Erasmus
School of
Economics

MOOC Econometrics

Lecture 5.1 on Binary Choice: Motivation

Richard Paap

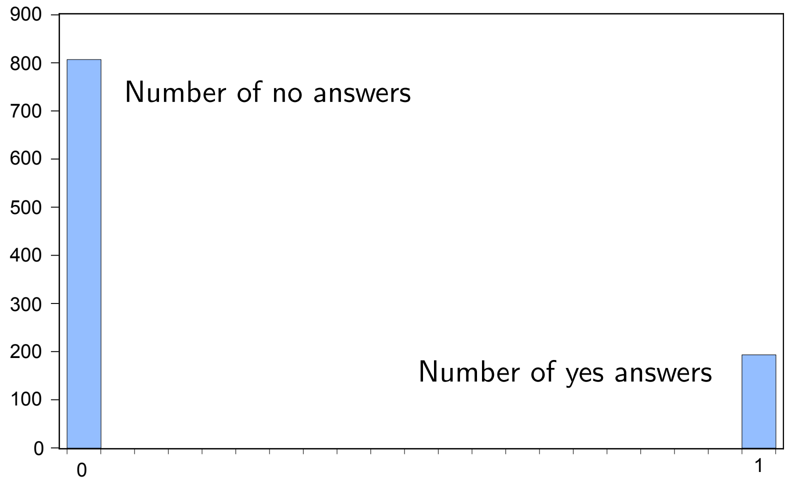
Erasmus University Rotterdam

Examples of binary dependent variables

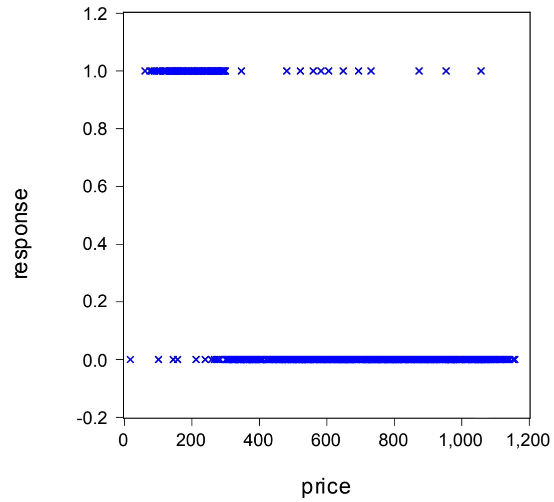
- Answers to "yes/no" questions
- Choice for private or public health care
- Vote decision for Democrat or Republican president (USA)
- Choice for private or public transport
- Choice to renew or cancel a mobile phone contract
- Business cycle indicator (expansion/recession)

and so forth.

Histogram of data



Scatter diagram



Linear regression model

Test

Suppose we model the binary variable using a regression model

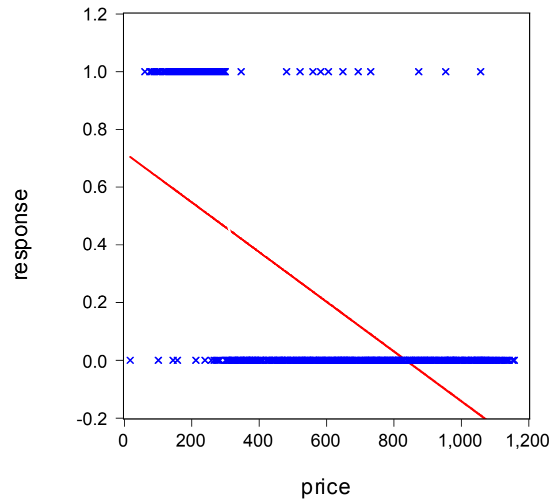
$$\text{response} = \beta_1 + \beta_2 \text{price} + \varepsilon$$

Is it possible to estimate the β parameters using least squares?

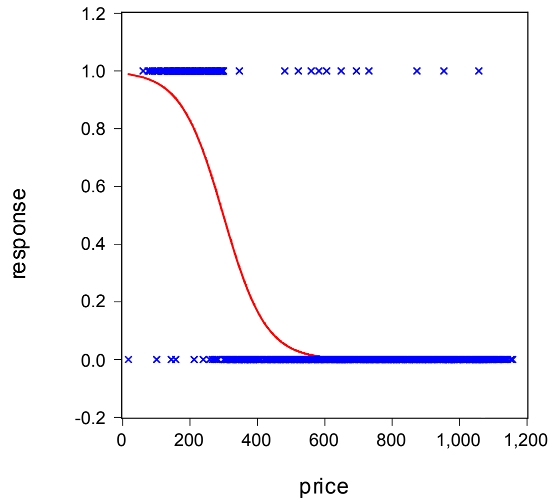
Yes, this is possible. Least squares estimation results in:

$$\text{response} = 0.720 - 0.861\text{price}/1000 + e$$

Scatter diagram with regression line



Model for binary dependent variable



Training Exercise 5.1

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus
School of
Economics

MOOC Econometrics

Lecture 5.2 on Binary Choice: Representation

Richard Paap

Erasmus University Rotterdam

Introduction

Let y_i be a binary variable with value 0 or 1 and assume

$$y_i \sim \mathit{Bernoulli}(\pi),$$

such that

$$\pi = \Pr[y_i = 1] \text{ with } 0 < \pi < 1$$

and hence

$$\Pr[y_i = 0] = 1 - \pi.$$

Individual-specific probabilities:

$$\Pr[y_i = 1] = \pi_i$$

Logit model

Individual-specific probabilities:

$$\Pr[y_i = 1] = \pi_i,$$

where the value of π_i depends on the explanatory variable x_i.

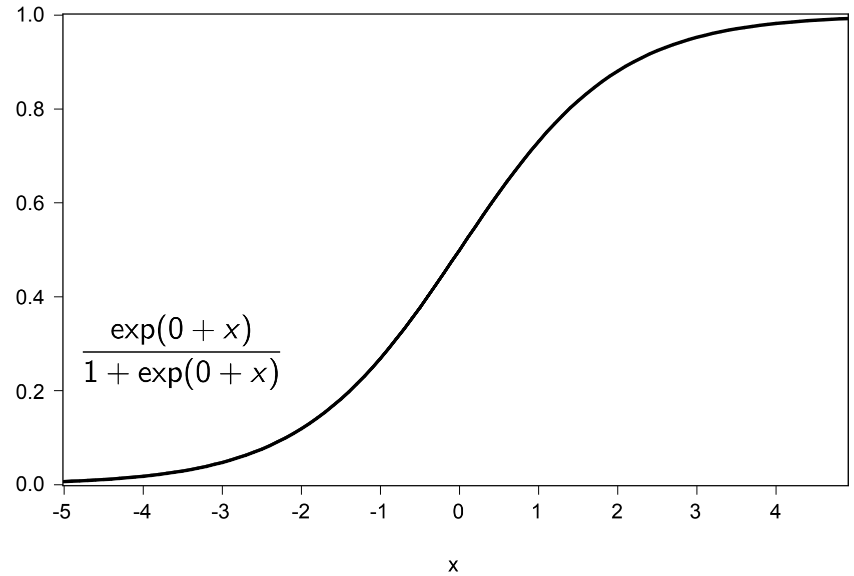
Logit model:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

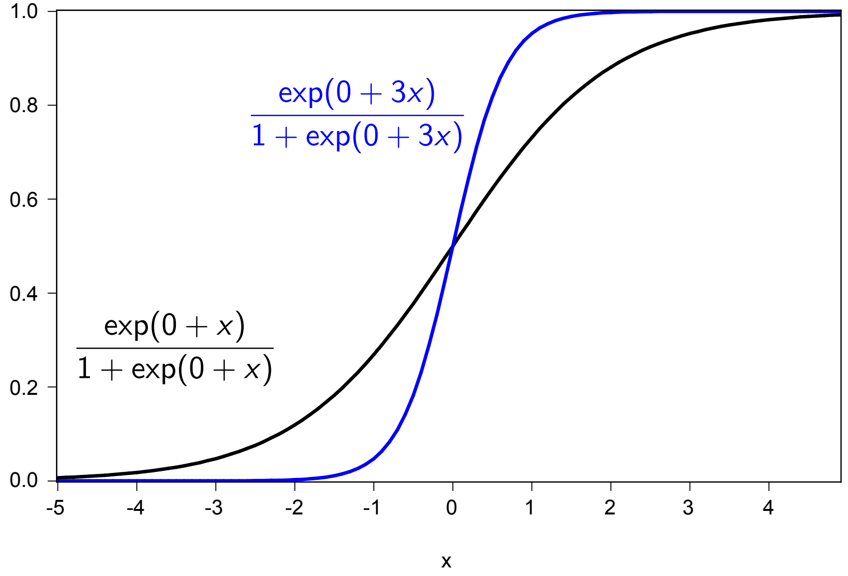
and

$$\Pr[y_i = 0] = 1 - \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)} = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

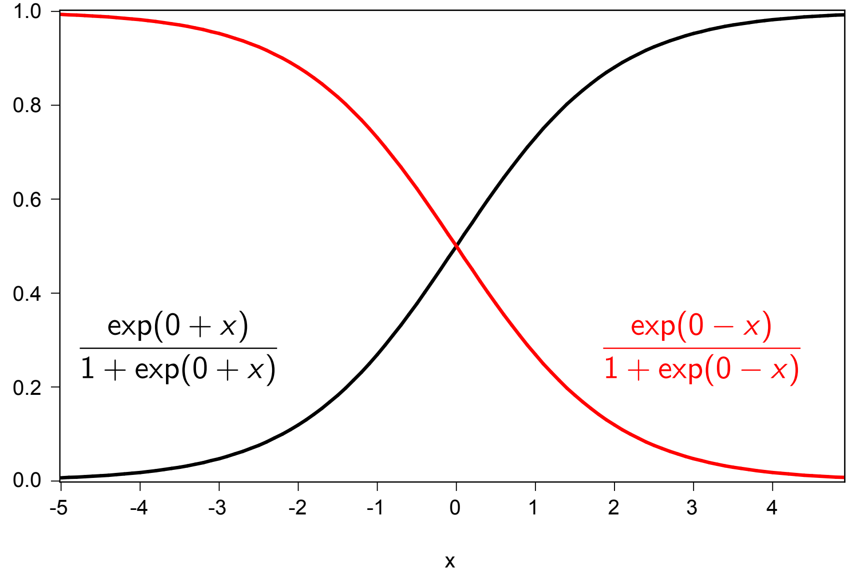
Graphical interpretation



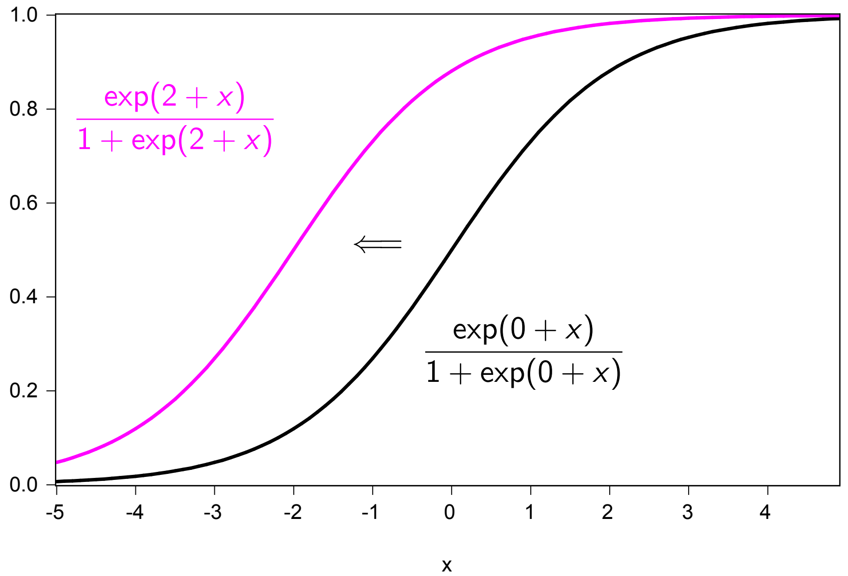
Graphical interpretation



Graphical interpretation



Graphical interpretation



Test question

Test

What happens to the location and shape of the logit function

$$\frac{\exp(\beta_1 + x)}{1 + \exp(\beta_1 + x)}$$

if you change the β_1 parameter from $\beta_1 = 0$ to $\beta_1 = -2$?

The logit function only shifts 2 units to the right.

Odds ratio

Logit model:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \beta_2 x_i)}{1 + \exp(\beta_1 + \beta_2 x_i)}$$
$$\Pr[y_i = 0] = \frac{1}{1 + \exp(\beta_1 + \beta_2 x_i)}$$

Odds ratio:

$$\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]} = \exp(\beta_1 + \beta_2 x_i)$$

Log odds ratio:

$$\log\left(\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]}\right) = \beta_1 + \beta_2 x_i$$

Marginal effect

Marginal effect:

$$\frac{d\Pr[y_i = 1]}{d\ x_i} = \Pr[y_i = 1]\Pr[y_i = 0]\beta_2$$

Change in probability that $y_i = 1$ due to change in x_i.

Average marginal effect:

$$\frac{1}{n}\sum_{i=1}^{n}\frac{d\Pr[y_i = 1]}{d\ x_i} = \left(\frac{1}{n}\sum_{i=1}^{n}\Pr[y_i = 1]\Pr[y_i = 0]\right)\beta_2$$

More explanatory variables

Logit specification with $x_{2i}, \ldots, x_{ki}$ as explanatory variables:

$$\Pr[y_i = 1] = \frac{\exp(\beta_1 + \sum_{j=2}^{k}\beta_j x_{ji})}{1 + \exp(\beta_1 + \sum_{j=2}^{k}\beta_j x_{ji})}$$

Log odds ratio:

$$\log\left(\frac{\Pr[y_i = 1]}{\Pr[y_i = 0]}\right) = \beta_1 + \sum_{j=2}^{k}\beta_j x_{ji}$$

Marginal effect:

$$\frac{\partial\Pr[y_i = 1]}{\partial x_{ji}} = \Pr[y_i = 1]\Pr[y_i = 0]\beta_j \text{ for } j = 2, \ldots, k.$$

Change in probability that $y_i = 1$ due to change in x_{ji}.

Training Exercise 5.2

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

Lecture 5.3 on Binary Choice: Estimation

Richard Paap

Erasmus University Rotterdam

Introduction

Logit model specification in vector notation:

$$\Pr[y_i = 1] = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)},$$

where $x_i = (1, x_{2i}, \ldots, x_{ki})'$ and $\beta = (\beta_1, \ldots, \beta_k)'$

It is not possible to write this model in regression notation

$$y_i = x_i'\beta + \varepsilon_i$$

We use maximum likelihood for parameter estimation.

Construction of likelihood function

Likelihood contribution for

- observation $y_i = 1$:

$$\Pr[y_i = 1] = \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}$$

- observation $y_i = 0$:

$$\Pr[y_i = 0] = \frac{1}{1 + \exp(x_i'\beta)}$$

Likelihood contribution for observation i:

$$\left(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}\right)^{y_i} \left(\frac{1}{1 + \exp(x_i'\beta)}\right)^{1-y_i}$$

(Log)-likelihood function

Likelihood function of n independent observations:

$$L(\beta) = \prod_{i=1}^{n} \left(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}\right)^{y_i} \left(\frac{1}{1 + \exp(x_i'\beta)}\right)^{1-y_i}$$

Log-likelihood function:

$$\begin{aligned}\log(L(\beta)) &= \sum_{i=1}^{n} y_i \log\left(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)}\right) + (1 - y_i) \log\left(\frac{1}{1 + \exp(x_i'\beta)}\right) \\ &= \sum_{i=1}^{n} y_i x_i'\beta - \log(1 + \exp(x_i'\beta)),\end{aligned}$$

where we use that $\log(ab) = \log(a) + \log(b)$, $\log(a^b) = b\log(a)$ and $\log(a/b) = \log(a) - \log(b)$.

Test question

Test

The maximum likelihood estimator [MLE] is the value of β that maximizes the log-likelihood function. Is the MLE also the value that maximizes the likelihood function?

As the log function is a monotonically increasing function in β, the answer is yes.

Maximum likelihood estimation

The MLE b is obtained by maximizing $\log(L(\beta))$ with respect to β. First-order conditions:

$$\frac{\partial \log(L(\beta))}{\partial \beta} = \frac{\partial \sum_{i=1}^{n} y_i x_i'\beta - \log(1+\exp(x_i'\beta))}{\partial \beta} = 0$$

$$= \sum_{i=1}^{n} y_i x_i' - \frac{\exp(x_i'\beta)x_i'}{1+\exp(x_i'\beta)} = 0$$

Use numerical methods to solve for β.

The first-order conditions imply that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\exp(x_i'b)}{1+\exp(x_i'b)} = \frac{1}{n}\sum_{i=1}^{n} y_i$$

Test question

Test

Suppose that all observations on y_i are 0, that is, $y_i = 0$ for $i = 1, \ldots, n$. What is the value of the maximum likelihood estimator in this case?

When all observations are 0, the first-order conditions imply that

$$\frac{1}{n}\sum_{i=1}^{n} \frac{\exp(x_i'b)}{1+\exp(x_i'b)} = \frac{1}{n}\sum_{i=1}^{n} y_i = 0.$$

As the logit function is always larger than 0, there is no value of b for which the first-order conditions holds. The MLE does not exist.

Properties of maximum likelihood estimator

It can be shown that under regularity conditions the maximum likelihood estimator [MLE] is

1. Consistent
2. Efficient for large n
3. Asymptotically normally distributed, and hence

$$b \approx N(\beta, V)$$

The (co)variance matrix V can be estimated by

$$\hat{V} = \left(\sum_{i=1}^{n} \left(\frac{\exp(x_i'b)}{1+\exp(x_i'b)}\right)\left(\frac{1}{1+\exp(x_i'b)}\right) x_i x_i'\right)^{-1}$$

Testing for single parameter restriction

We want to compare

- logit model without parameter restrictions
- logit model with a single $\beta_j = 0$

Hypothesis:

$$H_0\colon\ \beta_j = 0 \text{ versus } H_1\colon\ \beta_j \neq 0$$

You can use the t-test like in a linear regression: Test statistic:

$$z_j = \frac{b_j - 0}{\mathrm{SE}(b_j)} \approx N(0,1),$$

where $\mathrm{SE}(b_j)$ is the standard error of b_j.

Testing for a set of parameter restrictions

We want to compare

- logit model without parameter restrictions and estimates b_1
- logit model with m parameter restrictions and estimates b_0

The null hypothesis is that the m parameter restrictions are correct.

To compute the test statistic we need

- $L(b_1)$: maximum likelihood value in full model
- $L(b_0)$: maximum likelihood value in restricted model

Test statistic:

$$LR = -2(\log(L(b_0)) - \log(L(b_1))) \approx \chi^2(m),$$

where m is the number of restrictions.

Training Exercise 5.3

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

Lecture 5.4 on Binary Choice: Evaluation

Richard Paap

Erasmus University Rotterdam

Residuals

Logit residuals:

$$\begin{aligned} y_i - \mathrm{E}[y_i] &= y_i - (0 \times \Pr[y_i = 0] + 1 \times \Pr[y_i = 1]) \\ &= y_i - \Pr[y_i = 1] \\ &= y_i - \frac{\exp(x_i'b)}{1 + \exp(x_i'b)} \end{aligned}$$

Interesting cases:

- Lower bound: $y_i - \mathrm{E}[y_i] \approx -1$
- Upper bound: $y_i - \mathrm{E}[y_i] \approx 1$
- Perfect fit $y_i - \mathrm{E}[y_i] \approx 0$

Test question

Test

Suppose that we have perfect fit for all n observations, that is,

$$y_i - \frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \approx 0$$

for all i. What is the numerical value of the likelihood function

$$\prod_{i=1}^{n} \left(\frac{\exp(x_i'\beta)}{1 + \exp(x_i'\beta)} \right)^{y_i} \left(\frac{1}{1 + \exp(x_i'\beta)} \right)^{1-y_i} \quad ?$$

For all observations equal to 1 (or 0) the likelihood contribution is very close to 1. Hence, the likelihood function equals about 1.

Measures of fit

Define

- $L(b)$: the maximum value of the likelihood function of the model under consideration
- $L(b_1)$: maximum value of the likelihood function in case the model only contains an intercept.

Perfect fit corresponds to $L(b) \approx 1$ or $\log(L(b)) \approx 0$.

Two popular pseudo R^2 measures are:

- McFadden R^2:

$$R^2 = 1 - \frac{\log(L(b))}{\log(L(b_1))}$$

- Nagelkerke R^2:

$$R^2 = 1 - \frac{1 - \left(\frac{L(b_1)}{L(b)}\right)^{2/n}}{1 - L(b_1)^{2/n}}$$

Prediction probability

If the value of x_{n+1} is available, one can predict the value of y_{n+1} using

$$\begin{aligned} \mathrm{E}[y_{n+1}] &= 0 \times \Pr[y_{n+1} = 0] + 1 \times \Pr[y_{n+1} = 1] \\ &= \Pr[y_{n+1} = 1] \\ &= \frac{\exp(x'_{n+1}\beta)}{1 + \exp(x'_{n+1}\beta)} \end{aligned}$$

To estimate this probability we replace β by its estimate b and obtain $\widehat{\Pr}[y_{n+1} = 1]$.

0/1 Prediction

The prediction is a probability and never exactly equal to 0 or 1.

Transform the prediction probability into 0/1 forecast $\hat{y}_{n+1}$ by the rule:

$$\begin{aligned} \hat{y}_{n+1} &= 1 \text{ if } \widehat{\Pr}[y_{n+1} = 1] > c \\ \hat{y}_{n+1} &= 0 \text{ if } \widehat{\Pr}[y_{n+1} = 1] \leq c. \end{aligned}$$

Many statistical packages use $c = 0.5$. However, one may also consider

$$c = \frac{1}{n}\sum_{i=1}^{n} y_i,$$

that is, the fraction of observations in the sample equal to one.

Test question

Test

Does a higher value of the cut-off value c generate more, the same or less predictions which are equal to 1?

A higher value of c means that less (or the same number of) prediction probabilities are above c and hence you forecast less (or the same number of) ones.

Evaluation of predication accuracy

Suppose one has m out-of-sample predictions for y_i denoted by $\hat{y}_i$.

Count the number of correct and incorrect predictions:

$$\begin{aligned} m_{11} &= \textstyle\sum_{i=1}^{m} y_{n+i}\hat{y}_{n+i} && \text{data=1 \& prediction=1} \\ m_{00} &= \textstyle\sum_{i=1}^{m} (1 - y_{n+i})(1 - \hat{y}_{n+i}) && \text{data=0 \& prediction=0} \\ m_{10} &= \textstyle\sum_{i=1}^{m} y_{n+i}(1 - \hat{y}_{n+i}) && \text{data=1 \& prediction=0} \\ m_{01} &= \textstyle\sum_{i=1}^{m} (1 - y_{n+i})\hat{y}_{n+i} && \text{data=0 \& prediction=1} \end{aligned}$$

Prediction-realization table

Classify predictions in right and wrong:

observed	predicted $\hat{y}=0$	predicted $\hat{y}=1$	sum
$y=0$	m_{00}/m	m_{01}/m	$(m_{00}+m_{01})/m$
$y=1$	m_{10}/m	m_{11}/m	$(m_{10}+m_{11})/m$
sum	$(m_{00}+m_{10})/m$	$(m_{01}+m_{11})/m$	1

Prediction-realization table

Classify predictions in right and wrong:

observed	predicted $\hat{y}=0$	predicted $\hat{y}=1$	sum
$y=0$	m_{00}/m	m_{01}/m	$(m_{00}+m_{01})/m$
$y=1$	m_{10}/m	m_{11}/m	$(m_{10}+m_{11})/m$
sum	$(m_{00}+m_{10})/m$	$(m_{01}+m_{11})/m$	1

The fraction $m_{00}/m + m_{11}/m$ is called the hit rate.

Prediction-realization table

Classify predictions in right and wrong:

observed	predicted $\hat{y}=0$	predicted $\hat{y}=1$	sum
$y=0$	m_{00}/m	m_{01}/m	$(m_{00}+m_{01})/m$
$y=1$	m_{10}/m	m_{11}/m	$(m_{10}+m_{11})/m$
sum	$(m_{00}+m_{10})/m$	$(m_{01}+m_{11})/m$	1

$m_{01}/m + m_{10}/m$ denotes the fraction of incorrect forecasts.

Training Exercise 5.4

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).

Erasmus School of Economics

MOOC Econometrics

Lecture 5.5 on Binary Choice: Application

Richard Paap

Erasmus University Rotterdam

Response to direct mailing

Sample:

- 925 observations

Dependent variable:

- Resp: Response to direct mailing with $1 =$ yes and $0 =$ no

Potential explanatory variables:

- Male: $1 =$ Male and $0 =$ Female
- Age: Age of the customer in years
- Active: $1 =$ Active customer and $0 =$ Inactive customer

Data characteristics

Average values of the explanatory variables

Variable	resp = 0	resp = 1	all observations
Gender	0.624	0.823	0.725
Active	0.114	0.260	0.188
Age	50.813	50.553	50.681

Data characteristics

Average values of the explanatory variables

Variable	resp = 0	resp = 1	all observations
Gender	0.624	0.823	0.725
Active	0.114	0.260	0.188
Age	50.813	50.553	50.681

Model specification

Proposed logit model specification:

$$\Pr[\text{resp}_i = 1] = \frac{\exp(\beta_0 + \beta_1 \text{male}_i + \beta_2 \text{active}_i + \beta_3 \text{age}_i + \beta_4 (\text{age}_i/10)^2)}{1 + \exp(\beta_0 + \beta_1 \text{male}_i + \beta_2 \text{active}_i + \beta_3 \text{age}_i + \beta_4 (\text{age}_i/10)^2)}$$

for $i = 1, \ldots, 925$.

Estimation results logit model

Variable	Coefficient	Std. Error	t-value	*p*-value.
Intercept	-2.488	0.890	-2.796	0.005
Male	0.954	0.158	6.029	0.000
Active	0.914	0.185	4.945	0.000
Age	0.070	0.036	1.964	0.050
$(\text{Age}/10)^2$	-0.069	0.034	-2.015	0.044
McFadden R^2	0.061			
Nagelkerke R^2	0.892			
Log-likelihood	-601.862			

Odds ratio

$$\frac{\Pr[\text{resp}_i = 1]}{\Pr[\text{resp}_i = 0]}$$
$$= \exp(\beta_0 + \beta_1 \text{male}_i + \beta_2 \text{active}_i + \beta_3 \text{age}_i + \beta_4 (\text{age}_i/10)^2)$$
$$\approx \exp(-2.49 + 0.95\text{male}_i + 0.91\text{active}_i + 0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$
$$= 0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$

Odds ratio

$$\frac{\Pr[\text{resp}_i = 1]}{\Pr[\text{resp}_i = 0]}$$
$$= \exp(\beta_0 + \beta_1 \text{male}_i + \beta_2 \text{active}_i + \beta_3 \text{age}_i + \beta_4 (\text{age}_i/10)^2)$$
$$\approx \exp(-2.49 + 0.95\text{male}_i + 0.91\text{active}_i + 0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$
$$= 0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$

Odds ratio

$$\frac{\Pr[\text{resp}_i = 1]}{\Pr[\text{resp}_i = 0]}$$
$$= \exp(\beta_0 + \beta_1 \text{male}_i + \beta_2 \text{active}_i + \beta_3 \text{age}_i + \beta_4 (\text{age}_i/10)^2)$$
$$\approx \exp(-2.49 + 0.95\text{male}_i + 0.91\text{active}_i + 0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$
$$= 0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)$$

Test question

Test

For which value of age do we have the highest value of the odds ratio

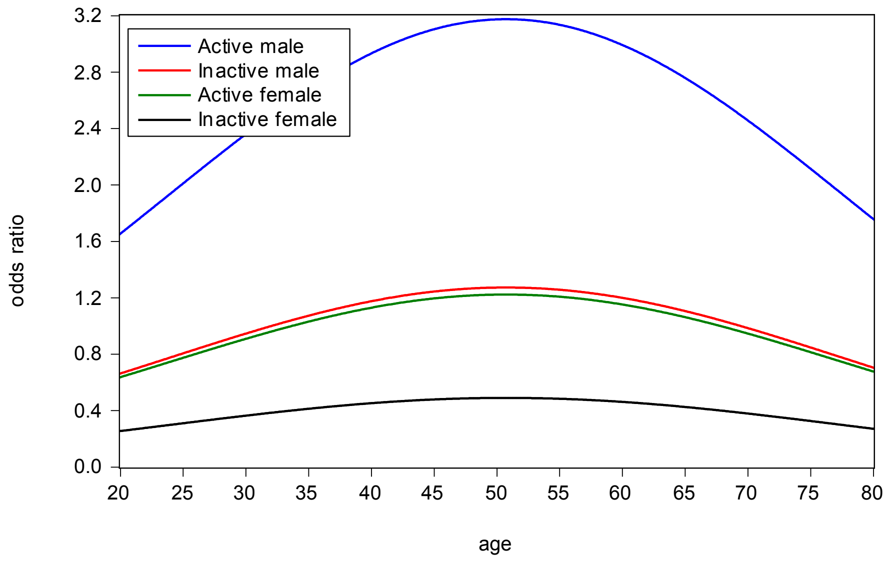
$$0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)?$$

The first-order condition is

$$[0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)] \times (0.07 - 2 \times 0.07(\text{age}_i/100)) = 0$$

The solution to this first-order condition is 50 years.

Odds ratio versus age



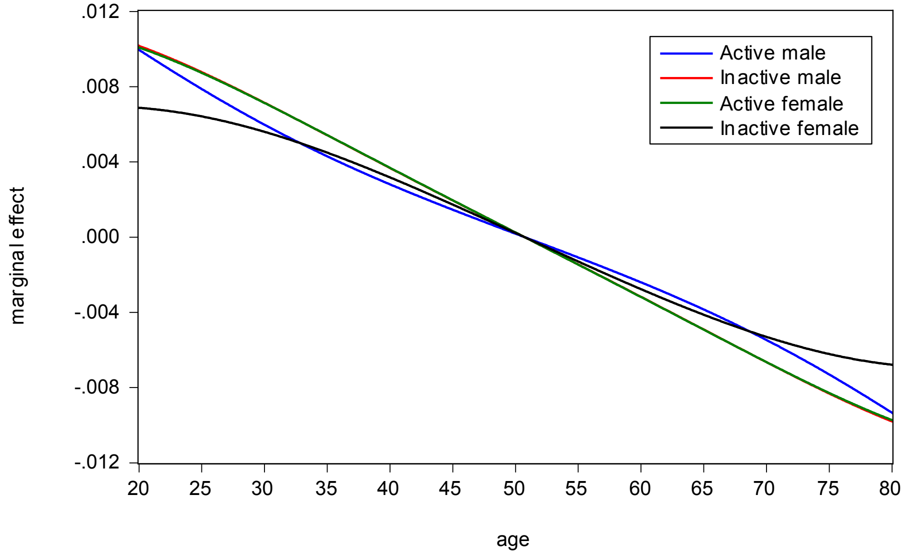
Marginal effect of age

$$\frac{\partial \Pr[\text{resp}_i = 1]}{\partial age_i} = \Pr[\text{resp}_i = 1]\Pr[\text{resp}_i = 0](\beta_3 + 2\beta_4(\text{age}_i/10)^2)$$
$$\approx \Pr[\text{resp}_i = 1]\Pr[\text{resp}_i = 0](0.07 - 2 \times 0.07\text{age}_i/100)$$

Marginal effect depends on

- age_i
- $\Pr[\text{resp}_i = 1]$ and $\Pr[\text{resp}_i = 0]$ and hence also on male and active dummy.

Marginal effect of age



In-sample prediction-realisation table

Cut-off value: 0.5

	predicted		
observed	$\hat{y} = 0$	$\hat{y} = 1$	sum
$y = 0$	0.212	0.280	0.492
$y = 1$	0.104	0.404	0.508
sum	0.316	0.684	1

Hit rate: 0.212+0.404=0.616.

Training Exercise 5.5

- Train yourself by making the training exercise (see the website).
- After making this exercise, check your answers by studying the webcast solution (also available on the website).