COLLABORATIVE MASTERS PROGRAMME IN ECONOMICS FOR ANGLOPHONE AFRICA (CMAP)

JOINT FACILITY FOR ELECTIVES

ECONOMETRICS THEORY AND PRACTICE PART TWO:TOPICS IN MICRO-ECONOMETRICS

AUGUST/SEPTEMBER 2011

DR. MOSES SICHEI*

* KSMS Research Centre
Mobile:+254 723383505;Email: sichei@yahoo.co.uk or Sicheimm@ksms.or.ke

Date: 16/8/2011

BINARY CHOICE PRACTICAL NOTES

1. OBJECTIVE:

Get hands-on-experience in specifying, estimating and interpreting results from major binary choice models. Specifically the practical focuses on;

(i) Specification of LPM, logit and probit models
(ii) Estimation of these models
(iii) Interpretation of the models
(iv) Diagnostic statistics
(v) Presentation of the results

2. DATA:

This data is the one used in Green (listed in appendix Table F14.1) on programme effectiveness covering 32 cross-sections. The data is taken from Spector and Mazzeo(1980). The details are

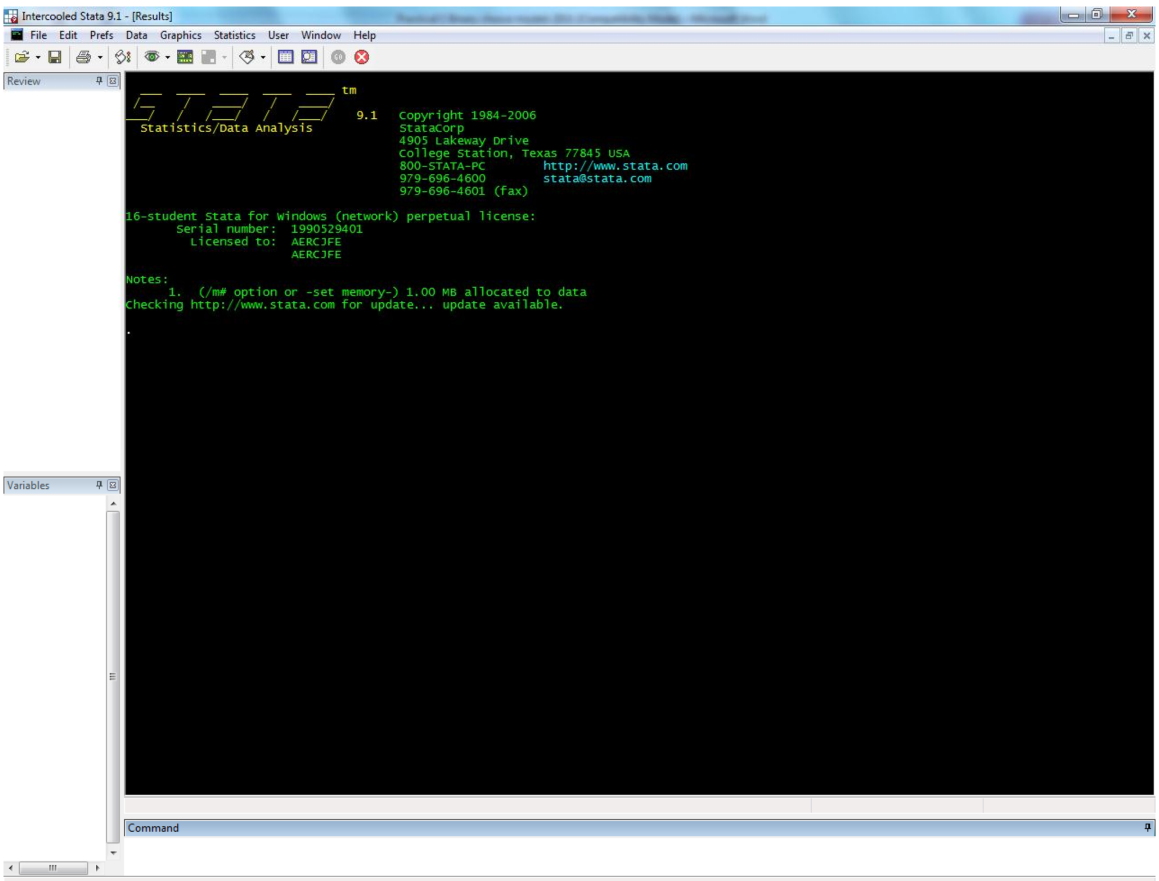
Obs=Observation
GRADE=grade improvement dummy
GPA =the studentôs grade point average
TUCE=previous knowledge of material
PSI=dummy for personalized system of instruction teaching method

- The main research problem is to ***examine whether a new method of teaching economics, the PSI significantly influence performance in later economics courses***.

- The dependent variable used is GRADE, which indicates whether a studentôs grade in intermediate macroeconomics course was higher than that in the principle course.
- The data is in an excel file called **binary choice data.xls**

3. SOFTWARE

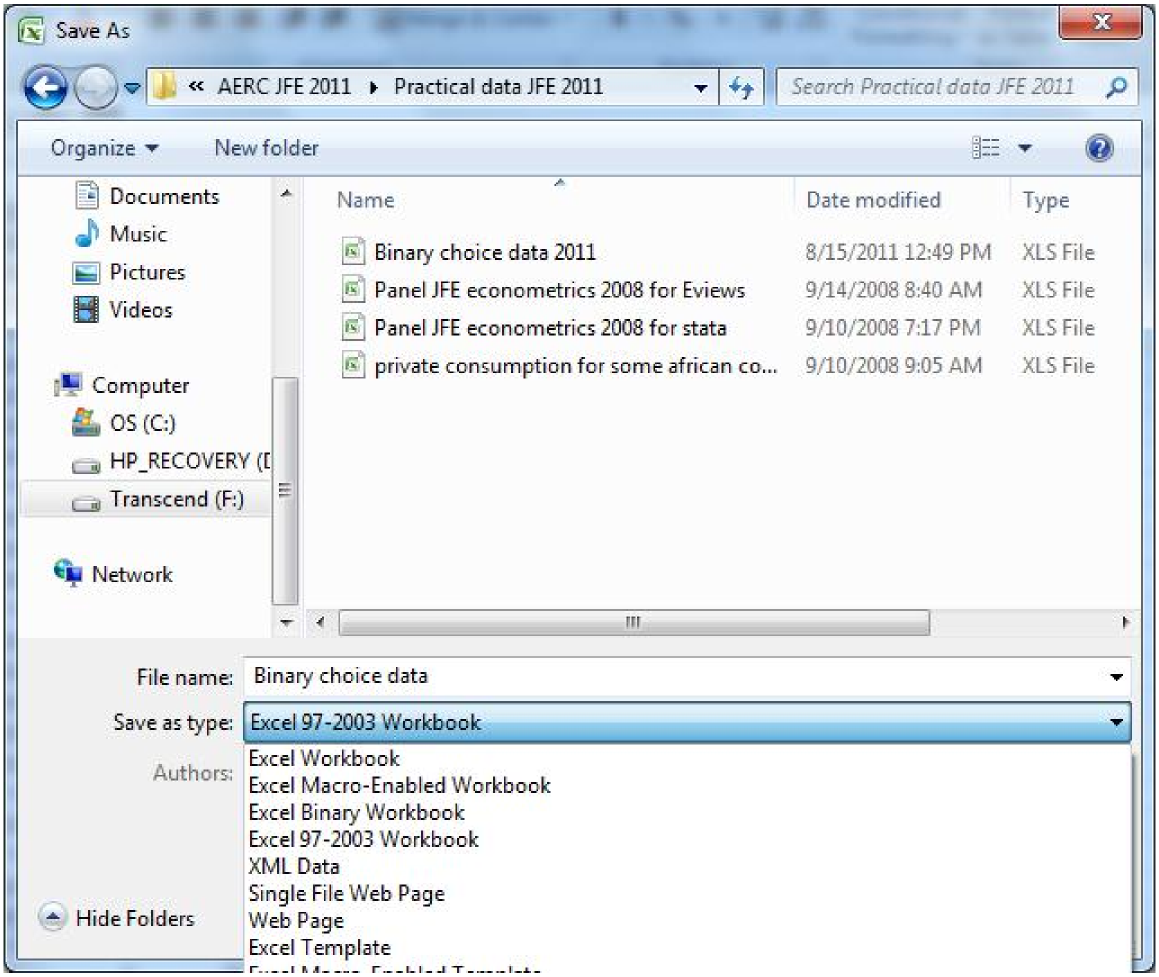
Use Stata Release 9 software. Introduction to stata 9 notes are provided separately.



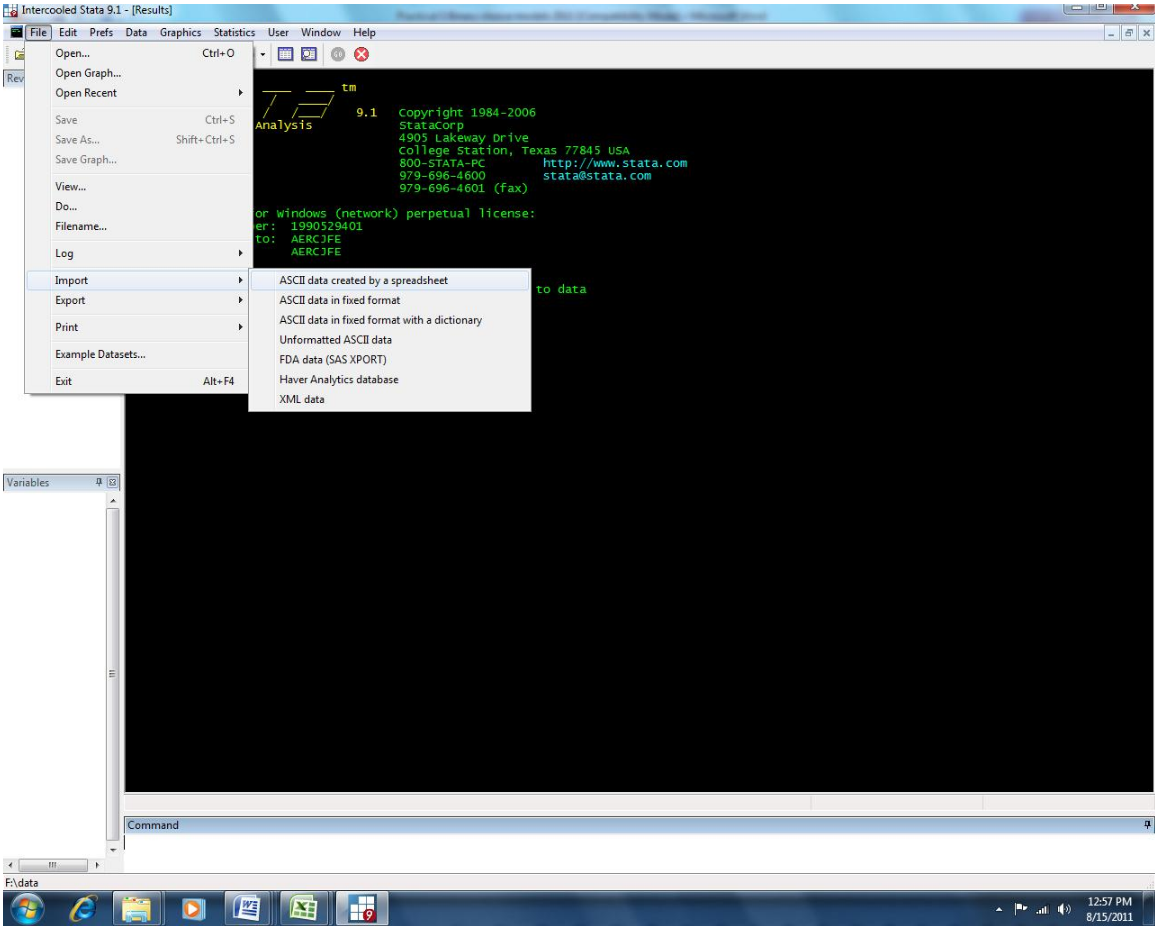
4. LOADING THE DATA

The steps to load the data are as follows:

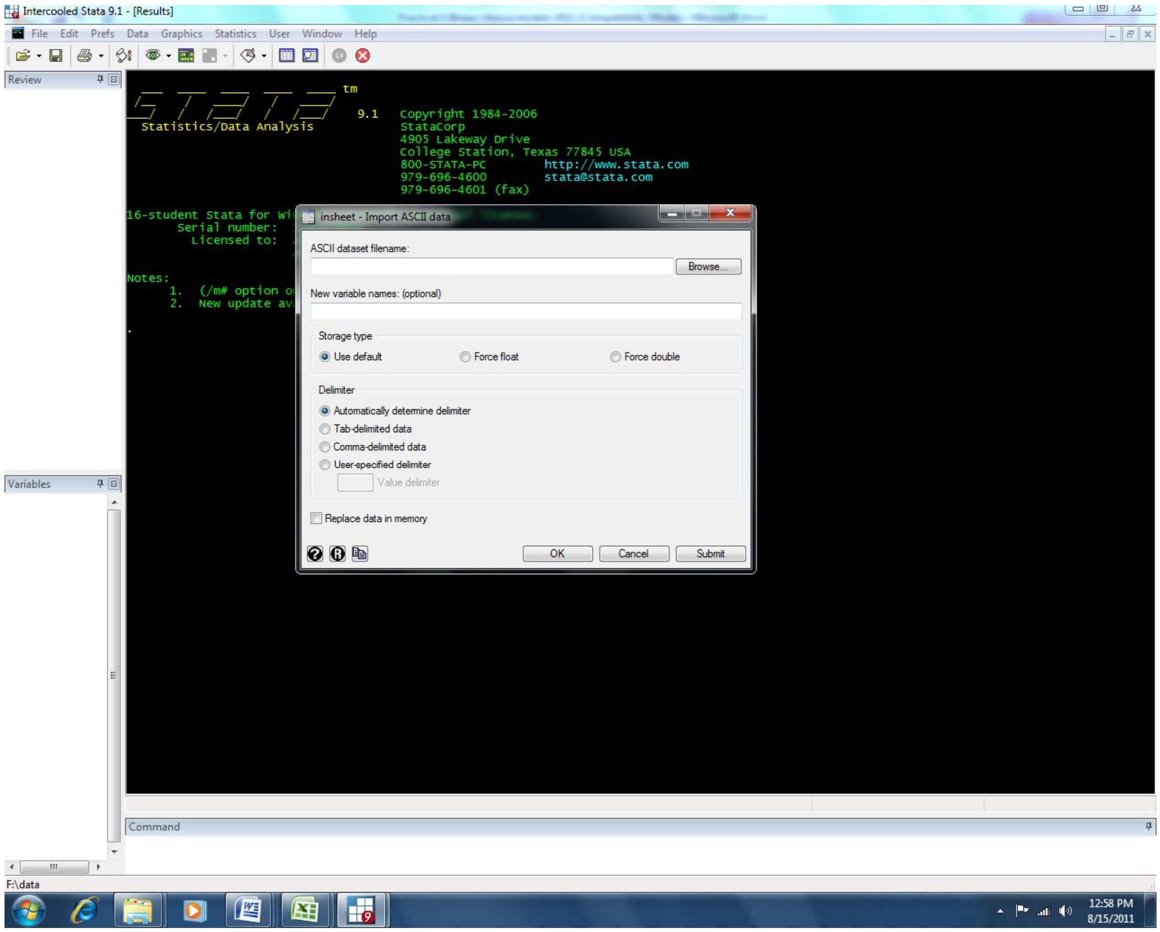
- Open the file
- Safe the data in an ASCII type, for instance the text **text(tab delimited)**



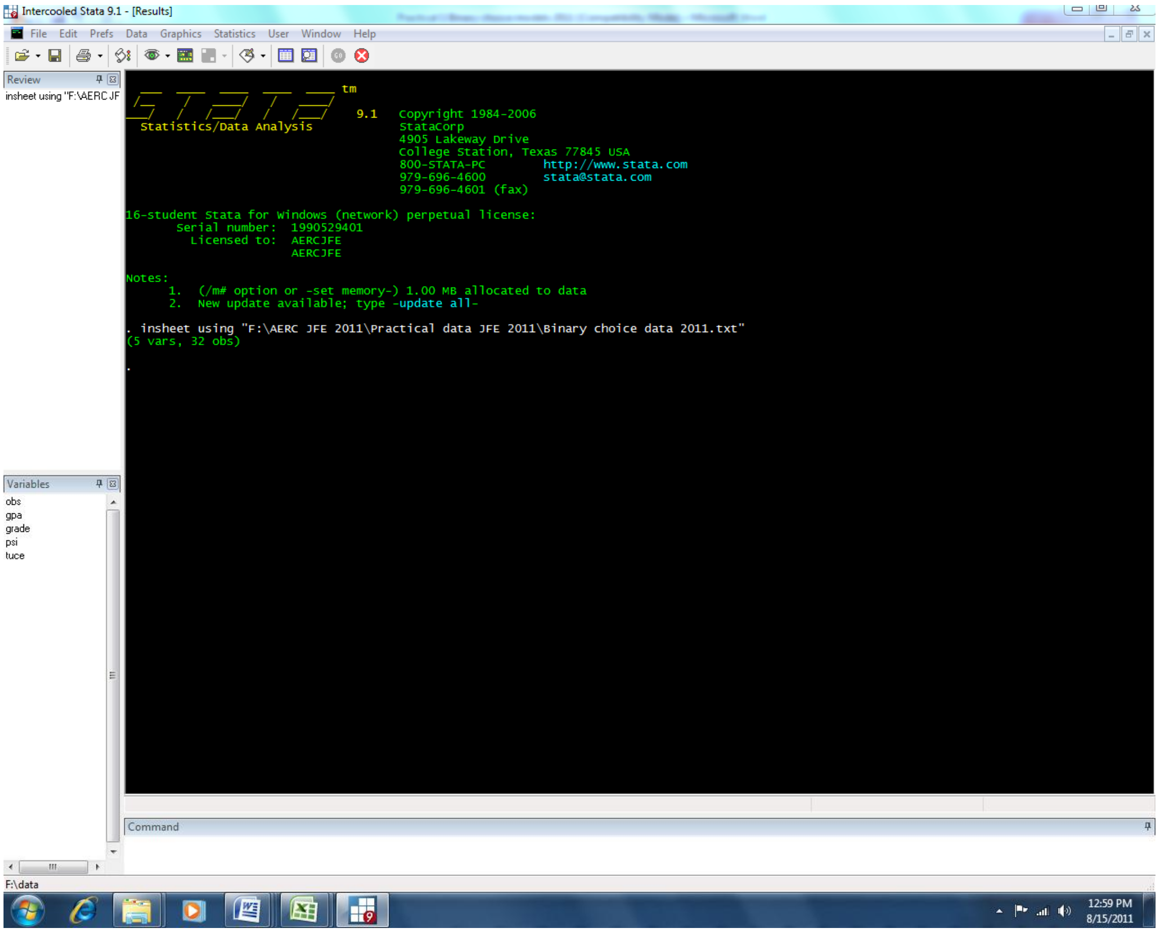
- Close the file in excel
- Start the stata
- Go to **File/import/ASCII data created by a spreadsheet**



- Browse to locate your file location



- Select **text files(*.txt)** in the file type
- Click OK
- The file will be loaded
- If you have done it well, you will see the following



The blue background (yours is black) is simply the setting in my machine. You could change yours by simply right clicking anywhere on the black screen and setting your preferences.

4. IMPLEMENTING LPM IN STATA

LPM is simply an OLS that is applied to Binary choice (response) model (BMR)

4.1 Specification of the LPM model

Specify the model for our data

$$grade_i = \beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i + \varepsilon_i$$

$$i = 1,2,\ldots,32$$

Note the following:

- The dependent variable GRADE is an interval/binary choice variable
- The independent variables GPA, PSI and TUCE are a linear combination
- The independent variables are non-stochastic
- The model is linear
- $E(grade_i) = \beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i + \varepsilon_i$
- $E(\varepsilon_i) = 0$
- We know from basic probability theory that

$$E\left[grade_i \middle| gpa_{i,} psi_i, tuce_i\right] = 1 \cdot \Pr[grade_i = 1] + 0 \cdot \Pr[grade_i = 0] = \Pr[grade_i = 1]$$

$$E(\varepsilon_i) = E[grade_i - E(grade_i)]^2 = E(grade_i^2) - [E(grade_i)]^2 = E(grade_i) - [E(grade_i)]^2$$

Factoring out $E(grade_i)$ we find that

$$E(\varepsilon_i) = E(grade_i)[1 - E(grade_i)] = \Pr[grade_i = 1](1 - \Pr[grade_i = 1])$$

This implies heteroscedasticity

4.2 Estimation of LPM Model

- The LPM uses the moment based estimation methods that do not require an assumption about the probability distribution
- Consequently, the LPM can work with small samples as compare to Logit or Probit models

- LPM estimation in stata is using OLS with standard ***regress*** command. In our case

Before we start letôs create a log file to track all the steps we follow for your review later

In the command space type

Regress GRADE GPA PSI TUCE

The results are

```
regress grade gpa psi tuce

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  3,    28) =    6.65
       Model |  3.00227631     3  1.00075877           Prob > F      =  0.0016
    Residual |  4.21647369    28  .150588346           R-squared     =  0.4159
-------------+------------------------------           Adj R-squared =  0.3533
       Total |     7.21875    31  .232862903           Root MSE      =  .38806

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   .4638517   .1619563     2.86   0.008     .1320992    .7956043
         psi |   .3785548   .1391727     2.72   0.011     .0934724    .6636372
        tuce |   .0104951   .0194829     0.54   0.594    -.0294137    .0504039
       _cons |  -1.498017   .5238886    -2.86   0.008    -2.571154   -.4248801
------------------------------------------------------------------------------
```

4.3 Interpretation of the LPM

- The R^2 and the F-statistic shows that the model fits the data well.

- We can interpret the coefficients in a straightforward manner

$$\frac{\partial grade}{\partial gpa} = 0.46 \, , \frac{\partial grade}{\partial psi} = 0.38 \, , \frac{\partial grade}{\partial tuce} = 0.01$$

- The gpa and psi are statistically significant
- Gpa increases the performance by 0.46
- The new method of teaching, psi, increases the performance by 0.38
- But
- **What do 0.46, 0.38 and 0.01 mean?**
- They are not the actual changes in predicted probability of improvement in grade

- What can we say about our initial research problem with these numbers?
- This is the fundamental problem with the linear probability model

4.4 Other limitations of the LPM

(i) Heteroscedasticity

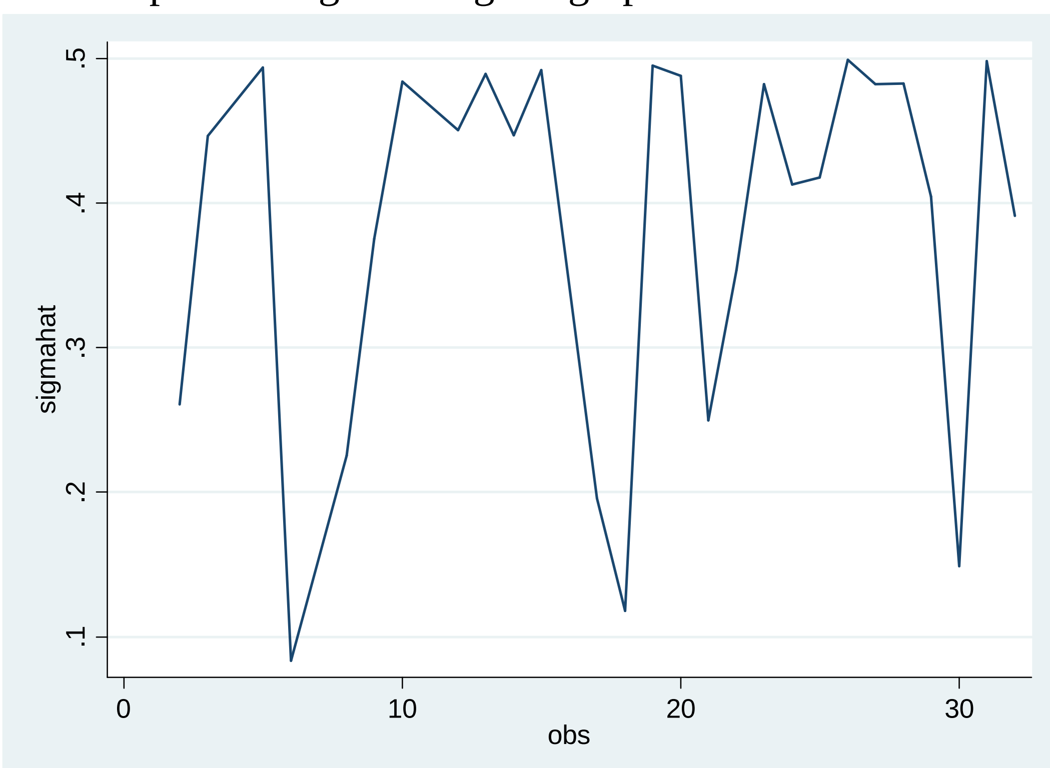
Recall the variance of a Bernoulli distribution

$$\hat{\sigma}_i = \sqrt{\widehat{grade}_i\left(1-\widehat{grade}_i\right)} = \sqrt{\left(\hat{\beta}_0 + \hat{\beta}_1 gpa_i + \hat{\beta}_2 psi_i + \hat{\beta}_3 tuce_i\right)\left(1-\hat{\beta}_0 - \hat{\beta}_1 gpa_i - \hat{\beta}_2 psi_i - \hat{\beta}_3 tuce_i\right)} = \sqrt{\operatorname{var}(\varepsilon_i)}$$

To implement this in stata the commands are the follows

Predict yhat
Generate sigmahat=sqrt(yhat*(1-yhat))

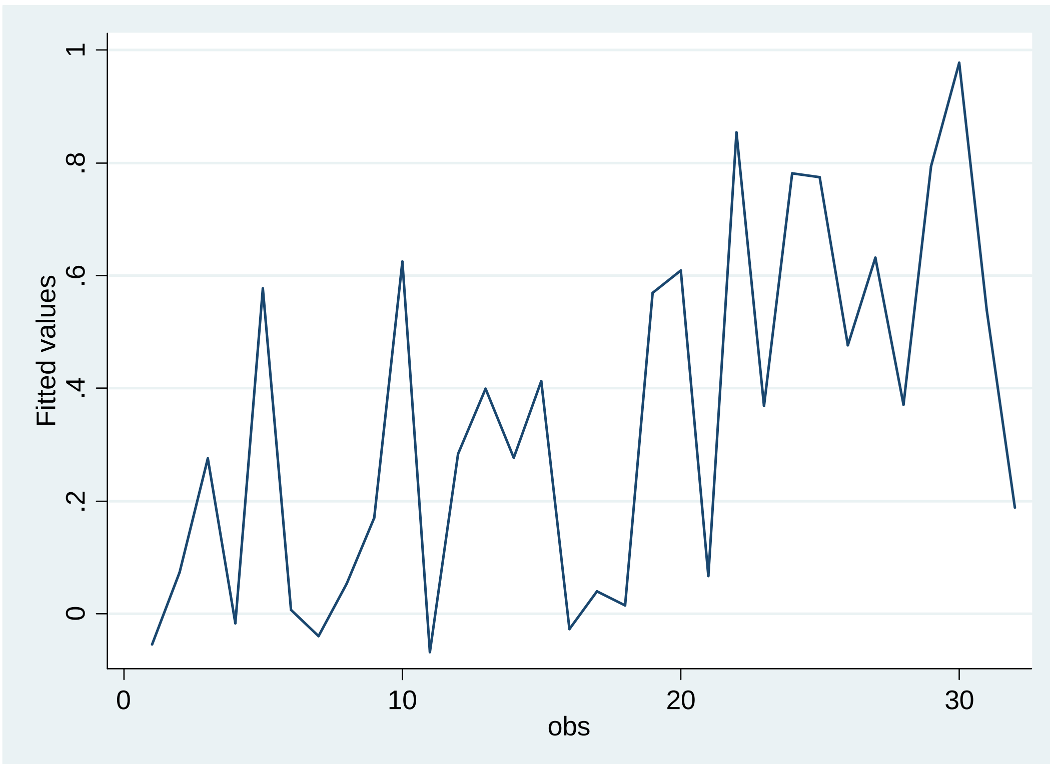
You can plot the sigma using the graphics menu



It is heteroscedastic

(ii) Predicted values outside the 0,1 range

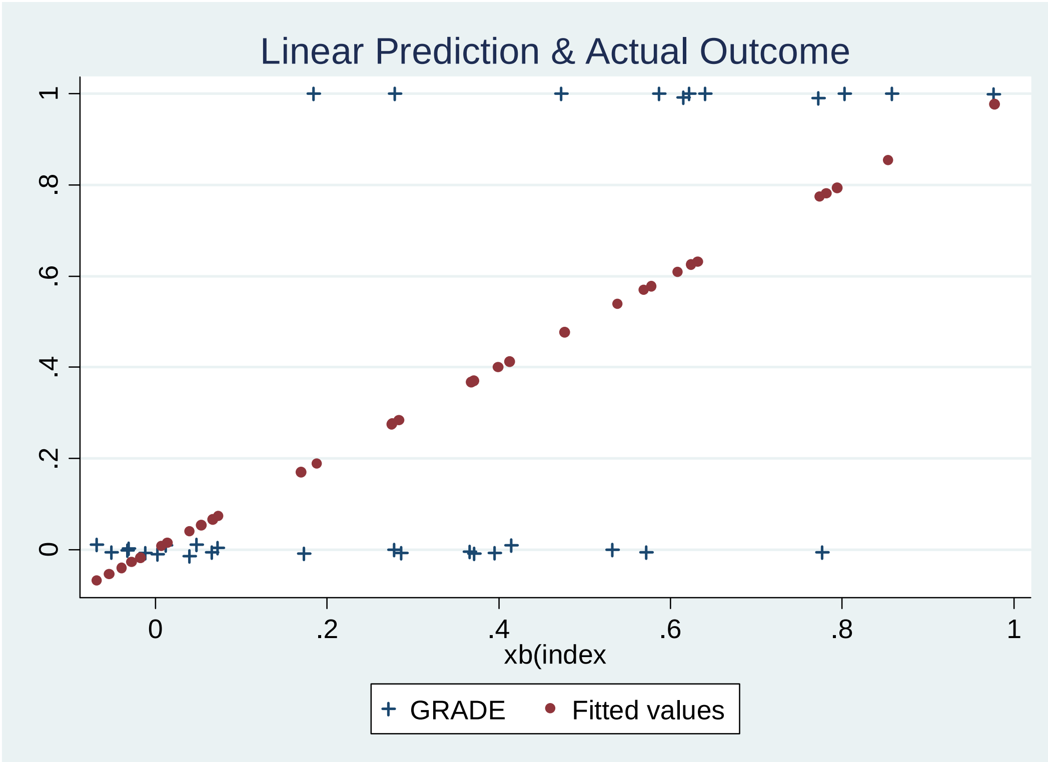
We can plot our **yhat**



You see that there are negative probabilities.

You can also see the same problem with a few steps
Predict xb, xb
Label var xb “xb(index)”
Draw the scatter graph using the command
scatter grade yhat xb, msymbol(+ o) jitter(2) title("Linear Prediction & Actual Outcome")

You will get



What you see is that there are negative probabilities

5. ESTIMATING LOGIT

5.1 Specification of the Logit model

Specify the model for our data

$$\Pr\left[grade_i = 1 \middle| gpa_i, psi_i, tuce_i\right] = \Lambda\left(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i\right) = \frac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i \beta_3 tuce_i}}$$

$$i = 1,2,\ldots,32$$

The probability density function for the logit is

$$\Lambda'\left(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i\right) = \frac{e^{u_i}}{\left(1 + e^{u_i}\right)^2}$$

$$\Pr\left[grade_i = 0 \middle| gpa_i, psi_i, tuce_i\right] = 1 - \frac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i \beta_3 tuce_i}} = \frac{1}{1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}$$

In effect there are **three possible specifications**

(i) $\Pr\left[grade_i = 1 \middle| gpa_i, psi_i, tuce_i\right] = \dfrac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i \beta_3 tuce_i}}$

(ii) $\dfrac{\Pr\left[grade_i = 1\right]}{\Pr\left[grade_i = 0\right]} = e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}$

Which is the odds-ratio

(iii) $\ln\left(\dfrac{\Pr\left[grade_i = 1\right]}{\Pr\left[grade_i = 0\right]}\right) = \beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i$

5.2 Estimation of Logit Model

- The Logit model uses the MLE that requires an assumption about the probability
- Consequently, it requires a large sample to take advantage of the asymptotic properties

In the command space type
logit grade gpa psi tuce

The results are

```
logit grade gpa psi tuce

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.496795
Iteration 2:   log likelihood = -12.929188
Iteration 3:   log likelihood = -12.889941
Iteration 4:   log likelihood = -12.889633
Iteration 5:   log likelihood = -12.889633

Logistic regression                               Number of obs   =         32
                                                  LR chi2(3)      =      15.40
                                                  Prob > chi2     =     0.0015
Log likelihood = -12.889633                       Pseudo R2       =     0.3740
```

```
------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |   2.826113   1.262941     2.24   0.025     .3507938    5.301432
         psi |   2.378688   1.064564     2.23   0.025       .29218    4.465195
        tuce |   .0951577   .1415542     0.67   0.501    -.1822835    .3725988
       _cons |  -13.02135   4.931325    -2.64   0.008    -22.68657    -3.35613
------------------------------------------------------------------------------
```

- These results are similar to the ones in table 17.1 of Greene (2011)

- Letôs substitute the results in the probability formulation of the logit model

$$\Pr\left[grade_i = 1 \middle| gpa_i, psi_i, tuce_i\right] = \frac{e^{-13.0+2.8gpa_i+2.4psi_i+0.09tuce_i}}{1+e^{-13.0+2.8gpa_i+2.4psi_i+0.09tuce_i}}$$

Can we say anything about the probability here? Not much

What about in the log-odds ratio version?
Letôs do it

$$\ln\left(\frac{\Pr[grade_i = 1]}{\Pr[grade_i = 0]}\right) = -13.0 + 2.8gpa_i + 2.4psi_i + 0.09tuce_i$$

We may be able to interpret the probability by arguing that since log is a monotonic transformation then

$$\frac{\partial \ln\left(\frac{\Pr[grade_i = 1]}{\Pr[grade_i = 0]}\right)}{\partial gpa_i} = 2.8$$ i.e. a student with a higher gpa increases the likelihood that such student will record a high performance by 2.8 times

The same can be done for the other variables

5.3 Interpretation of the Logit model

The basic **logit** commands reports coefficient estimates and the underlying standard errors.

- These coefficients are the index coefficients and do not correspond to the average partial effects

$$Logit \quad index = -13.0 + 2.8 gpa_i + 2.4 psi_i + 0.09 tuce_i$$

What we are looking for are the marginal effects

$$\frac{\partial \Pr[grade_i = 1]}{\partial gpa_i} = ?$$
$$\frac{\partial \Pr[grade_i = 1]}{\partial psi_i} = ?$$
$$\frac{\partial \Pr[grade_i = 1]}{\partial tuce_i} = ?$$

- Notice the use of the difference operator for the psi instead of the partial derivative

Theory tells us that

$$\frac{\partial \Pr[grade_i = 1]}{\partial gpa_i} = \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)[1 - \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)]\beta_1$$
$$= \frac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{\left(1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}\right)^2}\beta_1$$

$$\frac{\Delta \Pr[grade_i = 1]}{\Delta psi_i} = \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)[1 - \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)]\beta_2$$
$$= \frac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{\left(1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}\right)^2}\beta_2$$

$$\frac{\partial \Pr[grade_i = 1]}{\partial tuce_i} = \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)[1 - \Lambda(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)]\beta_3$$

$$= \frac{e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}}{\left(1 + e^{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}\right)^2}\beta_3$$

These marginal effects differ for each observation

In stata these marginal effects can be computed using the **mfx** command

```
mfx

Marginal effects after logit
      y  = Pr(grade) (predict)
         =  .25282025
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     gpa |   .5338589      .23704    2.25   0.024   .069273  .998445   3.11719
    psi*|   .4564984      .18105    2.52   0.012    .10164  .811357     .4375
    tuce |   .0179755      .02624    0.69   0.493  -.033448  .069399   21.9375
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- The marginal effects vary depending on the values of the independent variables.
- Consequently, it is appropriate to choose baseline for the independent and dependent variables
- Mean values are often used
- But median is more informative when variables are skewed
- The interpretation of the effects is as follows
- Recall that for one unit increase in the dependent variable from the baseline, the probability of an event is expected to increase/decrease by the magnitude of the marginal change holding other variables constant
- In our case one unit increase in GPA from the baseline mark of 3.11 increases the probability of grade improvement by 53.3%

- One unit increase in the previous knowledge of the material from the baseline (21.93) increases the probability of grade improvement by 1.8 %
- What about the psi?
- Letôs deal with it later

6. ESTIMATING PROBIT

6.1 Specification of the Probit model

Specify the model for our data

$$\Pr\left[grade_i = 1 \middle| gpa_i, psi_i, tuce_i\right] = \Phi\left(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i\right)$$

$$\Phi\left(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i\right) = \int_{-\infty}^{\left(\frac{\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i}{\sigma}\right)} \frac{1}{(2\pi)^{2}} e^{-\frac{t}{2}} dt$$

Notice there is no closed-form expression as in logit model since the function involves integrals

Recall that we can not be able to estimate each of the parameters and the variance separately, Hence we normalise the variance as $\sigma = 1$

6.2 Estimation of Probit Model

- The Probit model uses the MLE that requires an assumption about the probability
- Consequently, it requires a large sample to take advantage of the asymptotic properties

In the command space type

```
probit grade gpa psi tuce


Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.315851
Iteration 2:   log likelihood = -12.832843
Iteration 3:   log likelihood = -12.818826
Iteration 4:   log likelihood = -12.818803

Probit regression                                 Number of obs   =         32
                                                  LR chi2(3)      =      15.55
                                                  Prob > chi2     =     0.0014
Log likelihood = -12.818803                       Pseudo R2       =     0.3775

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         gpa |    1.62581   .6938818     2.34   0.019     .2658269    2.985794
         psi |   1.426332    .595037     2.40   0.017     .2600814    2.592583
        tuce |   .0517289   .0838901     0.62   0.537    -.1126927    .2161506
       _cons |   -7.45232   2.542467    -2.93   0.003    -12.43546   -2.469177
------------------------------------------------------------------------------
```

Letôs substitute the results in the probability formulation of the Probit model model

$$\Pr\left[grade_i = 1\right] = \Phi\left(-7.5 + 1.6\,gpa_i + 1.4\,psi_i + 0.05\,tuce_i\right) = \int_{-\infty}^{\left(\frac{-7.5+1.6\,gpa_i+1.4\,psi_i+0.05\,tuce_i}{\sigma}\right)} \frac{1}{(2\pi)^2} e^{-\frac{t}{2}}\,dt$$

What we can only say is the direction of the effect and partial effects on the Probit index/score

$$\Pr obit \quad index = -7.5 + 1.6gpa_i + 1.4psi_i + 0.05tuce_i$$

$$\frac{\partial \Pr obit \quad index}{\partial gpa_i} = 1.6$$

$$\frac{\partial \Pr obit \quad index}{\partial psi_i} = 1.6$$

$$\frac{\partial \Pr obit \quad index}{\partial tuce_i} = 0.05$$

6.3 Interpretation of the Probit Model

The basic **probit** commands report coefficient estimates and the underlying standard errors.

These coefficients are the index coefficients and do not correspond to the average partial effects

What we are looking for are the marginal effects

$$\frac{\partial \Pr[grade_i = 1]}{\partial gpa_i} = ?$$
$$\frac{\partial \Pr[grade_i = 1]}{\partial psi_i} = ?$$
$$\frac{\partial \Pr[grade_i = 1]}{\partial tuce_i} = ?$$

Theory tells us that for Probit model

$$\frac{\partial \Pr[grade_i = 1]}{\partial gpa_i} = \phi(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)\beta_1$$
$$\frac{\Delta \Pr[grade_i = 1]}{\Delta psi_i} = \phi(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)\beta_2$$
$$\frac{\partial \Pr[grade_i = 1]}{\partial tuce_i} = \phi(\beta_0 + \beta_1 gpa_i + \beta_2 psi_i + \beta_3 tuce_i)\beta_3$$

Notice the use of the difference operator (Δ) **for the discrete changes and not the partial derivative operator** (∂)

In stata these marginal effects can be computed using two methods

- **dprobit** command
- **Mfx** compute command

Lets compute and see what happens

(i) dprobit results

dprobit grade gpa psi tuce

```
. dprobit grade gpa psi tuce

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.315851
Iteration 2:   log likelihood = -12.832843
Iteration 3:   log likelihood = -12.818826
Iteration 4:   log likelihood = -12.818803

Probit regression, reporting marginal effects           Number of obs =     32
                                                        LR chi2(3)    =  15.55
                                                        Prob > chi2   = 0.0014
Log likelihood = -12.818803                             Pseudo R2     = 0.3775

------------------------------------------------------------------------------
   grade |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
     gpa |   .5333471   .2324639     2.34   0.019   3.11719   .077726  .988968
    psi*|    .464426   .1702806     2.40   0.017     .4375   .130682   .79817
    tuce |   .0169697   .0271198     0.62   0.537   21.9375  -.036184  .070123
---------+--------------------------------------------------------------------
  obs. P |     .34375
 pred. P |   .2658081  (at x-bar)
------------------------------------------------------------------------------
(*) dF/dx is for discrete change of dummy variable from 0 to 1
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

(ii) Marginal effects using mfx command

. mfx compute

```
.  mfx compute

Marginal effects after dprobit
      y  = Pr(grade) (predict)
         =  .26580809
------------------------------------------------------------------------------
variable |      dy/dx    Std. Err.     z    P>|z|  [    95% C.I.   ]      X
---------+--------------------------------------------------------------------
     gpa |   .5333471      .23246    2.29   0.022   .077726  .988968   3.11719
    psi*|    .464426      .17028    2.73   0.006   .130682   .79817     .4375
    tuce |   .0169697      .02712    0.63   0.531  -.036184  .070123   21.9375
------------------------------------------------------------------------------
(*) dy/dx is for discrete change of dummy variable from 0 to 1
```

- There are no major differences in the results except in the z score.
- The mfx uses X represent mean values whereas the dprobit uses x-bar

- The marginal effects for both models are essentially the same
- The probability at the mean values is 0.26580809
- The interpretations are the same as the one for Logit model

7. DISCRETE CHANGE VS MARGINAL EFFECTS

- Discrete changes are important in two respects

(i) Dummy variables

(ii) When we wish to focus on predicted probability changes for a particular range of independent variables

Discrete changes are computed as follows

$$\frac{\Delta\Lambda\left[Logit \quad index\right]}{\Delta X_k} = \Pr\left[grade_i = 1 \middle| X, X_k + \delta\right] - \Pr\left[grade_i = 1 \middle| X, X_k\right]$$

Notice the use of X vector for independent variables

- The discrete changes can be interpreted as follows
- For a change in X_k from X_k to $X_k + \delta$
- Notice that we use **discrete changes** for psi and not **marginal changes.**
- This is interpreted that the probability of grade improvement is expected to change by the magnitude of the indicated changes, holding all other variables at the given levels
- For our psi, the movement is from 0 (not exposed to the personalised system of instruction to 1 (exposed to personalised system of instruction)
- That is why you see the star (*) in the results to warn you that we are dealing with discrete changes
- The interpretation is that a student that exposed to the new method has a probability of grade improvement of 0.46 greater than another student who is not exposed to the same method

This applies to the logit model results as well

8. TESTING OF HYPOTHESIS

8.1 Test the Significance of the covariates

We could test the significance of each of the variables as well as the joint significance

Single covariate

test gpa

```
 ( 1)  gpa = 0

          chi2(  1) =    5.49
        Prob > chi2 =    0.0191
```

test psi

```
 ( 1)  psi = 0

          chi2(  1) =    5.75
        Prob > chi2 =    0.0165
```

test tuce

```
 ( 1)  tuce = 0

          chi2(  1) =    0.38
        Prob > chi2 =    0.5375
```

More than one variable (Wald test)

test gpa tuce psi

```
 ( 1)  gpa = 0
 ( 2)  tuce = 0
 ( 3)  psi = 0

          chi2(  3) =   10.39
        Prob > chi2 =    0.0155
```

Likelihood Ratio test

Estimate unrestricted equation

logit grade gpa psi tuce

Store the model parameters

Est store A

Restricted equation
logit grade gpa tuce

Store the results
est store B

Perform likelihood ratio test
lrtest A B, stats

```
lrtest A B, stats

Likelihood-ratio test                                 LR chi2(1)  =      6.20
(Assumption: B nested in A)                           Prob > chi2 =    0.0127

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           B |     32   -20.59173   -15.99148      3     37.98296    42.38017
           A |     32   -20.59173   -12.88963      4     33.77927    39.64221
-----------------------------------------------------------------------------
```

Letôs test for tuce, which is already statistically insignificant
We just need to change the restricted model as

logit grade gpa psi
est store B
lrtest A B, stats

```
lrtest A B, stats

Likelihood-ratio test                                 LR chi2(1)  =      0.47
(Assumption: B nested in A)                           Prob > chi2 =    0.4912

-----------------------------------------------------------------------------
       Model |    Obs    ll(null)   ll(model)     df          AIC         BIC
-------------+---------------------------------------------------------------
           B |     32   -20.59173   -13.12657      3     32.25315    36.65035
           A |     32   -20.59173   -12.88963      4     33.77927    39.64221
-----------------------------------------------------------------------------
```

As you can see this is not statistically significant

8.2 Heteroscedastic probit model

- Heteroscedasticity is an important statistical problem to deal with
- One way of dealing with it from a probit perspective is to relax the assumption that the error term is homoscedastic, by writing the variance of the error term as $[\exp(gx)]^2$ where x is any of the covariates.
- In this case g is a parameter to be estimated (note: if g=0 we have homoscedasticity)
- We can get this in stata using the **hetprob command**
- For instance if we would like to know if the variance of hte error term falls or rises with psi i.e.responsible for the heteroscedasticity, we can use the following command

hetprob grade gpa psi tuce,het(psi)

```
hetprob grade gpa psi tuce,het(psi)

Fitting probit model:

Iteration 0:   log likelihood =  -20.59173
Iteration 1:   log likelihood = -13.315851
Iteration 2:   log likelihood = -12.832843
Iteration 3:   log likelihood = -12.818826
Iteration 4:   log likelihood = -12.818803

Fitting full model:

Iteration 0:   log likelihood = -12.818803
Iteration 1:   log likelihood = -12.080094
Iteration 2:   log likelihood = -11.965838
Iteration 3:   log likelihood = -11.896545
Iteration 4:   log likelihood = -11.895852
Iteration 5:   log likelihood = -11.895851

Heteroskedastic probit model                      Number of obs     =         32
                                                  Zero outcomes     =         21
                                                  Nonzero outcomes  =         11

                                                  Wald chi2(3)      =       3.33
Log likelihood = -11.89585                        Prob > chi2       =     0.3438

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
grade        |
         gpa |    3.12155   1.760869     1.77   0.076    -.3296887    6.572789
         psi |    2.34322   1.670631     1.40   0.161    -.9311565    5.617597
```

```
        tuce |   .1237515   .2134227     0.58   0.562    -.2945493    .5420523
       _cons |  -14.28904   8.860899    -1.61   0.107    -31.65609    3.077997
-------------+----------------------------------------------------------------
lnsigma2     |
         psi |   1.093371   .8805796     1.24   0.214    -.6325333    2.819275
------------------------------------------------------------------------------
Likelihood-ratio test of lnsigma2=0: chi2(1) =     1.85   Prob > chi2 = 0.1743
```

.

- Clearly there is no evidence here that the variance of the error term rises with psi
- This is because it is not statically significant
- If it is significant, one should consider adding a squared psi to the model and check if the squared term is significant or not