Econometrics Test Exercise 5

Vishesh Gupta

Questions

Consider again the application in lecture 5.5, where we have analyzed response to a direct mailing using the following logit specification

$$Pr[resp_i = 1] = \frac{exp\left(\beta_0 + \beta_1 * male_i + \beta_2 * active_i + \beta_3 * age_i + \beta_4 * (age_i/10)^2\right)}{1 + exp\left(\beta_0 + \beta_1 * male_i + \beta_2 * active_i + \beta_3 * age_i + \beta_4 * (age_i/10)^2\right)}$$

for $i = 1, \ldots, 925$ The maximum likelihood estimates of the parameters are given by

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
(Age/10)^2	-0.069	0.034	-2.015	0.044

(a) Show that

$$\frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} + \frac{\partial\, Pr[resp_i = 0]}{\partial\, age_i} = 0\,.$$

(b) Assume that you recode the dependent variable as follows: $resp_i^{new} = -resp_i + 1$ Hence, positive response is now defined to be equal to zero and negative response to be equal to 1. Use the odds ratio to show that this transformation implies that the sign of all parameters change.

(c) Consider again the odds ratio positive response versus negative response:

$$\frac{Pr[resp_i = 1]}{Pr[resp_i = 0]} = exp\left(\beta_0 + \beta_1 * male_i + \beta_2 * active_i + \beta_3 * age_i + \beta_4 * (age_i/10)^2\right)\,.$$

During lecture 5.5 you have seen that this odds ratio obtains its maximum value for age equal to 50 years for males as well as females. Suppose now that you want to extend the logit model and allow that this age value is possibly different for males than for females. Discuss how you can extend the logit specification.

Exercise initialization

Loading data…

```
##
## Call:
## glm(formula = response ~ male + activity + age + age2, family = binomial,
##     data = dat)
##
## Deviance Residuals:
##     Min       1Q   Median       3Q      Max
## -1.6926  -1.2156   0.7389   1.0982   1.8473
##
## Coefficients:
##             Estimate Std. Error z value Pr(>|z|)
## (Intercept) -2.48836    0.88999  -2.796  0.00517 **
## male         0.95369    0.15818   6.029 1.65e-09 ***
## activity     0.91375    0.18478   4.945 7.61e-07 ***
## age          0.06995    0.03561   1.964  0.04948 *
## age2        -0.06869    0.03410  -2.015  0.04394 *
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##     Null deviance: 1282.1  on 924  degrees of freedom
## Residual deviance: 1203.7  on 920  degrees of freedom
## AIC: 1213.7
##
## Number of Fisher Scoring iterations: 4
```

```
library(pscl)
LogLikelihood   <- pR2(model)['llh']
Likelihood      <- exp(pR2(model)['llh'])
McFaddenR2      <- pR2(model)['McFadden']
NagelkerkeR2    <- pR2(model)['r2CU']

library(fmsb)
NagelkerkeR2 <- NagelkerkeR2(model)$R2
```

(a) Partial differentiations sum

(a) Show that

$$\frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} + \frac{\partial\, Pr[resp_i = 0]}{\partial\, age_i} = 0\,.$$

This can be shown mathematically:

$$\begin{aligned}
\frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} + \frac{\partial\, Pr[resp_i = 0]}{\partial\, age_i} &= \frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} + \frac{\partial\,(1 - Pr[resp_i = 1])}{\partial\, age_i} && Pr[resp_i = 0] = 1 - Pr[resp_i = 1] \\
&= \frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} + \frac{-\partial\,(Pr[resp_i = 1])}{\partial\, age_i} && \partial\, 1 = 0 \\
&= \frac{\partial\, Pr[resp_i = 1]}{\partial\, age_i} - \frac{\partial\,(Pr[resp_i = 1])}{\partial\, age_i} && terms\ cancelled\ out \\
&= 0
\end{aligned}$$

(b) Odds ratio usage

(b) Assume that you recode the dependent variable as follows: $resp_i^{new} = -resp_i + 1$. Hence, positive response is now defined to be equal to zero and negative response to be equal to 1. Use the odds ratio to show that this transformation implies that the sign of all parameters change.

The condition $resp_i^{new} = -resp_i + 1$means that the odds ratio of each new response is inversed, or:

$$\frac{Pr[resp_i{}^{new}=1]}{Pr[resp_i{}^{new}=0]}=\frac{1}{Pr[resp_i=1]/Pr[resp_i=0]} \qquad =\frac{Pr[resp_i=0]}{Pr[resp_i=1]}$$
$$\Rightarrow \frac{Pr[resp_i{}^{new}=0]}{Pr[resp_i{}^{new}=1]}=\frac{Pr[resp_i=1]}{Pr[resp_i=0]}$$

And the odds ratio formula is given as:

$$\frac{Pr[resp_i=1]}{Pr[resp_i=0]}=exp\left(\beta_0+\sum_{j=2}^{k}\beta_j*x_{ji}\right)$$
$$=exp\left(\beta_0+\beta_1*male_i+\beta_2*active_i+\beta_3*age_i+\beta_4*(age_i/10)^2\right)$$
$$\Rightarrow \frac{Pr[resp_i=0]}{Pr[resp_i=1]}=\frac{1}{exp\left(\beta_0+\sum_{j=2}^{k}\beta_j*x_{ji}\right)}$$
$$=\frac{1}{exp\left(\beta_0+\beta_1*male_i+\beta_2*active_i+\beta_3*age_i+\beta_4*(age_i/10)^2\right)}$$

Hence:

$$\frac{Pr[resp_i{}^{new}=1]}{Pr[resp_i{}^{new}=0]}=\frac{Pr[resp_i=0]}{Pr[resp_i=1]}$$
$$=\frac{1}{exp\left(\beta_0+\sum_{j=2}^{k}\beta_j*x_{ji}\right)}$$
$$=exp\left(-\beta_0-\sum_{j=2}^{k}\beta_j*x_{ji})\right) \qquad \frac{1}{exp\,(value)}=exp\,(-value)$$
$$=exp\left(-\beta_0-\beta_1*male_i-\beta_2*active_i-\beta_3*age_i-\beta_4*(age_i/10)^2\right)$$

Therefore, the transformation **indeed implies** that the sign of all parameters change.

(c) Extending the logit model

(c) Consider again the odds ratio positive response versus negative response:

$$\frac{Pr[resp_i=1]}{Pr[resp_i=0]}=exp\left(\beta_0+\beta_1*male_i+\beta_2*active_i+\beta_3*age_i+\beta_4*(age_i/10)^2\right)\ .$$

During lecture 5.5 you have seen that this odds ratio obtains its maximum value for age equal to 50 years for males as well as females. Suppose now that you want to extend the logit model and allow that this age value is possibly different for males than for females. Discuss how you can extend the logit specification.

One way is to split the sample in two ($male-only$ and $female-only$) groups, and estimate the logit model seperately for each group. In this case, the remaining $activity$ variable parameter β_2 could also differ for each one of the groups and should be estimated seperately.

Another way is to modify the logit model in order to use combinations of $male$ and age variables (and related parameters) and leave the sample united as is.

For example, we could add the variables $male_i*age_i$ and $male_i*(age_i/10)^2$ to the logit specification, thus extending it by 2 more variables with parameters β_5 and β_6.
Then, the model odds ratio would be:

$$\frac{Pr[resp_i=1]}{Pr[resp_i=0]}=exp\Big(\beta_0+\beta_1*male_i+\beta_2*active_i+\beta_3*age_i+\beta_4*(age_i/10)^2+$$
$$\beta_5*male_i*age_i+\beta_6*male_i*(age_i/10)^2\Big)\,.$$

Or, we could replace the age related variables and parameters

$$\beta_3*age_i+\beta_4*(age_i/10)^2$$

with their $male/female$ related combinations:

$$\beta_3 * male_i * age_i + \beta_4 * male_i * (age_i/10)^2 +$$
$$\beta_5 * (1 - male_i) * age_i + \beta_6 * (1 - male_i) * (age_i/10)^2$$

. Then, the model odds ratio would be:

$$\frac{Pr[resp_i = 1]}{Pr[resp_i = 0]} = exp\Big(\beta_0 + \beta_1 * male_i + \beta_2 * active_i +$$
$$\beta_3 * male_i * age_i + \beta_4 * male_i * (age_i/10)^2 +$$
$$\beta_5 * (1 - male_i) * age_i + \beta_6 * (1 - male_i) * (age_i/10)^2\Big).$$

These second way variations would allow for similar flexibility to study on the age effect throught the different variables and parameters they incorporate.

– End of document –