ADVANCED ECONOMETRICS I

Practice Exercises (2/2)

Instructor:	Joaquim J. S. Ramalho
E.mail:	jjsro@iscte-iul.pt
Personal Website:	http://home.iscte-iul.pt/~jjsro
Office:	D5.10
Course Website:	https://jjsramalho.wixsite.com/advecoi
Fénix:	https://fenix.iscte-iul.pt/disciplinas/03089

Illustration 5

Aims:

- Study the determinants of the choice between 2 tomato ketchup brands: Heinz and Hunts
- In particular, investigate whether the promotional activities developed by both brands have any impact on the probability of consumers choosing one instead of the other

Sample:

- 2798 purchases made during a period of about two years in Springfield, Missouri
- Purchase data were collected automatically using an optical scanner; the shelf price of the other brand was also stored
- It was also registered if at the time of purchase the brands had ongoing promotional activities

Illustration 5

Promotional activities:

- Display only
- Newspaper feature only
- Both

Model:

$$Pr(Heinz = 1| \dots) = G\left[\beta_0 + \beta_1 Dhei + \beta_2 Fhei + \beta_3 DFhei + \beta_4 Dhun + \beta_5 Fhun + \beta_6 DFhun + \beta_7 log\left(\frac{Phei}{Phun}\right)\right]$$

Details:

- Franses and Paap (2001), ch. 4.4

Illustration 5 – Question 1

```
. describe

--------------------------------------------------------------------------------
                  storage  display     value
variable name      type    format      label      variable label
--------------------------------------------------------------------------------
Family             int      %8.0g                 Family id
Heinz              byte     %8.0g                 =1 Heinz chosen
Hunts              byte     %8.0g                 =1 Hunts chosen
Phei               float    %9.0g                 Heinz price (US$/oz)
Phun               float    %9.0g                 Hunts price (US$/oz)
Dhei               byte     %8.0g                 =1 if Heinz was on display
                                                    but not featured
Dhun               byte     %8.0g                 =1 if Hunts was on display
                                                    but not featured
Fhei               byte     %8.0g                 =1 if Heinz was featured but
                                                    not on display
Fhun               byte     %8.0g                 =1 if Hunts was featured but
                                                    not on display
DFhei              byte     %8.0g                 =1 if Heinz was on display
                                                    and featured
DFhun              byte     %8.0g                 =1 if Hunts was on display
                                                    and featured
--------------------------------------------------------------------------------
```

Illustration 5 – Question 2

```
. summarize Heinz Hunts Dhei Dhun Fhei Fhun DFhei DFhun Phei Phun
```

Variable	Obs	Mean	Std. Dev.	Min	Max
Heinz	2798	.8902788	.3125978	0	1
Hunts	2798	.1097212	.3125978	0	1
Dhei	2798	.159757	.366446	0	1
Dhun	2798	.0353824	.1847774	0	1
Fhei	2798	.124732	.3304738	0	1
Fhun	2798	.0364546	.1874519	0	1
DFhei	2798	.0375268	.1900828	0	1
DFhun	2798	.0092924	.0959651	0	1
Phei	2798	.0348276	.0089737	.001	.061
Phun	2798	.0335547	.0053069	.003	.087

Illustration 5 – Question 3

```
. gen lpp=ln(Phei/Phun)
. logit Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp

Logistic regression                             Number of obs   =       2798
                                                LR chi2(7)      =     584.96
                                                Prob > chi2     =     0.0000
Log likelihood = -675.43973                     Pseudo R2       =     0.3022

------------------------------------------------------------------------------
       Heinz |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Dhei |   .5474406   .2427129     2.26   0.024      .071732    1.023149
        Dhun |  -.6207141   .2449583    -2.53   0.011    -1.100824   -.1406047
        Fhei |   .5785839   .3165889     1.83   0.068     -.041919    1.199087
        Fhun |  -.9971683   .3466031    -2.88   0.004    -1.676498   -.3178388
       DFhei |   .4452738   .4431288     1.00   0.315    -.4232427     1.31379
       DFhun |  -1.940127   .4802308    -4.04   0.000    -2.881362   -.9988919
         lpp |  -6.110695   .3858905   -15.84   0.000    -6.867026   -5.354363
       _cons |    3.27461   .1423418    23.01   0.000     2.995625    3.553595
------------------------------------------------------------------------------

. estimates store logit
```

Illustration 5 – Question 3 (cont.)

```
. probit Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp

Probit regression                          Number of obs   =       2798
                                           LR chi2(7)      =     591.39
                                           Prob > chi2     =     0.0000
Log likelihood = -672.22516                Pseudo R2       =     0.3055

------------------------------------------------------------------------------
       Heinz |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Dhei |   .2768603   .1226386     2.26   0.024     .0364931    .5172275
        Dhun |  -.3599038   .1455494    -2.47   0.013    -.6451753   -.0746322
        Fhei |   .2396884   .1549172     1.55   0.122    -.0639437    .5433206
        Fhun |  -.5526825   .1904251    -2.90   0.004    -.9259089   -.1794561
       DFhei |   .2486314   .2282668     1.09   0.276    -.1987634    .6960262
       DFhun |  -1.068493      .2749    -3.89   0.000    -1.607287   -.5296985
         lpp |  -3.348536   .2094459   -15.99   0.000    -3.759042   -2.938029
       _cons |   1.840782   .0723702    25.44   0.000     1.698939    1.982625
------------------------------------------------------------------------------

. estimates store probit
```

Illustration 5 – Question 3 (cont.)

```
. cloglog Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp

Complementary log-log regression                Number of obs     =      2798
                                                Zero outcomes     =       307
                                                Nonzero outcomes  =      2491

                                                LR chi2(7)        =    589.99
Log likelihood = -672.92517                     Prob > chi2       =    0.0000

------------------------------------------------------------------------------
       Heinz |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Dhei |   .2088229   .0936429     2.23   0.026     .0252863    .3923596
        Dhun |  -.3518195   .1501518    -2.34   0.019    -.6461116   -.0575275
        Fhei |   .1227095   .1145124     1.07   0.284    -.1017307    .3471497
        Fhun |  -.4820495    .164347    -2.93   0.003    -.8041638   -.1599353
       DFhei |    .209065   .1765308     1.18   0.236     -.136929    .5550589
       DFhun |  -1.077127   .3208661    -3.36   0.001    -1.706013    -.448241
         lpp |  -2.802396   .1883769   -14.88   0.000    -3.171608   -2.433184
       _cons |   1.270849   .0566031    22.45   0.000     1.159909    1.381789
------------------------------------------------------------------------------

. estimates store cloglog
```

Illustration 5 – Question 3 (cont.)

```
. estimates table logit probit cloglog, b star(0.1 0.05 0.01)

--------------------------------------------------------------
    Variable |     logit          probit          cloglog
-------------+------------------------------------------------
        Dhei |  .54744056**     .27686031**     .20882294**
        Dhun | -.62071414**    -.35990375**    -.35181954**
        Fhei |  .57858393*      .23968844       .12270948
        Fhun | -.99716835***   -.55268247***   -.48204954***
       DFhei |   .4452738       .24863139       .20906496
       DFhun | -1.9401271***   -1.0684927***   -1.0771271***
         lpp | -6.1106949***   -3.3485359***   -2.8023956***
       _cons |  3.2746098***    1.8407819***    1.2708491***
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

Illustration 5 – Question 3 (cont.)

Main conclusions:

- There is no clear evidence on whether the promotional activities undertaken by Heinz increase significantly the probability of consumers purchasing its ketchup
- The promotional activities made by Hunts decrease significantly the probability of consumers purchasing Heinz ketchup $\longrightarrow$ clearly, because Hunts is a smaller and less known brand, it benefits a lot from promotional activities, especially when display and feature activities are held at the same time
- Increasing the relative price of Heinz ketchup decreases the probability of consumers purchasing its ketchup

Illustration 5 – Question 4

```
. estimates restore logit
. predict XBl, xb
. gen XBl2=XBl^2
. quietly logit Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp XBl2
. test XBl2
 ( 1)  [Heinz]XBl2 = 0

           chi2(  1) =    9.35
         Prob > chi2 =    0.0022
```

The null hypothesis of a well-specified functional form is rejected

```
. estimates restore probit
. predict XBp, xb
. gen XBp2=XBp^2
. quietly probit Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp XBp2
. test XBp2
 ( 1)  [Heinz]XBp2 = 0

           chi2(  1) =    0.73
         Prob > chi2 =    0.3933
```

The null hypothesis of a well-specified functional form cannot be rejected

Illustration 5 – Question 4 (cont.)

```
. estimates restore cloglog
. predict XBcl, xb
. gen XBcl2=XBcl^2
. quietly cloglog Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp XBcl2
. test XBcl2
 ( 1)  [Heinz]XBcl2 = 0

          chi2(  1) =     9.63
        Prob > chi2 =    0.0019
```

The null hypothesis of a well-specified functional form is rejected

Illustration 5 – Question 5.1

```
. quietly probit Heinz Dhei Dhun Fhei Fhun DFhei DFhun lpp XBp2

. estimates store probitR

. lrtest probit probitR

Likelihood-ratio test                                 LR chi2(1)  =      0.73
(Assumption: probit nested in probitR)                Prob > chi2 =    0.3922
```

The null hypothesis of a well-specified functional form is not rejected

Illustration 5 – Question 5.2

```
. estimates restore probit
(results probit are active now)

. estat classification

Probit model for Heinz

              -------- True --------
Classified |         D            ~D  |      Total
-----------+--------------------------+-----------
     +     |      2462           232  |       2694
     -     |        29            75  |        104
-----------+--------------------------+-----------
   Total   |      2491           307  |       2798

Classified + if predicted Pr(D) >= .5
True D defined as Heinz != 0
```

(continues on the next slide)

Illustration 5 – Question 5.2 (cont.)

```
--------------------------------------------------
Sensitivity                     Pr( +| D)   98.84%   → % 1's correctly predicted
Specificity                     Pr( -|~D)   24.43%   → % 0's correctly predicted
Positive predictive value       Pr( D| +)   91.39%
Negative predictive value       Pr(~D| -)   72.12%
--------------------------------------------------
False + rate for true ~D        Pr( +|~D)   75.57%
False - rate for true D         Pr( -| D)    1.16%
False + rate for classified +   Pr(~D| +)    8.61%
False - rate for classified -   Pr( D| -)   27.88%
--------------------------------------------------
Correctly classified                        90.67%   → % correct predictions
--------------------------------------------------
```

Illustration 5 – Question 6.1

```
. summarize Phei Phun

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        Phei |      2798    .0348276    .0089737       .001       .061
        Phun |      2798    .0335547    .0053069       .003       .087

. scalar lppm=log(0.0348276/0.0335547)

. display normal(_b[_cons]+_b[lpp]*lppm)
.9569286

. display normal(_b[_cons]+_b[DFhei]+_b[lpp]*lppm)
.97527763

. display normal(_b[_cons]+_b[DFhun]+_b[lpp]*lppm)
.74138217

. display normal(_b[_cons]+_b[DFhei]+_b[DFhun]+_b[lpp]*lppm)
.81493872
```

Illustration 5 – Question 6.1 (cont.)

	I	II	III	IV
Dhei	0	0	0	0
Fhei	0	0	0	0
DFhei	0	1	0	1
Dhun	0	0	0	0
Fhun	0	0	0	0
DFhun	0	0	1	1
Phei	3.48	3.48	3.48	3.48
Phun	3.36	3.36	3.36	3.36
$Pr(Heinz = 1\| \dots)$	95.69	97.53	74.14	81.49

Illustration 5 – Question 6.2

```
. margins, dydx(_all)

Average marginal effects                          Number of obs    =       2798
Model VCE     : OIM

Expression    : Pr(Heinz), predict()
dy/dx w.r.t. : Dhei Dhun Fhei Fhun DFhei DFhun lpp

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        Dhei |   .0363993   .0161128     2.26   0.024     .0048189    .0679798
        Dhun |  -.0473172   .0190541    -2.48   0.013    -.0846626   -.0099719
        Fhei |   .0315123    .020376     1.55   0.122     -.008424    .0714486
        Fhun |  -.0726622    .024952    -2.91   0.004    -.1215672   -.0237572
       DFhei |    .032688   .0299977     1.09   0.276    -.0261065    .0914825
       DFhun |  -.1404768   .0358536    -3.92   0.000    -.2107485    -.070205
         lpp |  -.4402384    .024884   -17.69   0.000      -.48901   -.3914667
------------------------------------------------------------------------------
```

Illustration 5 – Question 6.3

```
. margins, dydx(DFhun) at(Dhei=0 Fhei=0 DFhei=0 Dhun=0 Fhun=0 DFhun=0 lpp=0)

Conditional marginal effects                      Number of obs     =       2798
Model VCE     : OIM

Expression    : Pr(Heinz), predict()
dy/dx w.r.t.  : DFhun
at            : Dhei            =           0
                Dhun            =           0
                Fhei            =           0
                Fhun            =           0
                DFhei           =           0
                DFhun           =           0
                lpp             =           0

------------------------------------------------------------------------------
             |            Delta-method
             |      dy/dx   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       DFhun |  -.0783219   .0220713    -3.55   0.000    -.1215808   -.0350629
------------------------------------------------------------------------------
```

Illustration 5 – Question 6.4

```
. gen P1=normal(_b[_cons]+_b[lpp]*lpp)

. gen P2=normal(_b[_cons]+_b[DFhei]+_b[lpp]*lpp)

. gen P3=normal(_b[_cons]+_b[DFhun]+_b[lpp]*lpp)

. line P1 P2 P3 lpp if lpp > -0.6 & lpp < 0.6, lpattern(solid dash dot)
```

Illustration 5 – Question 6.4 (cont.)

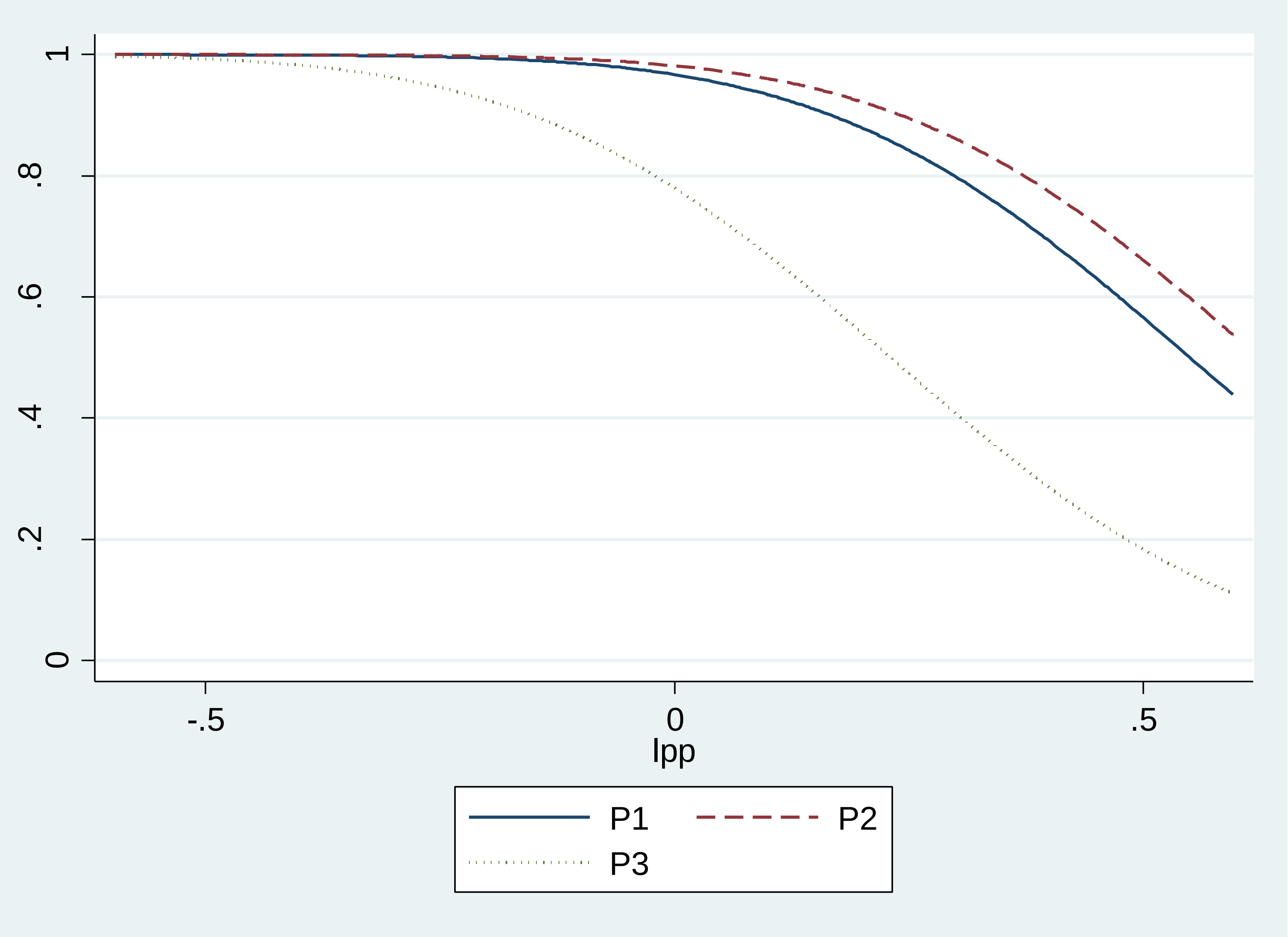

Illustration 5 – Question 6.4 (cont.)

Main conclusions:

- The higher the price of Heinz relative to Hunts, the less the probability of Heinz being purchased (and vice-versa)
- Heinz promotional activities do not allow this brand to substantially increase its price or market share
- Hunts promotional activities are more effective, with two possible effects:
  - If Hunts does not change its price, the probability of consumers purchasing its ketchup increases substantially
  - If Hunts opts for keeping its market share, it is able to sell the same quantity but at a higher price

Illustration 6

Aim:

- Studying the probability of an individual with health insurance have at least one contact with a medical doctor per year

Model:

$$Pr(dmdu = 1| \dots) \\ = G(\beta_0 + \beta_1 lcoins + \beta_2 ndisease + \beta_3 female \\ + \beta_4 age + \beta_5 lfam + \beta_6 child + \alpha_i)$$

Details:

- Cameron and Trivedi (2010), ch. 18.4

Illustration 6 – Question 1.1

```
. xtset id year
       panel variable:  id (unbalanced)
        time variable:  year, 1 to 5, but with gaps
                delta:  1 unit
. xtdescribe
      id:  125024, 125025, ..., 632167                          n =       5908
    year:  1, 2, ..., 5                                         T =          5
           Delta(year) = 1 unit
           Span(year)  = 5 periods
           (id*year uniquely identifies each observation)

Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       3         3         5       5       5
```

(continue in the next slide)

Illustration 6 – Question 1.1 (cont.)

Freq.	Percent	Cum.	Pattern
3710	62.80	62.80	111..
1584	26.81	89.61	11111
156	2.64	92.25	1....
147	2.49	94.74	11...
79	1.34	96.07	..1..
66	1.12	97.19	.11..
33	0.56	97.75	..111
33	0.56	98.31	.1111
29	0.49	98.80	...11
71	1.20	100.00	(other patterns)
5908	100.00		XXXXX

Illustration 6 – Question 1.1 (cont.)

```
. describe
  obs:        20,186
 vars:            13
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
coins           float  %9.0g                  coinsurance -- medical
year            float  %9.0g                  study year
id              float  %9.0g                  person id, leading digit is sit
age             float  %9.0g                  age that year
female          float  %9.0g                  =1 if female
med             float  %9.0g                  medical expenditures (dollars)
mdu             float  %9.0g                  number face-to-face md visits
num             float  %9.0g                  family size
ndisease        float  %9.0g                  count of chronic diseases
child           float  %9.0g                  =1 if child
lfam            float  %9.0g                  log of family size
lcoins          float  %9.0g                  log(coinsurance+1)
dmdu            float  %9.0g                  = 1 if mdu > 0
--------------------------------------------------------------------------------
```

Illustration 6 – Question 1.2

```
. xtsum lcoins ndisease female age lfam child
Variable         |      Mean   Std. Dev.       Min        Max |    Observations
-----------------+--------------------------------------------+----------------
lcoins   overall |  2.383588   2.041713          0   4.564348 |     N =   20186
         between |             2.045796          0   4.564348 |     n =    5908
         within  |                    0   2.383588   2.383588 | T-bar = 3.41672
ndisease overall |   11.2445   6.741647          0       58.6 |     N =   20186
         between |             6.712985          0       58.6 |     n =    5908
         within  |                    0    11.2445    11.2445 | T-bar = 3.41672
female   overall |  .5169424   .4997252          0          1 |     N =   20186
         between |             .4997557          0          1 |     n =    5908
         within  |                    0   .5169424   .5169424 | T-bar = 3.41672
age      overall |  25.71844   16.76759          0   64.27515 |     N =   20186
         between |             16.97265          0   63.27515 |     n =    5908
         within  |             1.086687   23.46844   27.96844 | T-bar = 3.41672
lfam     overall |  1.248404   .5390681          0   2.639057 |     N =   20186
         between |             .5372082          0   2.639057 |     n =    5908
         within  |             .0730824   .3242075    2.44291 | T-bar = 3.41672
child    overall |  .4014168   .4901972          0          1 |     N =   20186
         between |             .4820984          0          1 |     n =    5908
         within  |             .1096116  -.3985832   1.201417 | T-bar = 3.41672
```

Illustration 6 – Question 1.3

```
. xttab dmdu

                  Overall             Between            Within
    dmdu |    Freq.  Percent      Freq.  Percent        Percent
----------+-----------------------------------------------------
       0 |     6308    31.25      3193    54.05          58.61
       1 |    13878    68.75      5164    87.41          78.17
----------+-----------------------------------------------------
   Total |    20186   100.00      8357   141.45          70.70
                               (n = 5908)
```

- Overall, 68,75% of observations concern cases where occurred at least one annual contact with a medical doctor
- 87,41% of individuals (5164 out of 5908) had a contact with a medical doctor in at least one year
- 78,17% of those 5164 individuals had a contact with a medical doctor in every year

Illustration 6 – Question 2

```
. logit dmdu lcoins ndisease female age lfam child, vce(cluster id)
(…)

Logistic regression                             Number of obs   =      20186
                                                Wald chi2(6)    =     488.18
                                                Prob > chi2     =     0.0000
Log pseudolikelihood = -11973.392               Pseudo R2       =     0.0450

                                  (Std. Err. adjusted for 5908 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.1572107   .0109064   -14.41   0.000    -.1785869   -.1358345
    ndisease |    .050301   .0039657    12.68   0.000     .0425285    .0580735
      female |   .3091573   .0445772     6.94   0.000     .2217876     .396527
         age |   .0042689   .0022307     1.91   0.056    -.0001032     .008641
        lfam |  -.2047573   .0470287    -4.35   0.000    -.2969317   -.1125828
       child |   .0921709   .0728107     1.27   0.206    -.0505355    .2348773
       _cons |   .6039411   .1107712     5.45   0.000     .3868335    .8210486
------------------------------------------------------------------------------

. estimates store POOLED
```

Illustration 6 – Question 2 (cont.)

```
. xtlogit dmdu lcoins ndisease female age lfam child
Iteration 3:   log likelihood = -10878.687
(…)
                                                Wald chi2(6)       =    549.76
Log likelihood  = -10878.687                    Prob > chi2        =    0.0000

------------------------------------------------------------------------------
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.2403864   .0162836   -14.76   0.000    -.2723017    -.208471
    ndisease |    .078151   .0055456    14.09   0.000     .0672819    .0890201
      female |   .4631005   .0663209     6.98   0.000     .3331138    .5930871
         age |   .0073441   .0031508     2.33   0.020     .0011687    .0135194
        lfam |  -.3021841   .0644721    -4.69   0.000    -.4285471    -.175821
       child |   .1935357   .1002267     1.93   0.053     -.002905    .3899763
       _cons |   .8629898   .1568968     5.50   0.000     .5554778    1.170502
-------------+----------------------------------------------------------------
    /lnsig2u |   1.225652   .0490898                      1.129438    1.321866
-------------+----------------------------------------------------------------
     sigma_u |    1.84564    .045301                      1.758953    1.936599
         rho |   .5087003   .0122687                      .4846525     .532708
------------------------------------------------------------------------------
Likelihood-ratio test of rho=0: chibar2(01) =  2189.41 Prob >= chibar2 = 0.000

. estimates store RE
```

Illustration 6 – Question 2 (cont.)

```
. xtlogit dmdu lcoins ndisease female age lfam child, fe
note: 3459 groups (11161 obs) dropped because of all positive or
      all negative outcomes.
(…)
Conditional fixed-effects logistic regression   Number of obs      =      9025
Group variable: id                              Number of groups   =      2449

                                                Obs per group: min =         2
                                                               avg =       3.7
                                                               max =         5

                                                LR chi2(3)         =     10.74
Log likelihood  = -3395.5996                    Prob > chi2        =    0.0132

------------------------------------------------------------------------------
        dmdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |          0  (omitted)
    ndisease |          0  (omitted)
      female |          0  (omitted)
         age |  -.0341815   .0183827    -1.86   0.063     -.070211     .001848
        lfam |    .478755   .2597327     1.84   0.065    -.0303116    .9878217
       child |    .270458   .1684974     1.61   0.108    -.0597907    .6007068
------------------------------------------------------------------------------
. estimates store FE
```

Illustration 6 – Question 2 (cont.)

```
. estimates table POOLED RE FE, b star(0.1 0.05 0.01)

--------------------------------------------------------------
    Variable |    POOLED            RE               FE
-------------+------------------------------------------------
dmdu         |
      lcoins | -.15721065***   -.24038638***    (omitted)
    ndisease |  .05030099***    .07815098***    (omitted)
      female |  .30915732***    .46310045***    (omitted)
         age |   .0042689*      .00734407**    -.03418147*
        lfam | -.20475725***   -.30218405***    .47875505*
       child |  .09217092       .19353567*      .27045805
       _cons |  .60394105***    .86298976***
-------------+------------------------------------------------
lnsig2u      |
       _cons |                  1.2256522***
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

Illustration 7

Sample:

- Credit ratings assigned by Standard and Poor's to 921 North-American firms in 2005
- Two rating scales:
  - 1 (D – lowest rating) to 7 (AAA – highest rating)
  - 'investment grade' (rating 4 / BBB or superior) and 'speculative grade': (rating 3 / BB or inferior)

Aim:

- Explaining firm's credit rating as a function of a set of explanatory variables

Details:

- Verbeek (2008), ch. 7.2.3

Illustration 7 – Question 1

```
. describe

  obs:             921
 vars:               7
 size:          25,788
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
booklev         float  %9.0g                  Book leverage
ebit            float  %9.0g                  EBIT / Total assets
invgrade        float  %9.0g                  Investment grade
logsales        float  %9.0g                  Log(Sales)
rating          float  %9.0g                  Credit rating
reta            float  %9.0g                  Retained earnings / Total assets
wka             float  %9.0g                  Working capital / Total assets
--------------------------------------------------------------------------------
```

Illustration 7 – Question 2

```
. summarize

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
    booklev |        921    .2931868    .1735339          0   .9992067
        ebit |        921    .0938921    .0843643  -.3841692   .6515085
    invgrade |        921    .4723127    .4995041          0          1
    logsales |        921    7.995754    1.497413   1.100278   12.70142
      rating |        921    3.499457    1.134561          1          7
-------------+---------------------------------------------------------
        reta |        921    .1569942     .300684  -.9958922   .9799219
         wka |        921    .1404142    .1503398  -.4120839   .7480223
```

Illustration 7 – Question 3.1

```
. logit invgrade booklev ebit logsales reta wka

Logistic regression                             Number of obs   =        921
                                                LR chi2(5)      =     591.80
                                                Prob > chi2     =     0.0000
Log likelihood = -341.07758                     Pseudo R2       =     0.4645

------------------------------------------------------------------------------
    invgrade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     booklev |  -4.427266   .7714185    -5.74   0.000    -5.939218   -2.915313
        ebit |   4.354735   1.439922     3.02   0.002     1.532539    7.176931
    logsales |   1.081593   .0956839    11.30   0.000     .8940558     1.26913
        reta |   4.116108   .4885083     8.43   0.000     3.158649    5.073566
         wka |  -4.012492   .7479141    -5.36   0.000    -5.478377   -2.546607
       _cons |  -8.214321   .8668543    -9.48   0.000    -9.913324   -6.515317
------------------------------------------------------------------------------
```

Illustration 7 – Question 3.2

```
. ologit rating booklev ebit logsales reta wka
Ordered logistic regression                        Number of obs   =        921
                                                   LR chi2(5)      =     862.87
                                                   Prob > chi2     =     0.0000
Log likelihood = -965.30716                        Pseudo R2       =     0.3089

------------------------------------------------------------------------------
      rating |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     booklev |  -2.751598   .4773028    -5.76   0.000    -3.687094   -1.816101
        ebit |   4.731324    .944758     5.01   0.000     2.879632    6.583015
    logsales |   .9405595   .0588423    15.98   0.000     .8252308    1.055888
        reta |   3.559983    .302294    11.78   0.000     2.967498    4.152468
         wka |  -2.580078    .482632    -5.35   0.000    -3.526019   -1.634136
-------------+----------------------------------------------------------------
       /cut1 |  -.3694825   .6332452                      -1.61062    .8716554
       /cut2 |   4.880646   .5208171                      3.859864    5.901429
       /cut3 |   7.625999   .5511597                      6.545746    8.706252
       /cut4 |    9.88504   .5916376                      8.725452    11.04463
       /cut5 |   12.88296    .673386                      11.56315    14.20277
       /cut6 |   14.78264   .7841076                      13.24582    16.31947
------------------------------------------------------------------------------
```

Illustration 7 – Question 4

Binary logit model:

$$Pr(invgrade_i = 1|x_i) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$$

Ordered logit model:

$$Pr(rating_i \geq 4|x_i) = 1 - \frac{e^{\gamma_2 - x_i'\beta}}{1 + e^{\gamma_2 - x_i'\beta}}$$

```
. summarize booklev ebit logsales reta wka

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     booklev |        921    .2931868    .1735339          0   .9992067
        ebit |        921    .0938921    .0843643  -.3841692   .6515085
    logsales |        921    7.995754    1.497413   1.100278   12.70142
        reta |        921    .1569942     .300684  -.9958922   .9799219
         wka |        921    .1404142    .1503398  -.4120839   .7480223
```

Illustration 7 – Question 4 (cont.)

```
. quietly logit invgrade booklev ebit logsales reta wka

. scalar xblogit = _b[_cons] + _b[booklev]*0.2931868 + _b[ebit]*0.0938921 +
_b[logsales]*7.995754 + _b[reta]*0.1569942 + _b[wka]*0.1404142

. display exp(xblogit)/(1+exp(xblogit))
.40793294


. quietly ologit rating booklev ebit logsales reta wka

. scalar xborden = _b[booklev]*0.2931868 + _b[ebit]*0.0938921 +
_b[logsales]*7.995754 + _b[reta]*0.1569942+_b[wka]*0.1404142

. display 1-exp(_b[/cut3]-xborden)/(1+exp(_b[/cut3]-xborden))
.43256406
```

Illustration 8

Sample:

- 210 observations concerning travel mode choice for travel between Sydney and Melbourne

Travel modes:

- Air
- Train
- Bus
- Car

Details:

- Greene (2003), ch. 21

Illustration 8 – Question 1

```
. describe

 obs:             840
 vars:              6                          11 Apr 2014 14:04
 size:         19,320
--------------------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------------------
id              int    %8.0g                  Identification
Modealt         str5   %9s                    Travel mode alternatives
Mode            float  %9.0g                  =1 if travel mode chosen
Ttme            float  %9.0g                  Terminal time
GC              float  %9.0g                  Cost of travel (transport cost and amount of
                                                time spent travelling)
Hinc            float  %9.0g                  Household income
--------------------------------------------------------------------------------------------
Sorted by:  id  Modealt
```

Illustration 8 – Question 2.1

```
. summarize Mode Ttme GC Hinc if Modealt=="air"

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Mode |        210    .2761905    .4481805          0          1
        Ttme |        210    61.00952    15.71943          5         99
          GC |        210    102.6476    30.57503         56        197
        Hinc |        210    34.54762    19.71132          2         72


. summarize Mode Ttme GC Hinc if Modealt=="bus"

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Mode |        210    .1428571    .3507633          0          1
        Ttme |        210    41.65714    12.07737          5         60
          GC |        210    115.2571    44.93441         45        222
        Hinc |        210    34.54762    19.71132          2         72
```

Illustration 8 – Question 2.1 (cont.)

```
. summarize Mode Ttme GC Hinc if Modealt=="car"

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Mode |        210    .2809524    .4505383          0          1
        Ttme |        210           0           0          0          0
          GC |        210    95.41429    46.82743         30        238
        Hinc |        210    34.54762    19.71132          2         72


. summarize Mode Ttme GC Hinc if Modealt=="train"

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Mode |        210          .3    .4593526          0          1
        Ttme |        210    35.69048    12.27922          1         99
          GC |        210       130.2    58.23454         42        269
        Hinc |        210    34.54762    19.71132          2         72
```

Illustration 8 – Question 2.2

```
. summarize Ttme GC Hinc if Modealt=="air" & Mode==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Ttme |         58    46.53448    24.38882          5         99
          GC |         58    113.5517    33.19753         58        197
        Hinc |         58    41.72414    19.11534          4         70
```

```
. summarize Ttme GC Hinc if Modealt=="train" & Mode==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Ttme |         63    28.52381    19.35397          1         99
          GC |         63     106.619    49.60113         42        211
        Hinc |         63    23.06349    17.28683          4         72
```

Illustration 8 – Question 2.2 (cont.)

```
. summarize Ttme GC Hinc if Modealt=="bus" & Mode==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Ttme |         30        25.2    14.91909          5         60
          GC |         30    108.1333    43.24408         45        201
        Hinc |         30        29.7    16.85056          2         60


. summarize Ttme GC Hinc if Modealt=="car" & Mode==1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        Ttme |         59           0           0          0          0
          GC |         59    89.08475    49.83292         30        238
        Hinc |         59    42.22034    17.68505          4         70
```

Illustration 8 – Question 3

- 'Long form' database:

	id	Modealt	Mode	Ttme	GC	Hinc
1.	1	air	0	69	70	35
2.	1	bus	0	35	70	35
3.	1	car	1	0	30	35
4.	1	train	0	34	71	35
5.	2	air	0	64	68	30
6.	2	bus	0	53	85	30
7.	2	car	1	0	50	30
8.	2	train	0	44	84	30
9.	3	air	0	69	129	40
10.	3	bus	0	35	149	40
11.	3	car	1	0	101	40
12.	3	train	0	34	195	40

- This is the data format necessary to estimate a conditional logit model (it may include x_i-type variables) using Stata

Illustration 8 – Question 3 (cont.)

- ‘Wide form’ database:

	id	Modeair	Ttmeair	GCair	Modebus	Ttmebus	GCbus	Modecar	Ttmecar
1.	1	0	69	70	0	35	70	1	0
2.	2	0	64	68	0	53	85	1	0
3.	3	0	69	129	0	35	149	1	0
4.	4	0	64	59	0	53	81	1	0
5.	5	0	64	82	0	53	94	1	0
6.	6	0	69	70	0	35	58	0	0
7.	7	1	45	160	0	35	167	0	0
8.	8	0	69	137	0	35	146	1	0
9.	9	0	69	70	0	35	70	1	0
10.	10	0	69	65	0	35	68	1	0
11.	11	0	64	68	0	53	73	1	0
12.	12	0	64	79	0	53	91	1	0

- This is the data format necessary to estimate a multinomial logit model using Stata when there are only x_i-type variables

Illustration 8 – Question 3 (cont.)

- From 'long form' to 'wide form':

```
. reshape wide Ttme GC Mode, i(id) j(Modealt) string
```

- After reshaping the data once, it is possible to alternate between the two formats using the following commands:

```
. reshape long
```

```
. reshape wide
```

Illustration 8 – Question 3.1

```
. gen Y=0 if Modeair==1
(152 missing values generated)

. replace Y=1 if Modetrain==1
(63 real changes made)

. replace Y=2 if Modebus==1
(30 real changes made)

. replace Y=3 if Modecar==1
(59 real changes made)
```

Illustration 8 – Question 3.1 (cont.)

```
. mlogit Y Hinc, baseoutcome(0)

Multinomial logistic regression                 Number of obs     =        210
                                                LR chi2(3)        =      44.03
                                                Prob > chi2       =     0.0000
Log likelihood = -261.74506                     Pseudo R2         =     0.0776

------------------------------------------------------------------------------
           Y |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
0            |  (base outcome)
-------------+----------------------------------------------------------------
1            |
        Hinc |   -.059059   .0116784    -5.06   0.000    -.0819482   -.0361698
       _cons |   1.963424   .4195611     4.68   0.000     1.141099    2.785749
-------------+----------------------------------------------------------------
2            |
        Hinc |  -.0353522   .0128212    -2.76   0.006    -.0604814    -.010223
       _cons |   .5991669   .4909705     1.22   0.222    -.3631176    1.561452
-------------+----------------------------------------------------------------
3            |
        Hinc |   .0014204   .0098938     0.14   0.886    -.0179711    .0208118
       _cons |  -.0425218    .454562    -0.09   0.925    -.9334469    .8484034
------------------------------------------------------------------------------
```

Illustration 8 – Question 3.2

```
. reshape long

. asclogit Mode Ttme GC, case(id) alternatives(Modealt) casevars(Hinc) basealternative(air)
(…)
------------------------------------------------------------------------------
        Mode |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Modealt      |
        Ttme |  -.0954606   .0104732    -9.11   0.000    -.1159876   -.0749335
          GC |  -.0109274   .0045878    -2.38   0.017    -.0199192   -.0019355
-------------+----------------------------------------------------------------
air          |  (base alternative)
-------------+----------------------------------------------------------------
bus          |
        Hinc |  -.0232107   .0162306    -1.43   0.153     -.055022    .0086006
       _cons |  -1.744529   .6775004    -2.57   0.010    -3.072406   -.4166531
-------------+----------------------------------------------------------------
car          |
        Hinc |   .0053735   .0115294     0.47   0.641    -.0172237    .0279707
       _cons |  -5.874813   .8020903    -7.32   0.000    -7.446882   -4.302745
-------------+----------------------------------------------------------------
train        |
        Hinc |  -.0511884   .0147352    -3.47   0.001    -.0800689   -.0223079
       _cons |  -.3249561   .5763335    -0.56   0.573    -1.454549    .8046369
------------------------------------------------------------------------------
```

Illustration 9

Aim:

- Explaining annual health care expenditures (med) and the annual number of medical consultations (mdu) as a function of health insurance characteristics ($lcoins$), socio-economic factors ($female, age, lfam, child$) and health status variables ($ndisease$)

Details:

- Cameron and Trivedi (2005), ch. 16.6

Illustration 9 – Question 1

```
. summarize med

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         med |      20186    171.5892    698.2689          0   39182.02

. summarize med if med>0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
         med |      15733    220.1551    784.1543   .4276146   39182.02
```

- Non-negative outcome
- High proportion of zeros: 4453 observations (22.1% of the sample units)

Illustration 9 – Question 2.1

```
. drop if year!=1
(14548 observations deleted)

. poisson med lcoins ndisease female age lfam child, robust
(…)

------------------------------------------------------------------------------
             |               Robust
         med |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0478877   .0245805    -1.95   0.051    -.0960646    .0002892
    ndisease |   .0281757   .0060746     4.64   0.000     .0162696    .0400817
      female |   .2263111   .1192456     1.90   0.058     -.007406    .4600282
         age |   .0097846    .003701     2.64   0.008     .0025307    .0170384
        lfam |  -.0928784   .0882497    -1.05   0.293    -.2658446    .0800879
       child |  -.4112285   .2444742    -1.68   0.093    -.8903892    .0679321
       _cons |   4.631177   .2039135    22.71   0.000     4.231514     5.03084
------------------------------------------------------------------------------

. estimates store poisson
```

Illustration 9 – Question 2.2

```
. poisson med lcoins ndisease female age lfam child if med>0, robust
(…)

Poisson regression                                Number of obs   =      4451
                                                  Wald chi2(6)    =     66.48
                                                  Prob > chi2     =    0.0000
Log pseudolikelihood = -1087511.2                 Pseudo R2       =    0.0657

------------------------------------------------------------------------------
             |               Robust
         med |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0200287   .0246889    -0.81   0.417     -.068418    .0283606
    ndisease |   .0210867   .0062785     3.36   0.001     .0087812    .0333923
      female |   .1474704   .1177851     1.25   0.211    -.0833842    .3783249
         age |   .0087803   .0036967     2.38   0.018     .0015349    .0160258
        lfam |  -.0823594   .0902764    -0.91   0.362    -.2592978    .0945791
       child |  -.3678983   .2481063    -1.48   0.138    -.8541777    .1183811
       _cons |    4.91825   .2098069    23.44   0.000     4.507036    5.329464
------------------------------------------------------------------------------

. estimates store poisson0
```

Illustration 9 – Question 2.3

```
. gen lmed=log(med)
(1187 missing values generated)

. regress lmed lcoins ndisease female age lfam child if med>0, robust
(…)

------------------------------------------------------------------------------
             |               Robust
        lmed |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |   -.054901   .0099633    -5.51   0.000    -.0744341   -.0353679
    ndisease |    .030717   .0030927     9.93   0.000     .0246537    .0367803
      female |   .1199091   .0417646     2.87   0.004     .0380298    .2017885
         age |   .0072817   .0020791     3.50   0.000     .0032055    .0113578
        lfam |  -.1357759   .0442937    -3.07   0.002    -.2226136   -.0489382
       child |  -.4972463   .0737896    -6.74   0.000    -.6419107   -.3525819
       _cons |   3.954487   .1043663    37.89   0.000     3.749877    4.159096
------------------------------------------------------------------------------

. estimates store log0
```

Illustration 9 – Question 2.4

```
. gen lmed1=log(med+1)

. regress lmed1 lcoins ndisease female age lfam child, robust
(…)

------------------------------------------------------------------------------
             |               Robust
       lmed1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.1485372   .0126915   -11.70   0.000    -.1734174   -.1236569
    ndisease |   .0584591   .0039132    14.94   0.000     .0507878    .0661304
      female |   .3064252   .0535977     5.72   0.000      .201353    .4114974
         age |   .0089512   .0027704     3.23   0.001     .0035202    .0143821
        lfam |  -.1806985   .0556686    -3.25   0.001    -.2898305   -.0715666
       child |  -.4759264   .0946324    -5.03   0.000    -.6614425   -.2904103
       _cons |   3.000465   .1346348    22.29   0.000     2.736529    3.264401
------------------------------------------------------------------------------

. estimates store log1
```

Illustration 9 – Question 2 (summary)

```
. estimates table poisson poisson0, b star(0.1 0.05 0.01)
----------------------------------------------
    Variable |    poisson        poisson0
-------------+--------------------------------
      lcoins | -.04788772*      -.0200287
    ndisease |  .02817566***    .02108674***
      female |  .22631112*      .14747037
         age |  .00978456***    .00878032**
        lfam | -.09287836      -.08235938
       child | -.41122852*      -.3678983
       _cons |  4.6311773***    4.9182498***
----------------------------------------------
           legend: * p<.1; ** p<.05; *** p<.01

. estimates table log0 log1, b star(0.1 0.05 0.01)
----------------------------------------------
    Variable |     log0             log1
-------------+--------------------------------
      lcoins | -.05490096***   -.14853716***
    ndisease |  .03071698***     .0584591***
      female |  .11990912***     .3064252***
         age |  .00728165***    .00895117***
        lfam | -.13577591***   -.18069853***
       child |  -.4972463***   -.47592639***
       _cons |  3.9544865***    3.0004653***
----------------------------------------------
```

Illustration 9 – Question 3.1

```
. summarize mdu
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
         mdu |      5638    2.877971    4.332918          0         69

. tabulate mdu

     number |
face-to-fac |
e md visits |      Freq.     Percent        Cum.
------------+-----------------------------------
          0 |      1,729       30.67       30.67
          1 |      1,047       18.57       49.24
          2 |        814       14.44       63.68
          3 |        511        9.06       72.74
          4 |        385        6.83       79.57
          5 |        271        4.81       84.37
          6 |        188        3.33       87.71
          7 |        156        2.77       90.48
          8 |        134        2.38       92.85
          9 |         80        1.42       94.27
         10 |         56        0.99       95.26
         11 |         56        0.99       96.26
         12 |         32        0.57       96.83
```

(…)

Illustration 9 – Question 3.2.1

```
. poisson mdu lcoins ndisease female age lfam child

Poisson regression                                Number of obs   =       5638
                                                  LR chi2(6)      =    2296.50
                                                  Prob > chi2     =     0.0000
Log likelihood = -17173.058                       Pseudo R2       =     0.0627

------------------------------------------------------------------------------
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0682563   .0038214   -17.86   0.000    -.0757462   -.0607665
    ndisease |     .03538   .0010286    34.40   0.000     .0333639     .037396
      female |   .1276798   .0164078     7.78   0.000     .0955211    .1598386
         age |   .0041167   .0007841     5.25   0.000     .0025798    .0056536
        lfam |   -.141855   .0154716    -9.17   0.000    -.1721789   -.1115312
       child |    .056673   .0278998     2.03   0.042     .0019903    .1113557
       _cons |   .7463066   .0385653    19.35   0.000       .67072    .8218932
------------------------------------------------------------------------------

. estimates store cpoissonml
```

Illustration 9 – Question 3.2.2

```
. poisson mdu lcoins ndisease female age lfam child, robust

Poisson regression                              Number of obs   =       5638
                                                Wald chi2(6)    =     388.20
                                                Prob > chi2     =     0.0000
Log pseudolikelihood = -17173.058               Pseudo R2       =     0.0627

------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0682563   .0094538    -7.22   0.000    -.0867854   -.0497273
    ndisease |     .03538    .002696    13.12   0.000     .0300959    .0406641
      female |   .1276798     .04043     3.16   0.002     .0484385    .2069211
         age |   .0041167   .0019155     2.15   0.032     .0003624     .007871
        lfam |   -.141855   .0336317    -4.22   0.000     -.207772   -.0759381
       child |    .056673   .0625248     0.91   0.365    -.0658734    .1792194
       _cons |   .7463066   .0861806     8.66   0.000     .5773956    .9152175
------------------------------------------------------------------------------

. estimates store cpoissonqml
```

Illustration 9 – Question 3.2.3

```
. nbreg mdu lcoins ndisease female age lfam child, dispersion(constant)

Negative binomial regression                    Number of obs   =       5638
                                                LR chi2(6)      =     529.02
Dispersion     = constant                       Prob > chi2     =     0.0000
Log likelihood = -12128.041                     Pseudo R2       =     0.0213

------------------------------------------------------------------------------
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |   -.076432   .0069326   -11.02   0.000    -.0900197   -.0628442
    ndisease |   .0309127   .0019493    15.86   0.000     .0270922    .0347333
      female |   .1447002    .029679     4.88   0.000     .0865305    .2028699
         age |   .0027029   .0014494     1.86   0.062    -.0001378    .0055436
        lfam |  -.1235042   .0276393    -4.47   0.000    -.1776761   -.0693323
       child |  -.0049686   .0499784    -0.10   0.921    -.1029245    .0929873
       _cons |   .8537349   .0697031    12.25   0.000     .7171195    .9903504
-------------+----------------------------------------------------------------
    /lndelta |   1.278115   .0319841                      1.215427    1.340802
-------------+----------------------------------------------------------------
       delta |   3.589865   .1148187                      3.371733    3.822109
------------------------------------------------------------------------------
Likelihood-ratio test of delta=0:  chibar2(01) = 1.0e+04 Prob>=chibar2 = 0.000

. estimates store cnegbin1
```

Overdispersion test: the hypothesis of correct specification of the Poisson mode is rejected

Illustration 9 – Question 3.2.4

```
. nbreg mdu lcoins ndisease female age lfam child, dispersion(mean)

Negative binomial regression                    Number of obs   =       5638
                                                LR chi2(6)      =     459.10
Dispersion     = mean                           Prob > chi2     =     0.0000
Log likelihood = -12163.001                     Pseudo R2       =     0.0185

------------------------------------------------------------------------------
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0719827   .0083402    -8.63   0.000    -.0883292   -.0556362
    ndisease |   .0391488    .002618    14.95   0.000     .0340175    .0442801
      female |   .1229883   .0350213     3.51   0.000     .0543479    .1916288
         age |   .0035899   .0017692     2.03   0.042     .0001223    .0070574
        lfam |  -.1834739    .036481    -5.03   0.000    -.2549754   -.1119723
       child |   .0842864   .0621692     1.36   0.175    -.0375631    .2061359
       _cons |   .7638038    .087809     8.70   0.000     .5917012    .9359063
-------------+----------------------------------------------------------------
    /lnalpha |   .2193579   .0274111                      .1656333    .2730826
-------------+----------------------------------------------------------------
       alpha |   1.245277   .0341344                       1.18014    1.314009
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) = 1.0e+04 Prob>=chibar2 = 0.000

. estimates store cnegbin2
```

Overdispersion test: the hypothesis of correct specification of the Poisson mode is rejected

Illustration 9 – Question 3.2 (summary)

```
. estimates table cpoissonml cpoissonqml cnegbin1 cnegbin2, b star(0.1 0.05
0.01)

--------------------------------------------------------------------------
    Variable |  cpoissonml      cpoissonqml      cnegbin1        cnegbin2
-------------+------------------------------------------------------------
mdu          |
      lcoins | -.06825635***   -.06825635***   -.07643198***    -.0719827***
    ndisease |  .03537998***    .03537998***    .03091273***    .03914881***
      female |  .12767984***    .12767984***     .1447002***    .12298834***
         age |   .0041167***     .0041167**      .0027029*      .00358987**
        lfam | -.14185505***   -.14185505***    -.1235042***   -.18347387***
       child |  .05667301**     .05667301      -.00496864       .0842864
       _cons |  .74630658***    .74630658***    .85373493***    .76380377***
-------------+------------------------------------------------------------
lndelta      |
       _cons |                                  1.2781147***
-------------+------------------------------------------------------------
lnalpha      |
       _cons |                                                  .21935795***
--------------------------------------------------------------------------
                                           legend: * p<.1; ** p<.05; *** p<.01
```

Illustration 9 – Question 3.3

```
. estimates restore cpoissonml
(results cpoissonml are active now)
```

Formulas:

$$Pr(Y_i = y|x_i) = \frac{e^{-\lambda_i}\lambda_i{}^y}{y!}, \qquad \lambda_i = E(Y|X) = \exp(x_i'\beta)$$

- $Pr(Y_i = 0|x_i) = e^{-\lambda_i}$
- $Pr(Y_i = 1|x_i) = e^{-\lambda_i}\lambda_i$
- $Pr(Y_i \geq 2|x_i) = 1 - Pr(Y_i = 0|x_i) - Pr(Y_i = 1|x_i)$

```
. scalar lambda0 = exp(_b[_cons] + _b[lcoins]*log(1) + _b[age]*50 +
_b[lfam]*log(3))

. scalar lambda1 = exp(_b[_cons] + _b[lcoins]*log(51) + _b[age]*50 +
_b[lfam]*log(3))

. scalar lambda2 = exp(_b[_cons] + _b[lcoins]*log(101) + _b[age]*50 +
_b[lfam]*log(3))
```

Illustration 9 – Question 3.3 (cont.)

```
. display lambda0
2.2173168

. display lambda1
1.695412

. display lambda2
1.6181549

. display exp(-lambda0)
.10890092

. display exp(-lambda1)
.18352361

. display exp(-lambda2)
.19826418
```

```
. display exp(-lambda0)*lambda0
.24146784

. display exp(-lambda1)*lambda1
.31114812

. display exp(-lambda2)*lambda2
.32082215

. display 1-0.10890092-0.24146784
.64963124

. display 1-0.18352361-0.31114812
.50532827

. display 1-0.19826418-0.32082215
.48091367
```

Illustration 9 – Question 3.3 (cont.)

$coins$:	0	50	100
$E(mdu\| \ldots)$	2.2	1.7	1.6
$Pr(mdu = 0\| \ldots)$	10.9	18.4	19.8
$Pr(mdu = 1\| \ldots)$	24.1	31.1	32.1
$Pr(mdu \geq 2\| \ldots)$	65.0	50.5	48.1

Illustration 9 – Question 4.1

```
. xtset id year
       panel variable:  id (unbalanced)
        time variable:  year, 1 to 5, but with gaps
                delta:  1 unit
. xtdescribe
      id:  125024, 125025, ..., 632167                       n =       5908
    year:  1, 2, ..., 5                                      T =          5
           Delta(year) = 1 unit
           Span(year)  = 5 periods
           (id*year uniquely identifies each observation)
Distribution of T_i:   min      5%     25%       50%       75%     95%     max
                         1       2       3         3         5       5       5
     Freq.  Percent    Cum. |  Pattern
 ---------------------------+---------
     3710     62.80   62.80 |  111..
     1584     26.81   89.61 |  11111
      156      2.64   92.25 |  1....
      147      2.49   94.74 |  11...
       79      1.34   96.07 |  ..1..
       66      1.12   97.19 |  .11..
       33      0.56   97.75 |  ..111
       33      0.56   98.31 |  .1111
       29      0.49   98.80 |  ...11
       71      1.20  100.00 | (other patterns)
 ---------------------------+---------
     5908    100.00         |  XXXXX
```

Illustration 9 – Question 4.2.1

```
. poisson mdu lcoins ndisease female age lfam child, vce(cluster id)

Poisson regression                                Number of obs   =      20186
                                                  Wald chi2(6)    =     476.93
                                                  Prob > chi2     =     0.0000
Log pseudolikelihood = -62579.401                 Pseudo R2       =     0.0609

                              (Std. Err. adjusted for 5908 clusters in id)
------------------------------------------------------------------------------
             |               Robust
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0808023   .0080013   -10.10   0.000    -.0964846   -.0651199
    ndisease |   .0339334   .0026024    13.04   0.000     .0288328     .039034
      female |   .1717862   .0342551     5.01   0.000     .1046473    .2389251
         age |   .0040585   .0016891     2.40   0.016      .000748    .0073691
        lfam |  -.1481981   .0323434    -4.58   0.000      -.21159   -.0848062
       child |   .1030453   .0506901     2.03   0.042     .0036944    .2023961
       _cons |    .748789   .0785738     9.53   0.000     .5947872    .9027907
------------------------------------------------------------------------------

. estimates store pooled
```

Illustration 9 – Question 4.2.2

```
. xtpoisson mdu lcoins ndisease female age lfam child, re
(…)

------------------------------------------------------------------------------
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lcoins |  -.0878258   .0068682   -12.79   0.000    -.1012873   -.0743642
    ndisease |   .0387629   .0022046    17.58   0.000      .034442    .0430839
      female |   .1667192   .0286298     5.82   0.000     .1106058    .2228325
         age |   .0019159   .0011134     1.72   0.085    -.0002663    .0040982
        lfam |  -.1351786   .0260022    -5.20   0.000     -.186142   -.0842152
       child |   .1082678   .0341477     3.17   0.002     .0413396    .1751961
       _cons |   .7574177   .0618346    12.25   0.000     .6362241    .8786112
-------------+----------------------------------------------------------------
    /lnalpha |   .0251256   .0209586                     -.0159526    .0662038
-------------+----------------------------------------------------------------
       alpha |   1.025444   .0214919                       .984174    1.068444
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0: chibar2(01) =  3.9e+04 Prob>=chibar2 = 0.000

. estimates store RE
```

Illustration 9 – Question 4.2.3

```
. xtpoisson mdu lcoins ndisease female age lfam child, fe
note: 265 groups (265 obs) dropped because of only one obs per group
note: 666 groups (2130 obs) dropped because of all zero outcomes
note: lcoins dropped because it is constant within group
note: ndisease dropped because it is constant within group
note: female dropped because it is constant within group
(…)

                                                Wald chi2(3)       =     19.24
Log likelihood  = -24173.211                    Prob > chi2        =    0.0002

------------------------------------------------------------------------------
         mdu |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         age |  -.0112009   .0039024    -2.87   0.004    -.0188494   -.0035523
        lfam |   .0877134   .0554606     1.58   0.114    -.0209874    .1964141
       child |   .1059867   .0437744     2.42   0.015     .0201905    .1917829
------------------------------------------------------------------------------

. estimates store FE
```

Illustration 9 – Question 4.2 (summary)

```
. estimates table pooled RE FE, b star(0.1 0.05 0.01)

--------------------------------------------------------------
    Variable |    pooled            RE               FE
-------------+------------------------------------------------
mdu          |
      lcoins | -.08080226***   -.08782576***
    ndisease |   .0339334***    .03876295***
      female |  .17178621***    .16671918***
         age |  .00405853**     .00191594*     -.01120087***
        lfam |  -.1481981***   -.13517858***     .08771336
       child |  .10304526**     .10826785***     .10598669**
       _cons |  .74878896***     .7574177***
-------------+------------------------------------------------
lnalpha      |
       _cons |                  .02512562
--------------------------------------------------------------
                           legend: * p<.1; ** p<.05; *** p<.01
```

Illustration 9 – Question 4.3

```
. hausman FE RE

                 ---- Coefficients ----
             |      (b)          (B)            (b-B)     sqrt(diag(V_b-V_B))
             |       EF           EA         Difference          S.E.
-------------+----------------------------------------------------------------
         age |   -.0112009     .0019159       -.0131168        .0037402
        lfam |    .0877134    -.1351786        .2228919        .0489874
       child |    .1059867     .1082678       -.0022812        .0273886
------------------------------------------------------------------------------
                         b = consistent under Ho and Ha; obtained from xtpoisson
          B = inconsistent under Ha, efficient under Ho; obtained from xtpoisson

    Test:  Ho:  difference in coefficients not systematic

                  chi2(3) = (b-B)'[(V_b-V_B)^(-1)](b-B)
                          =       29.73
                Prob>chi2 =      0.0000
```

The hypothesis of random effects is rejected

Illustration 10 – Question 1

```
. drop if LE==1
(1922 observations deleted)

. summarize LEV_LT1

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     LEV_LT1 |      30304    .0734047    .1668595          0    .9982489

. count if LEV_LT1==0
23027

. display 23027/30304
.75986668
```

Illustration 10 – Question 2.1

```
. glm LEV_LT1 SIZE2 COLLAT2 PROF1 GROWTH2 AGE, family(binomial) link(logit)
vce(cluster id)
(…)

                                                       AIC             =  .4259979
Log pseudolikelihood =  -6448.71996                    BIC             = -304709.7

                                  (Std. Err. adjusted for 4995 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     LEV_LT1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       SIZE2 |   .3160809   .0150444    21.01   0.000     .2865946    .3455673
     COLLAT2 |    .453255   .1394787     3.25   0.001     .1798818    .7266282
       PROF1 |  -3.629699   .2667238   -13.61   0.000    -4.152469    -3.10693
     GROWTH2 |   .0004846   .0002476     1.96   0.050    -6.89e-07    .0009699
         AGE |  -.0045993   .0017012    -2.70   0.007    -.0079335   -.0012651
       _cons |  -6.823869   .2264392   -30.14   0.000    -7.267681   -6.380056
------------------------------------------------------------------------------

. estimates store LOGIT
```

Illustration 10 – Question 2.2

```
. gen DEBT=LEV_LT1>0

. probit DEBT SIZE2 COLLAT2 PROF1 GROWTH2 AGE, vce(cluster id)
(…)
Probit regression                               Number of obs     =      30304
                                                Wald chi2(5)      =    1125.30
                                                Prob > chi2       =     0.0000
Log pseudolikelihood = -14363.976               Pseudo R2         =     0.1401

                                 (Std. Err. adjusted for 4995 clusters in id)
------------------------------------------------------------------------------
             |               Robust
        DEBT |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       SIZE2 |   .3309445    .011001    30.08   0.000      .309383    .3525061
     COLLAT2 |    .754961     .07028    10.74   0.000     .6172147    .8927074
       PROF1 |  -2.114841    .176704   -11.97   0.000    -2.461175   -1.768508
     GROWTH2 |    .000226   .0001309     1.73   0.084    -.0000307    .0004826
         AGE |   .0011396   .0011065     1.03   0.303     -.001029    .0033082
       _cons |  -5.510938   .1548431   -35.59   0.000    -5.814425   -5.207451
------------------------------------------------------------------------------

. estimates store TP1
```

Illustration 10 – Question 2.2 (cont.)

```
. glm LEV_LT1 SIZE2 COLLAT2 PROF1 GROWTH2 AGE if LEV_LT>0, family(binomial)
link(logit) vce(cluster id)
(…)

                                                  AIC             =  .8776548
Log pseudolikelihood = -3187.347033               BIC             = -63243.49

                                  (Std. Err. adjusted for 2064 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     LEV_LT1 |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       SIZE2 |   -.185588   .0140524   -13.21   0.000    -.2131301   -.1580459
     COLLAT2 |  -.7261906   .0998168    -7.28   0.000    -.9218279   -.5305533
       PROF1 |  -1.523694   .2894609    -5.26   0.000    -2.091027   -.9563615
     GROWTH2 |   .0030029   .0004916     6.11   0.000     .0020395    .0039663
         AGE |  -.0071666   .0013872    -5.17   0.000    -.0098856   -.0044477
       _cons |   2.389845   .2057352    11.62   0.000     1.986611    2.793078
------------------------------------------------------------------------------

. estimates store TP2
```

Illustration 10 – Question 2.3

```
. tobit LEV_LT1 SIZE2 COLLAT2 PROF1 GROWTH2 AGE, ll(0) vce(cluster id)
(…)
                                  (Std. Err. adjusted for 4995 clusters in id)
------------------------------------------------------------------------------
             |               Robust
     LEV_LT1 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       SIZE2 |   .1228862   .0040377    30.43   0.000     .1149721    .1308004
     COLLAT2 |   .2341081   .0318943     7.34   0.000     .1715941    .2966222
       PROF1 |  -.9789402   .0758649   -12.90   0.000    -1.127639   -.8302418
     GROWTH2 |   .0001581   .0000812     1.95   0.052    -1.14e-06    .0003174
         AGE |  -.0003276   .0004359    -0.75   0.452    -.001182     .0005268
       _cons |  -2.064098   .0565277   -36.51   0.000    -2.174894   -1.953301
-------------+----------------------------------------------------------------
 var(e.LEV_LT1)|   .2180737   .0075559                     .2037555    .2333981
------------------------------------------------------------------------------


. estimates store TOBIT
```

Illustration 10 – Question 3.1

Fractional logit model:

$$E(LEV\_LT_i|x_i) = \frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$$

```
. estimates restore LOGIT
(results LOGIT are active now)

. scalar xb = _b[_cons] + _b[SIZE2]*13.54 + _b[COLLAT2]*0.41 + _b[PROF1]*0.07 +
_b[GROWTH2]*15.03 + _b[AGE]*19

. display exp(xb)/(1+exp(xb))
.06341858
```

Illustration 10 – Question 3.1 (cont.)

Two-part model:

$$E(LEV\_LT_i|x_i) = \Phi(x_i'\theta)\frac{e^{x_i'\beta}}{1+e^{x_i'\beta}}$$

```
. estimates restore TP1
(results DP1 are active now)

. scalar xb1 = _b[_cons] + _b[SIZE2]*13.54 + _b[COLLAT2]*0.41 + _b[PROF1]*0.07
+ _b[GROWTH2]*15.03 + _b[AGE]*19

. estimates restore TP2
(results DP2 are active now)

. scalar xb2 = _b[_cons] + _b[SIZE2]*13.54 + _b[COLLAT2]*0.41 + _b[PROF1]*0.07
+ _b[GROWTH2]*15.03 + _b[AGE]*19

. display normal(xb1)*exp(xb2)/(1+exp(xb2))
.06985232
```

Illustration 10 – Question 3.1 (cont.)

Tobit model:

$$E(LEV\_LT_i|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)x_i'\beta + \sigma\phi\left(\frac{x_i'\beta}{\sigma}\right)$$

```
. estimates restore TOBIT
(results TOBIT are active now)

. scalar xbt = _b[_cons] + _b[SIZE2]*13.54 + _b[COLLAT2]*0.41 + _b[PROF1]*0.07
+ _b[GROWTH2]*15.03 + _b[AGE]*19

. display normal(xbt/sqrt(_b[/var(e.LEV_LT1)]))*xbt+sqrt(_b[/var(e.LEV_LT1)])*
normalden(xbt/sqrt(_b[/var(e.LEV_LT1)]))
.05549759
```

Illustration 10 – Question 3.2

Fractional logit model:

- By definition, it is approximately equal to one

Two-part model:

$$Pr(LEV\_LT_i > 0|x_i) = \Phi(x_i'\theta)$$

```
. display normal(xb1)
.19950111
```

Tobit model:

$$Pr(LEV\_LT_i > 0|x_i) = \Phi\left(\frac{x_i'\beta}{\sigma}\right)$$

```
. estimates restore TOBIT
(results TOBIT are active now)

. display normal(xbt/sqrt(_b[/var(e.LEV_LT1)]))
.20998609
```

Illustration 10 – Question 3.3

Fractional logit model:

- By definition, $E(LEV\_LT_i|x_i, Y_i > 0) \approx E(LEV\_LT_i|x_i)$

Two-part model:

$$E(LEV\_LT_i|x_i, Y_i > 0) = \frac{e^{x_i'\beta}}{1 + e^{x_i'\beta}}$$

```
. display exp(xb2)/(1+exp(xb2))
.35013498
```

Tobit model:

$$E(Y_i|x_i, Y_i > 0) = x_i'\beta + \sigma\lambda\left(\frac{x_i'\beta}{\sigma}\right)$$

```
. estimates restore TOBIT
(results TOBIT are active now)

. display xbt+sqrt(_b[/var(e.LEV_LT1)])*normalden(xbt/sqrt(_b[/var(e.LEV_LT1)])
)/normal(xbt/sqrt(_b[/var(e.LEV_LT1)]))
.26429175
```

Illustration 10 – Question 4

```
. xtset id YEAR
       panel variable:  id (unbalanced)
        time variable:  YEAR, 1995 to 2001, but with gaps
                delta:  1 unit

. gen LEVt= LEV_LT1/(1-LEV_LT1)

. xtpoisson LEVt SIZE2 COLLAT2 PROF1 GROWTH2 AGE, fe vce(robust)
(…)

                                     (Std. Err. adjusted for clustering on id)
------------------------------------------------------------------------------
             |               Robust
        LEVt |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       SIZE2 |  -.2563746   .1914515    -1.34   0.181    -.6316126    .1188634
     COLLAT2 |   .7169194   .5366941     1.34   0.182    -.3349818    1.768821
       PROF1 |  -1.205703   2.620676    -0.46   0.645    -6.342134    3.930727
     GROWTH2 |   .0024675     .00165     1.50   0.135    -.0007665    .0057014
         AGE |   .0402399   .0259754     1.55   0.121     -.010671    .0911508
------------------------------------------------------------------------------
```