Intermediate Social Statistics
Lecture 7: Count and Duration Models

Raymond Duch
Nuffield College, Oxford University
www.raymondduch.com

February 28, 2008

1 Motivation

Today we'll consider two class of models:

1. where Y represents event counts such as the number of daily terrorist bombings in Iraq;
2. where Y represents the duration of events such as the tenure of an Italian cabinet as expressed in months

And we'll treat these separately.

All of our models will feel like probability models, of the sort:

Prob(event j occurs) = Prob(Y=j) = F[stochastic component, systematic component]

2 Event Count Data

Count data is data where the dependent variable assumes nonnegative integer values (0, 1, 2, . . .) for each of n observations. These values represent the number of times an event occurs within a fixed observation period. Examples of count data would include number of presidential vetos per congressional session, annual number of presidential nominations for the Supreme Court, and the number of military conflicts between countries. Events occur at an unobserved expected rate of event occurrence during the observation period. We get to see the number of events that occurred during the period only at the end of the period. Least squares regression does not handle these kinds of data very well:

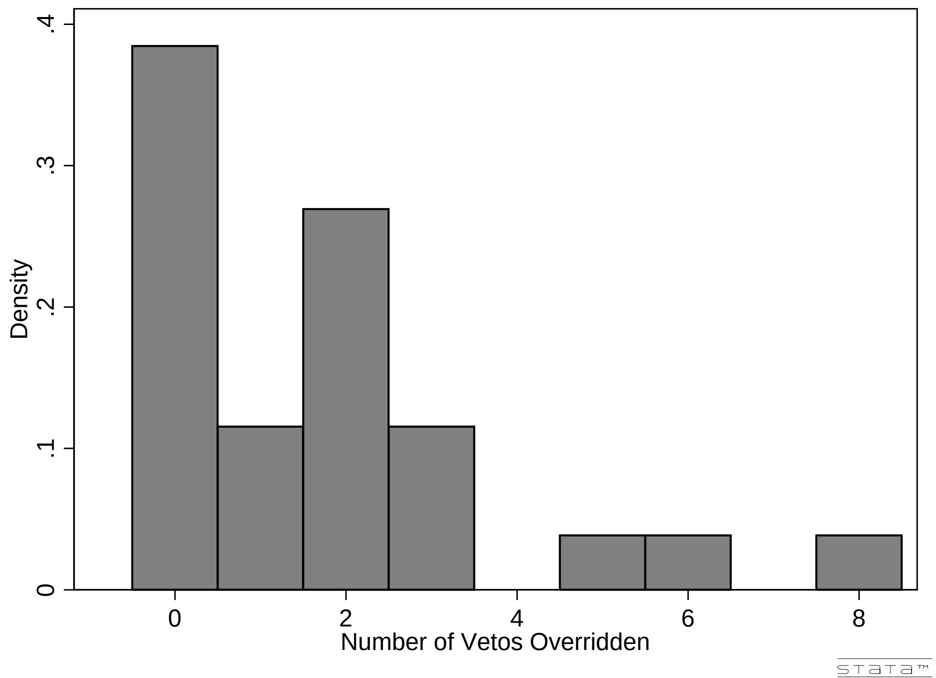
1. The linearity assumption is inappropriate for count data.
2. OLS does not constrain the expected number of events to be positive.
3. Count data typically come from distributions that are heteroskedastic. As a result least squares tends to be inefficient and it gives inconsistent standard errors.

2.1 Example of Count Data: Annual Presidential Veto Overrides

Table 1: Presidential Veto Overrides

	Frequency	Percentage
0	10	39
1	3	12
2	7	37
3	3	12
5	1	4
6	1	4
8	1	4

Figure 1: Number of Annual Presidential Veto Overrides



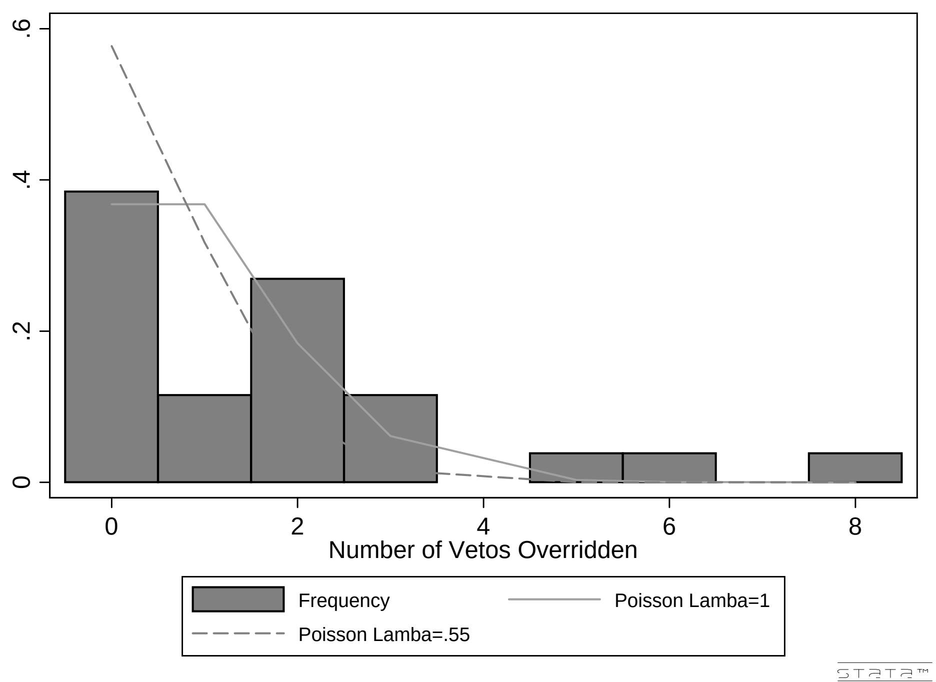
2.2 Poisson Distribution

The most basic data generating process we can use to model count data – and the veto override example above – is the Poisson distribution.

$$Prob(y_i|\lambda) = \frac{e^{(-\lambda)}\lambda^{y_i}}{y_i!} \text{ for } y_i = 0, 1, 2...$$

Modeling veto overrides as a simple poisson distribution with $\lambda = 1$ and $\lambda = .55$ does a reasonably good job of fitting the data.

Figure 2: Number of Annual Presidential Veto Overrides



```
. use "e:\Oxford08\Department08\ISS\Count_models\veto2.dta", clear

.
. stset nover

     failure event:  (assumed to fail at time=nover)
obs. time interval:  (0, nover]
 exit on or before:  failure

------------------------------------------------------------------------------
       26  total obs.
       10  obs. end on or before enter()
------------------------------------------------------------------------------
       16  obs. remaining, representing
       16  failures in single record/single failure data
       45  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =         8

.
. poisson nover

Iteration 0:   log likelihood = -52.513187
Iteration 1:   log likelihood = -52.513187

Poisson regression                                Number of obs   =         26
                                                  LR chi2(0)      =      -0.00
                                                  Prob > chi2     =          .
Log likelihood = -52.513187                       Pseudo R2       =    -0.0000

------------------------------------------------------------------------------
       nover |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |    .548566   .1490712     3.68   0.000     .2563918    .8407401
------------------------------------------------------------------------------

.
end of do-file
```

2.3 Poisson Regression

In the Poisson regression model the number of events, y_i, has a Poisson distribution, as above, except now it has a conditional mean that depends on an individual's characteristics – hence we have added a systematic component to the data generating process, or to the distribution function. The systematic component of the model can be specified as.

$$\lambda = E(y_i|\mathbf{x_i}) = exp(\mathbf{x_i}\beta)$$

The probability of particular events is calculated as

where $\lambda = E(y_i|\mathbf{x_i}) = exp(\mathbf{x_i}\beta)$

The likelihood for the Poisson Regression Model is the following.

$$L(\beta|\mathbf{y}, \mathbf{X}) = \prod_{i=1}^{N} P(y_i|\lambda_i) = \prod_{i=0}^{N} \frac{e^{(-\lambda)}\lambda^{y_i}}{y_i!}$$

where $\lambda = e^{\mathbf{X}\beta}$

We identify the βs that maximize this likelihood function. Here I have generated an example of such an estimation using Stata and the Presidential veto override data.

```
.
. poisson nover nveto janpop preshmaj pressmaj

Iteration 0:   log likelihood =  -37.96409
Iteration 1:   log likelihood = -37.908086
Iteration 2:   log likelihood = -37.907938
Iteration 3:   log likelihood = -37.907938

Poisson regression                                Number of obs   =         26
                                                  LR chi2(4)      =      29.21
                                                  Prob > chi2     =     0.0000
Log likelihood = -37.907938                       Pseudo R2       =     0.2781

------------------------------------------------------------------------------
       nover |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nveto |   .0406958   .0090975     4.47   0.000      .0228651    .0585265
      janpop |  -.0296944   .0158098    -1.88   0.060     -.0606811    .0012923
    preshmaj |  -1.167975   .6312272    -1.85   0.064     -2.405158    .0692071
    pressmaj |   .0084897   .4738788     0.02   0.986     -.9202958    .9372751
       _cons |   1.718194   .8871924     1.94   0.053     -.0206716    3.457059
------------------------------------------------------------------------------
```

2.4 Model Fit

What do we gain by incorporating this systematic component to the poisson distribution function? By adding this systematic component we are taking into account the possibility that λ varies across years (or more accurately political contexts). In other words the rate of presidential overrides varies by political contexts. This is referred to as capturing heterogeneity in the sample.

The log likelihood ratio test is simply:

$$LRI = 2 * (lnL - lnL_0) \tag{1}$$

where L_0 is the log-likelihood computed only with a constant, and ln L is log-likelihood with the systematic component included. This has a chi square distribution with degrees of freedom equal to the number of restrictions imposed. So in the Poisson example above the first Poisson estimate without any systematic component provides lnL_0 which is -52.51 and the estimate with the systematic component provides the unconstrained lnL which is -37.90. Hence the log likelihood ratio equals $-37.9 - (-52.5) = 29.2$ The Chi-square p value associated with 4 degrees and a value of 29.2 is 0.000. Hence adding this systematic component to the model significantly improves on the fit between the predictions from the Poisson distribution and the actual data.

2.5 Model Predictions

First of all what are the variables in the systematic component of the model.

```
. summarize  nover nveto janpop preshmaj pressmaj

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       nover |        26    1.730769    2.050516          0          8
       nveto |        26    15.61538    14.54119          0         70
      janpop |        26    58.23077    11.09705         36         74
    preshmaj |        26    .3846154    .4961389          0          1
    pressmaj |        26          .5     .509902          0          1
```

In order to generate a prediction we need to determine the values of the independent variable associated with the prediction we want to make. For example, what is the predicted rate of veto overrides if

1. the number of vetos equal the mean: 16
2. the president's popularity level is at its mean: 58
3. the president has a majority in the house: 1
4. the president has a majority in the senate: 1

First calculate the value of the systematic term in the Poisson Regression Model which is $\lambda = e^{1.72+.04(16)-.029(58)-1.17(1)+0.008(1)} = .616$

Then calculate the probability using the Poisson distribution including the the λ term we generated.

$$Prob(y_i = 2|\lambda = .616) = \frac{e^{(-.616)}(-.616)^2}{2!} = .10$$

```
.
. simqi, prval(0 1 2 3 4 5 6 7 8)

      Quantity of Interest |      Mean       Std. Err.    [95% Conf. Interval]
---------------------------+--------------------------------------------------
               Pr(nover=0) |   .5503067      .1307383      .2807208     .7684717
               Pr(nover=1) |   .3123263      .0492409      .2023782     .3677523
               Pr(nover=2) |   .1036074      .0541755      .0266483      .226528
               Pr(nover=3) |   .0265253      .0252612      .0023393     .0959267
               Pr(nover=4) |   .0058239      .0087028       .000154     .0304662
               Pr(nover=5) |   .0011542      .0025519      8.11e-06     .0077408
               Pr(nover=6) |   .0002122      .0006716      3.56e-07      .001639
               Pr(nover=7) |   .0000368      .0001617      1.34e-08     .0002975
               Pr(nover=8) |   6.06e-06      .0000358      4.41e-10     .0000472

.
```

We can also generate the expected value of y, or the incidence rate, for a given value of $\mathbf{x_i}$. where $\lambda = E(y|\mathbf{X}) = e^{\mathbf{X}\beta}$

So rather than calculating the probability that the number of veto overrides equal 2 for a given set of values on the independent variables $\mathbf{X}$ we can calculate the expected number of veto overrides associated with a political contexts corresponding to values on the independent variables that we used earlier, i.e.,

1. the number of vetos equal the mean: 16
2. the president's popularity level is at its mean: 58
3. the president has a majority in the house: 1
4. the president has a majority in the senate: 1

Recall that $\lambda = .616$ – hence the expected number of veto overrides is calculated as. where $E(y|\mathbf{X}) = e^{\mathbf{X}\beta} = e^{\lambda} = e^{.616} = 1.85$

2.6 Interpreting the effect of changes in the independent variable: factor change

The impact of a δ factor change in one of the independent variables ($\mathbf{x_k}$) on the expected incidence rate is calculated as follows

$$\frac{E(y|\mathbf{x}, \mathbf{x_k} + \delta)}{E(y|\mathbf{x}, \mathbf{x_k})} = exp(\beta_k * \delta)$$

So, for example, a $\delta = 1$ (or a unit change in) number of vetos (nveto) results in an increase in the expected veto override count by a factor of exp(.04 X 1) which is 1.04. You can request that Stata generate these factor changes associated with a 1 unit change in each of the independent variables by specifying the irr option.

```
. poisson nover nveto janpop preshmaj pressmaj, irr

Iteration 0:   log likelihood =  -37.96409
Iteration 1:   log likelihood = -37.908086
Iteration 2:   log likelihood = -37.907938
Iteration 3:   log likelihood = -37.907938


Poisson regression                                Number of obs   =         26
                                                  LR chi2(4)      =      29.21
                                                  Prob > chi2     =     0.0000
Log likelihood = -37.907938                       Pseudo R2       =     0.2781

------------------------------------------------------------------------------
       nover |        IRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nveto |   1.041535   .0094753     4.47   0.000     1.023128    1.060273
      janpop |   .9707421   .0153473    -1.88   0.060     .9411233    1.001293
    preshmaj |   .3109959   .1963091    -1.85   0.064     .0902512    1.071658
    pressmaj |   1.008526    .477919     0.02   0.986     .3984012    2.553015
------------------------------------------------------------------------------
```

2.7 Interpreting the effect of changes in the independent variable: discrete change

Another strategy for evaluating the effect of independent variables is to set the independent variables to meaningful values and then to change the value of one independent variable of interest. The change in the value of this independent variable of interest should also be a meaningful magnitude. Define the Δx_k as a change in x_k from x_s to x_e then the calculation for the change in the expected count is the following.

$$\frac{\Delta E(y|\mathbf{x}}{\Delta \mathbf{x_k}} = E(y|\mathbf{x}, x_k = x_e) - E(y|\mathbf{x}, x_k = x_s)$$

To illustate using the veto override data, we start by setting the independent variables to the values we used before:

1. the number of vetos equal the mean: 16
2. the president's popularity level is at its mean: 58
3. the president has a majority in the house: 1
4. the president has a majority in the senate: 1

And then we can evaluate the impact on the expected count of veto overrides:

```
. simqi, fd(ev) changex(janpop 50 70)

First Difference: janpop 50 70

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+--------------------------------------------------
                 dE(nover) |  -.3639504      .272256    -1.032616    .0220907

. simqi, fd(ev) changex(preshmaj 0 1)

First Difference: preshmaj 0 1

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+--------------------------------------------------
                 dE(nover) |  -1.458771     1.023935    -3.796993    .0521842

. simqi, fd(ev) changex(nveto 10 16)

First Difference: nveto 10 16

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+--------------------------------------------------
                 dE(nover) |   .1321806     .0499325     .0609719    .2558647

.
end of do-file
```

We can also examine the impact this has on the probability of different counts of veto overrides:

```
. simqi, fd(prval(1 2 3 4 5 6 7 8)) changex(janpop 50 70)

First Difference: janpop 50 70

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+------------------------------------------------
            dPr(nover = 1) |  -.0597374     .0549053    -.1566853     .067014
            dPr(nover = 2) |  -.0698484     .0442448    -.1631751    .0031131
            dPr(nover = 3) |  -.0320729     .0326001    -.1201177    .0006835
            dPr(nover = 4) |   -.011035     .0175319    -.0620989    .0001342
            dPr(nover = 5) |  -.0032987     .0077012    -.0256795    .0000171
            dPr(nover = 6) |  -.0009051     .0029273    -.0085707    1.49e-06
            dPr(nover = 7) |  -.0002327     .0009866    -.0024067    1.30e-07
            dPr(nover = 8) |  -.0000565     .0002986    -.0005754    9.08e-09

. simqi, fd(prval(1 2 3 4 5 6 7 8)) changex(preshmaj 0 1)

First Difference: preshmaj 0 1

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+------------------------------------------------
            dPr(nover = 1) |   .0507437     .1025863    -.1242928    .2668237
            dPr(nover = 2) |  -.1257866     .0727356    -.2324062    .0419945
            dPr(nover = 3) |  -.1284578     .0650636    -.2157733    .0071279
            dPr(nover = 4) |  -.0842315     .0594846    -.1907398    .0015623
            dPr(nover = 5) |  -.0463707     .0463272    -.1671727    .0004097
            dPr(nover = 6) |   -.023288     .0318944    -.1254237    .0000712
            dPr(nover = 7) |   -.011083     .0204767    -.0811602    .0000116
            dPr(nover = 8) |  -.0051225     .0130159    -.0454118    1.58e-06

. simqi, fd(prval(1 2 3 4 5 6 7 8)) changex(nveto 10 16)

First Difference: nveto 10 16

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+------------------------------------------------
            dPr(nover = 1) |   .0296663     .0176271    -.0163592    .0546356
            dPr(nover = 2) |   .0271057     .0098019     .0095376    .0473241
            dPr(nover = 3) |   .0104703     .0080145     .0012539    .0310343
            dPr(nover = 4) |   .0029935     .0039455     .0001043     .014723
            dPr(nover = 5) |   .0007312     .0014828     6.22e-06    .0050751
            dPr(nover = 6) |   .0001615     .0004643     2.99e-07    .0013541
            dPr(nover = 7) |    .000033     .0001253     1.20e-08     .000296
            dPr(nover = 8) |   6.26e-06     .0000298     4.15e-10    .0000551
```

3 Negative Binomial

In the Poisson Regression Model the variance $(Y|\mathbf{x}) = exp(\mathbf{x}\beta)$. The Negative Binomial Regression Model is an extension of the Poisson Regression Model that allows for the variance of y to exceed the conditional mean.

Recall that in the Poisson Regression Model the conditional mean of y given x is $exp(\mathbf{x}\beta)$. The essential extension to the Poisson is that λ is replaced with a random variable $\tilde{\lambda}$:

$$\tilde{\lambda_i} = exp(\mathbf{x_i}\beta + \epsilon_i)$$

The relationship between $\tilde{\lambda}$ and our original λ is the following:

$$\tilde{\lambda_i} = exp(\mathbf{x_i}\beta) * exp(\epsilon_i) = \lambda_i * exp(\epsilon_i) = \lambda_i\delta_i$$

where by assumption $E(\delta_i) = 1$

The conditional variance of y though is different.

$$Var(y_i|\mathbf{x}) = \lambda_i(1 + \frac{\lambda_i}{\alpha^{-1}} = \lambda_i(1 + \alpha\lambda_i) = \lambda_i + \alpha\lambda_i^2$$

Note that α in effect calibrate the extent to which there is dispersion – its a dispersion parameter. When it assumes a value of zero, then this is equivalent to the Poisson Regression Model

3.1 Estimation

see Text

3.2 Illustration: Replication of Veto Override Estimation

```
. nbreg nover nveto janpop preshmaj pressmaj

Fitting Poisson model:

Iteration 0:   log likelihood =  -37.96409
Iteration 1:   log likelihood = -37.908086
Iteration 2:   log likelihood = -37.907938
Iteration 3:   log likelihood = -37.907938

Fitting constant-only model:
```

```
Iteration 0:   log likelihood = -46.639949
Iteration 1:   log likelihood =  -46.62391
Iteration 2:   log likelihood = -46.623894
Iteration 3:   log likelihood = -46.623894

Fitting full model:

Iteration 0:   log likelihood = -41.236354
Iteration 1:   log likelihood = -38.114232
Iteration 2:   log likelihood = -37.659138
Iteration 3:   log likelihood = -37.611238
Iteration 4:   log likelihood = -37.609978
Iteration 5:   log likelihood = -37.609976

Negative binomial regression                      Number of obs   =         26
                                                  LR chi2(4)      =      18.03
Dispersion     = mean                             Prob > chi2     =     0.0012
Log likelihood = -37.609976                       Pseudo R2       =     0.1933

------------------------------------------------------------------------------
       nover |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nveto |   .0425453     .01128     3.77   0.000     .0204368    .0646538
      janpop |   -.035034   .0188896    -1.85   0.064     -.072057    .0019889
    preshmaj |  -1.149386   .6660737    -1.73   0.084    -2.454866    .1560944
    pressmaj |  -.0048036   .5338629    -0.01   0.993    -1.051156    1.041549
       _cons |   1.983813   1.044684     1.90   0.058    -.0637302    4.031356
-------------+----------------------------------------------------------------
    /lnalpha |  -2.083274   1.598841                     -5.216946    1.050398
-------------+----------------------------------------------------------------
       alpha |   .1245219   .1990907                      .0054239    2.858787
------------------------------------------------------------------------------
Likelihood-ratio test of alpha=0:  chibar2(01) =    0.60 Prob>=chibar2 = 0.220

.
end of do-file
```

3.3 Illustration: Comparing PRM and NBRM

Estimated probabilities from NBRM

```
. simqi, prval(0 1 2 3 4 5 6 7 8)

      Quantity of Interest |     Mean       Std. Err.    [95% Conf. Interval]
---------------------------+--------------------------------------------------
              Pr(nover=0) |   .5791194     .1331436     .2887952    .8047574
```

```
Pr(nover=1) |   .2792721     .0625722     .1194495    .3620463
Pr(nover=2) |   .0974029     .0498643     .0243773    .2173391
Pr(nover=3) |   .0298162     .0252833     .0024623    .1002751
Pr(nover=4) |   .0090083      .011114      .000193     .042707
Pr(nover=5) |   .0029427     .0050196     .0000129    .0173933
Pr(nover=6) |   .0011164     .0025887     7.54e-07    .0089101
Pr(nover=7) |   .0005063     .0015378     3.61e-08    .0056669
Pr(nover=8) |   .0002688      .001004     1.56e-09    .0036611
```

.

Table 2: Comparing Count Probabilities in Poisson and Negative Binomial Regression Model

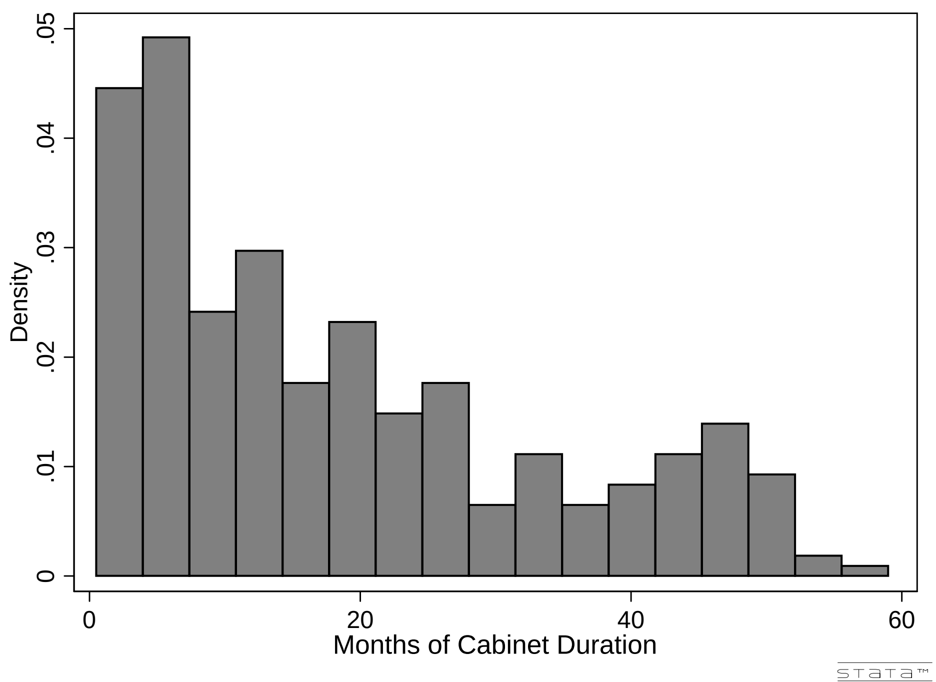
	PRM	NBRM
0	.55	.579
1	.312	.28
2	.104	.097
3	.027	.03
4	.005	.009
5	.001	.003
6	.0002	.001
7	.00003	.005
8	.000006	.0003

4 Introduction to Duration Models

1. Why OLS might not be appropriate for estimating duration models.
2. Censuring
3. Time varying independent variables

4.1 Typical Duration Data

Figure 3: Cabinet Duration Data from King et al



```
. list DURAT BELGIUM  POLAR  INVEST  NP if BELGIUM==1
```

	DURAT	BELGIUM	POLAR	INVEST	NP
1.	3	1	11	1	2.90698
2.	7	1	11	1	2.90698
3.	20	1	11	1	2.90698
4.	6	1	11	1	2.90698
5.	7	1	6	1	2.73224
6.	2	1	3	1	2.49377
7.	17	1	3	1	2.49377
8.	27	1	3	1	2.49377
9.	49	1	2	1	2.63158
10.	4	1	1	1	2.45098
11.	29	1	1	1	2.45098
12.	49	1	5	1	2.68817
13.	6	1	11	1	3.55872
14.	23	1	11	1	3.55872
15.	41	1	18	1	4.11523
16.	10	1	24	1	4.44444
17.	12	1	24	1	4.44444
18.	2	1	24	1	4.20168
19.	33	1	24	1	4.20168
20.	1	1	24	1	4.20168
21.	16	1	17	1	3.77359
22.	2	1	17	1	4.04858
23.	9	1	17	1	6.66667
24.	3	1	17	1	6.66667
25.	5	1	17	1	6.66667
26.	5	1	17	1	6.66667
27.	6	1	17	1	6.66667
28.	45	1	15	1	7.57576
29.	23	1	9	1	6.99301

```
. stset DURAT

     failure event:  (assumed to fail at time=DURAT)
obs. time interval:  (0, DURAT]
 exit on or before:  failure

------------------------------------------------------------------------------
      313  total obs.
        0  exclusions
------------------------------------------------------------------------------
      313  obs. remaining, representing
      313  failures in single record/single failure data
     5789  total analysis time at risk, at risk from t =         0
                             earliest observed entry t =         0
                                  last observed exit t =        59

.
. streg, nohr distribution(exponential)

         failure _d:  1 (meaning all fail)
   analysis time _t:  DURAT

Iteration 0:   log likelihood = -460.73222
Iteration 1:   log likelihood = -460.73222

Exponential regression -- log relative-hazard form

No. of subjects =          313                     Number of obs   =       313
No. of failures =          313
Time at risk    =         5789
                                                   LR chi2(0)      =      0.00
Log likelihood  =   -460.73222                     Prob > chi2     =         .

------------------------------------------------------------------------------
          _t |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       _cons |  -2.917512   .0565233   -51.62   0.000    -3.028295   -2.806728
------------------------------------------------------------------------------
```

5 Readings for Duration Models

Allison, Paul D. 1986. Event History Analysis Newbury Park: Sage. Chapters 1-2.

Klein, John P. and Melvin L. Moeschberger. 1997. Survival Analysis. New York: Springer. Chapters 2-3, 8, 12.

Blossfeld, Hans-Peter, Alfred Hamerle and Karl Ulrich Mayer. 1989. Event History Analysis: Statistical Theory and Applications in the Social Sciences. Hillsdale, NJ: Erlbaum. Chapters 1-4.

King, Gary, James Alt, Michael Laver and Nancy Burns. 1990. A unified model of cabinet dissolution in parliamentary democracies" American Journal of Political Science, 34: 847-71.

Homework Questions

(Due Friday of Week 9, Hilary Term)

Question 1

I have estimated an ordered multinomial logistic model to explore how characteristics of citizens impact their vote choice. I surveyed 1000 citizens before an election in which the Socialists, Conservatives, and Christian Democrats were running. The main explanatory variables are left/right, ideology and gender.

Data: surveys responses for 1000 respondents

Dependent Variable: coded 1 if voter voted for Socialists , 2 for Conservatives, 3 Christian Democrats

Independent variables:
ideology = scale of 1 to 10 with 1 being most left and 10 most right (mean = 5, sd = 2)
male = 1 identifies respondent was a male

Baseline group is Christian Democrats

1. What is the stochastic component of this model?
2. What is the systematic component of this model?

Table 3: Results

	Coefficient
Socialist	
Ideology	-0.3
Male	0.3
Constant	0.88
Conservative	
Ideology	0.2
Male	0.02
Constant	-0.9

3. What is the predicted probability of voting for each party for a male with ideology score 5?

Question 2

I have replicated below the poisson regression results for the veto override example employed in class.

```
Poisson regression                                Number of obs   =         26
                                                  LR chi2(4)      =      29.21
                                                  Prob > chi2     =     0.0000
Log likelihood = -37.907938                       Pseudo R2       =     0.2781

------------------------------------------------------------------------------
       nover |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       nveto |   .0406958   .0090975     4.47   0.000      .0228651    .0585265
      janpop |  -.0296944   .0158098    -1.88   0.060     -.0606811    .0012923
    preshmaj |  -1.167975   .6312272    -1.85   0.064     -2.405158    .0692071
    pressmaj |   .0084897   .4738788     0.02   0.986     -.9202958    .9372751
       _cons |   1.718194   .8871924     1.94   0.053     -.0206716    3.457059
------------------------------------------------------------------------------
```

1. What is the stochastic component of this model?
2. What is the systematic component of this model?
3. Here are the summary statistics for the independent variables employed in that example. enumerate

```
. summarize  nover nveto janpop preshmaj pressmaj

    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
       nover |        26    1.730769    2.050516          0          8
       nveto |        26    15.61538    14.54119          0         70
      janpop |        26    58.23077    11.09705         36         74
    preshmaj |        26    .3846154    .4961389          0          1
    pressmaj |        26          .5     .509902          0          1
```

What is the expected number of veto overrides for a President over the course of a Congressional term that had the following political characteristics:

enumerate

4. the president's party controls neither the House nor the Senate.
5. the president's approval rating stands at 35 percent.
6. congress has exercised 35 vetoes.

What is the probability of 1 veto override in this context?

Question 3

The dataset `conflict.dta` on the website is a subset of the Cross-National Time-Series Data Archive compiled by Arthur Banks. The dataset has the following variables:

year year

country country name

popdens population density

defexpgdp defense expenditure as a proportion of national expenditure

phonespc telephones per 100,000 people

tvpc televisions per 100,000 people

gdppcfc gdp/capita (factor costs)

conflictevent number of conflict events (riots, guerrilla wars, revolutions, assassinations, coups and government crises)

1. The variable `conflictevent` will be your dependent variable. Without running a regression, assess if the Poisson model will be sufficient or if you will need the negative binomial model to get efficient results.

2. We want to test three hypotheses.

   $\mathbf{H}_1$ Structure in the way of gdp and population density are significant predictors of conflict

   $\mathbf{H}_2$ Defense spending should decrease conflict because citizens then know that the government is well-armed.

   $\mathbf{H}_3$ People really engage in conflict because there is nothing better to do so giving them phones and televisions will reduce the number of conflict events.

   Run a Poisson model with `conflictevent` as the dependent variable and `defexpgdp, phonespc, tvbpc, gdppcfc` as the independent variables and test these three hypotheses above. Do you believe the Poisson model? Without running a negative binomial, how could you test to see whether your dependent variable is still overdispersed? Interpret your results in terms of percent changes holding all other variables constant.

3. Run a negative binomial regression similar to the one above. How do the results change and why do they change this way? What does the test of overdispersion in the NBR tell you?

4. Using the `prcounts` routine in stata, generate a graph that shows, on average, how close each model gets to the observed probability of the counts for the values 0-10.