UHCDA15 A3: ORM+MNLM

Your Name: Points received: ____ out of 120

In this assignment, you will interpret, discuss & compare estimates from an ordinal logit model & a multinomial logit model run on the same data.

The outcome variable for both models is a four-category indicator of self-reported health, with the following marginal distribution:

```
. tab health, m

  Condition |
  of health |      Freq.     Percent        Cum.
------------+-----------------------------------
      1Poor |         57        1.58        1.58
      2Fair |        488       13.56       15.15
      3Good |      1,799       50.00       65.15
   4Excllnt |      1,254       34.85      100.00
------------+-----------------------------------
      Total |      3,598      100.00
```

Five predictor variables are included in the model. These variables are summarized below:

```
. codebook male age jobsei respinc ms_mar, compact

Variable    Obs Unique      Mean  Min  Max  Label
-------------------------------------------------------------------------------------------
male       3598      2  .5027793    0    1  Is R male? (1=yes)
age        3598     67  40.08449   18   89  Respondent's age
jobsei     3598     58  44.46276   17   86  Occupational prestige
respinc    3598     25  27.84512   .5  110  R's income in thousands
ms_mar     3598      2  .5166759    0    1  R is married (1=yes)
-------------------------------------------------------------------------------------------

. sum male age jobsei respinc ms_mar

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        male |      3,598    .5027793    .5000618          0          1
         age |      3,598    40.08449    12.21044         18         89
      jobsei |      3,598    44.46276      13.591         17         86
     respinc |      3,598    27.84512      21.388         .5        110
      ms_mar |      3,598    .5166759    .4997913          0          1
```

Next, I estimate my ordinal logit model (OLM) of gender, age, occupational prestige, income and marital status on self-reported health:

```
. ologit health i.male age jobsei respinc i.ms_mar, nolog

Ordered logistic regression                     Number of obs     =      3,598
                                                LR chi2(5)        =     172.75
                                                Prob > chi2       =     0.0000
Log likelihood = -3693.5679                     Pseudo R2         =     0.0229

------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
        male |
      1_Yes  |  -.1025382   .0679107    -1.51   0.131    -.2356408    .0305643
```

```
         age |  -.0182334   .0027769    -6.57   0.000     -.023676   -.0127908
      jobsei |   .0163425   .0026004     6.28   0.000     .0112459    .0214391
     respinc |   .0123102   .0017846     6.90   0.000     .0088125    .0158079
             |
      ms_mar |
      1_Yes  |   .1532105   .0657576     2.33   0.020      .024328    .2820931
-------------+----------------------------------------------------------------
       /cut1 |  -3.848634   .2006875                     -4.241974   -3.455293
       /cut2 |  -1.410277   .1564028                     -1.716821   -1.103733
       /cut3 |   1.024395   .1549382                      .7207214    1.328068
------------------------------------------------------------------------------
```

The output from `listcoef` gives me the following results:

```
. listcoef, help

ologit (N=3598): Factor change in odds

  Odds of: >m vs <=m

------------------------------------------------------------------------
             |         b         z     P>|z|       e^b   e^bStdX    SDofX
-------------+----------------------------------------------------------
        male |
      1_Yes  |   -0.1025    -1.510    0.131     0.903     0.950     0.500
         age |   -0.0182    -6.566    0.000     0.982     0.800    12.210
      jobsei |    0.0163     6.285    0.000     1.016     1.249    13.591
     respinc |    0.0123     6.898    0.000     1.012     1.301    21.388
             |
      ms_mar |
      1_Yes  |    0.1532     2.330    0.020     1.166     1.080     0.500
------------------------------------------------------------------------
       b = raw coefficient
       z = z-score for test of b=0
   P>|z| = p-value for z-test
     e^b = exp(b) = factor change in odds for unit increase in X
 e^bStdX = exp(b*SD of X) = change in odds for SD increase in X
   SDofX = standard deviation of X
```

1. ___ of 5: Using the factor change coefficient, describe the effect of one of the continuous variables in the model as though it is part of a published paper. Use either standardized or unstandardized coefficients, and include information on statistical significance.

2. ___ of 5: Using the factor change coefficient, describe the effect of one of the dummy (binary) variables as though it is part of a published paper. Include information on statistical significance.

Next, I use `mchange` to compute my discrete change coefficients:

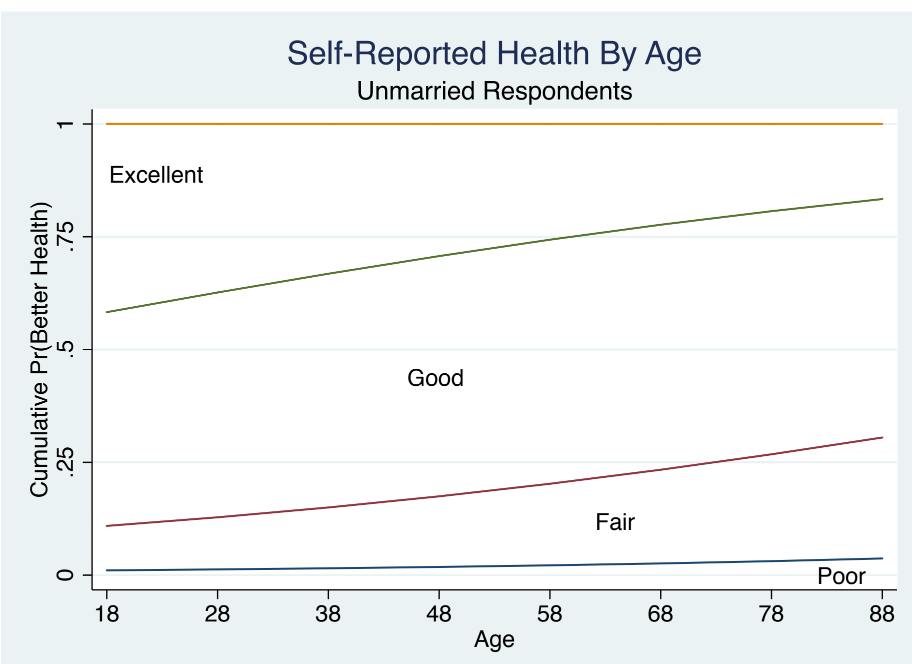
```
. mchange, atmeans

ologit: Changes in Pr(y) | Number of obs = 3598

Expression: Pr(health), predict(outcome())

               |     1Poor      2Fair      3Good   4Exclnt
---------------+--------------------------------------------
male           |
 1 Yes vs 0 No |     0.001      0.011      0.010    -0.023
       p-value |     0.138      0.132      0.132     0.131
```

```
age                 |
         +1 cntr |     0.000       0.002       0.002      -0.004
         p-value |     0.000       0.000       0.000       0.000
        +SD cntr |     0.003       0.024       0.022      -0.050
         p-value |     0.000       0.000       0.000       0.000
        Marginal |     0.000       0.002       0.002      -0.004
         p-value |     0.000       0.000       0.000       0.000
jobsei              |
         +1 cntr |    -0.000      -0.002      -0.002       0.004
         p-value |     0.000       0.000       0.000       0.000
        +SD cntr |    -0.003      -0.024      -0.022       0.050
         p-value |     0.000       0.000       0.000       0.000
        Marginal |    -0.000      -0.002      -0.002       0.004
         p-value |     0.000       0.000       0.000       0.000
respinc             |
         +1 cntr |    -0.000      -0.001      -0.001       0.003
         p-value |     0.000       0.000       0.000       0.000
        +SD cntr |    -0.004      -0.029      -0.026       0.059
         p-value |     0.000       0.000       0.000       0.000
        Marginal |    -0.000      -0.001      -0.001       0.003
         p-value |     0.000       0.000       0.000       0.000
ms mar              |
 1 Yes vs 0 No |    -0.002      -0.017      -0.015       0.034
         p-value |     0.026       0.020       0.020       0.020
```

3. ___ of 5: Describe the discrete change for marital status as though it were part of a published paper. It is not necessary to interpret the discrete change for all categories of the dependent variable (although you should interpret more than just one). Instead of listing the effect of marital status on each outcome category, paint an overall picture of the effect. Include information on statistical significance in your interpretation.

4. ___ of 5: Choose **one** appropriate form of the discrete change coefficients (e.g., change from min to max, unit change, standard deviation change) for age from the output in 7 above. Describe the discrete change for C as though it were part of a published paper. Include information on statistical significance in your interpretation.

Finally, I plot my cumulative probabilities of self-reported health across age, for non-married individuals:



5. ___ of 10: Using the plot above, describe the effect of age on Pr(Y). This should read as though it were part of a published paper and should include the magnitude of effects. It is not necessary to list or interpret the effect of age on each category of the dependent variable. Rather, paint an overall picture of the effect.

Finally, before I send my article off for publication, I run a brant test of the parallel regression assumption:

```
. brant

Brant test of parallel regression assumption

              |       chi2     p>chi2      df
 -------------+------------------------------
          All |      38.91      0.000      10
 -------------+------------------------------
       1.male |      14.46      0.001       2
          age |       9.75      0.008       2
       jobsei |       3.92      0.141       2
      respinc |      14.52      0.001       2
     1.ms_mar |       3.36      0.186       2

A significant test statistic provides evidence that the parallel
regression assumption has been violated.
```

6. ___ of 10: Describe the results of the brant test as though it were part of a published paper. Report on both the omnibus test & specific variables.

7. ___ of 5: In substantive terms, explain what it means to violate the parallel regression/proportional odds assumption.

8. ___ of 15: In 2002 the International Social Survey Program fielded the "Family and Changing Gender Roles III" module that included the following question: "A working mother can establish just as warm and secure a relationship with her children as a mother who does not work." The response categories are: 1. Strongly agree; 2. Agree; 3. Neither agree nor disagree; 4. Disagree; 5. Strongly disagree. However, the codebook reports that in the US the category '4. Disagree' is omitted. Give simple and direct answers to the following:

a) Assume that the question was asked with all categories presented, but that a programming error led to categories 4 and 5 being combined. Would it be reasonable to compare the factor change coefficients from a model estimated in the US to those from other countries where the categories were not collapsed? What about the discrete change coefficients?

b) Now assume that the error was made by forgetting to include 'disagree' on the survey itself (i.e., respondents did not get a chance to choose disagree.). Would it be reasonable to compare the factor change coefficients from a model estimated in the US to those from other countries? What about the discrete change coefficients?

As the model shows a violation of the parallel regression assumption, I decide to estimate the same model using the MNLM. Results of this model are shown below:

```
. mlogit health i.male age jobsei respinc i.ms_mar, base(1) nolog

Multinomial logistic regression                 Number of obs     =      3,598
                                                LR chi2(15)       =     210.77
                                                Prob > chi2       =     0.0000
Log likelihood = -3674.5582                     Pseudo R2         =     0.0279
```

```
------------------------------------------------------------------------------
      health |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1Poor        |  (base outcome)
-------------+----------------------------------------------------------------
2Fair        |
        male |
      1_Yes  |   .4226944   .2974852     1.42   0.155     -.160366    1.005755
         age |  -.0224624   .0095428    -2.35   0.019     -.041166   -.0037588
      jobsei |  -.0186605   .0112409    -1.66   0.097    -.0406923    .0033713
     respinc |   .0259101   .0109954     2.36   0.018     .0043594    .0474607
             |
      ms_mar |
      1_Yes  |   .5232183   .2937238     1.78   0.075    -.0524698    1.098906
       _cons |   2.967975   .6462752     4.59   0.000     1.701299    4.234651
-------------+----------------------------------------------------------------
3Good        |
        male |
      1_Yes  |  -.0184407   .2872515    -0.06   0.949    -.5814434    .5445619
         age |  -.0399795   .0091812    -4.35   0.000    -.0579742   -.0219848
      jobsei |  -.0061899   .0107623    -0.58   0.565    -.0272836    .0149038
     respinc |   .0405326   .0107142     3.78   0.000      .019533    .0615321
             |
      ms_mar |
      1_Yes  |   .6347105   .2835946     2.24   0.025     .0788753    1.190546
       _cons |   4.272586   .6221958     6.87   0.000     3.053105    5.492068
-------------+----------------------------------------------------------------
4Excllnt     |
        male |
      1_Yes  |   .0776762   .2899189     0.27   0.789    -.4905544    .6459069
         age |   -.048591   .0093743    -5.18   0.000    -.0669642   -.0302177
      jobsei |   .0080143   .0108537     0.74   0.460    -.0132586    .0292872
     respinc |   .0466032   .0107534     4.33   0.000     .0255269    .0676794
             |
      ms_mar |
      1_Yes  |   .7145632   .2861763     2.50   0.013      .153668    1.275459
       _cons |    3.34013    .629062     5.31   0.000     2.107191    4.573069
------------------------------------------------------------------------------
```

9. ___ of 10: Using this minimal set of coefficients, compute the coefficients for the following effects:

a) The effect of male on Excellent compared to Good.

b) The effect of age on Fair compared to Excellent

c) The effect of jobsei on Good compared to Fair.

Next, I use `mlogtest, wald` to test the significance of the variables in my model:

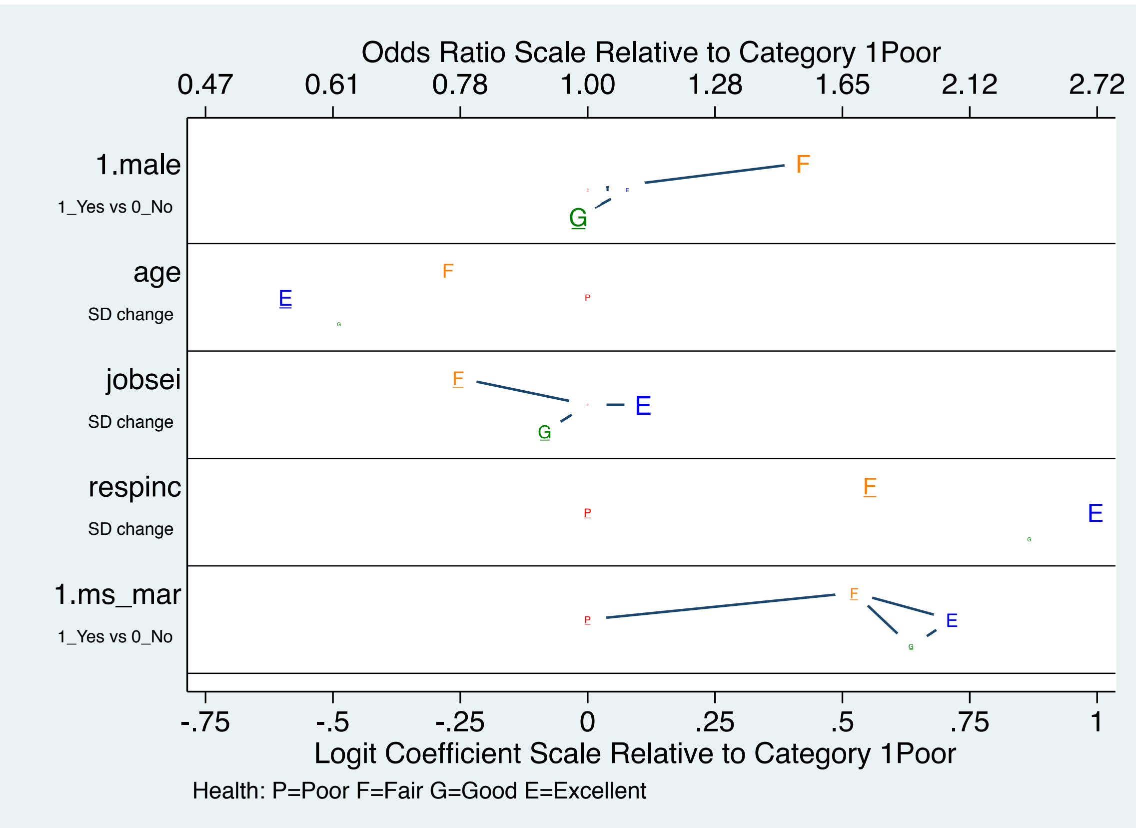
```
. mlogtest, wald

Wald tests for independent variables (N=3598)

  Ho: All coefficients associated with given variable(s) are 0

                 |       chi2     df   P>chi2
-----------------+-------------------------
          1.male |     16.468      3    0.001
             age |     52.354      3    0.000
          jobsei |     41.629      3    0.000
```

```
     respinc |     52.900      3     0.000
    1.ms_mar |      8.201      3     0.042
```

10. ___ of 5: Write out the null hypothesis tested for the effect of age by the test above.

11. ___ of 5: Write up the effect of age as though it were part of a published paper.

Finally, I create my mlogitplot (including discrete change) for my model (note some letters are *very* small!)



12. ___ of 10: Using the plot above, describe the impact of two variables (your choice!) on self-reported health. Try to include information on both changes in odds and discrete changes.

13. ___ of 10: What do you learn substantively from the information on discrete changes that was not clear from the information on odds ratios? What do you learn substantively from the information on odds ratios that was not clear from the information on discrete changes?

14. ___ of 10: Answer the following questions, in 2-3 sentences each:
    a) Based on your results from the OLM, MNLM and Brant test, which model specification for self-reported health do you prefer? Why?
    b) How did the MNLM's relaxing of the proportional odds assumption change understanding of the relationship between your X variables and self-reported health?

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