8.2) Logit and Probit

Vitor Kamada

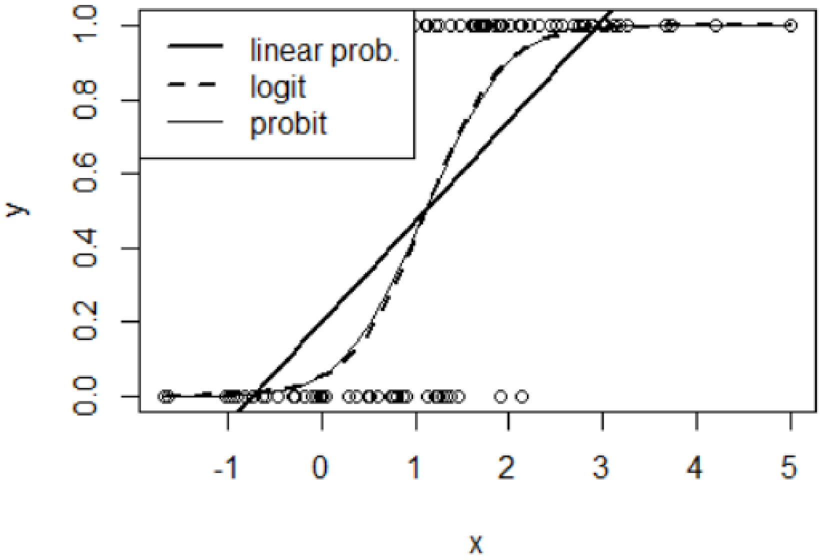
December 2018

Reference

Tables, Graphics, and Figures from:

Wooldridge (2010). **Econometric Analysis of Cross Section and Panel Data.** Ch 15.

Logit and Probit



Specifying Logit and Probit Models

$$P(y = 1|x) = G(\beta_0 + x\beta)$$

$$0 < G(z) < 1$$

$$G(z) = \frac{exp^z}{1+exp^z} = \Lambda(z)$$

$$G(z) = \Phi(z) = \int_{-\infty}^{z} \phi(v)dv$$

$$\phi(v) = (2\pi)^{-\frac{1}{2}} exp(-\frac{z^2}{2})$$

Latent Variable Model

$$y^* = \beta_0 + x\beta + e$$

$$y = 1[y^* > 0]$$

$$e \perp x$$

$$e \sim N(0,1) \text{ or } Logistic(0,1)$$

$$\therefore 1 - G(-z) = G(z)$$

Maximum Likelihood Estimation (MLE)

$$f(y|x_i;\beta) = [G(x_i\beta)]^y[1 - G(x_i\beta)]^{1-y}$$

$$\ell_i(\beta) = y_i log[G(x_i\beta)] + (1 - y_i) log[1 - G(x_i\beta)]$$

$$\mathcal{L}(\beta) = \sum_{i=1}^{n} \ell_i(\beta)$$

Logistic Regression

$$\ell_i(\beta) = y_i log[G(x_i\beta)] + (1 - y_i)log[1 - G(x_i\beta)]$$

$$\frac{\partial log \mathcal{L}(\beta)}{\partial \beta} = 0$$

$$\sum_{i=1}^{n} [\frac{y_i - G(x_i\beta)}{G(x_i\beta)(1 - G(x_i\beta))} g(x_i\beta)]x_i = 0$$

$$\sum_{i=1}^{n} [y_i - \frac{e^{x_i\beta}}{1 + e^{x_i\beta}}]x_i = 0$$

Response Probability for y

$$P(y=1|x) = P(y^* > 0|x) = P[e > -(\beta_0 + x\beta)|x]$$

$$= 1 - G[-(\beta_0 + x\beta)] = G(\beta_0 + x\beta)$$

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j$$

$$g(z) = \frac{dG}{dz}(z) > 0$$

$G(\cdot)$ is a strictly increasing cdf

$$\phi(0) = \frac{1}{\sqrt{2\pi}} \approx .4, \qquad \lambda(0) = \frac{exp(0)}{[1+exp(0)]^2} \approx .25$$

Partial Effect of Continuous Variables on the Response Probability

$$\frac{\partial p(x)}{\partial x_j} = g(\beta_0 + x\beta)\beta_j$$

$$\frac{\partial p(x)}{\partial x_j} \Big/ \frac{\partial p(x)}{\partial x_h} = \frac{\beta_j}{\beta_h}$$

Binary: $\frac{\Delta p(x)}{\Delta x_1} = G(\beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k)$

$$-G(\beta_0 + \beta_2 x_2 + \ldots + \beta_k x_k)$$

Mroz (1987)

library(foreign);library(car); library(lmtest)

data(mroz, package='wooldridge')

Statistic	N	Mean	St. Dev.	Min	Max
inlf	753	0.568	0.496	0	1
kidslt6	753	0.238	0.524	0	3
kidsge6	753	1.353	1.320	0	8
age	753	42.538	8.073	30	60
educ	753	12.287	2.280	5	17
wage	428	4.178	3.310	0.128	25.000
nwifeinc	753	20.129	11.635	−0.029	96.000
lwage	428	1.190	0.723	−2.054	3.219

Linear Probablity Model

linprob <- lm(inlf~nwifeinc+educ+exper+I(exper^2)+ age+kidslt6+kidsge6, data=mroz)

coeftest(linprob,vcov=hccm)

```
              Estimate   Std. Error t value  Pr(>|t|)    
(Intercept)  0.58551922  0.15358032  3.8125  0.000149 ***
nwifeinc    -0.00340517  0.00155826 -2.1852  0.029182 *  
educ         0.03799530  0.00733982  5.1766 2.909e-07 ***
exper        0.03949239  0.00598359  6.6001 7.800e-11 ***
I(exper^2)  -0.00059631  0.00019895 -2.9973  0.002814 ** 
age         -0.01609081  0.00241459 -6.6640 5.183e-11 ***
kidslt6     -0.26181047  0.03215160 -8.1430 1.621e-15 ***
kidsge6      0.01301223  0.01366031  0.9526  0.341123    
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’
```

Logit Model

logitres<-glm(inlf~nwifeinc+educ+exper+I(exper^2)
+age+kidslt6+kidsge6, family=binomial(link=logit),data=mroz)

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.425452   0.860365   0.495  0.62095
nwifeinc    -0.021345   0.008421  -2.535  0.01126 *
educ         0.221170   0.043439   5.091 3.55e-07 ***
exper        0.205870   0.032057   6.422 1.34e-10 ***
I(exper^2)  -0.003154   0.001016  -3.104  0.00191 **
age         -0.088024   0.014573  -6.040 1.54e-09 ***
kidslt6     -1.443354   0.203583  -7.090 1.34e-12 ***
kidsge6      0.060112   0.074789   0.804  0.42154
---
```

Probit Model

probitres<-glm(inlf~nwifeinc+educ+exper+I(exper^2)+
age+kidslt6+kidsge6, family=binomial(link=probit),data=mroz)

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.2700736  0.5080782   0.532  0.59503
nwifeinc    -0.0120236  0.0049392  -2.434  0.01492 *
educ         0.1309040  0.0253987   5.154 2.55e-07 ***
exper        0.1233472  0.0187587   6.575 4.85e-11 ***
I(exper^2)  -0.0018871  0.0005999  -3.145  0.00166 **
age         -0.0528524  0.0084624  -6.246 4.22e-10 ***
kidslt6     -0.8683247  0.1183773  -7.335 2.21e-13 ***
kidsge6      0.0360056  0.0440303   0.818  0.41350
```

Pseudo R-squared

$$\text{Pseudo } R^2 = 1 - \frac{\mathcal{L}_{ur}}{\mathcal{L}_0} = 1 - \frac{D_{ur}}{D_0}$$

$\mathcal{L}_0$ the model with only an intercept

$$|\mathcal{L}_{ur}| \leq |\mathcal{L}_0|$$

$$D = -2\mathcal{L}$$

McFadden's Pseudo R^2

logLik(probitres)

$$-401.3022$$

1 - probitres$deviance/probitres$null.deviance

$$0.22$$

Likelihood Ratio (LR) Test

$$LR = 2(\mathcal{L}_{ur} - \mathcal{L}_{r})$$

$$LR \overset{a}{\sim} \chi^2_q$$

q # of exclusion restrictions

lrtest(restr,probitres)

```
Model 1: inlf ~ nwifeinc + educ + exper + I(exper^2) + age + kidslt6 +
    kidsge6
Model 2: inlf ~ 1
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   8 -401.30
2   1 -514.87 -7 227.14  < 2.2e-16 ***
```

restr <- glm(inlf~nwifeinc+educ+ kidslt6+kidsge6, family=binomial(link=logit),data=mroz)

lrtest(restr,probitres)

```
Model 1: inlf ~ nwifeinc + educ + kidslt6 + kidsge6
Model 2: inlf ~ nwifeinc + educ + exper + I(exper^2) + age + kidslt6 +
    kidsge6
  #Df  LogLik Df  Chisq Pr(>Chisq)
1   5 -464.92
2   8 -401.30  3 127.25  < 2.2e-16 ***
```

Partial Effect at the Average (PEA)

$$\Delta\hat{P}(y=1|x) \approx [g(\hat{\beta}_0 + x\hat{\beta})\hat{\beta}_j]\Delta x_j$$

$$g(\hat{\beta}_0 + \bar{x}\hat{\beta}) = g(\hat{\beta}_0 + \hat{\beta}_1\bar{x}_1 + \hat{\beta}_2\bar{x}_2 + ... + \hat{\beta}_k\bar{x}_k)$$

Partial Effect of x_j for the "average" person in the sample.

Marginal Effect at the Mean (MEM)

library(mfx)

logitmfx(inlf~nwifeinc+educ+exper+I(exper^2)
+age+kidslt6+kidsge6, data=mroz, atmean=TRUE)

```
Marginal Effects:
                  dF/dx   Std. Err.       z     P>|z|
nwifeinc    -0.00519005  0.00204820 -2.5340  0.011278 *
educ         0.05377731  0.01056074  5.0922 3.539e-07 ***
exper        0.05005693  0.00782462  6.3974 1.581e-10 ***
I(exper^2)  -0.00076692  0.00024768 -3.0965  0.001959 **
age         -0.02140302  0.00353973 -6.0465 1.480e-09 ***
kidslt6     -0.35094982  0.04963897 -7.0700 1.549e-12 ***
kidsge6      0.01461621  0.01818832  0.8036  0.421625
---
```

Average Partial Effect (APE) or Average Marginal Effect (AME)

$$n^{-1} \sum_{i=1}^{n} [g(\hat{\beta}_0 + x_i \hat{\beta})\hat{\beta}_j]$$

AME is the average of the nonlinear function rather than the nonlinear function of the average

$$g[E(x\beta)] \neq E[g(x\beta)]$$

Average Marginal Effect (AME)

logitmfx(inlf~nwifeinc+educ+exper+I(exper^2)+age +kidslt6+kidsge6, data=mroz, atmean=FALSE)

```
Marginal Effects:
                  dF/dx   Std. Err.       z      P>|z|
nwifeinc    -0.00381181  0.00153898 -2.4769  0.013255 *
educ         0.03949652  0.00846811  4.6641 3.099e-06 ***
exper        0.03676411  0.00655577  5.6079 2.048e-08 ***
I(exper^2)  -0.00056326  0.00018795 -2.9968  0.002728 **
age         -0.01571936  0.00293269 -5.3600 8.320e-08 ***
kidslt6     -0.25775366  0.04263493 -6.0456 1.489e-09 ***
kidsge6      0.01073482  0.01339130  0.8016  0.422769
---
```

xpred <- list(nwifeinc=c(100,0),educ=c(5,17),exper= c(0,30), age=c(20,52),kidslt6=c(2,0),kidsge6=c(0,0))

predict(linprob, xpred,type = "response")

1	2
-0.41	1.04

predict(logitres, xpred,type = "response")

1	2
0.005	0.95

predict(probitres,xpred,type = "response")

1	2
0.001	0.96