

**Advanced Methods in Health Services Research: Analysis - 309.716**  
**Tuesday and Thursday 10:30-11:50**  
**Instructors: Darrell J. Gaskin, Ph.D.**  
**Roland J. Thorpe, Ph.D.**  
**250 Hampton House**

**Computer Exercise #5: Poisson and Negative Binomial Models**  
**Due: October 4, 2016**

**Answer Guide**

Use the analysis file built for computer exercise #1 and the subpop command to reduce the sample to those persons 18 years to 64 years old.

Use the Poisson and negative binomial regression techniques to estimate the following model.

Number of hospital stays = f(health status, age, race, sex, insurance status, education, poverty status, and location). *Use the following reference categories: excellent health, age category 25-44, male, white non-Hispanic, HS diploma/GED, privately insured, high income, msa, East*). Conduct a Poisson Regression model check.

1. Display coefficients and IRR in table 1. Interpret the impact of insurance status, health status and location on hospital use.

TABLE 1. Poisson and Negative Binomial Regression Models

	(1) Poisson Coef	(2) Poisson IRR	(3) NB Coef	(4) NB IRR
ipdis08				
Very Good	0.136 (0.154)	1.146 (0.177)	0.170 (0.147)	1.185 (0.174)
Good	0.537*** (0.142)	1.712*** (0.243)	0.581*** (0.133)	1.789*** (0.238)
Fair	1.044*** (0.167)	2.841*** (0.475)	1.081*** (0.157)	2.949*** (0.462)
Poor	2.010*** (0.188)	7.465*** (1.402)	2.161*** (0.175)	8.682*** (1.520)
South	0.177 (0.129)	1.194 (0.154)	0.165 (0.135)	1.179 (0.159)
Midwest	0.240	1.271	0.198	1.219

	(0.138)	(0.175)	(0.143)	(0.175)
West	0.0162 (0.156)	1.016 (0.158)	0.0237 (0.166)	1.024 (0.170)
Public Ins	0.578*** (0.136)	1.782*** (0.242)	0.665*** (0.135)	1.944*** (0.263)
Uninsured	-0.775*** (0.142)	0.461*** (0.0652)	-0.790*** (0.147)	0.454*** (0.0669)
Medicare	0.735*** (0.199)	2.085*** (0.415)	0.717*** (0.184)	2.049*** (0.377)
<i>N</i>	30158	30158	30158	30158
<i>R</i> <sup>2</sup>				

Standard errors in parentheses

\*  $p < 0.05$ , \*\*  $p < 0.01$ , \*\*\*  $p < 0.001$

### POISSON

From the table above we see that having good, fair, and poor health increases the expected number of hospital stays by factors of 1.712, 2.841, 7.465, respectively, holding all other variables at their mean. The coefficient and IRR from having very good health is not statistically significant. Location is not associated with changes in the expected number of hospital stays. Having public insurance and Medicare increases the expected number of hospital stays by factors of 1.782 and 2.085, respectively, holding all other variables at their mean. On the other hand, being uninsured, decreases the expected number of hospital stays by a factor of 0.461, again holding all other variables at their mean.

### NEGATIVE BINOMIAL

From the table above we see that having good, fair, and poor health increases the expected number of hospital stays by factors of 1.789, 2.949, 8.682, respectively, holding all other variables at their mean. The coefficient and IRR from having very good health is not statistically significant. Location is not associated with changes in the expected number of hospital stays. Having public insurance and Medicare increases the expected number of hospital stays by factors of 1.944 and 2.049, respectively, holding all other variables at their mean. On the other hand, being uninsured, decreases the expected number of hospital stays by a factor of 0.454, again holding all other variables at their mean.

2. Test for over-dispersion. Do you have over-dispersion? Which model should you use?

We first check if the Poisson model could be an appropriate model to use given our data by looking at the mean and the variance of our dependent variable.

```
. summ ipdis08, detail
```

# hospital discharges, 2008					
Percentiles		Smallest			
1%	0	0			
5%	0	0			
10%	0	0	Obs		33066
25%	0	0	Sum of Wgt.		33066
50%	0		Mean		.0853142
		Largest	Std. Dev.		.3665397
75%	0	6			
90%	0	6	Variance		.1343514
95%	1	6	Skewness		6.501082
99%	2	10	Kurtosis		69.60763

We see here that the variance (0.134) is greater than the mean (0.085) of the number of hospital stays/discharges. This suggests that there may be over-dispersion. We check with the negative binomial regression.

```
nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance if adult == 1
```

Negative binomial regression	Number of obs	=	16462
	LR chi2(26)	=	997.76
Dispersion = mean	Prob > chi2	=	0.0000
Log likelihood = -4377.582	Pseudo R2	=	0.1023

ipdis08	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
healthstatus						
Very Good	.2561705	.1026902	2.49	0.013	.0549015	.4574395
Good	.6020772	.099928	6.03	0.000	.4062219	.7979324
Fair	1.16913	.113451	10.31	0.000	.94677	1.39149
Poor	2.135988	.1359027	15.72	0.000	1.869623	2.402352
agecat						
18-24	.0040355	.0923493	0.04	0.965	-.1769657	.1850368
45-64	-.034171	.0718499	-0.48	0.634	-.1749942	.1066521
race						
Black	-.1831203	.0836236	-2.19	0.029	-.3470196	-.019221
Hispanic	-.2783769	.0871194	-3.20	0.001	-.4491278	-.107626
Asian	-1.062207	.2167938	-4.90	0.000	-1.487115	-.6372988
Other	-.0521937	.189124	-0.28	0.783	-.42287	.3184825
2.sex	.6681996	.0675341	9.89	0.000	.5358351	.8005641
fpl						
Poor	.6386398	.1123707	5.68	0.000	.4183973	.8588823
Near Poor	.3327764	.1495629	2.22	0.026	.0396386	.6259142
Low Income	.3427518	.1097915	3.12	0.002	.1275643	.5579392
Middle Income	.1211003	.0935389	1.29	0.195	-.0622326	.3044332
education						

No High School		-.2257622	.1271187	-1.78	0.076	-.4749102	.0233859
Some High		-.2734027	.0939895	-2.91	0.004	-.4576187	-.0891868
Some College/Tech School/AA degree		.0218915	.0866826	0.25	0.801	-.1480033	.1917863
College		.2084079	.0973672	2.14	0.032	.0175716	.3992442
msa							
Not in MSA		-.0828144	.088161	-0.94	0.348	-.2556068	.089978
region							
South		.147477	.0945778	1.56	0.119	-.037892	.3328461
Midwest		.1202743	.1039238	1.16	0.247	-.0834127	.3239613
West		-.0712041	.104121	-0.68	0.494	-.2752775	.1328693
insurance							
Public Insurance		.5968584	.0941504	6.34	0.000	.412327	.7813898
Uninsured		-.7074249	.1024949	-6.90	0.000	-.9083112	-.5065387
Medicare		.709979	.1327196	5.35	0.000	.4498533	.9701047
_cons		-3.651248	.1461914	-24.98	0.000	-3.937778	-3.364718
/lnalpha		.6547484	.1014965			.4558189	.8536779
alpha		1.924658	.1953461			1.577465	2.348268
Likelihood-ratio test of alpha=0: chibar2(01) = 278.06 Prob>=chibar2 = 0.000							

We see from the likelihood ratio test of  $\alpha = 0$  that we reject the null hypothesis that  $\alpha = 0$  ( $p < 0.001$ ). Thus the alpha parameter in the negative binomial distribution is non-zero and it is not equivalent to the Poisson distribution. Combining this with our previous check of the mean and variance of our dependent variable tells us that we should use the negative binomial model.

## **Do-file**

\* Advanced methods in health services research: analysis - 309.716  
\* October 4, 2016  
\* Assignment # 5: Suggested code

```
set varabbrev off
set more off, permanently
log using "C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods -
Analysis\Answers\Assignment4_2016.log", replace
use "C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods -
Analysis\meps08.dta"
cd "C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods -
Analysis\Answers"
svyset varpsu [pweight = perwt08f], strata(varstr)

** Data preparation
* Trimming the data
summ totexp08
summ totexp08, detail

gen healthexp = totexp08
replace healthexp = . if totexp08 > 100000
summ healthexp
summ healthexp, detail

* Create the log of health expenditure
gen lnhealthexp = ln(1+healthexp)

summ ttlp08x
summ ttlp08x, detail

gen income = ttlp08x
replace income = . if ttlp08x < 0
replace income = . if ttlp08x > 170000

summ income
summ income, detail

* Create independent variables
summ age08x
gen age = age08x
replace age = . if age08x < 0
summ age

gen agecat = 1 if age08x >= 0 & age08x <= 17
replace agecat = 2 if age08x >= 18 & age08x <= 24
replace agecat = 3 if age08x >= 25 & age08x <= 44
replace agecat = 4 if age08x >= 45 & age08x <= 64
replace agecat = 5 if age08x >= 65 & age08x <= 74
replace agecat = 6 if age08x >= 75

label define agecats 1 "0-18" 2 "18-24" 3 "25-44" 4 "45-64" 5 "65-74" 6 "75+"
label values agecat agecats

gen female = 0 if sex == 1
replace female = 1 if sex == 2

label define sexn 0 "Male" 1 "Female"
label values female sexn

label define raceethn 1 "Hispanic" 2 "Black non-Hispanic" 3 "Asian non-Hispanic" ///
4 "Other race/not Hispanic"
label values raceethn raceethn

label define racexn 1 "White" 2 "Black" 3 "Amer Indian/Alaska Native" 4 "Asian" ///
5 "Native Hawaiian/Pacific Islander" 6 "Multiple races reported"
label values racexn racexn

gen race = 1 if racex == 1 & raceethn != 1
```

```

replace race = 2 if racex == 2 & racethnx != 1
replace race = 3 if racethnx == 1 == 1
replace race = 4 if racex == 4 & racethnx != 1
replace race = 5 if (racex == 3 | racex == 5 | racex == 6) & racethnx != 1

label define racexn2 1 "White" 2 "Black" 3 "Hispanic" 4 "Asian" 5 "Other"
label values race racexn2

gen education = 1 if educyr >= 0 & educyr <= 8
replace education = 2 if educyr >= 9 & educyr <= 12 & hideg == 1
replace education = 2 if educyr >= 9 & educyr <= 11 & hideg < 0
replace education = 3 if hideg == 2 | hideg == 3
replace education = 4 if educyr >= 13 & educyr <= 17 & hideg == 3
replace education = 5 if hideg == 4
replace education = 6 if educyr == 7 & hideg == 7

label define educn 1 "No High School" 2 "Some High" 3 "High School/GED" ///
4 "Some College/Tech School/AA degree" 5 "College" 6 "Advanced Degree"
label values education educn

gen insurance = 1 if inscov08 == 1
replace insurance = 2 if inscov08 == 2 & mcrev08 == 2
replace insurance = 3 if inscov08 == 3
replace insurance = 4 if inscov08 == 2 & mcrev08 == 1

label define insr 1 "Private Insurance" 2 "Public Insurance" 3 "Uninsured" ///
4 "Medicare"
label values insurance insr

gen msa = 0 if msa08 == 0
replace msa = 1 if msa08 == 1

label define msan 0 "Not in MSA" 1 "In MSA"
label values msa msan

gen region = 1 if region08 == 1
replace region = 2 if region08 == 3
replace region = 3 if region08 == 2
replace region = 4 if region08 == 4

label define regi 1 "North East" 2 "South" 3 "Midwest" 4 "West"
label values region regi

gen healthstatus = 1 if rthlth42 == 1
replace healthstatus = 2 if rthlth42 == 2
replace healthstatus = 3 if rthlth42 == 3
replace healthstatus = 4 if rthlth42 == 4
replace healthstatus = 5 if rthlth42 == 5
label define health 1 "Excellent" 2 "Very Good" 3 "Good" 4 "Fair" 5 "Poor"
label values healthstatus health

gen fpl = 1 if povcat08 == 1
replace fpl = 2 if povcat08 == 2
replace fpl = 3 if povcat08 == 3
replace fpl = 4 if povcat08 == 4
replace fpl = 5 if povcat08 == 5

label define fplstat 1 "Poor" 2 "Near Poor" 3 "Low Income" 4 "Middle Income" ///
5 "High Income"
label values fpl fplstat

* Indicator variable for adults 18-64
gen adult = 0
replace adult = 1 if age08x > 17 & age08x < 65

* Generate indicator variable for office visit
gen visit = 0
replace visit = 1 if obtotv08 > 0

*****
** Poisson Regression

```

```

* Poisson Model Check
svy, subpop(adult): poisson ipdis08

svy, subpop(adult): poisson ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl
ib3.education ///
ib1.msa i.region i.insurance

eststo poisson1

svy, subpop(adult): poisson ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl
ib3.education ///
ib1.msa i.region i.insurance, nolog irr

eststo poisson2

* Negative Binomial Regression
svy, subpop(adult): nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl
ib3.education ///
ib1.msa i.region i.insurance

eststo nbreg1

svy, subpop(adult): nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl
ib3.education ///
ib1.msa i.region i.insurance, nolog irr

eststo nbreg2

* Table 1
esttab poisson1 poisson2 nbreg1 nbreg2, se r2 ///
keep(2.insurance 3.insurance 4.insurance 2.healthstatus 3.healthstatus 4.healthstatus ///
5.healthstatus 2.region 3.region 4.region) ///
mtitles("Poisson Coef" "Poisson IRR" "NB Coef" "NB IRR") ///
coeflabels(2.insurance "Public Ins" 3.insurance Uninsured 4.insurance Medicare ///
2.healthstatus "Very Good" 3.healthstatus Good 4.healthstatus Fair ///
5.healthstatus Poor 2.region South 3.region Midwest 4.region West) ///
title(TABLE 1. Poisson and Negative Binomial Regression Models) eform(0 1 0 1)

esttab poisson1 poisson2 nbreg1 nbreg2 using hw5table1.rtf, se r2 replace ///
keep(2.insurance 3.insurance 4.insurance 2.healthstatus 3.healthstatus 4.healthstatus ///
5.healthstatus 2.region 3.region 4.region) ///
mtitles("Poisson Coef" "Poisson IRR" "NB Coef" "NB IRR") ///
coeflabels(2.insurance "Public Ins" 3.insurance Uninsured 4.insurance Medicare ///
2.healthstatus "Very Good" 3.healthstatus Good 4.healthstatus Fair ///
5.healthstatus Poor 2.region South 3.region Midwest 4.region West) ///
title(TABLE 1. Poisson and Negative Binomial Regression Models) eform(0 1 0 1)

** Number 2
* Test of Overdispersion
summ ipdis08, detail

nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
ib1.msa i.region i.insurance if adult == 1

nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
ib1.msa i.region i.insurance if adult == 1, nolog irr

```

## Log file

```
-----
name: <unnamed>
log: C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods -
Analysis\Answers\Assignment4_2016.log
log type: text
opened on: 4 Oct 2016, 22:22:18

. use "C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods - Analysis\meps08.dta"

. cd "C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods - Analysis\Answers"
C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods - Analysis\Answers

. svyset varpsu [pweight = perwt08f], strata(varstr)

pweight: perwt08f
VCE: linearized
Single unit: missing
Strata 1: varstr
SU 1: varpsu
FPC 1: <zero>

.
. ** Data preparation
. * Trimming the data
. summ totexp08

Variable | Obs Mean Std. Dev. Min Max
-----+-----
totexp08 | 33066 3142.069 9786.619 0 553493

. summ totexp08, detail

total health care exp 08
-----
Percentiles Smallest
1% 0 0
5% 0 0
10% 0 0 Obs 33066
25% 67 0 Sum of Wgt. 33066

50% 528.5 Mean 3142.069
75% 2425 Largest Std. Dev. 9786.619
90% 7453 238659
95% 13582 264510 Variance 9.58e+07
99% 40763 373799 Skewness 14.40863
Kurtosis 469.447

.
. gen healthexp = totexp08

. replace healthexp = . if totexp08 > 100000
(42 real changes made, 42 to missing)

. summ healthexp

Variable | Obs Mean Std. Dev. Min Max
-----+-----
healthexp | 33024 2937.216 7412.106 0 99988

. summ healthexp, detail

healthexp
-----
Percentiles Smallest
1% 0 0
5% 0 0
10% 0 0 Obs 33024
25% 66 0 Sum of Wgt. 33024

50% 526 Mean 2937.216
75% 2413.5 Largest Std. Dev. 7412.106
90% 7353 98210
95% 13387 99251 Variance 5.49e+07
99% 38550 99988 Skewness 5.819768
Kurtosis 48.38959

.
. * Create the log of health expenditure
. gen lnhealthexp = ln(1+healthexp)
```



(42 missing values generated)

```
.  
. summ ttlp08x
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
ttl08x	33066	20369.66	29399.89	-186193	313026

```
. summ ttlp08x, detail
```

person s total income					
Percentiles		Smallest			
1%	0	-186193			
5%	0	-68779			
10%	0	-40779	Obs	33066	
25%	0	-39888	Sum of Wgt.	33066	
50%	10359.5		Mean	20369.66	
		Largest	Std. Dev.	29399.89	
75%	29700	238419			
90%	54505	251918	Variance	8.64e+08	
95%	75000	294216	Skewness	2.738784	
99%	143840	313026	Kurtosis	14.04593	

```
.  
. gen income = ttlp08x  
  
. replace income = . if ttlp08x < 0  
(25 real changes made, 25 to missing)  
  
. replace income = . if ttlp08x > 170000  
(325 real changes made, 325 to missing)
```

```
.  
. summ income
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
income	32716	18715.83	24109.08	0	169564

```
. summ income, detail
```

income					
Percentiles		Smallest			
1%	0	0			
5%	0	0			
10%	0	0	Obs	32716	
25%	0	0	Sum of Wgt.	32716	
50%	10000		Mean	18715.83	
		Largest	Std. Dev.	24109.08	
75%	28838	164895			
90%	52000	164895	Variance	5.81e+08	
95%	70000	165391	Skewness	1.717878	
99%	103084	169564	Kurtosis	6.147364	

```
.  
. * Create independent variables  
. summ age08x
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
age08x	33066	33.69395	22.38051	-1	85

```
. gen age = age08x
```

```
. replace age = . if age08x < 0  
(251 real changes made, 251 to missing)
```

```
. summ age
```

Variable	Obs	Mean	Std. Dev.	Min	Max
-----+-----					
age	32815	33.95932	22.25851	0	85

```
.  
. gen agecat = 1 if age08x >= 0 & age08x <= 17  
(23434 missing values generated)
```

```

. replace agecat = 2 if age08x >= 18 & age08x <= 24
(3372 real changes made)

. replace agecat = 3 if age08x >= 25 & age08x <= 44
(8813 real changes made)

. replace agecat = 4 if age08x >= 45 & age08x <= 64
(7614 real changes made)

. replace agecat = 5 if age08x >= 65 & age08x <= 74
(1867 real changes made)

. replace agecat = 6 if age08x >= 75
(1517 real changes made)

.
. label define agecats 1 "0-18" 2 "18-24" 3 "25-44" 4 "45-64" 5 "65-74" 6 "75+"

. label values agecat agecats

.
. gen female = 0 if sex == 1
(17181 missing values generated)

. replace female = 1 if sex == 2
(17181 real changes made)

.
. label define sexn 0 "Male" 1 "Female"

. label values female sexn

.
. label define raceethn 1 "Hispanic" 2 "Black non-Hispanic" 3 "Asian non-Hispanic" ///
> 4 "Other race/not Hispanic"

. label values raceethn raceethn

.
. label define racexn 1 "White" 2 "Black" 3 "Amer Indian/Alaska Native" 4 "Asian" ///
> 5 "Native Hawaiian/Pacific Islander" 6 "Multiple races reported"

. label values racexn racexn

.
. gen race = 1 if racex == 1 & raceethn != 1
(18775 missing values generated)

. replace race = 2 if racex == 2 & raceethn != 1
(6476 real changes made)

. replace race = 3 if raceethn == 1 == 1
(9392 real changes made)

. replace race = 4 if racex == 4 & raceethn != 1
(1997 real changes made)

. replace race = 5 if (racex == 3 | racex == 5 | racex == 6) & raceethn != 1
(910 real changes made)

.
. label define racexn2 1 "White" 2 "Black" 3 "Hispanic" 4 "Asian" 5 "Other"

. label values race racexn2

.
. gen education = 1 if educyr >= 0 & educyr <= 8
(25465 missing values generated)

. replace education = 2 if educyr >= 9 & educyr <= 12 & hideg == 1
(3732 real changes made)

. replace education = 2 if educyr >= 9 & educyr <= 11 & hideg < 0
(3 real changes made)

. replace education = 3 if hideg == 2 | hideg == 3
(11173 real changes made)

. replace education = 4 if educyr >= 13 & educyr <= 17 & hideg == 3
(3629 real changes made)

```

```

. replace education = 5 if hideg == 4
(3245 real changes made)

. replace education = 6 if educyr == 7 & hideg == 7
(0 real changes made)

.
. label define educn 1 "No High School" 2 "Some High" 3 "High School/GED" ///
> 4 "Some College/Tech School/AA degree" 5 "College" 6 "Advanced Degree"

. label values education educn

.
. gen inscov08 = 1 if inscov08 == 1
(14773 missing values generated)

. replace insurance = 2 if inscov08 == 2 & mcrev08 == 2
(6642 real changes made)

. replace insurance = 3 if inscov08 == 3
(5662 real changes made)

. replace insurance = 4 if inscov08 == 2 & mcrev08 == 1
(2469 real changes made)

.
. label define insr 1 "Private Insurance" 2 "Public Insurance" 3 "Uninsured" ///
> 4 "Medicare"

. label values insurance insr

.
. gen msa = 0 if msa08 == 0
(28406 missing values generated)

. replace msa = 1 if msa08 == 1
(28155 real changes made)

.
. label define msan 0 "Not in MSA" 1 "In MSA"

. label values msa msan

.
. gen region = 1 if region08 == 1
(28080 missing values generated)

. replace region = 2 if region08 == 3
(12424 real changes made)

. replace region = 3 if region08 == 2
(6499 real changes made)

. replace region = 4 if region08 == 4
(8906 real changes made)

.
. label define regi 1 "North East" 2 "South" 3 "Midwest" 4 "West"

. label values region regi

.
. gen healthstatus = 1 if rthlth42 == 1
(23215 missing values generated)

. replace healthstatus = 2 if rthlth42 == 2
(10188 real changes made)

. replace healthstatus = 3 if rthlth42 == 3
(8628 real changes made)

. replace healthstatus = 4 if rthlth42 == 4
(2777 real changes made)

. replace healthstatus = 5 if rthlth42 == 5
(815 real changes made)

. label define health 1 "Excellent" 2 "Very Good" 3 "Good" 4 "Fair" 5 "Poor"

. label values healthstatus health

```

```

.
. gen fpl = 1 if povcat08 == 1
(26099 missing values generated)

. replace fpl = 2 if povcat08 == 2
(2171 real changes made)

. replace fpl = 3 if povcat08 == 3
(5667 real changes made)

. replace fpl = 4 if povcat08 == 4
(9595 real changes made)

. replace fpl = 5 if povcat08 == 5
(8666 real changes made)

.
. label define fplstat 1 "Poor" 2 "Near Poor" 3 "Low Income" 4 "Middle Income" ///
> 5 "High Income"

. label values fpl fplstat

.
. * Indicator variable for adults 18-64
. gen adult = 0

. replace adult = 1 if age08x > 17 & age08x < 65
(19799 real changes made)

.
. * Generate indicator variable for office visit
. gen visit = 0

. replace visit = 1 if obtotv08 > 0
(22054 real changes made)

.
. *****
. ** Poisson Regression
. * Poisson Model Check
. svy, subpop(adult): poisson ipdis08
(running poisson on estimation sample)

Survey: Poisson regression

Number of strata   =      165          Number of obs       =      33066
Number of PSUs     =      370          Population size      =     304375942
                                          Subpop. no. of obs   =      18619
                                          Subpop. size         =     189180298
                                          Design df            =        205
                                          F(    0,    205)     =          .
                                          Prob > F              =          .

-----
            |               Linearized
            |               Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      _cons |    -2.506156   .039934    -62.76   0.000    -2.58489    -2.427422
-----+-----

.
. svy, subpop(adult): poisson ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance
(running poisson on estimation sample)

Survey: Poisson regression

Number of strata   =      165          Number of obs       =      30158
Number of PSUs     =      370          Population size      =     269717779
                                          Subpop. no. of obs   =      15711
                                          Subpop. size         =     154522136
                                          Design df            =        205
                                          F(   26,   180)     =      29.31
                                          Prob > F              =      0.0000

-----
            |               Linearized
            |               Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-----+-----
      ipdis08 |
      healthstatus |

```

	Very Good		.13611	.154231	0.88	0.379	-.1679724	.4401924
	Good		.5374676	.1417559	3.79	0.000	.2579811	.8169541
	Fair		1.044139	.1670905	6.25	0.000	.7147033	1.373576
	Poor		2.010193	.1878259	10.70	0.000	1.639875	2.380511
	agecat							
	18-24		-.1647947	.1168722	-1.41	0.160	-.3952204	.065631
	45-64		-.0694087	.091689	-0.76	0.450	-.2501831	.1113656
	race							
	Black		-.1340813	.1085609	-1.24	0.218	-.3481202	.0799577
	Hispanic		-.2939982	.1086169	-2.71	0.007	-.5081477	-.0798487
	Asian		-1.321949	.2358151	-5.61	0.000	-1.786882	-.8570145
	Other		-.3730918	.2002208	-1.86	0.064	-.7678479	.0216642
	2.sex		.5017556	.08776	5.72	0.000	.3287276	.6747836
	fpl							
	Poor		.573867	.1465463	3.92	0.000	.2849358	.8627981
	Near Poor		.2636066	.185484	1.42	0.157	-.1020943	.6293075
	Low Income		.3901411	.1599408	2.44	0.016	.0748012	.705481
	Middle Income		.1655807	.1227448	1.35	0.179	-.0764234	.4075849
	education							
	No High School		-.2571697	.1523875	-1.69	0.093	-.5576174	.0432781
	Some High		-.2481007	.1245399	-1.99	0.048	-.4936439	-.0025575
Some College/Tech School/AA degree			.0259212	.1028877	0.25	0.801	-.1769325	.2287749
College			.1637648	.1308417	1.25	0.212	-.0942031	.4217326
	msa							
	Not in MSA		-.097197	.1130691	-0.86	0.391	-.3201245	.1257305
	region							
	South		.1771329	.1290782	1.37	0.171	-.077358	.4316239
	Midwest		.2395409	.1380703	1.73	0.084	-.0326789	.5117607
	West		.016211	.1555544	0.10	0.917	-.2904807	.3229026
	insurance							
	Public Insurance		.5777011	.1355833	4.26	0.000	.3103847	.8450175
	Uninsured		-.7749347	.141621	-5.47	0.000	-1.054155	-.4957142
	Medicare		.73453	.1990296	3.69	0.000	.3421225	1.126937
	_cons		-3.485319	.2185371	-15.95	0.000	-3.916188	-3.054451

```
.
. eststo poisson1
```

```
.
. svy, subpop(adult): poisson ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance, nolog irr
(running poisson on estimation sample)
```

Survey: Poisson regression

Number of strata	=	165	Number of obs	=	30158
Number of PSUs	=	370	Population size	=	269717779
			Subpop. no. of obs	=	15711
			Subpop. size	=	154522136
			Design df	=	205
			F( 26, 180)	=	29.31
			Prob > F	=	0.0000

	ipdis08	IRR	Std. Err.	t	P> t	[95% Conf. Interval]
healthstatus						
Very Good		1.145808	.1767191	0.88	0.379	.8453772 1.553006
Good		1.711667	.2426389	3.79	0.000	1.294314 2.263595
Fair		2.840953	.4746962	6.25	0.000	2.04358 3.949448
Poor		7.464756	1.402074	10.70	0.000	5.154523 10.81042
agecat						
18-24		.8480678	.0991156	-1.41	0.160	.6735316 1.067833
45-64		.9329453	.0855408	-0.76	0.450	.7786582 1.117803
race						
Black		.874519	.0949385	-1.24	0.218	.706014 1.083241
Hispanic		.7452778	.0809498	-2.71	0.007	.6016089 .9232561
Asian		.2666153	.0628719	-5.61	0.000	.1674815 .4244273

Other		.688602	.1378724	-1.86	0.064	.4640106	1.021901
2.sex		1.651618	.1449461	5.72	0.000	1.389199	1.963608
fpl							
Poor		1.775118	.260137	3.92	0.000	1.329677	2.369782
Near Poor		1.301616	.2414289	1.42	0.157	.9029444	1.876311
Low Income		1.477189	.2362629	2.44	0.016	1.07767	2.02482
Middle Income		1.180078	.1448485	1.35	0.179	.9264238	1.503183
education							
No High School		.773237	.1178317	-1.69	0.093	.5725717	1.044228
Some High		.7802814	.0971761	-1.99	0.048	.6103981	.9974458
Some College/Tech School/AA degree		1.02626	.1055895	0.25	0.801	.8378363	1.257059
College		1.177937	.1541233	1.25	0.212	.9100979	1.524601
msa							
Not in MSA		.9073773	.1025964	-0.86	0.391	.7260586	1.133977
region							
South		1.19379	.1540922	1.37	0.171	.9255584	1.539756
Midwest		1.270666	.1754411	1.73	0.084	.9678493	1.668226
West		1.016343	.1580967	0.10	0.917	.747904	1.381131
insurance							
Public Insurance		1.781937	.2416008	4.26	0.000	1.36395	2.328019
Uninsured		.4607339	.0652496	-5.47	0.000	.3484867	.6091357
Medicare		2.084502	.4148776	3.69	0.000	1.407933	3.08619
_cons		.030644	.0066968	-15.95	0.000	.0199169	.0471486

. eststo poisson2

. \* Negative Binomial Regression  
. svy, subpop(adult): nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///  
> ib1.msa i.region i.insurance  
(running nbreg on estimation sample)

Survey: Negative binomial regression

Number of strata	=	165	Number of obs	=	30158
Number of PSUs	=	370	Population size	=	269717779
			Subpop. no. of obs	=	15711
			Subpop. size	=	154522136
			Design df	=	205
			F( 26, 180)	=	25.52
			Prob > F	=	0.0000

ipdis08	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
healthstatus						
Very Good	.1701555	.1471643	1.16	0.249	-.1199943	.4603052
Good	.5814995	.1331453	4.37	0.000	.3189898	.8440093
Fair	1.081331	.1566505	6.90	0.000	.7724783	1.390184
Poor	2.161205	.1751328	12.34	0.000	1.815913	2.506498
agecat						
18-24	-.2023371	.1202609	-1.68	0.094	-.4394439	.0347696
45-64	-.094602	.0852853	-1.11	0.269	-.2627508	.0735468
race						
Black	-.1255076	.1158948	-1.08	0.280	-.3540061	.1029909
Hispanic	-.2922451	.1133647	-2.58	0.011	-.5157554	-.0687348
Asian	-1.315051	.2396286	-5.49	0.000	-1.787504	-.8425986
Other	-.3370561	.2027886	-1.66	0.098	-.7368748	.0627626
2.sex	.5769324	.0843118	6.84	0.000	.410703	.7431618
fpl						
Poor	.6296796	.1423286	4.42	0.000	.349064	.9102953
Near Poor	.2952571	.1763905	1.67	0.096	-.052515	.6430292
Low Income	.3758715	.1512662	2.48	0.014	.0776345	.6741084
Middle Income	.1701499	.1227542	1.39	0.167	-.0718727	.4121724
education						
No High School	-.2286532	.1575493	-1.45	0.148	-.539278	.0819716

Some College/Tech School/AA degree	Some High	-.2719432	.1243588	-2.19	0.030	-.5171294	-.026757
	degree	.062248	.1061282	0.59	0.558	-.1469947	.2714906
	College	.1883055	.1330021	1.42	0.158	-.0739219	.4505328
	msa						
	Not in MSA	-.0673544	.1188589	-0.57	0.572	-.3016972	.1669883
	region						
	South	.1645419	.1351018	1.22	0.225	-.1018253	.4309091
	Midwest	.1978021	.1432064	1.38	0.169	-.0845441	.4801483
	West	.0236759	.1658462	0.14	0.887	-.3033071	.3506588
	insurance						
	Public Insurance	.664882	.1350795	4.92	0.000	.3985588	.9312051
	Uninsured	-.7900412	.1473936	-5.36	0.000	-1.080643	-.4994395
	Medicare	.7174657	.1841538	3.90	0.000	.3543875	1.080544
	_cons	-3.58248	.2114656	-16.94	0.000	-3.999406	-3.165554
-----							
	/lnalpha	.8115126	.161885			.4923395	1.130686
-----							
	alpha	2.251311	.3644535			1.63614	3.09778
-----							

```
. eststo nbreg1
```

```
. svy, subpop(adult): nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance, nolog irr
(running nbreg on estimation sample)
```

Survey: Negative binomial regression

Number of strata	=	165	Number of obs	=	30158
Number of PSUs	=	370	Population size	=	269717779
			Subpop. no. of obs	=	15711
			Subpop. size	=	154522136
			Design df	=	205
			F( 26, 180)	=	25.52
			Prob > F	=	0.0000

		Linearized					
ipdis08		IRR	Std. Err.	t	P> t	[95% Conf. Interval]	
healthstatus							
Very Good		1.185489	.1744617	1.16	0.249	.8869255	1.584557
Good		1.788719	.2381595	4.37	0.000	1.375737	2.325673
Fair		2.948601	.4618999	6.90	0.000	2.165125	4.015587
Poor		8.681595	1.520432	12.34	0.000	6.146684	12.26191
agecat							
18-24		.8168195	.0982314	-1.68	0.094	.6443947	1.035381
45-64		.9097349	.077587	-1.11	0.269	.7689335	1.076319
race							
Black		.8820491	.1022249	-1.08	0.280	.7018707	1.108481
Hispanic		.7465855	.0846365	-2.58	0.011	.5970494	.9335742
Asian		.2684606	.0643308	-5.49	0.000	.1673775	.4305901
Other		.7138688	.1447644	-1.66	0.098	.4786073	1.064774
2.sex		1.780568	.1501229	6.84	0.000	1.507877	2.102573
fpl							
Poor		1.877009	.2671522	4.42	0.000	1.41774	2.485056
Near Poor		1.343472	.2369756	1.67	0.096	.9488401	1.902234
Low Income		1.45626	.2202829	2.48	0.014	1.080728	1.962283
Middle Income		1.185482	.1455229	1.39	0.167	.9306494	1.510095
education							
No High School		.7956044	.125347	-1.45	0.148	.5831691	1.085425
Some High		.7618975	.0947486	-2.19	0.030	.5962296	.9735978
Some College/Tech School/AA degree		1.064226	.1129444	0.59	0.558	.8632985	1.311919
College		1.207202	.1605604	1.42	0.158	.9287442	1.569148
msa							
Not in MSA		.9348638	.1111169	-0.57	0.572	.739562	1.18174
region							
South		1.178853	.1592652	1.22	0.225	.9031873	1.538656

	Midwest		1.218721	.1745286	1.38	0.169	.9189311	1.616314
	West		1.023958	.1698196	0.14	0.887	.7383723	1.420003
	insurance							
	Public Insurance		1.944261	.2626297	4.92	0.000	1.489676	2.537565
	Uninsured		.4538261	.066891	-5.36	0.000	.3393773	.6068707
	Medicare		2.049233	.3773741	3.90	0.000	1.425307	2.946282
	_cons		.0278067	.0058802	-16.94	0.000	.0183265	.0421908
-----								
	/lnalpha		.8115126	.161885			.4923395	1.130686
-----								
	alpha		2.251311	.3644535			1.63614	3.09778
-----								

```
. eststo nbreg2
```

```
. * Table 1
. esttab poisson1 poisson2 nbreg1 nbreg2, se r2 ///
> keep(2.insurance 3.insurance 4.insurance 2.healthstatus 3.healthstatus 4.healthstatus ///
> 5.healthstatus 2.region 3.region 4.region) ///
> mtitles("Poisson Coef" "Poisson IRR" "NB Coef" "NB IRR") ///
> coelabels(2.insurance "Public Ins" 3.insurance Uninsured 4.insurance Medicare ///
> 2.healthstatus "Very Good" 3.healthstatus Good 4.healthstatus Fair ///
> 5.healthstatus Poor 2.region South 3.region Midwest 4.region West) ///
> title(TABLE 1. Poisson and Negative Binomial Regression Models) eform(0 1 0 1)
```

TABLE 1. Poisson and Negative Binomial Regression Models

	(1) Poisson Coef	(2) Poisson IRR	(3) NB Coef	(4) NB IRR
-----				
ipdis08				
Very Good	0.136 (0.154)	1.146 (0.177)	0.170 (0.147)	1.185 (0.174)
Good	0.537*** (0.142)	1.712*** (0.243)	0.581*** (0.133)	1.789*** (0.238)
Fair	1.044*** (0.167)	2.841*** (0.475)	1.081*** (0.157)	2.949*** (0.462)
Poor	2.010*** (0.188)	7.465*** (1.402)	2.161*** (0.175)	8.682*** (1.520)
South	0.177 (0.129)	1.194 (0.154)	0.165 (0.135)	1.179 (0.159)
Midwest	0.240 (0.138)	1.271 (0.175)	0.198 (0.143)	1.219 (0.175)
West	0.0162 (0.156)	1.016 (0.158)	0.0237 (0.166)	1.024 (0.170)
Public Ins	0.578*** (0.136)	1.782*** (0.242)	0.665*** (0.135)	1.944*** (0.263)
Uninsured	-0.775*** (0.142)	0.461*** (0.0652)	-0.790*** (0.147)	0.454*** (0.0669)
Medicare	0.735*** (0.199)	2.085*** (0.415)	0.717*** (0.184)	2.049*** (0.377)
-----				
N	30158	30158	30158	30158
R-sq				

Standard errors in parentheses  
\* p<0.05, \*\* p<0.01, \*\*\* p<0.001

```
. esttab poisson1 poisson2 nbreg1 nbreg2 using hw5table1.rtf, se r2 replace ///
> keep(2.insurance 3.insurance 4.insurance 2.healthstatus 3.healthstatus 4.healthstatus ///
> 5.healthstatus 2.region 3.region 4.region) ///
> mtitles("Poisson Coef" "Poisson IRR" "NB Coef" "NB IRR") ///
> coelabels(2.insurance "Public Ins" 3.insurance Uninsured 4.insurance Medicare ///
> 2.healthstatus "Very Good" 3.healthstatus Good 4.healthstatus Fair ///
> 5.healthstatus Poor 2.region South 3.region Midwest 4.region West) ///
> title(TABLE 1. Poisson and Negative Binomial Regression Models) eform(0 1 0 1)
(output written to hw5table1.rtf)
```



```
.
. ** Number 2
. * Test of Overdispersion
. summ ipdis08, detail
```

```
-----# hospital discharges, 2008-----
Percentiles      Smallest
1%              0          0
5%              0          0
10%             0          0      Obs          33066
25%             0          0      Sum of Wgt.    33066

50%             0          0      Mean          .0853142
75%             0          6      Std. Dev.      .3665397
90%             0          6      Variance       .1343514
95%             1          6      Skewness       6.501082
99%             2          10     Kurtosis      69.60763
```

```
. nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance if adult == 1
```

Fitting Poisson model:

```
Iteration 0:  log likelihood = -5998.5006
Iteration 1:  log likelihood = -4890.4048
Iteration 2:  log likelihood = -4530.7785
Iteration 3:  log likelihood = -4516.6701
Iteration 4:  log likelihood = -4516.6127
Iteration 5:  log likelihood = -4516.6127
```

Fitting constant-only model:

```
Iteration 0:  log likelihood = -4991.9964 (not concave)
Iteration 1:  log likelihood = -4889.5789
Iteration 2:  log likelihood = -4876.4644
Iteration 3:  log likelihood = -4876.4627
Iteration 4:  log likelihood = -4876.4627
```

Fitting full model:

```
Iteration 0:  log likelihood = -4509.1479
Iteration 1:  log likelihood = -4389.7638
Iteration 2:  log likelihood = -4377.67
Iteration 3:  log likelihood = -4377.582
Iteration 4:  log likelihood = -4377.582
```

```
Negative binomial regression      Number of obs   =      16462
                                LR chi2(26)           =      997.76
                                Prob > chi2            =      0.0000
                                Pseudo R2              =      0.1023

Dispersion      = mean
Log likelihood = -4377.582
```

	ipdis08	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
healthstatus							
Very Good		.2561705	.1026902	2.49	0.013	.0549015	.4574395
Good		.6020772	.099928	6.03	0.000	.4062219	.7979324
Fair		1.16913	.113451	10.31	0.000	.94677	1.39149
Poor		2.135988	.1359027	15.72	0.000	1.869623	2.402352
agecat							
18-24		.0040355	.0923493	0.04	0.965	-.1769657	.1850368
45-64		-.034171	.0718499	-0.48	0.634	-.1749942	.1066521
race							
Black		-.1831203	.0836236	-2.19	0.029	-.3470196	-.019221
Hispanic		-.2783769	.0871194	-3.20	0.001	-.4491278	-.107626
Asian		-1.062207	.2167938	-4.90	0.000	-1.487115	-.6372988
Other		-.0521937	.189124	-0.28	0.783	-.42287	.3184825
2.sex							
		.6681996	.0675341	9.89	0.000	.5358351	.8005641
fpl							
Poor		.6386398	.1123707	5.68	0.000	.4183973	.8588823
Near Poor		.3327764	.1495629	2.22	0.026	.0396386	.6259142
Low Income		.3427518	.1097915	3.12	0.002	.1275643	.5579392
Middle Income		.1211003	.0935389	1.29	0.195	-.0622326	.3044332

education							
No High School	-.2257622	.1271187	-1.78	0.076	-.4749102	.0233859	
Some High	-.2734027	.0939895	-2.91	0.004	-.4576187	-.0891868	
Some College/Tech School/AA degree	.0218915	.0866826	0.25	0.801	-.1480033	.1917863	
College	.2084079	.0973672	2.14	0.032	.0175716	.3992442	
msa							
Not in MSA	-.0828144	.088161	-0.94	0.348	-.2556068	.089978	
region							
South	.147477	.0945778	1.56	0.119	-.037892	.3328461	
Midwest	.1202743	.1039238	1.16	0.247	-.0834127	.3239613	
West	-.0712041	.104121	-0.68	0.494	-.2752775	.1328693	
insurance							
Public Insurance	.5968584	.0941504	6.34	0.000	.412327	.7813898	
Uninsured	-.7074249	.1024949	-6.90	0.000	-.9083112	-.5065387	
Medicare	.709979	.1327196	5.35	0.000	.4498533	.9701047	
_cons	-3.651248	.1461914	-24.98	0.000	-3.937778	-3.364718	
-----							
/lnalpha	.6547484	.1014965			.4558189	.8536779	
-----							
alpha	1.924658	.1953461			1.577465	2.348268	
-----							

Likelihood-ratio test of alpha=0: chibar2(01) = 278.06 Prob>=chibar2 = 0.000

```
. nbreg ipdis08 i.healthstatus ib3.agecat i.race i.sex ib5.fpl ib3.education ///
> ib1.msa i.region i.insurance if adult == 1, nolog irr
```

Negative binomial regression	Number of obs	=	16462
	LR chi2(26)	=	997.76
Dispersion = mean	Prob > chi2	=	0.0000
Log likelihood = -4377.582	Pseudo R2	=	0.1023

	ipdis08	IRR	Std. Err.	z	P> z	[95% Conf. Interval]	
	healthstatus						
	Very Good	1.291973	.1326729	2.49	0.013	1.056437	1.580023
	Good	1.825908	.1824593	6.03	0.000	1.501136	2.220944
	Fair	3.21919	.3652202	10.31	0.000	2.577371	4.020835
	Poor	8.465403	1.150471	15.72	0.000	6.485852	11.04913
	agecat						
	18-24	1.004044	.0927227	0.04	0.965	.8378085	1.203263
	45-64	.9664062	.0694361	-0.48	0.634	.8394619	1.112547
	race						
	Black	.832668	.0696307	-2.19	0.029	.7067915	.9809625
	Hispanic	.7570115	.0659504	-3.20	0.001	.6381845	.8979634
	Asian	.3456921	.0749439	-4.90	0.000	.2260238	.5287187
	Other	.949145	.1795061	-0.28	0.783	.6551638	1.37504
	2.sex	1.950722	.1317404	9.89	0.000	1.708875	2.226797
	fpl						
	Poor	1.893903	.2128192	5.68	0.000	1.519524	2.360521
	Near Poor	1.394835	.2086156	2.22	0.026	1.040435	1.869955
	Low Income	1.408819	.1546764	3.12	0.002	1.136058	1.747068
	Middle Income	1.128738	.1055809	1.29	0.195	.9396643	1.355856
	education						
	No High School	.7979078	.101429	-1.78	0.076	.6219409	1.023661
	Some High	.7607863	.0715059	-2.91	0.004	.6327887	.9146747
Some College/Tech School/AA degree		1.022133	.0886011	0.25	0.801	.8624283	1.211412
	College	1.231715	.1199287	2.14	0.032	1.017727	1.490698
	msa						
	Not in MSA	.920522	.0811541	-0.94	0.348	.7744464	1.09415
	region						
	South	1.158907	.1096068	1.56	0.119	.9628169	1.394933
	Midwest	1.127806	.117206	1.16	0.247	.9199714	1.382594
	West	.9312718	.096965	-0.68	0.494	.7593614	1.142101
	insurance						
	Public Insurance	1.816403	.1710151	6.34	0.000	1.510328	2.184506
	Uninsured	.4929119	.0505209	-6.90	0.000	.4032046	.6025777

Medicare		2.033948	.2699449	5.35	0.000	1.568082	2.638221
_cons		.0259587	.0037949	-24.98	0.000	.0194915	.0345718
-----							
/lnalpha		.6547484	.1014965			.4558189	.8536779
-----							
alpha		1.924658	.1953461			1.577465	2.348268
-----							

Likelihood-ratio test of alpha=0: chibar2(01) = 278.06 Prob>=chibar2 = 0.000

.  
end of do-file

. log close  
name: <unnamed>  
log: C:\Users\bdela\OneDrive\Documents\JHU\TA Folder\Advanced HSR Methods -  
Analysis\Answers\Assignment4\_2016.log  
log type: text  
closed on: 5 Oct 2016, 22:22:42

-----  
-----