S. N. Smith
ICPSRCDA-02BRM-Pt1
Last updated: 4 July 2017

1. Choose one of the data sets available for the class. Pick a binary dependent variable and three independent variables. Include output that verifies that all cases with missing data for any of the variables in the model have been dropped.

**Select dataset

```
cda_scireview3
```

**Select variables

```
*DV: faculty
*Continuous IV (C): mcit3
*Dichotomous IV (D): fellow
*Other IV (X): phd
```

2. Demonstrate that the data are clean.

```
. codebook faculty fellow phd mcit3 mnas, compact

Variable   Obs Unique      Mean  Min   Max  Label
-------------------------------------------------------------------------------
faculty    264      2  .5340909    0     1  Faculty in Univ? (1=yes)
fellow     264      2  .4128788    0     1  Postdoctoral fellow? (1=yes)
phd        264     79  3.181894    1  4.66  Prestige of Ph.D. department.
mcit3      264     59  20.71591    0   129  Mentor's 3 yr citation.
mnas       264      2  .0833333    0     1  Was mentor in NAS? (1=yes)
-------------------------------------------------------------------------------

. sum faculty fellow phd mcit3 mnas

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
     faculty |       264    .5340909    .4997839          0          1
      fellow |       264    .4128788    .4932865          0          1
         phd |       264    3.181894     1.00518          1       4.66
       mcit3 |       264    20.71591    25.44536          0        129
        mnas |       264    .0833333    .2769103          0          1
```

3. Estimate a logit of Y on C, D, and X.

```
. logit faculty mcit3 i.fellow phd

Iteration 0:   log likelihood = -182.37674
Iteration 1:   log likelihood = -163.99038
Iteration 2:   log likelihood = -163.77501
Iteration 3:   log likelihood = -163.77427
Iteration 4:   log likelihood = -163.77427

Logistic regression                             Number of obs   =        264
                                                LR chi2(3)      =      37.20
                                                Prob > chi2     =     0.0000
Log likelihood = -163.77427                     Pseudo R2       =     0.1020
```

```
------------------------------------------------------------------------------
     faculty |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       mcit3 |   .0212656   .0071144     2.99   0.003     .0073216    .0352097
             |
      fellow |
      1_Yes  |   1.265773   .2758366     4.59   0.000     .7251437    1.806403
         phd |  -.0439657    .144072    -0.31   0.760    -.3263416    .2384102
       _cons |  -.6344166   .4425034    -1.43   0.152    -1.501707     .232874
------------------------------------------------------------------------------
```

4. Compute the factor change coefficients & construct a table with the logit coefficients, the appropriate standardized and unstandardized factor change coefficients, and the z-values.

```
. listcoef, help

logit (N=264): Factor change in odds

  Odds of: 1_Yes vs 0_No

------------------------------------------------------------------------------
             |          b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
       mcit3 |     0.0213    2.989    0.003     1.021     1.718    25.445
             |
      fellow |
      1_Yes  |     1.2658    4.589    0.000     3.546     1.867     0.493
         phd |    -0.0440   -0.305    0.760     0.957     0.957     1.005
    constant |    -0.6344   -1.434    0.152         .         .         .
------------------------------------------------------------------------------
       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
```

Table 1: Logit coefficients for variables of interest on total number of publications (N=264)

	B	**E^{B}**	**E^{Bx}**	**Z**
Had a fellowship?	1.258***	3.546	--	4.589
PhD prestige	-0.044	0.957	0.957	-0.305
Mentor's citations	0.021***	1.021	1.718	2.989
Intercept	-0.634	--	--	--

Note: * p < 0.05; ** p < 0.01; *** p < 0.001

5. Interpret the appropriate standardized and unstandardized factor change coefficient(s) for C; include statistical significance in your interpretation.

```
------------------------------------------------------------------------------
             |          b        z    P>|z|       e^b   e^bStdX     SDofX
-------------+----------------------------------------------------------------
       mcit3 |     0.0213    2.989    0.003     1.021     1.718    25.445
             |
      fellow |
      1_Yes  |     1.2658    4.589    0.000     3.546     1.867     0.493
         phd |    -0.0440   -0.305    0.760     0.957     0.957     1.005
------------------------------------------------------------------------------
```

Unstandardized: For each additional mentor citation, the odds of obtaining a faculty position increase by a factor of 1.02, holding all other variables constant. This effect is significant at the p<0.01 level (z = 2.989).

Standardized: For a standard deviation increase in mentor's citations (about 25 citations), the odds of obtaining a faculty position increase by a factor of 1.72, holding all other variables constant. This effect is significant at the p<0.01 level (z = 2.989).

6. Interpret the appropriate standardized and unstandardized factor change coefficient(s) for B; include statistical significance in your interpretation.

	b	z	P>|z|	e^b	e^bStdX	SDofX
mcit3	0.0213	2.989	0.003	1.021	1.718	25.445
fellow						
1_Yes	1.2658	4.589	0.000	**3.546**	1.867	0.493
phd	-0.0440	-0.305	0.760	0.957	0.957	1.005

Unstandardized: All else being equal, individuals who have held a fellowship have odds of holding a faculty position that are 3.55 higher than non-fellows. This effect is significant at the p<.001 level (z = 4.589).

7.

(a) Use `margins` to compute the Pr(y=1) when B is equal to 1 and the other variables are held at their means.

```
. margins, at(fellow=1) atmeans

Adjusted predictions                              Number of obs   =        264
Model VCE    : OIM

Expression   : Pr(faculty), predict()
at           : mcit3           =    20.71591 (mean)
               fellow          =           1
               phd             =    3.181894 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |   .7174807   .0443355    16.18   0.000     .6305847    .8043767
------------------------------------------------------------------------------
```

(b) Use `margins` to compute the Pr(y=1) when B is equal to 0 and the other variables are held at their means.

```
. margins, at(fellow=0) atmeans

Adjusted predictions                              Number of obs   =        264
Model VCE    : OIM

Expression   : Pr(faculty), predict()
at           : mcit3           =    20.71591 (mean)
```

```
               fellow          =           0
               phd             =    3.181894  (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .4173226   .0411165    10.15   0.000     .3367356    .4979095
------------------------------------------------------------------------------
```

(c) Calculate the factor change coefficient for B using these values (you can do this by hand or by using the display command in Stata).

```
. display ((0.7175/(1-0.7175))/(0.4173/(1-0.4173)))
3.546501
```

(d) Repeat steps a-c holding the other variables at some value other than their mean.

```
. **B=1; max of other vars
. margins, at(fellow=1 phd=4.66 mcit3=129)

Adjusted predictions                              Number of obs   =        264
Model VCE    : OIM

Expression   : Pr(faculty), predict()
at           : mcit3           =         129
               fellow          =           1
               phd             =        4.66

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .9596795   .0299394    32.05   0.000     .9009994     1.01836
------------------------------------------------------------------------------


. **B=0; max of other vars
. margins, at(fellow=0 phd=4.66 mcit3=129)

Adjusted predictions                              Number of obs   =        264
Model VCE    : OIM

Expression   : Pr(faculty), predict()
at           : mcit3           =         129
               fellow          =           0
               phd             =        4.66

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .8703399   .0853233    10.20   0.000     .7031093     1.03757
------------------------------------------------------------------------------


. **compute factor change
. display ((0.9597/(1-0.9597)) / (0.8703/(1-0.8703)))
3.5489628
```

(e) What do you find?

<<Your original & appropriate text here>>

: : : Assignment continues from here : : :