**Running Poisson & Negative Binomial Regression with Stata**

In this Computer Lab, we will run various examples to understand (1) the setup of a Poisson & negative binomial regression models, (2) comparisons between the Poisson regression and the negative binomial regression, (3) producing marginal effects (AME, MEM, and MER) with Stata and producing MEMs and MERs with Excel, and (4) interpretation of results from the negative binomial regression model.

**The Stata Syntax for All Examples:**

**cd C:\6910\Lab10**

**set more off**

**log using lab10, replace**

**//Example 1: Doctoral publications**

**//Compare Poisson with negbin**

**use couart4, clear**

**poisson art i.female i.married kid5 phd mentor**

**estimate store prm**

**nbreg art i.female i.married kid5 phd mentor**

**estimate store negbin**

**display 2\*((-1560.9583)-(-1651.0563))**

**display chi2tail(1,180.196)**

**estimate table prm negbin, b(%9.3f) se p(%9.3f) varlabel eform**

**//Example 2: Robust SE and exp(B)**

**// Always use robust SE for the final model**

**nbreg art i.female i.married kid5 phd mentor, vce(robust)**

**//Obtain irr or exp(B)**

**listcoef, help**

**listcoef, percent help**

**//Example3: Predicted probabilities (MEMs, MERs, AMEs)**

**//Obtain means to create predicted probabilities**

**sum female married kid5 phd mentor**

**//MEMs**

**mtable, atmeans pr(0/5)**

**//MERs to see the effect of mentor's publications**

**mtable, at(mentor=(1 2 3 4 5 6)) atmeans pr(0/5)**

**//AMEs**

**predict prob0, pr(0)**

**predict prob1, pr(1)**

**predict prob2, pr(2)**

**predict prob3, pr(3)**

**predict prob4, pr(4)**

**sum prob0-prob4**

**mchange, pr(0/4)**

**//Example 4: Draw a line chart to show group differences on a continuous variable**

**tab mentor**

**margins i.female, at(mentor=(0(10)100)) atmeans noatlegend**

**marginsplot, noci title("The Impact of Mentor's publications on the Expected" "Number of Publications by Gender") ///**

**ytitle("Predicted/Expected Number of Publications")**

**//Example 5: Test an interaction model**

**tab mentor**

**tab phd**

**gen mentor\_h=0**

**replace mentor\_h=1 if mentor > 20**

**label define h 0 "Mentor Pub below 20" 1 "Mentor Pub > 20"**

**label value mentor\_h h**

**nbreg art i.female i.married kid5 i.mentor\_h i.mentor\_h##c.phd, vce(robust)**

**margins i.mentor\_h, at(phd=(0(.5)5)) atmeans noatlegend**

**marginsplot, noci title("The Interactive Effect of" "Mentor's Productivity & PhD Prestige") ///**

**ytitle("Predicted/Expected Number of Publications")**

**log close**

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

**[Example 1] Doctoral publications**

This example (Long, 1990) studies the number of articles published by biochemists in the 3 years prior to receiving their doctorate. Below we run two models, the Poisson regression and the negbin regression. We than compare the two models.

. //Example 1: Doctoral publications

. //Compare Poisson with negbin

. use couart4, clear

(couart4.dta | Long data on Ph.D. biochemists | 2014-04-24)

. poisson art i.female i.married kid5 phd mentor

Iteration 0: log likelihood = -1651.4574

Iteration 1: log likelihood = -1651.0567

Iteration 2: log likelihood = -1651.0563

Iteration 3: log likelihood = -1651.0563

Poisson regression Number of obs = 915

LR chi2(5) = 183.03

Prob > chi2 = 0.0000

Log likelihood = -1651.0563 Pseudo R2 = 0.0525

------------------------------------------------------------------------------

art | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

female |

Female | -.2245942 .0546138 -4.11 0.000 -.3316352 -.1175532

|

married |

Married | .1552434 .0613747 2.53 0.011 .0349512 .2755356

kid5 | -.1848827 .0401272 -4.61 0.000 -.2635305 -.1062349

phd | .0128226 .0263972 0.49 0.627 -.038915 .0645601

mentor | .0255427 .0020061 12.73 0.000 .0216109 .0294746

\_cons | .3046168 .1029822 2.96 0.003 .1027755 .5064581

------------------------------------------------------------------------------

. estimate store prm

. nbreg art i.female i.married kid5 phd mentor

Fitting Poisson model:

Iteration 0: log likelihood = -1651.4574

Iteration 1: log likelihood = -1651.0567

Iteration 2: log likelihood = -1651.0563

Iteration 3: log likelihood = -1651.0563

Fitting constant-only model:

Iteration 0: log likelihood = -1625.4242

Iteration 1: log likelihood = -1609.9746

Iteration 2: log likelihood = -1609.9368

Iteration 3: log likelihood = -1609.9367

Fitting full model:

Iteration 0: log likelihood = -1565.6652

Iteration 1: log likelihood = -1561.0095

Iteration 2: log likelihood = -1560.9583

Iteration 3: log likelihood = -1560.9583

Negative binomial regression Number of obs = 915

LR chi2(5) = 97.96

Dispersion = mean Prob > chi2 = 0.0000

Log likelihood = -1560.9583 Pseudo R2 = 0.0304

------------------------------------------------------------------------------

art | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

female |

Female | -.2164184 .0726724 -2.98 0.003 -.3588537 -.0739832

|

married |

Married | .1504895 .0821063 1.83 0.067 -.0104359 .3114148

kid5 | -.1764152 .0530598 -3.32 0.001 -.2804105 -.07242

phd | .0152712 .0360396 0.42 0.672 -.0553652 .0859075

mentor | .0290823 .0034701 8.38 0.000 .0222811 .0358836

\_cons | .256144 .1385604 1.85 0.065 -.0154294 .5277174

-------------+----------------------------------------------------------------

/lnalpha | -.8173044 .1199372 -1.052377 -.5822318

-------------+----------------------------------------------------------------

alpha | .4416205 .0529667 .3491069 .5586502

------------------------------------------------------------------------------

LR test of alpha=0: chibar2(01) = 180.20 Prob >= chibar2 = 0.000

. estimate store negbin

The results show that the estimated overdispersion parameter alpha is .4416. Note that the ***nbreg*** output reports results of the LR test of alpha=0, which can be verified as follows:

. display 2\*((-1560.9583)-(-1651.0563))

180.196

. display chi2tail(1,180.196)

4.392e-41

This means that the negbin regression is better than the Poisson regression, we should accept the negbin results other than the Poisson regression.

Below we further compare the estimates between the two models. As the results show, the Poisson regression estimates SEs that are always smaller than those from the negbin. This implies that the Poisson regression leads to biased significance tests, and tends to make non-significant predictors significant.

. estimate table prm negbin, b(%9.3f) se p(%9.3f) varlabel eform

--------------------------------------------------

Variable | prm negbin

-------------------------+------------------------

art |

female |

Female | 0.799 0.805

| 0.044 0.059

| 0.000 0.003

|

married |

Married | 1.168 1.162

| 0.072 0.095

| 0.011 0.067

# of kids < 6 | 0.831 0.838

| 0.033 0.044

| 0.000 0.001

PhD prestige | 1.013 1.015

| 0.027 0.037

| 0.627 0.672

Mentor's # of articles | 1.026 1.030

| 0.002 0.004

| 0.000 0.000

Constant | 1.356 1.292

| 0.140 0.179

| 0.003 0.065

-------------------------+------------------------

lnalpha |

Constant | 0.442

| 0.053

| 0.000

--------------------------------------------------

legend: b/se/p

**[Example 2] Use robust SE and obtain IRRs**

To follow the convention in running negbin, we always use robust estimator of standard errors. This should also be done when you run the Poisson regression. You do this by using ***vce(robust)*** command.

. //Example 2: Robust SE and exp(B)

. // Always use robust SE for the final model

. nbreg art i.female i.married kid5 phd mentor, vce(robust)

Fitting Poisson model:

Iteration 0: log pseudolikelihood = -1651.4574

Iteration 1: log pseudolikelihood = -1651.0567

Iteration 2: log pseudolikelihood = -1651.0563

Iteration 3: log pseudolikelihood = -1651.0563

Fitting constant-only model:

Iteration 0: log pseudolikelihood = -1625.4242

Iteration 1: log pseudolikelihood = -1609.9746

Iteration 2: log pseudolikelihood = -1609.9368

Iteration 3: log pseudolikelihood = -1609.9367

Fitting full model:

Iteration 0: log pseudolikelihood = -1565.6652

Iteration 1: log pseudolikelihood = -1561.0095

Iteration 2: log pseudolikelihood = -1560.9583

Iteration 3: log pseudolikelihood = -1560.9583

Negative binomial regression Number of obs = 915

Wald chi2(5) = 85.49

Dispersion = mean Prob > chi2 = 0.0000

Log pseudolikelihood = -1560.9583 Pseudo R2 = 0.0304

------------------------------------------------------------------------------

| Robust

art | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

female |

Female | -.2164184 .0704667 -3.07 0.002 -.3545306 -.0783063

|

married |

Married | .1504895 .0805541 1.87 0.062 -.0073937 .3083726

kid5 | -.1764152 .0531021 -3.32 0.001 -.2804935 -.0723369

phd | .0152712 .037523 0.41 0.684 -.0582725 .0888148

mentor | .0290823 .0038835 7.49 0.000 .0214709 .0366938

\_cons | .256144 .1402295 1.83 0.068 -.0187007 .5309887

-------------+----------------------------------------------------------------

/lnalpha | -.8173044 .1249058 -1.062115 -.5724935

-------------+----------------------------------------------------------------

alpha | .4416205 .055161 .3457237 .564117

------------------------------------------------------------------------------

Note that Stata does not report the LR test about the alpha when you run ***vce(robust)***. So you need to get the test first by running the model without using ***vce(robust)***.

Below are the syntax and output bout obtaining incidence-rate ratios.

. //Obtain irr or exp(B)

. listcoef, help

nbreg (N=915): Factor change in expected count

Observed SD: 1.9261

------------------------------------------------------------------------

| b z P>|z| e^b e^bStdX SDofX

-------------+----------------------------------------------------------

female |

Female | -0.2164 -3.071 0.002 0.805 0.898 0.499

|

married |

Married | 0.1505 1.868 0.062 1.162 1.074 0.473

kid5 | -0.1764 -3.322 0.001 0.838 0.874 0.765

phd | 0.0153 0.407 0.684 1.015 1.015 0.984

mentor | 0.0291 7.489 0.000 1.030 1.318 9.484

constant | 0.2561 1.827 0.068 . . .

-------------+----------------------------------------------------------

alpha |

lnalpha | -0.8173 . . . . .

alpha | 0.4416 . . . . .

------------------------------------------------------------------------

LR test of alpha=0: . Prob>=LRX2 = .

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 expected count for unit increase in X

e^bStdX = exp(b\*SD of X) = change in expected count for SD increase in X

SDofX = standard deviation of X

. listcoef, percent help

nbreg (N=915): Percentage change in expected count

Observed SD: 1.9261

------------------------------------------------------------------------

| b z P>|z| % %StdX SDofX

-------------+----------------------------------------------------------

female |

Female | -0.2164 -3.071 0.002 -19.5 -10.2 0.499

|

married |

Married | 0.1505 1.868 0.062 16.2 7.4 0.473

kid5 | -0.1764 -3.322 0.001 -16.2 -12.6 0.765

phd | 0.0153 0.407 0.684 1.5 1.5 0.984

mentor | 0.0291 7.489 0.000 3.0 31.8 9.484

constant | 0.2561 1.827 0.068 . . .

-------------+----------------------------------------------------------

alpha |

lnalpha | -0.8173 . . . . .

alpha | 0.4416 . . . . .

------------------------------------------------------------------------

LR test of alpha=0: . Prob>=LRX2 = .

b = raw coefficient

z = z-score for test of b=0

P>|z| = p-value for z-test

% = percent change in expected count for unit increase in X

%StdX = percent change in expected count for SD increase in X

SDofX = standard deviation of X

**Interpretations:**

For a categorical variable (“female”):

* Being a female scientist decreases the expected number of articles by a factor of 0.805, holding other variables constant.
* Being a female scientist decreases the expected number of articles by 19.5%, holding other variables constant.

For a continuous variable (“mentor”):

* For a standard deviation increase in the mentor’s productivity, roughly 10 articles, a scientist’s expected productivity increases by a factor of 1.318, holding other variables constant.
* For every additional article by the mentor, a scientist’s expected productivity increase by 3.0%, holding other variables constant.
* For a standard deviation increase in the mentor’s productivity, roughly 10 articles, a scientist’s expected productivity increases by 31.8%, holding other variables constant.

**[Example 3] Predicted probabilities (MEMs, MERs, & AMEs)**

Below I show the commands and output of obtaining predicted probabilities. There are three types of them: MEMs, MERs, and AMEs.

. //Example3: Predicted probabilities (MEMs, MERs, AMEs)

. //Obtain means to create predicted probabilities

. sum female married kid5 phd mentor

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

female | 915 .4601093 .4986788 0 1

married | 915 .6622951 .473186 0 1

kid5 | 915 .495082 .76488 0 3

phd | 915 3.103109 .9842491 .755 4.62

mentor | 915 8.767213 9.483916 0 77

.

. //MEMs

. mtable, atmeans pr(0/5)

Expression: Pr(art), predict(pr())

0 1 2 3 4 5

----------------------------------------------------------

0.298 0.279 0.189 0.111 0.061 0.031

Specified values of covariates

| 1. 1.

| female married kid5 phd mentor

----------+--------------------------------------------------

Current | .46 .662 .495 3.1 8.77

. //MERs to see the effect of mentor's publications

. mtable, at(mentor=(1 2 3 4 5 6)) atmeans pr(0/5)

Expression: Pr(art), predict(pr())

| mentor 0 1 2 3 4 5

----------+---------------------------------------------------------------------

1 | 1 0.363 0.297 0.175 0.090 0.043 0.019

2 | 2 0.354 0.295 0.177 0.092 0.045 0.021

3 | 3 0.346 0.293 0.179 0.095 0.047 0.022

4 | 4 0.337 0.291 0.181 0.098 0.049 0.024

5 | 5 0.329 0.289 0.183 0.101 0.052 0.025

6 | 6 0.321 0.287 0.185 0.104 0.054 0.027

Specified values of covariates

| 1. 1.

| female married kid5 phd

----------+----------------------------------------

Current | .46 .662 .495 3.1

**Guo developed an Excel file to obtain model predicted probabilities, primarily for MEMs and MERs. Results confirm that the two programs provide exactly the same results.**



**AMEs**

The ***predict*** commands create predicted probabilities for the count equal to 1, 2, … m for each case. Averaging the sample all cases gives the AME, as:

. //AMEs

. predict prob0, pr(0)

. predict prob1, pr(1)

. predict prob2, pr(2)

. predict prob3, pr(3)

. predict prob4, pr(4)

. sum prob0-prob4

Variable | Obs Mean Std. Dev. Min Max

-------------+---------------------------------------------------------

prob0 | 915 .3035957 .0781645 .015145 .4801816

prob1 | 915 .2722666 .037721 .0289042 .3030041

prob2 | 915 .1800492 .0152423 .0397624 .1938989

prob3 | 915 .1063388 .0206307 .0476373 .1439137

prob4 | 915 .0598104 .0214007 .0194697 .1147368

In practice, the easiest way to obtain AMEs is to use ***mchange***, as:

. mchange, pr(0/4)

nbreg: Changes in Pr(y) | Number of obs = 915

Expression: Pr(art), predict(pr())

| 0 1 2 3 4

-------------------+-------------------------------------------------------

female |

Female vs Male | 0.059 0.019 -0.009 -0.018 -0.016

p-value | 0.002 0.002 0.004 0.002 0.002

married |

Married vs Single | -0.041 -0.013 0.007 0.012 0.011

p-value | 0.063 0.049 0.092 0.064 0.057

kid5 |

+1 | 0.049 0.013 -0.010 -0.015 -0.013

p-value | 0.001 0.000 0.009 0.001 0.000

+SD | 0.037 0.010 -0.007 -0.011 -0.010

p-value | 0.001 0.000 0.008 0.001 0.000

Marginal | 0.048 0.015 -0.007 -0.014 -0.013

p-value | 0.001 0.001 0.003 0.001 0.001

phd |

+1 | -0.004 -0.001 0.001 0.001 0.001

p-value | 0.684 0.689 0.678 0.684 0.686

+SD | -0.004 -0.001 0.001 0.001 0.001

p-value | 0.684 0.689 0.678 0.684 0.686

Marginal | -0.004 -0.001 0.001 0.001 0.001

p-value | 0.684 0.685 0.687 0.686 0.685

mentor |

+1 | -0.008 -0.003 0.001 0.002 0.002

p-value | 0.000 0.000 0.000 0.000 0.000

+SD | -0.071 -0.030 0.005 0.019 0.020

p-value | 0.000 0.000 0.003 0.000 0.000

Marginal | -0.008 -0.003 0.001 0.002 0.002

p-value | 0.000 0.000 0.000 0.000 0.000

Average predictions

| 0 1 2 3 4

-------------+-------------------------------------------------------

Pr(y|base) | 0.304 0.272 0.180 0.106 0.060

**[Example 4] Graphic representation of the findings**

Below I show how to present a line chart depicting the impact of a continuous variable on the expected number of articles by group. I use the number of mentor’s publication as the continuous variable, and gender as the group. In order to determine the scale of x-axis in the chart, I first take a look at the distribution of the continuous variable using ***tab mentor***.

. //Example 4: Draw a line chart to show group differences on a continuous variable

. tab mentor

Mentor's # |

of articles | Freq. Percent Cum.

------------+-----------------------------------

0 | 90 9.84 9.84

1 | 52 5.68 15.52

2 | 79 8.63 24.15

3 | 70 7.65 31.80

4 | 72 7.87 39.67

5 | 77 8.42 48.09

6 | 52 5.68 53.77

7 | 44 4.81 58.58

8 | 47 5.14 63.72

9 | 32 3.50 67.21

10 | 32 3.50 70.71

11 | 34 3.72 74.43

12 | 27 2.95 77.38

13 | 21 2.30 79.67

14 | 23 2.51 82.19

15 | 19 2.08 84.26

16 | 16 1.75 86.01

17 | 9 0.98 86.99

18 | 14 1.53 88.52

19 | 14 1.53 90.05

20 | 4 0.44 90.49

21 | 12 1.31 91.80

22 | 6 0.66 92.46

23 | 5 0.55 93.01

24 | 8 0.87 93.88

25 | 7 0.77 94.64

26 | 4 0.44 95.08

27 | 4 0.44 95.52

29 | 3 0.33 95.85

30 | 5 0.55 96.39

31 | 2 0.22 96.61

32 | 1 0.11 96.72

34 | 2 0.22 96.94

35 | 3 0.33 97.27

36 | 2 0.22 97.49

37 | 3 0.33 97.81

38 | 2 0.22 98.03

39 | 2 0.22 98.25

42 | 2 0.22 98.47

45 | 1 0.11 98.58

46 | 1 0.11 98.69

47 | 4 0.44 99.13

48 | 1 0.11 99.23

49 | 1 0.11 99.34

53 | 2 0.22 99.56

55 | 1 0.11 99.67

57 | 1 0.11 99.78

66 | 1 0.11 99.89

77 | 1 0.11 100.00

------------+-----------------------------------

Total | 915 100.00

. margins i.female, at(mentor=(0(10)100)) atmeans noatlegend

Adjusted predictions Number of obs = 915

Model VCE : Robust

Expression : Predicted number of events, predict()

------------------------------------------------------------------------------

| Delta-method

| Margin Std. Err. z P>|z| [95% Conf. Interval]

-------------+----------------------------------------------------------------

\_at#female |

1#Male | 1.371439 .0831134 16.50 0.000 1.20854 1.534338

1#Female | 1.104555 .0701954 15.74 0.000 .966974 1.242135

2#Male | 1.834339 .0886856 20.68 0.000 1.660518 2.008159

2#Female | 1.477373 .0751181 19.67 0.000 1.330144 1.624602

3#Male | 2.45348 .15553 15.77 0.000 2.148647 2.758313

3#Female | 1.976028 .127268 15.53 0.000 1.726588 2.225469

4#Male | 3.281599 .3063646 10.71 0.000 2.681135 3.882063

4#Female | 2.642994 .2469077 10.70 0.000 2.159064 3.126924

5#Male | 4.389232 .5625984 7.80 0.000 3.286559 5.491904

5#Female | 3.535079 .4516418 7.83 0.000 2.649878 4.420281

6#Male | 5.870722 .9674971 6.07 0.000 3.974462 7.766981

6#Female | 4.728269 .7759876 6.09 0.000 3.207361 6.249177

7#Male | 7.852257 1.588123 4.94 0.000 4.739594 10.96492

7#Female | 6.324194 1.273696 4.97 0.000 3.827795 8.820593

8#Male | 10.50262 2.522107 4.16 0.000 5.559378 15.44586

8#Female | 8.458789 2.023161 4.18 0.000 4.493466 12.42411

9#Male | 14.04755 3.909316 3.59 0.000 6.38543 21.70967

9#Female | 11.31387 3.136756 3.61 0.000 5.165942 17.4618

10#Male | 18.78899 5.948762 3.16 0.002 7.129636 30.44835

10#Female | 15.13262 4.774422 3.17 0.002 5.774927 24.49032

11#Male | 25.13081 8.922406 2.82 0.005 7.643219 42.61841

11#Female | 20.24031 7.1628 2.83 0.005 6.201483 34.27914

------------------------------------------------------------------------------

. marginsplot, noci title("The Impact of Mentor's publications on the Expected" "Number of Publications by Gender") ///

> ytitle("Predicted/Expected Number of Publications")

Variables that uniquely identify margins: mentor female



**[Example 5] Test interaction**

Testing interaction is an important procedure in statistical modeling. This often determines by a study’s research questions or conceptual model. Researchers also sometime test significant interaction through a data-driven procedure. In this example, I am interested in the joint effect of PhD program’s prestige and mentor’s number of publications on the study student’s number of publications. Since both variables are continuous, we need to categorize one of the two variables first so that the interaction effects show the impact of a continuous variable on the DV by the level of the categorical variable. We first tabulate the two continuous variables. From the distribution, I choose mentor’s publications as a variable on which I dichotomous it. I use 20 articles as a cutoff: those who published more than 20 articles as high. So the study investigates how the impact of PhD program’s prestige on students’ publication varies by mentor’s high- versus low- publication status. Results of this interaction is very interesting, and important. Note that the interaction is statistically significant.

. //Example 5: Test an interaction model

. tab mentor

Mentor's # |

of articles | Freq. Percent Cum.

------------+-----------------------------------

0 | 90 9.84 9.84

1 | 52 5.68 15.52

2 | 79 8.63 24.15

3 | 70 7.65 31.80

4 | 72 7.87 39.67

5 | 77 8.42 48.09

6 | 52 5.68 53.77

7 | 44 4.81 58.58

8 | 47 5.14 63.72

9 | 32 3.50 67.21

10 | 32 3.50 70.71

11 | 34 3.72 74.43

12 | 27 2.95 77.38

13 | 21 2.30 79.67

14 | 23 2.51 82.19

15 | 19 2.08 84.26

16 | 16 1.75 86.01

17 | 9 0.98 86.99

18 | 14 1.53 88.52

19 | 14 1.53 90.05

20 | 4 0.44 90.49

21 | 12 1.31 91.80

22 | 6 0.66 92.46

23 | 5 0.55 93.01

24 | 8 0.87 93.88

25 | 7 0.77 94.64

26 | 4 0.44 95.08

27 | 4 0.44 95.52

29 | 3 0.33 95.85

30 | 5 0.55 96.39

31 | 2 0.22 96.61

32 | 1 0.11 96.72

34 | 2 0.22 96.94

35 | 3 0.33 97.27

36 | 2 0.22 97.49

37 | 3 0.33 97.81

38 | 2 0.22 98.03

39 | 2 0.22 98.25

42 | 2 0.22 98.47

45 | 1 0.11 98.58

46 | 1 0.11 98.69

47 | 4 0.44 99.13

48 | 1 0.11 99.23

49 | 1 0.11 99.34

53 | 2 0.22 99.56

55 | 1 0.11 99.67

57 | 1 0.11 99.78

66 | 1 0.11 99.89

77 | 1 0.11 100.00

------------+-----------------------------------

Total | 915 100.00

. tab phd

PhD |

prestige | Freq. Percent Cum.

------------+-----------------------------------

.755 | 2 0.22 0.22

.92 | 3 0.33 0.55

1.005 | 1 0.11 0.66

1.18 | 5 0.55 1.20

1.22 | 6 0.66 1.86

1.25 | 6 0.66 2.51

1.255 | 2 0.22 2.73

1.28 | 3 0.33 3.06

1.38 | 1 0.11 3.17

1.4 | 9 0.98 4.15

1.42 | 2 0.22 4.37

1.45 | 1 0.11 4.48

1.48 | 1 0.11 4.59

1.505 | 3 0.33 4.92

1.52 | 8 0.87 5.79

1.63 | 4 0.44 6.23

1.64 | 1 0.11 6.34

1.655 | 1 0.11 6.45

1.67 | 1 0.11 6.56

1.68 | 7 0.77 7.32

1.72 | 2 0.22 7.54

1.74 | 4 0.44 7.98

1.75 | 5 0.55 8.52

1.76 | 9 0.98 9.51

1.78 | 7 0.77 10.27

1.79 | 3 0.33 10.60

1.8 | 5 0.55 11.15

1.81 | 8 0.87 12.02

1.83 | 1 0.11 12.13

1.86 | 12 1.31 13.44

1.89 | 1 0.11 13.55

1.95 | 5 0.55 14.10

1.97 | 5 0.55 14.64

2 | 18 1.97 16.61

2.05 | 8 0.87 17.49

2.1 | 25 2.73 20.22

2.12 | 15 1.64 21.86

2.14 | 3 0.33 22.19

2.15 | 2 0.22 22.40

2.2 | 2 0.22 22.62

2.21 | 10 1.09 23.72

2.25 | 3 0.33 24.04

2.26 | 18 1.97 26.01

2.32 | 15 1.64 27.65

2.36 | 2 0.22 27.87

2.39 | 12 1.31 29.18

2.5 | 12 1.31 30.49

2.51 | 9 0.98 31.48

2.52 | 8 0.87 32.35

2.54 | 8 0.87 33.22

2.55 | 5 0.55 33.77

2.56 | 6 0.66 34.43

2.58 | 17 1.86 36.28

2.61 | 8 0.87 37.16

2.76 | 15 1.64 38.80

2.83 | 8 0.87 39.67

2.86 | 26 2.84 42.51

2.87 | 14 1.53 44.04

2.96 | 31 3.39 47.43

3.09 | 12 1.31 48.74

3.15 | 13 1.42 50.16

3.19 | 22 2.40 52.57

3.21 | 8 0.87 53.44

3.32 | 3 0.33 53.77

3.34 | 3 0.33 54.10

3.36 | 25 2.73 56.83

3.4 | 15 1.64 58.47

3.41 | 10 1.09 59.56

3.42 | 1 0.11 59.67

3.47 | 17 1.86 61.53

3.54 | 16 1.75 63.28

3.59 | 39 4.26 67.54

3.62 | 6 0.66 68.20

3.69 | 26 2.84 71.04

3.75 | 21 2.30 73.33

3.85 | 8 0.87 74.21

3.92 | 14 1.53 75.74

4.14 | 6 0.66 76.39

4.25 | 8 0.87 77.27

4.29 | 114 12.46 89.73

4.34 | 10 1.09 90.82

4.54 | 49 5.36 96.17

4.62 | 35 3.83 100.00

------------+-----------------------------------

Total | 915 100.00

. gen mentor\_h=0

. replace mentor\_h=1 if mentor > 20

(87 real changes made)

. label define h 0 "Mentor Pub below 20" 1 "Mentor Pub > 20"

. label value mentor\_h h

. nbreg art i.female i.married kid5 i.mentor\_h i.mentor\_h##c.phd, vce(robust)

Fitting Poisson model:

Iteration 0: log pseudolikelihood = -1662.1716

Iteration 1: log pseudolikelihood = -1660.9099

Iteration 2: log pseudolikelihood = -1660.9072

Iteration 3: log pseudolikelihood = -1660.9072

Fitting constant-only model:

Iteration 0: log pseudolikelihood = -1625.4242

Iteration 1: log pseudolikelihood = -1609.9746

Iteration 2: log pseudolikelihood = -1609.9368

Iteration 3: log pseudolikelihood = -1609.9367

Fitting full model:

Iteration 0: log pseudolikelihood = -1574.3578

Iteration 1: log pseudolikelihood = -1571.294

Iteration 2: log pseudolikelihood = -1571.2632

Iteration 3: log pseudolikelihood = -1571.2632

Negative binomial regression Number of obs = 915

Wald chi2(6) = 75.92

Dispersion = mean Prob > chi2 = 0.0000

Log pseudolikelihood = -1571.2632 Pseudo R2 = 0.0240

----------------------------------------------------------------------------------

| Robust

art | Coef. Std. Err. z P>|z| [95% Conf. Interval]

-----------------+----------------------------------------------------------------

female |

Female | -.2388442 .0711542 -3.36 0.001 -.3783039 -.0993846

|

married |

Married | .1118127 .0815917 1.37 0.171 -.048104 .2717294

kid5 | -.1535556 .051634 -2.97 0.003 -.2547563 -.0523548

|

mentor\_h |

Mentor Pub > 20 | 1.841293 .4312309 4.27 0.000 .9960958 2.68649

phd | .1032549 .0366309 2.82 0.005 .0314596 .1750502

|

mentor\_h#c.phd |

Mentor Pub > 20 | -.3445193 .1156133 -2.98 0.003 -.5711171 -.1179215

|

\_cons | .2161072 .141716 1.52 0.127 -.0616511 .4938656

-----------------+----------------------------------------------------------------

/lnalpha | -.7784598 .1235778 -1.020668 -.5362517

-----------------+----------------------------------------------------------------

alpha | .4591126 .0567361 .3603542 .5849367

----------------------------------------------------------------------------------

**Note that the following syntax shows how to generate the interaction chart.**

. margins i.mentor\_h, at(phd=(0(.5)5)) atmeans noatlegend

Adjusted predictions Number of obs = 915

Model VCE : Robust

Expression : Predicted number of events, predict()

-----------------------------------------------------------------------------------------

| Delta-method

| Margin Std. Err. z P>|z| [95% Conf. Interval]

------------------------+----------------------------------------------------------------

\_at#mentor\_h |

1#Mentor Pub below 20 | 1.10987 .135091 8.22 0.000 .8450968 1.374644

1#Mentor Pub > 20 | 6.997382 2.895624 2.42 0.016 1.322063 12.6727

2#Mentor Pub below 20 | 1.168675 .1220843 9.57 0.000 .9293941 1.407956

2#Mentor Pub > 20 | 6.202199 2.23415 2.78 0.006 1.823346 10.58105

3#Mentor Pub below 20 | 1.230595 .1078516 11.41 0.000 1.01921 1.441981

3#Mentor Pub > 20 | 5.497381 1.68819 3.26 0.001 2.188589 8.806172

4#Mentor Pub below 20 | 1.295796 .0927231 13.97 0.000 1.114063 1.47753

4#Mentor Pub > 20 | 4.872658 1.241167 3.93 0.000 2.440016 7.3053

5#Mentor Pub below 20 | 1.364452 .0775531 17.59 0.000 1.212451 1.516453

5#Mentor Pub > 20 | 4.318928 .8797577 4.91 0.000 2.594635 6.043222

6#Mentor Pub below 20 | 1.436745 .0644409 22.30 0.000 1.310444 1.563047

6#Mentor Pub > 20 | 3.828125 .5947181 6.44 0.000 2.662499 4.993751

7#Mentor Pub below 20 | 1.512869 .0578806 26.14 0.000 1.399425 1.626313

7#Mentor Pub > 20 | 3.393096 .3840242 8.84 0.000 2.640423 4.14577

8#Mentor Pub below 20 | 1.593026 .0634552 25.10 0.000 1.468656 1.717396

8#Mentor Pub > 20 | 3.007505 .2607232 11.54 0.000 2.496496 3.518513

9#Mentor Pub below 20 | 1.67743 .0818431 20.50 0.000 1.51702 1.837839

9#Mentor Pub > 20 | 2.665731 .2421747 11.01 0.000 2.191078 3.140385

10#Mentor Pub below 20 | 1.766306 .1095296 16.13 0.000 1.551632 1.98098

10#Mentor Pub > 20 | 2.362797 .2896051 8.16 0.000 1.795182 2.930413

11#Mentor Pub below 20 | 1.859891 .1438381 12.93 0.000 1.577973 2.141808

11#Mentor Pub > 20 | 2.094289 .3493437 5.99 0.000 1.409588 2.77899

-----------------------------------------------------------------------------------------

. marginsplot, noci title("The Interactive Effect of" "Mentor's Productivity & PhD Prestige") ///

> ytitle("Predicted/Expected Number of Publications")

Variables that uniquely identify margins: phd mentor\_h



Results show that PhD program’s impact on students’ publications varies by the mentor’s productivity rate. **For a mentor who is not productive, the impact of PhD program’s prestige has a positive impact on students’ expected number of publications: the higher the prestige, the more articles the student produced. However, if the mentor is very productive, the impact of PhD program’s prestige on students’ productivity is negative: the higher the prestige, the lower the students’ productivity. This finding is very interesting!** Am I right that working with a very productive professor in a highly prestigious program is not necessarily productive? **The chart shows that working with a professor who is very productive in a program with the highest prestige is the same as working with a not-productive professor in the same program. Don’t generalize the findings without caution. Note that this study is about biochemist PhD students.**