UHCDA15: Binary Regression Model + Testing & Fit

Your Name: Points received: ____ out of 120

The following questions query information & interpretations from logit model output, post-estimation commands, and tests of significance. Please answer in complete sentences, to the best of your ability.

Model overview:

Using data from the U.S. General Social Survey, I am examining predictors of confidence in business. The GSS asked respondents to report on their level of confidence in major corporations. Response options were as follows: (1) Great Deal; (2) Some; (3) Hardly Any. I opt to dichotomize this measure, separating respondents into those who have a great deal of confidence in corporations, and those who do not. I also select three variables I believe will distinguish between those with a great deal of confidence in corporations & those without; all four are summarized below:

```
. codebook conbusB male faminc age satjob, compact

Variable    Obs Unique      Mean  Min  Max  Label
-------------------------------------------------------------------------------
conbusB    5412      2  .1108647    0    1  Great confidence in business?
male       5412      2  .4929786    0    1  Is R male? (1=yes)
faminc     5412     25  41.40812   .5  110  Family income in thousands
age        5412     69  40.03548   18   89  Respondent's age
satjob     5189      4   3.29948    1    4  Satisfied with job or housework?
-------------------------------------------------------------------------------

. sum conbusB male faminc age

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
     conbusB |      5,412    .1108647    .3139936          0          1
        male |      5,412    .4929786    .4999969          0          1
      faminc |      5,412    41.40812    25.47944         .5        110
         age |      5,412    40.03548      12.345         18         89
```

I run a `logit` model predicting great confidence in business:

```
. logit conbusB i.male faminc age, nolog

Logistic regression                             Number of obs     =      5,412
                                                LR chi2(3)        =      47.53
                                                Prob > chi2       =     0.0000
Log likelihood = -1861.3423                     Pseudo R2         =     0.0126

------------------------------------------------------------------------------
     conbusB |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |
      1_Yes  |   .2762145   .0877608     3.15   0.002     .1042065    .4482224
      faminc |  -.0114012   .0018924    -6.02   0.000    -.0151102   -.0076922
         age |   .0084153   .0034076     2.47   0.014     .0017365    .0150942
       _cons |  -2.119459   .1574115   -13.46   0.000     -2.42798   -1.810938
------------------------------------------------------------------------------
```

I also use `listcoef` to compute the factor change coefficients:

```
. listcoef

logit (N=5412): Factor change in odds

  Odds of: 1Great vs 0NotSoGreat

------------------------------------------------------------------------------
             |         b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
        male |
      1_Yes  |    0.2762    3.147    0.002     1.318     1.148     0.500
      faminc |   -0.0114   -6.025    0.000     0.989     0.748    25.479
         age |    0.0084    2.470    0.014     1.008     1.109    12.345
    constant |   -2.1195  -13.464    0.000         .         .         .
------------------------------------------------------------------------------
```

1. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for age. Use the z-statistic from the logit output to test if age significantly impacts conbusB; include this information in your interpretation.
2. ___ of 10: Interpret the appropriate standardized and unstandardized factor change coefficient(s) for male. Use the z-statistic from the logit output to test if male significantly impacts conbusB. Include this information in your write-up.
3. ___ of 10: Imagine I had estimated a probit model for this same model rather than a logit model.
    a) Which unstandardized coefficients would have been larger: logit or probit? How much larger would they have been & why?
    b) What would the ratio between the z-statistics for logit:probit likely have been? Why?

I also test the significance of male on confidence in business using a Wald test:

```
. test 1.male

 ( 1)  [conbusB]1.male = 0

           chi2(  1) =    9.91
         Prob > chi2 =    0.0016
```

4. ___ of 5:
    a) Write up a sentence with your conclusion about the significance of male on conbusB.
    b) How is the specific value of the Wald test related to the z-test in question 2?

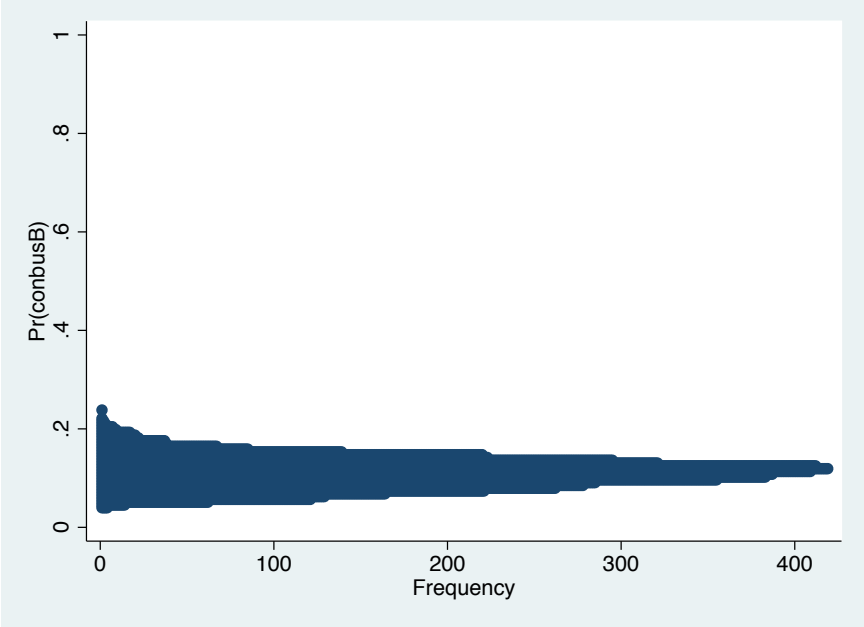
I also decide to test that all coefficients in my model are simultaneously equal to 0, using a likelihood-ratio test.

```
. qui logit conbusB i.male faminc age

. est sto full

. qui logit conbusB

. est sto empty

. lrtest empty full

Likelihood-ratio test                                 LR chi2(3)  =     47.53
(Assumption: empty nested in full)                    Prob > chi2 =    0.0000
```

5. ___ of 10:
    a) Using formal notation, write out the null & alternative hypotheses tested by this LR test.
    b) Write a sentence indicating your conclusion based on the LR test.

Using the predict command, I now estimate my in-sample predicted probabilities & plot them with a dotplot, shown below:

```
. predict pr
(option pr assumed; Pr(conbusB))


. dotplot pr, ylabel(0(.2)1)
```



6. ___of 10: What substantive insights do you gain from this dotplot? What portions of the S-shaped curve are represented in our data? What challenges might this model hold?

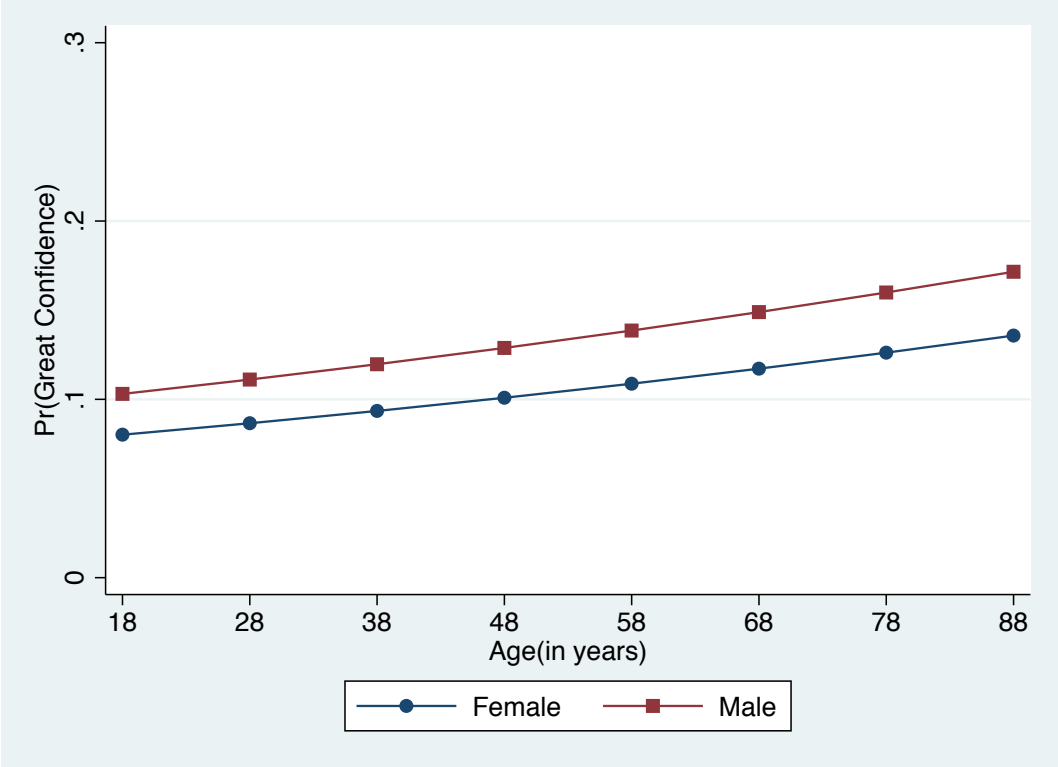
Next, I examine my discrete change coefficients using `mchange`:

```
. mchange, atmeans

logit: Changes in Pr(y) | Number of obs = 5412

Expression: Pr(conbusB), predict(pr)

               |    Change    p-value
---------------+----------------------
male           |
 1 Yes vs 0 No |     0.027      0.002
faminc         |
       +1 cntr |    -0.001      0.000
      +SD cntr |    -0.028      0.000
      Marginal |    -0.001      0.000
age            |
       +1 cntr |     0.001      0.013
      +SD cntr |     0.010      0.013
      Marginal |     0.001      0.013

Predictions at base value

             | 0NotSoG~t     1Great
-------------+----------------------
  Pr(y|base) |     0.893      0.107

Base values of regressors

             |         1.
             |      male     faminc        age
-------------+----------------------------------
          at |      .493       41.4         40

1: Estimates with margins option atmeans.
```

7. ___ of 10: Choose an appropriate discrete change coefficient for male from the `mchange` output and interpret it. Include information on significance.

8. ___ of 10: Choose an appropriate discrete change coefficient for age from the `mchange` output and interpret it. Include information on significance.

Finally, I graph my predicted probabilities over age, for both males & females.



9. ___ of 10: Write a paragraph telling the **story** of your results. This should read as though it were part of a journal article. Incorporate the magnitude of the effects. Also make sure to indicate the levels of any other variables in your model.

10. ___ of 5: After seeing the results plotted, are there other modeling techniques you might consider using for your model?

11. ___ of 10: Looking back on this assignment & course lectures, which method(s) of interpretation did you find most useful (factor change, discrete change, plotting, some combination)? Why? Which do you find least useful?

Finally: in an attempt to explain more variation in my dependent variable, I estimate a series of models with the same dependent variable. Four such models are presented below:

```
. esttab m*, mti aic bic
```

	(1) m1	(2) m2	(3) m3	(4) m4
conbusB				
1.male	0.287**	0.256**	0.257**	0.299***
	(3.26)	(2.87)	(2.82)	(3.38)
faminc	-0.0131***	-0.0113***	-0.0111***	-0.0126***
	(-6.68)	(-5.93)	(-5.58)	(-6.00)
age	0.0909***	0.00984**	0.00885*	0.0877***
	(4.12)	(2.83)	(2.42)	(3.94)

agesq	-0.000928***			-0.000884***
	(-3.75)			(-3.54)
2.relig		-0.204	-0.195	
		(-1.79)	(-1.67)	
3.relig		0.613*	0.672*	
		(2.17)	(2.36)	
4.relig		0.385**	0.436***	
		(2.98)	(3.33)	
5.relig		0.360	0.357	
		(1.86)	(1.74)	
1.didvote			-0.0127	
			(-0.13)	
1.white				-0.522***
				(-5.02)
1.ms_mar				0.0895
				(0.93)
_cons	-3.741***	-2.214***	-2.184***	-3.345***
	(-8.14)	(-13.27)	(-12.42)	(-7.15)
N	5412	5412	5412	5412
AIC	3717.0	3681.0	3521.2	3696.6
BIC	3750.0	3733.7	3580.1	3742.8

t statistics in parentheses
* p<0.05, ** p<0.01, *** p<0.001

12. __ of 10:

    a) Based on the BIC statistic, which model is preferred and how strong is the evidence?

    b) Does AIC give you the same conclusion? If not, why might this be?

13. ___ of 10: My overall evaluation of your work.