual 1:} Prob(y = 3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \text{ Individual 2 :} \qquad Prob(y = 1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

Probability of observing a series of events by all individuals

The probability that we observe the collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call llikelihood function.

Probability of observing the number of events we observed

$$\text{Individual 1:} Prob(y=3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \text{ Individual 2 :} \qquad Prob(y=1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

Probability of observing a series of events by all individuals

The probability that we observe the collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call llikelihood function.

Log-likelihood function

$$LL = log(L) = log(\frac{\lambda_1^3 e^{-\lambda_1}}{3!}) + log(\frac{\lambda_2^1 e^{-\lambda_2}}{1!})$$

(Remember $\lambda_i = exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k})$)

Probability of observing the number of events we observed

$$\text{Individual 1:} Prob(y = 3|\mathbf{x_1}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \text{ Individual 2 :} \qquad Prob(y = 1|\mathbf{x_2}) = \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

Probability of observing a series of events by all individuals

The probability that we observe the collection of choices made by them (if their events are independent)

$$L = Prob(y_1 = 3|\mathbf{x_1}) \times Prob(y_2 = 1|\mathbf{x_2}) = \frac{\lambda_1^3 e^{-\lambda_1}}{3!} \times \frac{\lambda_2^1 e^{-\lambda_2}}{1!}$$

which we call llikelihood function.

Log-likelihood function

$$LL = log(L) = log(\frac{\lambda_1^3 e^{-\lambda_1}}{3!}) + log(\frac{\lambda_2^1 e^{-\lambda_2}}{1!})$$

(Remember $\lambda_i = exp(\beta_0 + \beta_1 x_{i,1} + \cdots + \beta_k x_{i,k})$)

$$Max_{\beta_1,...,\beta_k} \; LL$$

Implementation in R with an example

The number of times a man is arrested during 1986:

$$Pr(narr86|\mathbf{x}) = G(z)$$

where

$$z = \beta_0 + \beta_1 pcnv + \beta_2 tottime + \beta_3 qemp86 + \beta_4 inc86 \qquad + \beta_5 black + \beta_6 hispan$$

- $narr86$: # of times arrested in 1986
- $pcnv$: proportion of prior conviction
- $tottime$: time in prison since 18
- $qemp86$: # quarters employed in 1986
- $inc86$: legal income in 1986 (in $\$100$)

R code: importing the data

```r
#--- import the data ---#
data <- read.dta13("CRIME1.dta")

#--- take a look ---#
dplyr::select(data, narr86, pcnv, qemp86, inc86) %>%
  head()
```

```
##   narr86 pcnv qemp86 inc86
## 1      0 0.38      0   0.0
## 2      2 0.44      1   0.8
## 3      1 0.33      0   0.0
## 4      2 0.25      2   8.8
## 5      1 0.00      2   8.1
## 6      0 1.00      4  97.6
```

R code: Poisson model estimation using glm()

```
pois_lf <- glm(

  #--- formula ---#
  narr86 ~ pcnv + tottime + qemp86 + inc86 + black + hispan,

  #--- data ---#
  data = data,

  #--- models ---#
  family = poisson(link = "log")
)
```

`family` option

- `poisson()`: tells R that your dependent variable is Poisson distributed
- `link = "log"`: tells R that you want to use $exp()$ (the inverse of $log()$) as $G()$ in $E(y = 1|\mathbf{x}) = G(z)$

```r
msummary(
  pois_lf,
  # keep these options as they ar
  stars = TRUE,
  gof_omit = "IC|Log|Adj|F|Pseudc
)
```

	Model 1
(Intercept)	-0.666
	(0.064)
pcnv	-0.432
	(0.085)
tottime	-0.001
	(0.006)
qemp86	-0.010
	(0.029)
inc86	-0.008
	(0.001)
black	0.644
	(0.074)
hispan	0.473
	(0.074)
Num.Obs.	2725
RMSE	1.02

+ p < 0.1, p < 0.05, p < 0.01, * p < 0.001

Calculate average marginal effects

Just like the binomial regressions we saw earlier, we can use `margins::margins()` function to get the average marginal effects of covariates.

```
pois_marginal_e <- margins(pois_lf, type = "response")
```