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	**0.105**			
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

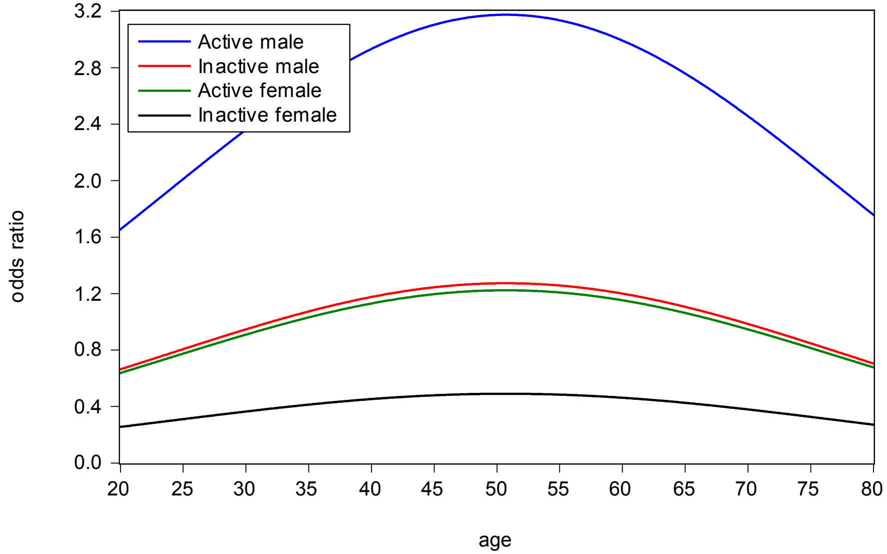
$$0.08 \times 2.57^{\text{male}_i} \times 2.50^{\text{active}_i} \times \exp(0.07\text{age}_i - 0.07(\text{age}_i/10)^2)?$$

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



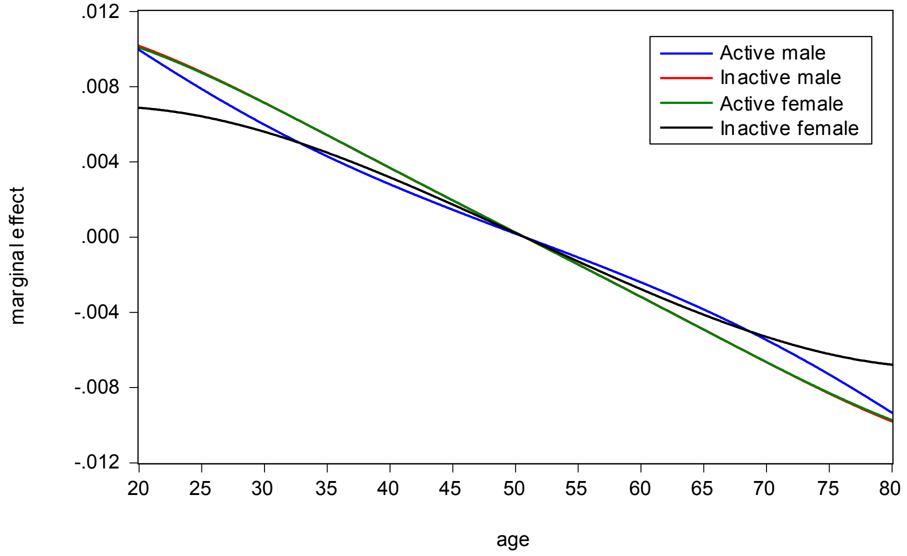
Marginal effect of age

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

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

observed	predicted $\hat{y}=0$	predicted $\hat{y}=1$	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).