```
------------------------------------------------------------------------------
      name:  PK
       log:
/Users/priyakoirala/Desktop/school/econometrics/projects/project5/koirala_project5.log
  log type:  text
 opened on:  18 May 2023, 19:39:43

. 
. /*===========================================================================
> The purpose of this exercise is to understand how borrower characteristics
> relate to the probability of home loan approval.
> 
> Open the LOANAPP.dta data set. It is a random sample of people who applied for
> home mortgage loans. It contains information about the people's characteristics
> and a binary variable for whether their loan application got approved or not.
> ===========================================================================*/
. 
. use "/Users/priyakoirala/Desktop/school/econometrics/projects/project5/loanapp.dta"
. 
```

```
. /*===============================================================================
> (Q1): Use a linear probability model to estimate the relationship between the
> probability of loan approval (approve) and the following variables:
>
> X1=bankruptcy
> X2=hh_expenditures
> X3=term
> X4=apr
> X5=Black
> X6=Hispanic
>
> What type of standard errors should you use?
> ===============================================================================*/
.
. sum

    Variable |        Obs        Mean    Std. dev.       Min        Max
-------------+---------------------------------------------------------
       gdlin |      1,983    .9152799     .278535          0          1
  bankruptcy |      1,983     .068583    .2528074          0          1
hh_expendi~s |      1,983    24.80028    7.122213          1         72
        term |      1,983    341.0671    64.52418          6        480
         apr |      1,983    205.2156    156.3076         25       4316
-------------+---------------------------------------------------------
       black |      1,983    .0983359    .2978433          0          1
      hispan |      1,983    .0559758     .229933          0          1
        male |      1,968    .8130081    .3900041          0          1
     approve |      1,983     .878467    .3268281          0          1

.
. reg approve i.bankruptcy c.hh_expenditures c.term c.apr i.black i.hisp, robust

Linear regression                               Number of obs     =      1,983
                                                F(6, 1976)        =      20.71
                                                Prob > F          =     0.0000
                                                R-squared         =     0.1102
                                                Root MSE          =     .30877

------------------------------------------------------------------------------
             |               Robust
     approve | Coefficient  std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
  1.bankruptcy |  -.3052161   .0430494    -7.09   0.000    -.3896431   -.2207891
hh_expenditures |  -.0027087   .0011926    -2.27   0.023    -.0050476   -.0003699
        term |   .0001036    .000112     0.92   0.355    -.0001161    .0003232
         apr |    .000101   .0000428     2.36   0.018     .0000171    .0001849
      1.black |  -.1840989    .032968    -5.58   0.000    -.2487546   -.1194433
     1.hispan |  -.1175171   .0398279    -2.95   0.003    -.1956262    -.039408
        _cons |   .9352098   .0461626    20.26   0.000     .8446774    1.025742
------------------------------------------------------------------------------

.
. /* For a Linear Probability Model, we should always use heteroskedasticity robust
> standard errors as the variance in these models are not constant, therefore, they
> are naturally heteroskedastic. */
.
```

```
. /*===============================================================================
> (Q2): What is the marginal effect of household expenditures on loan approval
> probability?
> ===============================================================================*/
.
. /* Marginal Effect: -0.0027087.
>
> For every $1,000 increase in household expenditure per year, the probability of
> loan approval decreases by 0.0027. */
.
```

```
. /*===============================================================================
> (Q3): Estimate the same model as in (Q1) but use the probit functional form.
>
> What is the marginal effect of household expenditures on loan approval
> probability when all variables are held at their mean values?
> ===============================================================================*/
.
. probit approve i.bankruptcy c.hh_expenditures c.term c.apr i.black i.hisp, robust

Iteration 0:   log pseudolikelihood = -733.64726
Iteration 1:   log pseudolikelihood = -650.20317
Iteration 2:   log pseudolikelihood = -648.12078
Iteration 3:   log pseudolikelihood = -648.11547
Iteration 4:   log pseudolikelihood = -648.11547

Probit regression                                     Number of obs =  1,983
                                                      Wald chi2(6)  = 160.06
                                                      Prob > chi2   = 0.0000
Log pseudolikelihood = -648.11547                     Pseudo R2     = 0.1166

------------------------------------------------------------------------------
             |               Robust
     approve | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  1.bankruptcy |   -1.04896   .1214742    -8.64   0.000    -1.287046   -.8108753
hh_expenditures |   -.014929   .0060127    -2.48   0.013    -.0267136   -.0031443
        term |   .0005265   .0006085     0.87   0.387    -.0006661    .0017192
         apr |   .0012518   .0004483     2.79   0.005     .0003731    .0021304
      1.black |  -.7127674   .1087012    -6.56   0.000    -.9258178    -.499717
      1.hispan |  -.5004237   .1437615    -3.48   0.000    -.7821911   -.2186563
       _cons |    1.37818   .2605525     5.29   0.000     .8675065    1.888853
------------------------------------------------------------------------------
Note: 0 failures and 1 success completely determined.

.
. margins, dydx(_all) at((means)_all)

Conditional marginal effects                          Number of obs = 1,983
Model VCE: Robust

Expression: Pr(approve), predict()
dy/dx wrt:  1.bankruptcy hh_expenditures term apr 1.black 1.hispan
At: 0.bankruptcy    =  .931417 (mean)
    1.bankruptcy    =  .068583 (mean)
    hh_expenditures = 24.80028 (mean)
    term            = 341.0671 (mean)
    apr             = 205.2156 (mean)
    0.black         = .9016641 (mean)
    1.black         = .0983359 (mean)
    0.hispan        = .9440242 (mean)
    1.hispan        = .0559758 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  1.bankruptcy |  -.2939985   .0443986    -6.62   0.000    -.3810181   -.2069789
hh_expenditures |  -.0026442   .0010571    -2.50   0.012     -.004716   -.0005723
        term |   .0000933   .0001077     0.87   0.386    -.0001178    .0003043
         apr |   .0002217   .0000779     2.85   0.004      .000069    .0003744
```

```
        1.black |   -.174394   .0335305    -5.20   0.000    -.2401126   -.1086755
       1.hispan |  -.1148957   .0404966    -2.84   0.005    -.1942676   -.0355238
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.
. /* Marginal Effect: -0.0026442.
>
> For every $1,000 increase in household expenditure per year (holding the mean
> value for all of the variables in the sample) the probability of loan approval
> decreases by 0.0026.*/
.
```

```
. /*==========================================================================
> (Q4): Test whether loan term and loan APR are jointly statistically significant.
> Make sure to write the null and alternative hypotheses (not in Stata format).
> ==========================================================================*/
.
. test term apr

 ( 1)  [approve]term = 0
 ( 2)  [approve]apr = 0

           chi2(  2) =    8.39
         Prob > chi2 =    0.0150

.
. /* H_0: beta3 = beta4 = 0
>    H_1: beta3 != 0 and/or beta4 != 0
>
>    p-value of Chi-squared statistic: |0.0150| < 0.05
>
>    We reject the null hypothesis of no joint statistical significance.
>
> We accept the alternate hypothesis and conclude that loan term and loan apr are
> jointly statistically significant. */
.
```

```
. /*===============================================================================
> (Q5): What is the predicted probability of loan approval for someone who has
> never filed for bankruptcy, has $25,000 in yearly household expenditures,
> wants a 360 month term mortgage at $205 APR, is not Black, and is not Hispanic?
> ===============================================================================*/
.
. probit approve i.bankruptcy c.hh_expenditures c.term c.apr i.black i.hisp, robust

Iteration 0:   log pseudolikelihood = -733.64726
Iteration 1:   log pseudolikelihood = -650.20317
Iteration 2:   log pseudolikelihood = -648.12078
Iteration 3:   log pseudolikelihood = -648.11547
Iteration 4:   log pseudolikelihood = -648.11547

Probit regression                                     Number of obs =  1,983
                                                      Wald chi2(6)  = 160.06
                                                      Prob > chi2   = 0.0000
Log pseudolikelihood = -648.11547                     Pseudo R2     = 0.1166

------------------------------------------------------------------------------
             |               Robust
     approve | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  1.bankruptcy |   -1.04896   .1214742    -8.64   0.000    -1.287046   -.8108753
hh_expenditures |   -.014929   .0060127    -2.48   0.013    -.0267136   -.0031443
        term |   .0005265   .0006085     0.87   0.387    -.0006661    .0017192
         apr |   .0012518   .0004483     2.79   0.005     .0003731    .0021304
      1.black |  -.7127674   .1087012    -6.56   0.000    -.9258178    -.499717
     1.hispan |  -.5004237   .1437615    -3.48   0.000    -.7821911   -.2186563
       _cons |    1.37818   .2605525     5.29   0.000     .8675065    1.888853
------------------------------------------------------------------------------
Note: 0 failures and 1 success completely determined.

.
. scalar yhat_nonblack = _b[_cons] + _b[hh_expenditures]*25 + _b[term]*360 + _b[apr
> ]*205


.
. display yhat_nonblack
1.4511299

.
. /* The standard normal distribution for Phi coefficient 1.45 is 0.92647. */
.
. display normprob(_b[_cons] +_b[hh_expenditures]*25 + _b[term]*360 + _b[apr]*205)
.92662816

.
. /* The predicted probability of loan approval for someone with these
> characteristics is 0.926. */
.
```

```
. /*=============================================================================
> (Q6) What is the relationship between being Black and home loan approval for an
> applicant with otherwise the same characteristics as the applicant in (Q5)?
>
> Calculate the result by hand first, and then use Stata to confirm.
> =============================================================================*/
.
. scalar yhat_nonblack = _b[_cons] + _b[hh_expenditures]*25 + _b[term]*360 + _b[apr
> ]*205 + _b[1.black]
.
. display yhat_nonblack
.73836257
.
. /* The standard normal distribution for Phi coefficient 0.73 is 0.76730.
> The predicted probability of home loan approval for a Black applicant with
> otherwise the same characteristics in (Q5) is 0.767.*/
.
. display normprob(_b[_cons] + _b[hh_expenditures]*25 + _b[term]*360 + _b[apr]*205
> + _b[1.black])
.76985292
.
. /* Black applicants: 0.767
>    Non-Black applicants: 0.926
>    0.767 - 0.926 = -0.159
>
> The difference in loan approval probability for two applicants with the same
> characteristics where one person is Black and the other is not Black is 0.159. */
.
. margins, dydx(_all) at(bankruptcy=0 hh_expenditures=25 term=360 apr=205 black=0 h
> isp=0)

Conditional marginal effects                              Number of obs = 1,983
Model VCE: Robust

Expression: Pr(approve), predict()
dy/dx wrt:  1.bankruptcy hh_expenditures term apr 1.black 1.hispan
At: bankruptcy      =   0
    hh_expenditures =  25
    term            = 360
    apr             = 205
    black           =   0
    hispan          =   0

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
  1.bankruptcy |  -.2704078   .0431548    -6.27   0.000    -.3549896    -.185826
hh_expenditures |  -.0020781   .0008337    -2.49   0.013    -.0037122   -.0004441
          term |   .0000733   .0000832     0.88   0.379    -.0000898    .0002364
           apr |   .0001743   .0000619     2.81   0.005     .0000528    .0002957
       1.black |  -.1567752   .0310937    -5.04   0.000    -.2177178   -.0958327
      1.hispan |  -.0975049   .0353522    -2.76   0.006    -.1667939   -.0282159
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

. /* dy/dx for Black applicant: -0.1567752
>
> There seems to be a negative relationship with being Black and approved for a home
> loan. The probability of a Black applicant being approved for a loan is 0.157
> lower than a non-Black applicant with the same characteristics. */
```

```
. /*===============================================================================
> (Q7) Now estimate a logit model where Y=approve and include the following Xs:
> X1=gdlin
> X2=hh_expenditures
> X3=black
> X4=Hispanic
>
> What is the marginal effect of household expenditures on loan approval
> probability for someone who meets the loan guidelines, has $30,000 per year
> in expenditures, is not Black, and is Hispanic?
> ===============================================================================*/
.
. logit approve i.gdlin c.hh_expenditures i.black i.hisp, robust

Iteration 0:   log pseudolikelihood = -733.64726
Iteration 1:   log pseudolikelihood = -566.40839
Iteration 2:   log pseudolikelihood = -498.48485
Iteration 3:   log pseudolikelihood =  -495.4255
Iteration 4:   log pseudolikelihood = -495.42102
Iteration 5:   log pseudolikelihood = -495.42102

Logistic regression                                     Number of obs =  1,983
                                                        Wald chi2(4)  = 346.80
                                                        Prob > chi2   = 0.0000
Log pseudolikelihood = -495.42102                       Pseudo R2     = 0.3247

------------------------------------------------------------------------------
             |               Robust
     approve | Coefficient  std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
      1.gdlin |   3.804443   .2142151    17.76   0.000     3.384589    4.224297
hh_expenditures |  -.0181353   .0144019    -1.26   0.208    -.0463624    .0100918
      1.black |  -.9566613   .2329437    -4.11   0.000    -1.413223   -.5000999
     1.hispan |   -.975948   .3167225    -3.08   0.002    -1.596713   -.3551833
        _cons |  -.4367085   .4143761    -1.05   0.292    -1.248871    .3754538
------------------------------------------------------------------------------

.
. margins, dydx(hh_expenditures) at (gdlin=1 hh_expenditures=30 black=0 hisp=1)

Conditional marginal effects                            Number of obs = 1,983
Model VCE: Robust

Expression: Pr(approve), predict()
dy/dx wrt:  hh_expenditures
At: gdlin          =  1
    hh_expenditures = 30
    black          =  0
    hispan         =  1

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
hh_expenditures |  -.0021328    .001865    -1.14   0.253    -.0057882    .0015226
------------------------------------------------------------------------------

. /* Marginal Effect:-0.0021328
>
> For every $1,000 increase in household expenditure per year, the probability of
> loan approval decreases by 0.0021 for an applicant with these characteristics. */
```

```
. /*==============================================================================
> (Q8) Is the marginal effect of household expenditures in (Q7) statistically
> significant?
> ==============================================================================*/
.
. /* H_0: beta2 = 0
>    H_1: beta2 != 0
>
>    t-test = |-1.14| < 1.96
>
> We fail to reject the null hypothesis at the 5% significance level. We conclude
> that the marginal effect of household expenditures in (Q7) is not statistically
> significant. */
.
```

```
. /*==========================================================================
> (Q9) What is the predicted probability of loan approval for a person with the
> characteristics described in (Q7)?
>
> Please calculate by hand. Then you may use Stata to confirm.
> ==========================================================================*/
.
. scalar yhat_hisp = _b[_cons] + _b[1.gdlin] + _b[hh_expenditures]*30 + _b[1.hisp]

.
. display yhat_hisp
1.8477279

.
. /* 1/(1+e^-1.85) = 0.86
>
> The predicted probability of loan approval for an applicant with the
> characteristics described in (Q7) is 0.86. */
.
. display 1/[1+exp(-1*(_b[_cons]+_b[1.gdlin]+_b[hh_expenditures]*30+_b[1.hisp]))]
.86386011

.
```

```
. /*==========================================================================
> (Q10) What is the predicted probability of loan approval for a person with the
> same characteristics as described in (Q7), except the person is non-Hispanic?
>
> Please calculate by hand.
> ==========================================================================*/
.
. scalar yhat_nonhisp = _b[_cons] + _b[1.gdlin] + _b[hh_expenditures]*30

.
. display yhat_nonhisp
2.8236759

.
. /* 1/(1+e^-2.82) = 0.94
>
> The predicated probability of loan approval for a Non-Hispanic person with the
> same characteristics described in (Q7) is 0.94. */
.
. display 1/[1+exp(-1*(_b[_cons]+_b[1.gdlin]+_b[hh_expenditures]*30))]
.9439419

.
. /* Hispanic applicant: 0.86
>    Non-Hispanic Applicant: 0.94
>
>    0.86 - 0.94 = -0.08
>
> The difference in loan approval probability between two applicants with the same
> characteristics where one person is Hispanic and the other one is Non-Hispanic is
> 0.08. */
.
. margins, dydx(hisp) at(gdlin=1 hh_expenditures=30 black=0 hisp=1)

Conditional marginal effects                              Number of obs = 1,983
Model VCE: Robust

Expression: Pr(approve), predict()
dy/dx wrt:  1.hispan
At: gdlin           =  1
    hh_expenditures = 30
    black           =  0
    hispan          =  1

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   std. err.      z    P>|z|     [95% conf. interval]
-------------+----------------------------------------------------------------
    1.hispan |  -.0800818   .0359376    -2.23   0.026    -.1505182   -.0096454
------------------------------------------------------------------------------
Note: dy/dx for factor levels is the discrete change from the base level.

.
. /* dy/dx for Hispanic applicant: -.0800818
>
> Therefore, according to our estimation model, the probability of a Hispanic
> applicant being approved for a loan is 0.08 lower than a non-Hispanic.*/
.
```

```
. /*===============================================================================
> (Q11) Based on your answers to the questions in this project, do you think there
> is discrimination in the market for home loans?
> ===============================================================================*/
.
> /* Based on the answers to the questions in this project, it seems that there is
> some discrimination in the market for home loans. Primarily, there seems to be a
> bias against Hispanic and Black applicants.
>
> We simulated models by changing the characteristics of applicants such as if the
> applicant has filed bankruptcy before, their household expenditures and the
> borrowing guidelines. We then compared the models from Hispanic applicants to non-
> Hispanic applicants, and black applicants to non-Black applicants and the
> difference suggests that Hispanic and Black applicants are less likely to be
> approved for a loan possessing the same characteristics as their non-Black or
> non-Hispanic counterparts.
>
> However, we should also note that there are other variables we may not have
> accounted for in our sample, and therefore, this result could also be the cause of
> omitted variable bias. For example, credit history, capital, and work history may
> be some other factors that lenders investigate when approving someone for a loan.
> In order to get a more accurate estimate from out models we may need more
> information regarding the applicants.
>
> Based on the project alone, it seems to be that Black and Hispanic applicants face
> discrimination in the loan approval process. */
>
> /*===============================================================================
> ===============================================================================*/
>

end of do-file
```