Economics 312 — Spring 2020
Daily Problem #37 — April 22

The output below was generated using a dataset on Reed students in the 1988–2002 period. The variables are defined as:

- grad = 1 if the student received a BA degree, zero otherwise.
- satm100 and satv100 are SAT scores divided by 100.
- hsgpa = high-school GPA.
- irdr = inverted admission office reader rating, on a scale with 5 being the top and 1 the bottom. (Note that this measure includes their assessment of SAT and high-school record.)
- humfresh = 1 if the student enrolled in Hum 110 as a freshman during his or her first full-time semester (to exclude transfers, part-time students, visiting students, etc.)

The Stata probit command reports the β coefficients of the probit function $\Pr[y=1|x]=G[\beta_1+\beta_2 x_2+\ldots]$, where G is the cumulative normal probability function. The dprobit command reports the "partial effects" $\left.\frac{\partial \Pr[y=1]}{\partial x_j}\right|_{x_j=\bar{x}_j} = \beta_j G'[\beta_1+\beta_2\bar{x}_2+\ldots]$.

```
. summarize grad satm100 satv100 hsgpa irdr if e(sample)

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
        grad |      2230    .6941704    .4608613          0          1
     satm100 |      2230    6.445291     .717163        3.8          8
     satv100 |      2230    6.824126    .7158408        3.6          8
       hsgpa |      2230    3.665368    .3823644      2.294      4.933
        irdr |      2230    3.274865     .500862        1.5          5

. probit grad satm100 satv100 hsgpa irdr if humfresh

Iteration 0:   log likelihood = -1373.0625
Iteration 1:   log likelihood = -1340.2387
Iteration 2:   log likelihood = -1340.1944
Iteration 3:   log likelihood = -1340.1944

Probit regression                                 Number of obs   =       2230
                                                  LR chi2(4)      =      65.74
                                                  Prob > chi2     =     0.0000
Log likelihood = -1340.1944                       Pseudo R2       =     0.0239

------------------------------------------------------------------------------
        grad |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     satm100 |   .0818228   .0429514     1.91   0.057    -.0023604    .1660059
     satv100 |   .0240013   .0431685     0.56   0.578    -.0606074     .10861
       hsgpa |   .3064659   .0893162     3.43   0.001     .1314094    .4815224
        irdr |   .2144542   .0735848     2.91   0.004     .0702307    .3586777
       _cons |  -1.995731   .3894392    -5.12   0.000    -2.759018   -1.232445
```

```
------------------------------------------------------------------------------


. dprobit grad satm100 satv100 hsgpa irdr if humfresh

Iteration 0:   log likelihood = -1373.0625
Iteration 1:   log likelihood = -1340.2387
Iteration 2:   log likelihood = -1340.1944
Iteration 3:   log likelihood = -1340.1944

Probit regression, reporting marginal effects           Number of obs =   2230
                                                        LR chi2(4)    =  65.74
                                                        Prob > chi2   = 0.0000
Log likelihood = -1340.1944                             Pseudo R2     = 0.0239

------------------------------------------------------------------------------
    grad |      dF/dx   Std. Err.      z    P>|z|     x-bar  [    95% C.I.   ]
---------+--------------------------------------------------------------------
 satm100 |   .0284991   .0149558     1.91   0.057   6.44529  -.000814  .057812
 satv100 |   .0083597   .0150356     0.56   0.578   6.82413   -.02111  .037829
   hsgpa |   .1067429   .0310972     3.43   0.001   3.66537   .045793  .167692
    irdr |    .074695   .0256149     2.91   0.004   3.27487   .024491  .124899
---------+--------------------------------------------------------------------
  obs. P |   .6941704
 pred. P |   .6988331  (at x-bar)
------------------------------------------------------------------------------
    z and P>|z| correspond to the test of the underlying coefficient being 0
```

1. The summarize command at the top was run after the probit commands. What does the if e(sample) clause in this statement do?

2. Why does Stata report *z* statistics rather than *t* statistics in these regressions? What distribution do they follow? Why are they identical between the probit and dprobit commands?

3. Other things being equal, how much would an increase of 0.5 units (about one standard deviation) in reader rating change the probability of a student completing a Reed degree? Around what values of the regressors are you evaluating this change? Would this effect be larger or smaller at the tails than at the means of the distribution? Why?

4. Assess the overall regression. What can we infer from it about what admission variables help predict Reed students' graduation rate during this period?