

Probit and Logit Models| R programming language

Preliminaries

Script Editor

```
# setup
rm(list = ls())
directory <- "C:/Users/amalz/OneDrive/Desktop/"
# Install packages
PackageNames <- c("tidyverse", "stargazer", "magrittr", "margins", "caret")
for(i in PackageNames){
  if(!require(i, character.only = T)){
    install.packages(i, dependencies = T)
    require(i, character.only = T)
  }
}

MROZ <- read.csv(paste0(directory, 'MROZ.csv'))
MROZ %>% select(inlf, nwifeinc, educ, exper, age, kidslt6) %>% str
MROZ %>% select(inlf, nwifeinc, educ, exper, age, kidslt6) %>% head(10)
MROZ %>% select(inlf, nwifeinc, educ, exper, age, kidslt6) %>% stargazer(type = "text")
count(MROZ, inlf)
```

LPM, Probit and Logit models

Script Editor

```
LPM <- lm(inlf ~ nwifeinc + educ + exper + age + kidslt6, MROZ)
summary(LPM)

Probit <- glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
              family = binomial(link = "probit"),
              data = MROZ)
summary(Probit)

Logit <- glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
             family = binomial(link = "logit"),
             data = MROZ)
summary(Logit)
```

Console

```
> summary(LPM)

Call:
lm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
    data = MROZ)

Residuals:
    Min       1Q   Median       3Q      Max
-1.06854 -0.38575  0.08284  0.35737  0.95335

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.769833   0.135007   5.702 1.70e-08 ***
nwifeinc     -0.003259   0.001456  -2.239  0.0255 *
educ         0.039129   0.007364   5.314 1.42e-07 ***
exper        0.022211   0.002144  10.358 < 2e-16 ***
age          -0.018508   0.002299  -8.052 3.22e-15 ***
kidslt6      -0.275306   0.033366  -8.251 7.08e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4298 on 747 degrees of freedom
Multiple R-squared:  0.253,    Adjusted R-squared:  0.248
F-statistic: 50.61 on 5 and 747 DF, p-value: < 2.2e-16
```

Probit and Logit Models| R programming language

Console ctd...

```
> summary(Probit)

Call:
glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
     family = binomial(link = "probit"), data = MROZ)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5942  -0.9371   0.4342   0.8934   2.3229

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.764541   0.439490   1.740   0.0819 .
nwifeinc     -0.011371   0.004857  -2.341   0.0192 *
educ         0.131532   0.025082   5.244 1.57e-07 ***
exper        0.069148   0.007556   9.151 < 2e-16 ***
age          -0.057919   0.007790  -7.435 1.04e-13 ***
kidslt6      -0.886208   0.116696  -7.594 3.10e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.75  on 752  degrees of freedom
Residual deviance:  813.08  on 747  degrees of freedom
AIC: 825.08

Number of Fisher Scoring iterations: 4
> summary(Logit)

Call:
glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
     family = binomial(link = "logit"), data = MROZ)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5175  -0.9174   0.4441   0.8841   2.2974

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.153219   0.742068   1.554   0.1202
nwifeinc     -0.019900   0.008268  -2.407   0.0161 *
educ         0.223366   0.042969   5.198 2.01e-07 ***
exper        0.117887   0.013386   8.807 < 2e-16 ***
age          -0.095141   0.013439  -7.080 1.45e-12 ***
kidslt6      -1.463577   0.200353  -7.305 2.77e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.75  on 752  degrees of freedom
Residual deviance:  812.92  on 747  degrees of freedom
AIC: 824.92

Number of Fisher Scoring iterations: 4
```

Probit and Logit Models| R programming language

Predicted probabilities

Script Editor

```
# Predicted probabilities for LPM
MROZ %>% mutate(inlfhat_lpm = fitted(LPM))
# Predicted probabilities for probit model
MROZ %>% mutate(inlfhat_probit = fitted(Probit))
# Predicted probabilities for logit model
MROZ %>% mutate(inlfhat_logit = fitted(Logit))
# Summarize predicted values for inlf
MROZ %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit) %>%
  stargazer(type = "text")
MROZ %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit) %>%
  head(5)
MROZ %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit) %>%
  .[601:605, ]
```

Console

```
=====
Statistic      N  Mean  St. Dev.  Min    Max
-----
inlf            753 0.568   0.496     0      1
inlfhat_lpm     753 0.568   0.249   -0.322  1.118
inlfhat_probit  753 0.569   0.252    0.004   0.976
inlfhat_logit   753 0.568   0.253    0.011   0.965
-----
> MROZ %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit) %>%
+   head(5)
   inlf inlfhat_lpm inlfhat_probit inlfhat_logit
1     1    0.6472128    0.6726774    0.6810984
2     1    0.7316488    0.7671096    0.7650845
3     1    0.6102180    0.6292476    0.6384870
4     1    0.7212125    0.7615327    0.7632913
5     1    0.5585574    0.5713338    0.5726343
> MROZ %>% select(inlf, inlfhat_lpm, inlfhat_probit, inlfhat_logit) %>%
+   .[601:605, ]
   inlf inlfhat_lpm inlfhat_probit inlfhat_logit
601    0    0.3076519    0.27197649    0.26245605
602    0    0.3475045    0.30681459    0.29163341
603    0    0.3877844    0.35342592    0.34347639
604    0    0.6664316    0.69497534    0.69682317
605    0   -0.1871596    0.01378312    0.02533543
```

Marginal effects

Script Editor

```
# Linear probability model (LPM)
summary(LPM)
coef(LPM)
# Coefficients are marginal effects in a linear model
# Probit model
summary(Probit)
Mean <- model.frame(Probit) %>%
  map_df(mean) # mean of independent variables
# Probit - marginal effect at the mean
Probit.atmean <- margins(Probit, at = Mean)
summary(Probit.atmean)
# Probit - average marginal effect
Probit.AME <- margins(Probit)
summary(Probit.AME)
# Logit model
summary(Logit)
# Logit - marginal effect at the mean
```

Probit and Logit Models| R programming language

Script Editor ctd...

```
# Logit - marginal effect at the mean
Logit.atmean <- margins(Logit, at = Mean)
summary(Logit.atmean)
# Logit - average marginal effect
Logit.AME <- margins(Logit)
summary(Logit.AME)
```

Console

```
Call:
lm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
    data = MROZ)

Residuals:
    Min       1Q   Median       3Q      Max
-1.06854 -0.38575  0.08284  0.35737  0.95335

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.769833   0.135007   5.702 1.70e-08 ***
nwifeinc     -0.003259   0.001456  -2.239  0.0255 *
educ         0.039129   0.007364   5.314 1.42e-07 ***
exper        0.022211   0.002144  10.358 < 2e-16 ***
age          -0.018508   0.002299  -8.052 3.22e-15 ***
kidslt6      -0.275306   0.033366  -8.251 7.08e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4298 on 747 degrees of freedom
Multiple R-squared:  0.253,    Adjusted R-squared:  0.248
F-statistic: 50.61 on 5 and 747 DF,  p-value: < 2.2e-16
> coef(LPM)
(Intercept)    nwifeinc      educ      exper      age    kidslt6
0.769832596 -0.003258636  0.039128569  0.022210504 -0.018507859 -0.275306267
> # Coefficients are marginal effects in a linear model
> summary(Probit)

Call:
glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
     family = binomial(link = "probit"), data = MROZ)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5942  -0.9371   0.4342   0.8934   2.3229

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.764541   0.439490   1.740  0.0819 .
nwifeinc     -0.011371   0.004857  -2.341  0.0192 *
educ         0.131532   0.025082   5.244 1.57e-07 ***
exper        0.069148   0.007556   9.151 < 2e-16 ***
age          -0.057919   0.007790  -7.435 1.04e-13 ***
kidslt6      -0.886208   0.116696  -7.594 3.10e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.75  on 752  degrees of freedom
Residual deviance:  813.08  on 747  degrees of freedom
AIC: 825.08

Number of Fisher Scoring iterations: 4
> Mean <- model.frame(Probit) %>%
+   map_df(mean) # mean of independent variables
> # Probit - marginal effect at the mean
> Probit.atmean <- margins(Probit, at = Mean)
> summary(Probit.atmean)
  factor    inlf nwifeinc    educ    exper    age kidslt6    AME    SE      z      p    lower
age 0.5684 20.1290 12.2869 10.6308 42.5378  0.2377 -0.0226 0.0030 -7.4250 0.0000 -0.0286
educ 0.5684 20.1290 12.2869 10.6308 42.5378  0.2377  0.0513 0.0098  5.2428 0.0000  0.0321
exper 0.5684 20.1290 12.2869 10.6308 42.5378  0.2377  0.0270 0.0029  9.1961 0.0000  0.0212
kidslt6 0.5684 20.1290 12.2869 10.6308 42.5378  0.2377 -0.3457 0.0457 -7.5682 0.0000 -0.4352
nwifeinc 0.5684 20.1290 12.2869 10.6308 42.5378  0.2377 -0.0044 0.0019 -2.3407 0.0192 -0.0081
upper
-0.0166
0.0705
0.0327
-0.2561
-0.0007
> # Probit - average marginal effect
> Probit.AME <- margins(Probit)
> summary(Probit.AME)
  factor    AME    SE      z      p    lower    upper
age -0.0177 0.0021 -8.3532 0.0000 -0.0219 -0.0136
educ  0.0402 0.0073  5.5269 0.0000  0.0260  0.0545
exper  0.0211 0.0019 10.9991 0.0000  0.0174  0.0249
kidslt6 -0.2709 0.0316 -8.5768 0.0000 -0.3328 -0.2090
nwifeinc -0.0035 0.0015 -2.3646 0.0181 -0.0064 -0.0006
> # Logit model
> summary(Logit)
```

Probit and Logit Models| R programming language

Console ctd...

```
Call:
glm(formula = inlf ~ nwifeinc + educ + exper + age + kidslt6,
    family = binomial(link = "logit"), data = MROZ)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.5175  -0.9174   0.4441   0.8841   2.2974

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.153219   0.742068   1.554   0.1202
nwifeinc     -0.019900   0.008268  -2.407   0.0161 *
educ         0.223366   0.042969   5.198 2.01e-07 ***
exper        0.117887   0.013386   8.807 < 2e-16 ***
age          -0.095141   0.013439  -7.080 1.45e-12 ***
kidslt6     -1.463577   0.200353  -7.305 2.77e-13 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1029.75  on 752  degrees of freedom
Residual deviance:  812.92  on 747  degrees of freedom
AIC: 824.92

Number of Fisher Scoring iterations: 4
> # Logit - marginal effect at the mean
> Logit.atmean <- margins(Logit, at = Mean)
> summary(Logit.atmean)
  factor   inlf nwifeinc   educ   exper   age kidslt6    AME    SE      z      p   lower
age 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0231 0.0033 -7.0706 0.0000 -0.0294
educ 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0541 0.0104  5.1986 0.0000  0.0337
exper 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377  0.0286 0.0032  8.9082 0.0000  0.0223
kidslt6 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.3546 0.0488 -7.2723 0.0000 -0.4501
nwifeinc 0.5684  20.1290 12.2869 10.6308 42.5378  0.2377 -0.0048 0.0020 -2.4066 0.0161 -0.0087
upper
-0.0167
 0.0745
 0.0348
-0.2590
-0.0009
> # Logit - average marginal effect
> Logit.AME <- margins(Logit)
> summary(Logit.AME)
  factor    AME    SE      z      p   lower  upper
age -0.0173 0.0021 -8.0925 0.0000 -0.0214 -0.0131
educ  0.0405 0.0073  5.5481 0.0000  0.0262  0.0548
exper  0.0214 0.0020 10.9388 0.0000  0.0176  0.0252
kidslt6 -0.2655 0.0315 -8.4298 0.0000 -0.3273 -0.2038
nwifeinc -0.0036 0.0015 -2.4388 0.0147 -0.0065 -0.0007
```

Pseudo R-squared

Script Editor

```
# Probit model - unrestricted model with all variables
ur.model <- Probit
# Log-likelihood for unrestricted model
(LLur <- logLik(ur.model))
# Probit restricted model with only constant
r.model <- glm(inlf ~ 1, binomial(link = "probit"), data = MROZ)
# Log-likelihood for model with only constant
(LL0 <- logLik(r.model))
# Calculate pseudo R-squared
(pseudo_r2 <- 1 - LLur/LL0)
```

Probit and Logit Models| R programming language

Console

```
> ur.model <- Probit
> # Log-likelihood for unrestricted model
> (LLur <- logLik(ur.model))
'log Lik.' -406.5421 (df=6)
> # Probit restricted model with only constant
> r.model <- glm(inlf ~ 1, binomial(link = "probit"), data = MROZ)
> # Log-likelihood for model with only constant
> (LL0 <- logLik(r.model))
'log Lik.' -514.8732 (df=1)
> # Calculate pseudo R-squared
> (pseudo_r2 <- 1 - LLur/LL0)
'log Lik.' 0.2104035 (df=6)
```



ECONIMIUM
ECONOMICS FOR ALL