```
# Probit and Logit Models in R
# Copyright 2013 by Ani Katchova

mydata<- read.csv("C:/Econometrics/Data/probit_insurance.csv")
attach(mydata)

# Define variables
Y <- cbind(ins)
X <- cbind(retire, age, hstatusg, hhincome, educyear, married, hisp)

# Descriptive statistics
summary(Y)
summary(X)

table(Y)
table(Y)/sum(table(Y))

# Regression coefficients
olsreg <- lm(Y ~ X)
summary(olsreg)

# Logit model coefficients
logit<- glm(Y ~ X, family=binomial (link = "logit"))
summary(logit)

# Logit model odds ratios
exp(logit$coefficients)

# Probit model coefficients
probit<- glm(Y ~ X, family=binomial (link="probit"))
summary(probit)


# Regression marginal effects
coef(olsreg)

# Logit model average marginal effects
LogitScalar <- mean(dlogis(predict(logit, type = "link")))
LogitScalar * coef(logit)

# Probit model average marginal effects
ProbitScalar <- mean(dnorm(predict(probit, type = "link")))
ProbitScalar * coef(probit)


# Regression predicted probabilities
polsreg<- predict(olsreg)
summary(polsreg)

# Logit model predicted probabilities
plogit<- predict(logit, type="response")
summary(plogit)

# Probit model predicted probabilities
pprobit<- predict(probit, type="response")
```

```
summary(pprobit)


# Percent correctly predicted values
table(true = Y, pred = round(fitted(probit)))
table(true = Y, pred = round(fitted(logit)))

# McFadden's Pseudo R-squared
probit0<-update(probit, formula= Y ~ 1)
McFadden<- 1-as.vector(logLik(probit)/logLik(probit0))
McFadden
```

```
> # Probit and Logit Models in R
> # Copyright 2013 by Ani Katchova
>
> mydata<- read.csv("C:/Econometrics/Data/probit_insurance.csv")
> attach(mydata)
>
> # Define variables
> Y <- cbind(ins)
> X <- cbind(retire, age, hstatusg, hhincome, educyear, married, hisp)
>
> # Descriptive statistics
> summary(Y)
      ins
 Min.   :0.0000
 1st Qu.:0.0000
 Median :0.0000
 Mean   :0.3871
 3rd Qu.:1.0000
 Max.   :1.0000
> summary(X)
     retire            age           hstatusg         hhincome
 Min.   :0.0000   Min.   :52.00   Min.   :0.0000   Min.   :   0.00
 1st Qu.:0.0000   1st Qu.:65.00   1st Qu.:0.0000   1st Qu.:  17.00
 Median :1.0000   Median :67.00   Median :1.0000   Median :  31.10
 Mean   :0.6248   Mean   :66.91   Mean   :0.7046   Mean   :  45.26
 3rd Qu.:1.0000   3rd Qu.:69.00   3rd Qu.:1.0000   3rd Qu.:  52.80
 Max.   :1.0000   Max.   :86.00   Max.   :1.0000   Max.   :1312.12
    educyear       married           hisp
 Min.   : 0.0   Min.   :0.000   Min.   :0.00000
 1st Qu.:10.0   1st Qu.:0.000   1st Qu.:0.00000
 Median :12.0   Median :1.000   Median :0.00000
 Mean   :11.9   Mean   :0.733   Mean   :0.07268
 3rd Qu.:14.0   3rd Qu.:1.000   3rd Qu.:0.00000
 Max.   :17.0   Max.   :1.000   Max.   :1.00000
>
> table(Y)
Y
   0    1
1965 1241
> table(Y)/sum(table(Y))
Y
        0         1
0.6129133 0.3870867
>
> # Regression coefficients
> olsreg <- lm(Y ~ X)
> summary(olsreg)

Call:
lm(formula = Y ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-1.1233 -0.4065 -0.2291  0.5298  1.0341
```

```
Coefficients:
             Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.1270857  0.1605628   0.792  0.42871
Xretire      0.0408508  0.0182197   2.242  0.02502 *
Xage        -0.0028955  0.0024189  -1.197  0.23138
Xhstatusg    0.0655583  0.0194531   3.370  0.00076 ***
Xhhincome    0.0004921  0.0001375   3.579  0.00035 ***
Xeducyear    0.0233686  0.0028672   8.150 5.15e-16 ***
Xmarried     0.1234699  0.0193618   6.377 2.07e-10 ***
Xhisp       -0.1210059  0.0336660  -3.594  0.00033 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

Residual standard error: 0.4671 on 3198 degrees of freedom
Multiple R-squared: 0.08262,     Adjusted R-squared: 0.08061
F-statistic: 41.14 on 7 and 3198 DF,  p-value: < 2.2e-16

>
> # Logit model coefficients
> logit<- glm(Y ~ X, family=binomial (link = "logit"))
> summary(logit)

Call:
glm(formula = Y ~ X, family = binomial(link = "logit"))

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-2.456  -1.009  -0.703   1.224   2.373

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.715578   0.748622  -2.292 0.021926 *
Xretire      0.196930   0.084207   2.339 0.019354 *
Xage        -0.014596   0.011287  -1.293 0.195969
Xhstatusg    0.312265   0.091674   3.406 0.000659 ***
Xhhincome    0.002304   0.000762   3.023 0.002503 **
Xeducyear    0.114263   0.014201   8.046 8.55e-16 ***
Xmarried     0.578636   0.093320   6.201 5.63e-10 ***
Xhisp       -0.810306   0.195751  -4.139 3.48e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4279.5  on 3205  degrees of freedom
Residual deviance: 3989.8  on 3198  degrees of freedom
AIC: 4005.8

Number of Fisher Scoring iterations: 4

>
> # Logit model odds ratios
> exp(logit$coefficients)
(Intercept)     Xretire        Xage   Xhstatusg   Xhhincome   Xeducyear
  0.1798597   1.2176584   0.9855105   1.3665173   1.0023063   1.1210464
```

```
   Xmarried       Xhisp
  1.7836040   0.4447220
>
> # Probit model coefficients
> probit<- glm(Y ~ X, family=binomial (link="probit"))
> summary(probit)

Call:
glm(formula = Y ~ X, family = binomial(link = "probit"))

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.4220  -1.0136  -0.6981   1.2236   2.4745

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept) -1.0693423  0.4553103  -2.349 0.018844 *
Xretire      0.1183526  0.0513402   2.305 0.021152 *
Xage        -0.0088694  0.0068630  -1.292 0.196235
Xhstatusg    0.1977411  0.0554905   3.564 0.000366 ***
Xhhincome    0.0012327  0.0004371   2.820 0.004798 **
Xeducyear    0.0707492  0.0084925   8.331  < 2e-16 ***
Xmarried     0.3623376  0.0560521   6.464 1.02e-10 ***
Xhisp       -0.4731143  0.1102319  -4.292 1.77e-05 ***
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 4279.5  on 3205  degrees of freedom
Residual deviance: 3987.2  on 3198  degrees of freedom
AIC: 4003.2

Number of Fisher Scoring iterations: 6

>
>
> # Regression marginal effects
> coef(olsreg)
  (Intercept)       Xretire          Xage     Xhstatusg     Xhhincome
 0.1270856958  0.0408508171 -0.0028955466  0.0655583417  0.0004920877
    Xeducyear      Xmarried         Xhisp
 0.0233686296  0.1234698801 -0.1210059350
>
> # Logit model average marginal effects
> LogitScalar <- mean(dlogis(predict(logit, type = "link")))
> LogitScalar * coef(logit)
  (Intercept)       Xretire          Xage     Xhstatusg     Xhhincome
-0.3725232720  0.0427616020 -0.0031692953  0.0678057706  0.0005002078
    Xeducyear      Xmarried         Xhisp
 0.0248111432  0.1256458981 -0.1759510453
>
> # Probit model average marginal effects
> ProbitScalar <- mean(dnorm(predict(probit, type = "link")))
> ProbitScalar * coef(probit)
```

```
  (Intercept)       Xretire          Xage     Xhstatusg     Xhhincome
-0.3792718940  0.0419770495 -0.0031457860  0.0701343649  0.0004371967
    Xeducyear      Xmarried         Xhisp
 0.0250931581  0.1285130972 -0.1678031247
>
>
> # Regression predicted probabilities
> polsreg<- predict(olsreg)
> summary(polsreg)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
-0.1557  0.3055  0.4074  0.3871  0.4736  1.1970
>
> # Logit model predicted probabilities
> plogit<- predict(logit, type="response")
> summary(plogit)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.03402 0.28490 0.39940 0.38710 0.47780 0.96500
>
> # Probit model predicted probabilities
> pprobit<- predict(probit, type="response")
> summary(pprobit)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
0.02064 0.28660 0.40170 0.38610 0.47680 0.96470
>
>
> # Percent correctly predicted values
> table(true = Y, pred = round(fitted(probit)))
    pred
true    0    1
   0 1660  305
   1  906  335
> table(true = Y, pred = round(fitted(logit)))
    pred
true    0    1
   0 1657  308
   1  896  345
>
> # McFadden's Pseudo R-squared
> probit0<-update(probit, formula= Y ~ 1)
> McFadden<- 1-as.vector(logLik(probit)/logLik(probit0))
> McFadden
[1] 0.06830054
```