Mixed Models for Longitudinal Count Data

Don Hedeker
Department of Public Health Sciences
University of Chicago

hedeker@uchicago.edu

https://voices.uchicago.edu/hedeker/

Dependent variable is a count

- number of hospitalizations

- number of service uses

- number of headaches (or some kind of disease symptom)

- number of times that an event occurs

Poisson distribution is often used to model count data

$$\Pr(y \mid \mu) = \frac{\exp(-\mu)\mu^{y}}{y!} \quad \text{for } y = 0, 1, 2, \ldots$$

μ is the expected count (per unit of time)

Can't I just analyze as continuous normal?

- count distribution is too skewed to satisfy normality (incorrect test results)
- normal model does not necessarily prevent negative estimated counts

Can't I just dichotomize count (0 vs >0) and analyze using logistic regression?

- loss of information resulting in under-powered tests
- is 1 event really equal to 100 events?
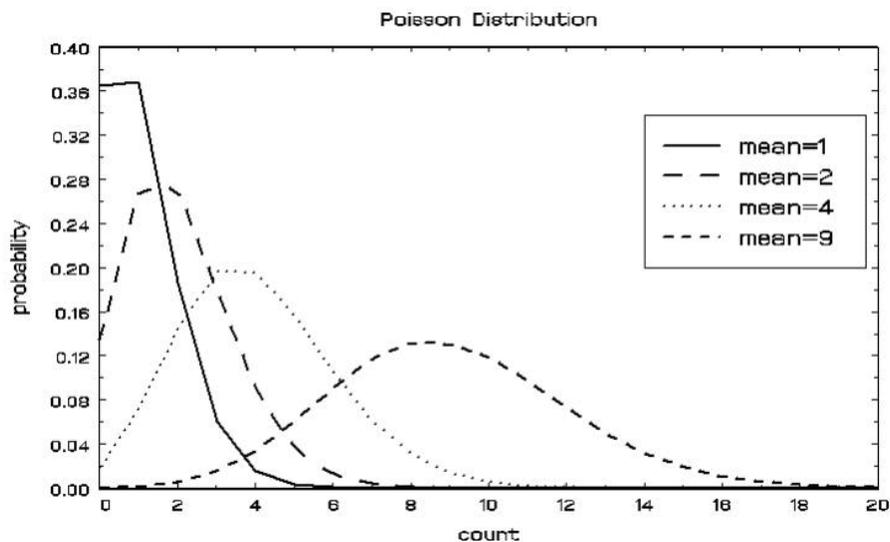
- expected number of counts (per unit of time), strictly positive
- as mean increases, probability of 0s decreases, distribution approximates normal
- mean equals the variance (if variance is greater, then overdispersion)

Reading materials and examples

- Cameron & Trivedi (1998) *Regression analysis of count data*, Cambridge Univ Press

- Long (1997) *Regression models for categorical and limited dependent variables*, Sage

- Elhai, Calhoun, & Ford (2008) Statistical procedures for analyzing mental health services data. *Psychiatry Research, 160*, 129-136.

- Walters (2007) Using Poisson Class Regression To Analyze Count Data in Correctional and Forensic Psychology. *Criminal Justice and Behavior, 34*, 1659-1674.

- Gagnon, Doron-LaMarca, Bell, O'Farrell, & Taft (2008) Poisson regression for modeling count and frequency outcomes in trauma research. *Journal of Traumatic Stress, 21*, 448-454.

- Supermix (for Windows):
  www.ssicentral.com/index.php/products/supermix/downloads

  - www.ssicentral.com/index.php/products/supermix/examples
  - in Supermix (even the free student version), from Help menu, select "Contents," "Examples from SMIX primer," "Count outcomes"

Notation is our friend!

- $i = 1, \ldots, N$ level-2 units (clusters or subjects)

- $j = 1, \ldots, n_i$ level-1 units (subjects or multiple observations)

- y_{ij} is the value of the count outcome, the number of events (y_{ij} can equal $0, 1, \ldots$)

- t_{ij} is the length of time during which the events are recorded
  - can be equal ($t_{ij} = t$): all observations are based on the same period of time, and the number of events within that same time period is of interest
  - can vary (t_{ij}): observations are based on varying periods of time; this should be accounted for when modeling the number of events within the varying time periods

Right-hand side of model

$$\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$$

- $\boldsymbol{x}_{ij}$ are covariates
  - at level-1, level-2, or cross-level interactions
  - can include polynomials, dummy variables, interactions, ...
- $\boldsymbol{\beta}$ are the regression coefficients for the covariates
- $\boldsymbol{z}_{ij}$ are the random effect variable(s)
  - usually just an intercept for clustered data
  - often an intercept and time for longitudinal data
- $\boldsymbol{v}_i$ are the random effects $\sim \mathrm{N}(0, \boldsymbol{\Sigma}_{v})$
  - how cluster i influences the observations within the cluster
  - how a subject starts and progresses across time

Mixed-effects Poisson Regression Model

The mixed-effects Poisson regression model indicates the expected number of counts in t_{ij} as:

$$E(y_{ij}) = \mu_{ij} = t_{ij} \exp\left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i\right]$$

or

$$\log(\mu_{ij}) = \log(t_{ij}) + \left[\boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i\right]$$

$$\log(\mu_{ij}) - \log(t_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$$
$$\log\left[\mu_{ij}/t_{ij}\right] = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$$

- link function for Poisson regression is the log link
- t_{ij} is sometimes called an offset or exposure variable
- $\exp\beta = $ incidence or event rate ratio

Simplest Poisson Regression example

(no random effects, no offset, dichotomous regressor)

Data and description: https://stats.idre.ucla.edu/stata/dae/negative-binomial-regression/

- School attendance data on 316 high school juniors

- Response variable is days absent (`daysabs`, range is 0 to 45)

- `male` is an indicator of student gender (0=F, 1=M)
  - `daysabs` mean for females = 6.6975
  - `daysabs` mean for males = 4.4877 (M to F ratio = .7281)

- $E(\texttt{daysabs}_i) = \mu_i = \exp(\beta_0 + \beta_1 \texttt{male}_i)$

$\exp \beta_0 = \text{mean for females} = 6.6975 \ (\hat{\beta}_0 = \log 6.6975 = 1.9017)$

$\exp(\beta_0 + \beta_1) = \exp(\beta_0) \times \exp(\beta_1) = \text{mean for males} = 4.4877$

$\exp(\hat{\beta}_1) = 4.4877/\exp(\hat{\beta}_0) = 4.4877/6.6975 = .7281$ (M to F ratio)

$(\hat{\beta}_1 = \log .7281 = -.3173)$

Longitudinal example - Aspartame and headaches

- data from McKnight and Van Den Eeden (1993) *Statistics in Medicine*, also Van Den Eeden *et al.*, (1994) *Neurology*

- number of headaches in a two treatment, multiple period crossover trial

- number of headaches per week was repeatedly measured for 27 subjects

- Following a seven-day placebo run-in period, subjects received either aspartame or placebo in four seven-day treatment periods according to a double-blind crossover treatment design

- Each treatment period was separated by a washout day

Table 4. The number of headaches, treatment days, and belief about aspartame effects by subject and treatment period

Treatment order*	ID	Run-in HA	Run-in Days	Treatment period† 1 HA	1 Days	2 HA	2 Days	3 HA	3 Days	4 HA	4 Days	Belief‡
APAP	2	0	7	5	7	2	7	—	—	—	—	V
	5	3	7	0	7	2	7	0	7	0	7	N
	13	7	7	7	7	7	7	6	7	7	7	N
	16	1	7	3	7	1	7	—	—	—	—	V
	19	0	7	1	7	1	7	0	7	1	7	V
	23	7	7	2	7	3	7	3	7	2	7	S
	25	1	7	6	7	1	7	7	7	0	7	V
APPA	1	3	7	0	7	3	7	1	7	0	7	N
	3	2	7	2	7	3	7	2	7	2	7	V
	6	1	7	1	7	0	7	3	7	1	7	S
	9	2	7	2	7	0	7	1	5	—	—	V
	17	4	7	1	1	—	—	—	—	—	—	S
	18	0	7	1	7	1	7	1	7	0	7	V
	21	1	7	2	7	3	7	3	7	6	7	V
	22	2	7	1	7	0	7	0	7	1	7	S
PAPA	7	1	7	1	7	4	7	2	7	3	7	N
	10	0	7	0	7	0	7	0	7	0	7	S
	11	0	7	3	7	1	7	0	7	1	3	S
	14	2	7	2	7	1	7	0	7	2	7	N
	24	1	7	0	7	1	7	0	7	2	7	V
	27	3	7	3	7	3	7	0	7	2	4	S
PAAP	4	0	7	0	7	0	7	0	7	0	7	V
	8	1	7	1	7	0	7	1	7	1	2	S
	12	0	7	0	7	5	7	0	7	0	7	S
	15	0	7	3	7	2	7	1	7	1	7	S
	20	1	7	6	7	1	2	—	—	—	—	V
	26	0	7	1	7	—	—	—	—	—	—	S

* A = aspartame; P = placebo. Run-in period used placebo capsules.
† HA = number of days on which subject reported a headache; Days = number of days subject participated during that period.
‡ Belief about how strongly subject felt aspartame caused headaches was asked prior to the start of the study. V = very sure; S = somewhat sure; N = not very sure or do not know.

	A	B	C	D	E	F	G	H	I
1	ID	Headache	Period1	Period2	Period3	Period4	DrugAsp	NPeriods	NTDays
2	2	0	0	0	0	0	0	3	7
3	2	5	1	0	0	0	1	3	7
4	2	2	0	1	0	0	0	3	7
5	5	3	0	0	0	0	0	5	7
6	5	0	1	0	0	0	1	5	7
7	5	2	0	1	0	0	0	5	7
8	5	0	0	0	1	0	1	5	7
9	5	0	0	0	0	1	0	5	7
10	13	7	0	0	0	0	0	5	7
11	13	7	1	0	0	0	1	5	7
12	13	7	0	1	0	0	0	5	7
13	13	6	0	0	1	0	1	5	7
14	13	7	0	0	0	1	0	5	7
15	16	1	0	0	0	0	0	3	7
16	16	3	1	0	0	0	1	3	7
17	16	1	0	1	0	0	0	3	7
18	19	0	0	0	0	0	0	5	7
19	19	1	1	0	0	0	1	5	7
20	19	1	0	1	0	0	0	5	7
21	19	0	0	0	1	0	1	5	7
22	19	1	0	0	0	1	0	5	7
23	23	7	0	0	0	0	0	5	7
24	23	2	1	0	0	0	1	5	7
25	23	3	0	1	0	0	0	5	7
26	23	3	0	0	1	0	1	5	7
27	23	2	0	0	0	1	0	5	7
28	25	1	0	0	0	0	0	5	7
29	25	6	1	0	0	0	1	5	7

- ID = patient ID (27 patients in total)

- HeadAche = number of headaches during the week (0 to 7)

- Period1 = period 1 indicator (1 = first tx period, else 0)

- Period2 = period 2 indicator (1 = second tx period, else 0)

- Period3 = period 3 indicator (1 = third tx period, else 0)

- Period4 = period 4 indicator (1 = fourth tx period, else 0)

- DrugAsp = period-specific drug (0 = placebo, 1 = aspartame)

- Nperiods = number of periods person was observed (2 to 5)

- NTDays = number of treatment days in the period (1 to 7)

Stata example: aspart.do

```
cd "u:\Stata_long\"
log using aspart.log, replace
import excel using aspart.xlsx, sheet("aspartamin") firstrow clear

* summary statistics on all variables
summ

* Headache descriptives by DrugAsp
format Headache %8.4f
tabulate DrugAsp, summarize(Headache)

* some graphs
format Headache %2.0f
twoway histogram Headache, discrete frequency
graph box Headache, over(DrugAsp, relabel(1 "Placebo" 2 "Aspartame"))
graph bar Headache, over(DrugAsp, relabel(1 "Placebo" 2 "Aspartame"))

* random intercept Poisson regression model
mepoisson Headache Period1 Period2 Period3 Period4 DrugAsp || ID:
```

```
. import excel using aspart.xlsx, sheet("aspartamin") firstrow clear

. summ

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
          ID |        122     13.7377    7.783011          1         27
    Headache |        122    1.680328    1.886299          0          7
     Period1 |        122    .2213115    .4168416          0          1
     Period2 |        122     .204918    .4053062          0          1
     Period3 |        122    .1803279    .3860457          0          1
-------------+--------------------------------------------------------
     Period4 |        122    .1721311    .3790511          0          1
     DrugAsp |        122    .3852459    .4886602          0          1
    NPeriods |        122    4.721311    .7414783          2          5
      NTDays |        122    6.795082    .9529296          1          7
```

```
. format Headache %8.4f

. tabulate DrugAsp, summarize(Headache)

             |      Summary of Headache
    DrugAsp  |      Mean   Std. Dev.      Freq.
-------------+------------------------------------
          0  |    1.5333      1.8034         75
          1  |    1.9149      2.0090         47
-------------+------------------------------------
      Total  |    1.6803      1.8863        122
```

- Placebo mean = 1.5333

- Aspartame mean = 1.9149

- Rate ratio = 1.9149 / 1.5333 = 1.25

- log(Rate ratio) = 0.223

```
. format Headache %2.0f
. twoway histogram Headache, discrete frequency
```
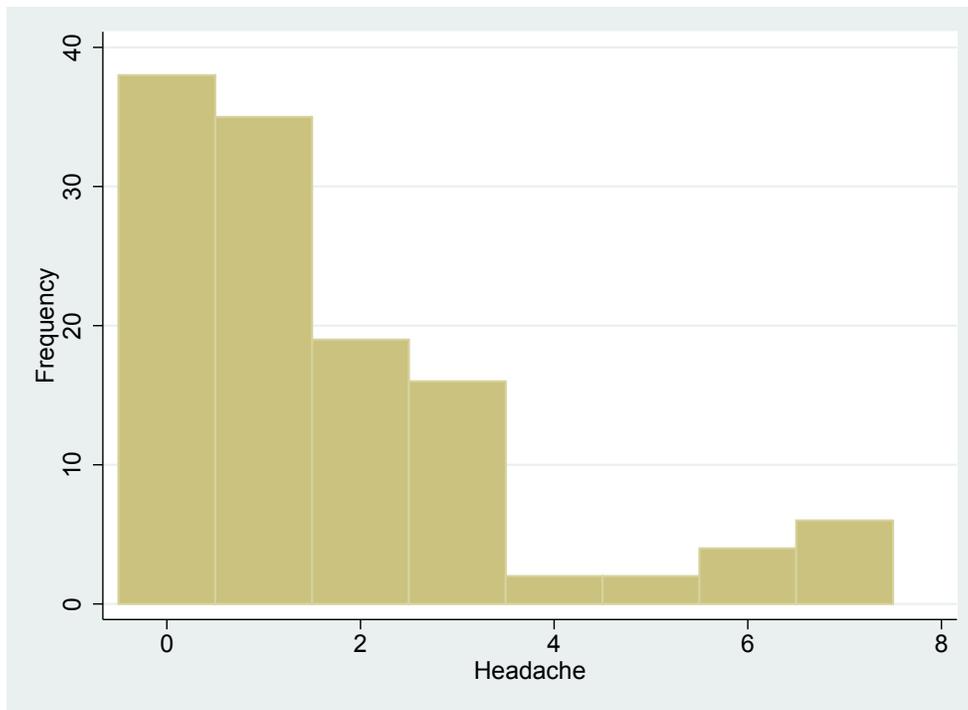
```
. graph box Headache, over(DrugAsp, relabel(1 "Placebo" 2 "Aspartame"))
```
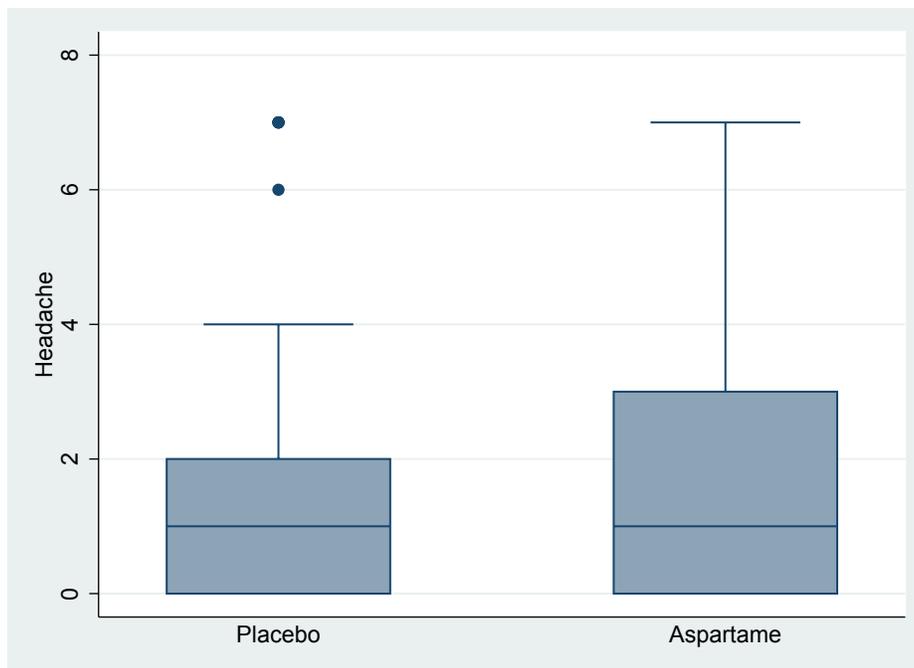

note: 1 and 2 refer to 1st and 2nd levels of the variable DrugAsp (which are coded 0 and 1)

. graph bar Headache, over(DrugAsp, relabel(1 "Placebo" 2 "Aspartame"))
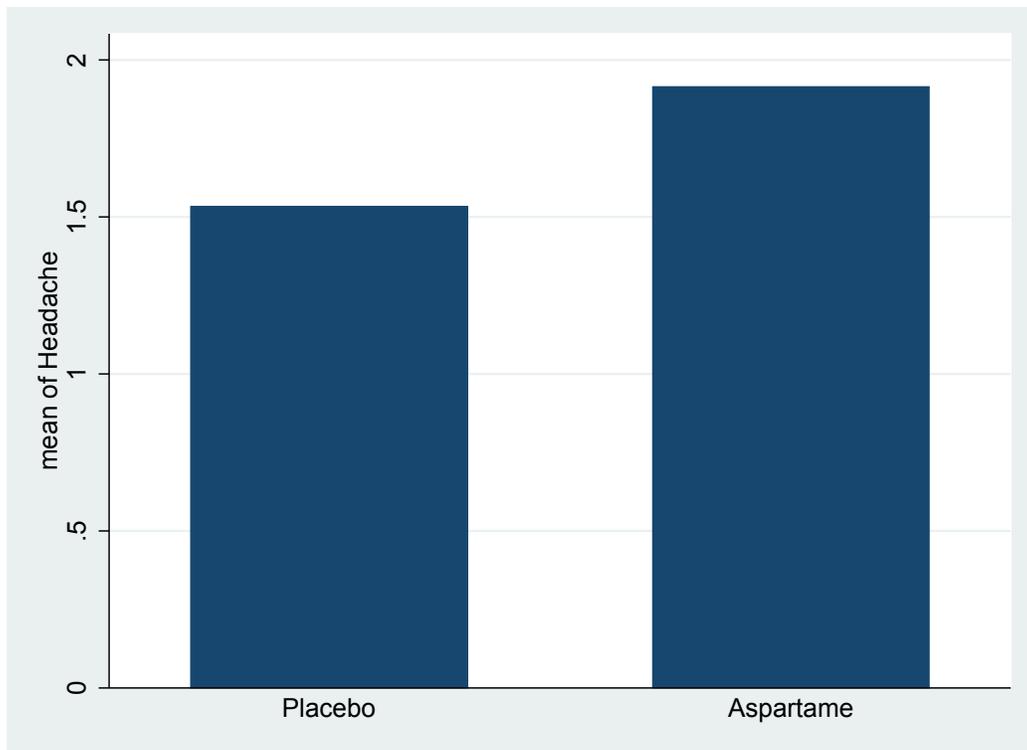
. mepoisson Headache Period1 Period2 Period3 Period4 DrugAsp || ID:


```
Mixed-effects Poisson regression              Number of obs      =         122
Group variable:                    ID         Number of groups   =          27

                                              Obs per group:
                                                           min =           2
                                                           avg =         4.5
                                                           max =           5

Integration method: mvaghermite                Integration pts.  =           7

                                              Wald chi2(5)        =        4.79
Log likelihood = -203.17671                   Prob > chi2         =      0.4426
------------------------------------------------------------------------------
    Headache |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Period1 |   .0805749    .235016     0.34   0.732     -.380048    .5411979
     Period2 |   .0344691   .2236416     0.15   0.878    -.4038603    .4727985
     Period3 |  -.2272522   .2545615    -0.89   0.372    -.7261835    .2716791
     Period4 |   -.159424   .2528384    -0.63   0.528    -.6549782    .3361301
```

```
     DrugAsp |    .2150593    .1638877      1.31   0.189     -.1061546     .5362733
       _cons |    .2507236    .2050712      1.22   0.221     -.1512086     .6526557
-------------+----------------------------------------------------------------
ID           |
   var(_cons)|    .4330674    .1741208                        .1969328     .9523419
-------------+----------------------------------------------------------------
LR test vs. Poisson model: chibar2(01) = 55.97         Prob >= chibar2 = 0.0000
```

Comparing mixed Poisson to ordinary (fixed) Poisson regression

```
LR test vs.  Poisson model:  chibar2(01) = 55.97 Prob>=chibar2 = 0.0000
```

$H_0 : \sigma_v^2 = 0, \quad H_A : \sigma_v^2 > 0 \quad \Rightarrow$ one-sided test

`chibar2(01)` refers to a 50:50 mixture of a χ_0^2 and a χ_1^2 distribution; chi-bar square distribution; p-value is obtained from χ_1^2, but is halved

Interpretation of Drug Effect

- $\hat{\beta}_{\text{DrugAsp}} = .2151$

- $\exp(\hat{\beta}_{\text{DrugAsp}}) = 1.24$

- Aspartame increases the expected number of headaches (per week) by 24%, controlling for the period and random subject effects

- However, this is NOT a significant effect (p-value $= .19$)

- For Poisson random-intercept model, this is also the marginal effect (except for intercept β_0, conditional β = marginal β)

  Aspartame increases the expected number of headaches by 24% controlling for the period effects

Observed means: Headaches across time by drug

drug	baseline	period 1	period 2	period 3	period 4
placebo	1.593	1.667	1.929	1.000	1.333
	(n=27)	(n=12)	(n=14)	(n=13)	(n=9)
aspartame		2.267	1.636	2.000	1.667
		(n=15)	(n=11)	(n=9)	(n=12)
rate ratio		1.36	0.85	2.00	1.25

Estimated means $= \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}_v^2/2)$

drug	baseline	period 1	period 2	period 3	period 4
placebo	1.603	1.737	1.659	1.278	1.367
aspartame		2.154	2.057	1.584	1.695
rate ratio		1.24	1.24	1.24	1.24

SAS for mixed model analysis of count outcomes

PROC GLIMMIX (version 9.1.3 and thereafter)

- Multiple levels of nesting, crossed random effects

- Pseudo-likelihood estimation (by default)

  - Linearization to avoid integration over the random effects
  - Produces biased estimates if number of level-1 or level-2 units is small and/or ICC is large

- Full likelihood estimation using numerical quadrature for integration over the random effects `METHOD=QUAD`; however for 3-level models can only use `METHOD=QUAD(QPOINTS=1)` or `METHOD=LAPLACE` (these are equivalent)

SAS code: `aspart.sas`

```
PROC IMPORT DATAFILE = "U:\Stata_long\aspart.xlsx"
DBMS=xlsx REPLACE OUT = one;  GETNAMES=YES;

PROC MEANS;

PROC FORMAT;
VALUE Dgrp 0='Placebo' 1='Aspartame';

PROC SORT; BY DrugAsp;
/* summary stats for Headache by DrugAsp groups */
PROC MEANS; VAR Headache; BY DrugAsp; FORMAT DrugAsp Dgrp.;

/* Histogram of Headache */
PROC SGPLOT;
VBAR Headache;

/* Box-plots of Headache by DrugAsp groups */
PROC SGPLOT;
VBOX Headache / GROUP=DrugAsp;
FORMAT DrugAsp Dgrp.;
RUN;
```

The MEANS Procedure

Variable	Label	N	Mean	Std Dev	Minimum	Maximum
ID	ID	122	13.7377049	7.7830110	1.0000000	27.0000000
Headache	Headache	122	1.6803279	1.8862985	0	7.0000000
Period1	Period1	122	0.2213115	0.4168416	0	1.0000000
Period2	Period2	122	0.2049180	0.4053062	0	1.0000000
Period3	Period3	122	0.1803279	0.3860457	0	1.0000000
Period4	Period4	122	0.1721311	0.3790511	0	1.0000000
DrugAsp	DrugAsp	122	0.3852459	0.4886602	0	1.0000000
NPeriods	NPeriods	122	4.7213115	0.7414783	2.0000000	5.0000000
NTDays	NTDays	122	6.7950820	0.9529296	1.0000000	7.0000000

```
------------------------------------- DrugAsp=Placebo -------------------------------------

                                  The MEANS Procedure

                          Analysis Variable : Headache Headache

        N             Mean          Std Dev          Minimum          Maximum

        75         1.5333333       1.8034002                0        7.0000000


------------------------------------- DrugAsp=Aspartame -------------------------------------

                          Analysis Variable : Headache Headache

        N             Mean          Std Dev          Minimum          Maximum

        47         1.9148936       2.0089992                0        7.0000000
```

$\Rightarrow$ Rate ratio $= 1.915/1.533 = 1.249$
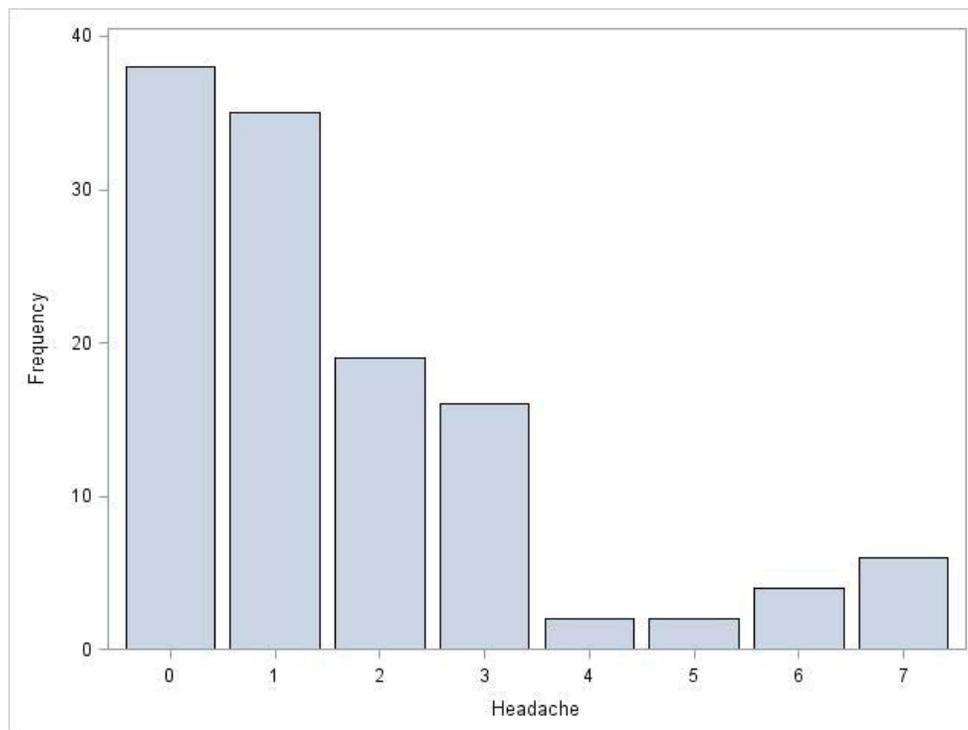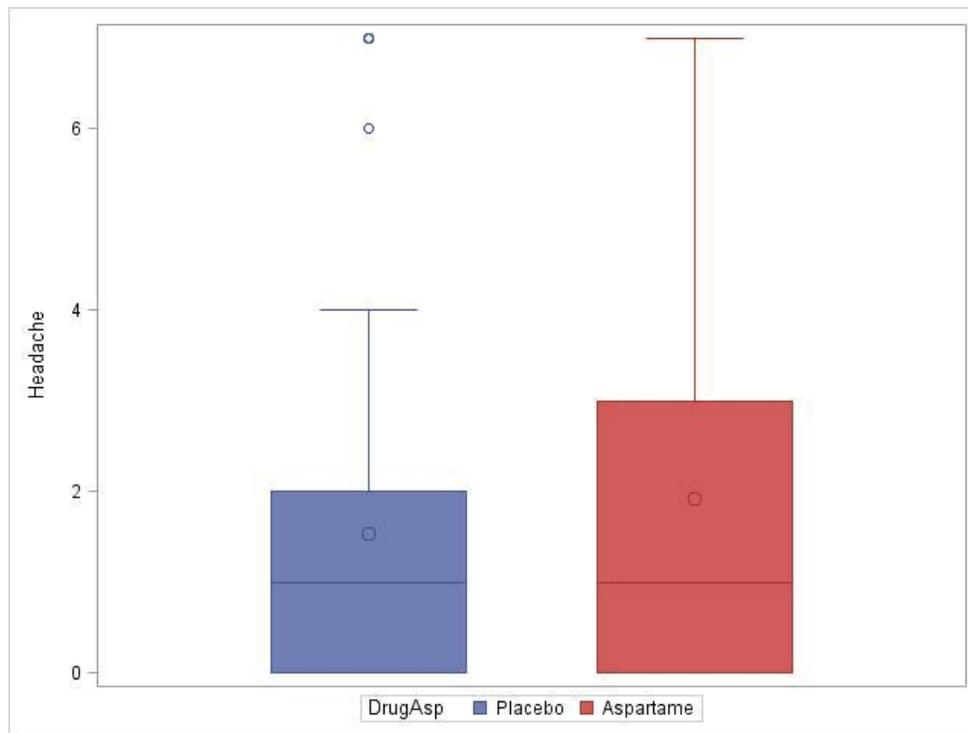
SAS code: `aspart.sas` (continued)

```
/* random intercept Poisson model */
PROC GLIMMIX METHOD=QUAD;
CLASS id;
MODEL Headache = Period1 Period2 Period3 Period4 DrugAsp
        / LINK=LOG DIST=POISSON;
RANDOM INT / SUBJECT=id;
COVTEST 'test of random intercept var' GLM;
RUN;
```

- `METHOD=QUAD` requests full-likelihood estimation (using numerical quadrature)

- `LINK=LOG` and `DIST=POISSON` for Poisson regression

- `COVTEST 'test of random intercept' GLM;`
  statement yields a likelihood ratio test of
  $H_0 : \sigma^2_{v_0} = 0, \quad H_A : \sigma^2_{v_0} > 0 \quad \Rightarrow$ one-sided test

The GLIMMIX Procedure

Model Information

```
Data Set                     WORK.ONE
Response Variable            Headache
Response Distribution        Poisson
Link Function                Log
Variance Function            Default
Variance Matrix Blocked By   ID
Estimation Technique         Maximum Likelihood
Likelihood Approximation     Gauss-Hermite Quadrature
Degrees of Freedom Method    Containment
```

Class Level Information

Class	Levels	Values
ID	27	1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

```
Number of Observations Read        122
Number of Observations Used        122
```

Dimensions

```
G-side Cov. Parameters         1
Columns in X                   6
Columns in Z per Subject       1
Subjects (Blocks in V)         27
Max Obs per Subject            5
```

Optimization Information

```
Optimization Technique       Dual Quasi-Newton
Parameters in Optimization   7
Lower Boundaries             1
Upper Boundaries             0
Fixed Effects                Not Profiled
Starting From                GLM estimates
Quadrature Points            5
```

Iteration History

Convergence criterion (GCONV=1E-8) satisfied.

```
                              Fit Statistics

                  -2 Log Likelihood               406.36
                  AIC  (smaller is better)        420.36
                  AICC (smaller is better)        421.34
                  BIC  (smaller is better)        429.43
                  CAIC (smaller is better)        436.43
                  HQIC (smaller is better)        423.05


                  Fit Statistics for Conditional Distribution

                  -2 log L(Headache | r. effects)     351.67
                  Pearson Chi-Square                   114.59
                  Pearson Chi-Square / DF               0.94


                       Covariance Parameter Estimates

                                                  Standard
                  Cov Parm     Subject    Estimate      Error

                  Intercept    ID           0.4324     0.1734



                         Solutions for Fixed Effects
```

```
                              Standard
              Effect       Estimate     Error      DF    t Value   Pr > |t|

              Intercept     0.2507     0.2051      26      1.22     0.2324
              Period1       0.08057    0.2350      90      0.34     0.7325
              Period2       0.03446    0.2236      90      0.15     0.8779
              Period3      -0.2273     0.2546      90     -0.89     0.3744
              Period4      -0.1594     0.2528      90     -0.63     0.5299
              DrugAsp       0.2151     0.1639      90      1.31     0.1927


                         Tests of Covariance Parameters
                             Based on the Likelihood

         Label                        DF    -2 Log Like     ChiSq   Pr > ChiSq   Note

         test of random intercept var  1       462.32       55.97     <.0001      MI

     MI: P-value based on a mixture of chi-squares.
```

Spearman correlations of repeated counts : aspart_corr.do

```
cd "u:\Stata_long\"
log using aspart_corr.log, replace
import excel using aspart.xlsx, sheet("aspartamin") firstrow clear
gen time=0
replace time=1 if Period1==1
replace time=2 if Period2==1
replace time=3 if Period3==1
replace time=4 if Period4==1
* summary statistics on all variables
summ
* reshape data to wide format
reshape wide Headache DrugAsp Period1 Period2 Period3 Period4 ///
           NTDays, i(ID) j(time)
* pairwise deleted Spearman correlation matrix
spearman Headache0 Headache1 Headache2 Headache3 Headache4, pw
```

```
. spearman Headache0 Headache1 Headache2 Headache3 Headache4, pw
(obs=varies)

            | Headac~0 Headac~1 Headac~2 Headac~3 Headac~4
-------------+---------------------------------------------
   Headache0 |   1.0000
   Headache1 |   0.1404    1.0000
   Headache2 |   0.3106    0.2066    1.0000
   Headache3 |   0.3029    0.4864    0.3013    1.0000
   Headache4 |   0.4420    0.5206    0.4304    0.3303    1.0000
```

Spearman correlations of repeated counts :

aspart_corr.sas

```
PROC IMPORT DATAFILE = "U:\Stata_long\aspart.xlsx"
DBMS=xlsx REPLACE OUT = one;
GETNAMES=YES;
PROC MEANS;
RUN;


/* Get Spearman correlations of repeated counts */
PROC SORT; BY id;
DATA baseline; SET one;
if Period1=0 and Period2=0 and Period3=0 and Period4=0; Headache0=Headache;
DATA t1; SET one; if Period1=1; Headache1 = Headache;
DATA t2; SET one; if Period2=1; Headache2 = Headache;
DATA t3; SET one; if Period3=1; Headache3 = Headache;
DATA t4; SET one; if Period4=1; Headache4 = Headache;
DATA wide (KEEP=id Headache0-Headache4);
MERGE baseline t1 t2 t3 t4; BY id;


PROC CORR SPEARMAN; VAR Headache0-Headache4;
RUN;
```

```
                            The CORR Procedure

            5  Variables:    Headache0 Headache1 Headache2 Headache3 Headache4


                            Simple Statistics

   Variable          N        Mean      Std Dev      Median     Minimum      Maximum

   Headache0        27     1.59259     1.92672     1.00000           0      7.00000
   Headache1        27     2.00000     1.98068     1.00000           0      7.00000
   Headache2        25     1.80000     1.75594     1.00000           0      7.00000
   Headache3        22     1.40909     1.96781     1.00000           0      7.00000
   Headache4        21     1.52381     1.88730     1.00000           0      7.00000
```

```
                    Spearman Correlation Coefficients
                       Prob > |r| under H0: Rho=0
                          Number of Observations

                Headache0      Headache1      Headache2      Headache3      Headache4

Headache0       1.00000        0.14039        0.31064        0.30294        0.44197
                               0.4849         0.1307         0.1706         0.0449
                   27             27             25             22             21


Headache1       0.14039        1.00000        0.20663        0.48640        0.52065
                0.4849                        0.3217         0.0217         0.0155
                   27             27             25             22             21


Headache2       0.31064        0.20663        1.00000        0.30127        0.43038
                0.1307         0.3217                        0.1730         0.0515
                   25             25             25             22             21


Headache3       0.30294        0.48640        0.30127        1.00000        0.33033
                0.1706         0.0217         0.1730                        0.1436
                   22             22             22             22             21


Headache4       0.44197        0.52065        0.43038        0.33033        1.00000
                0.0449         0.0155         0.0515         0.1436
                   21             21             21             21             21
```

Prior analysis assumed that all subjects were assessed for 7 days for each period. Is this true? Does it matter?

From `Aspart.do`:
`twoway histogram NTDays, discrete frequency`
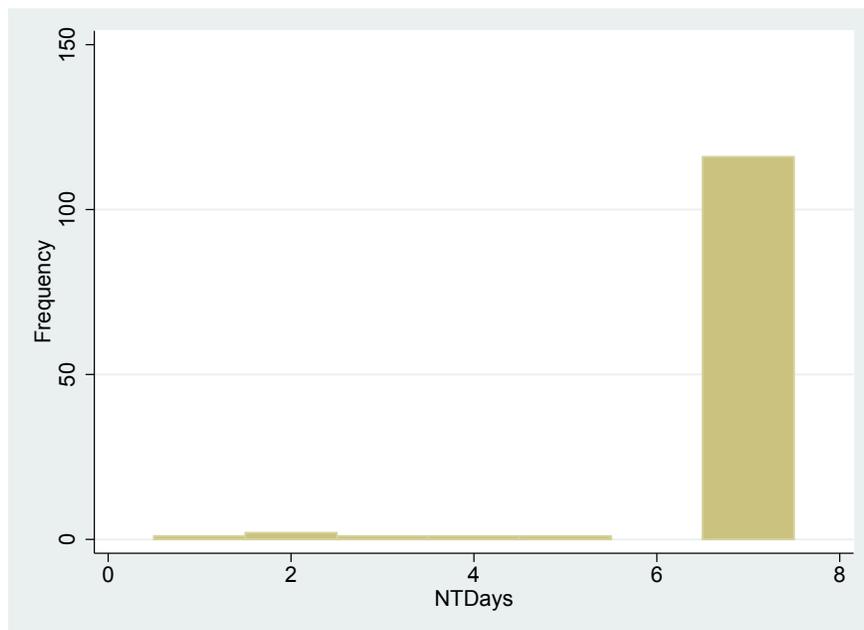
From `Aspart.sas`:

```
PROC SGPLOT;
VBAR NTDays;  RUN;
```
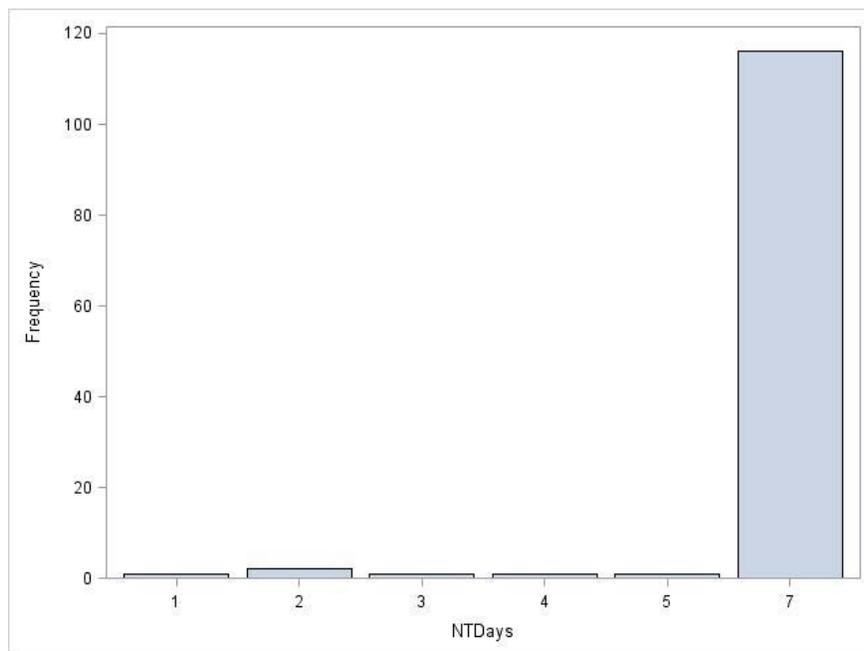
Observations with less than 7 days in a period

Period	ID	Headaches	NTDays	DrugAsp
1	17	1	1	1
2	20	1	2	1
3	9	1	5	0
4	11	1	3	1
4	27	2	4	1
4	8	1	2	0

Question

Including this information about NTDays in the model, will the drug effect be greater or smaller? or the same?

Hint: what did the prior analysis assume about NTDays?

Mixed-effects Poisson Regression Models

The mixed-effects Poisson regression model without an offest or exposure variable:

$$\log(\mu_{ij}) = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$$

The mixed-effects Poisson regression model WITH an offest or exposure variable t_{ij}:

$$\log(\mu_{ij}) = \log(t_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \boldsymbol{z}'_{ij}\boldsymbol{v}_i$$

- the (log of the) offset is like a regressor with a slope=1
- for `Stata`, identify t_{ij} as the exposure variable (it will take the log of this variable internally)

From `aspart.do`:

```
mepoisson Headache Period1 Period2 Period3 Period4 DrugAsp,  ///
exposure(NTDays) || ID:
scalar m1=e(ll)
```

```
Mixed-effects Poisson regression              Number of obs     =        122
Group variable:              ID               Number of groups  =         27

                                              Obs per group:
                                                          min =          2
                                                          avg =        4.5
                                                          max =          5

Integration method: mvaghermite               Integration pts.  =          7

                                              Wald chi2(5)      =       6.43
Log likelihood = -202.43392                   Prob > chi2       =     0.2668
------------------------------------------------------------------------------
    Headache |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Period1 |   .1001073   .2357201     0.42   0.671    -.3618957    .5621102
     Period2 |    .087958   .2249783     0.39   0.696    -.3529913    .5289074
```

```
    Period3 |  -.2115442    .2566767    -0.82   0.410    -.7146212    .2915328
    Period4 |  -.0786684    .2544603    -0.31   0.757    -.5774014    .4200645
    DrugAsp |   .2796849    .1640922     1.70   0.088    -.0419299    .6012997
      _cons |  -1.712675    .2104049    -8.14   0.000    -2.125061   -1.300289
  ln(NTDays) |          1   (exposure)
------------+------------------------------------------------------------------
ID          |
  var(_cons)|   .4779805    .1930506                      .2165808    1.054873
--------------------------------------------------------------------------------
LR test vs. Poisson model: chibar2(01) = 57.92          Prob >= chibar2 = 0.0000
```

Interpretation of Drug Effect

- $\hat{\beta}_{\text{DrugAsp}} = .2797$

- $\exp(\hat{\beta}_{\text{DrugAsp}}) = 1.32$

- Aspartame increases the expected rate of headaches (# of headaches per day) by 32%, controlling for the period and random subject effects

- 2-tailed non-significant (p-value $= .088$), but significant by a 1-tailed test (p-value $= .088/2 = .044$)

- For Poisson random-intercept model, this is also the marginal effect (except for intercept β_0, conditional $\beta =$ marginal β)

  Aspartame increases the expected rate of headaches by 32% controlling for the period effects

Observed means: Headaches/Day across time by drug

drug	baseline	period 1	period 2	period 3	period 4
placebo	.228	.238	.276	.147	.230
	(n=27)	(n=12)	(n=14)	(n=13)	(n=9)
aspartame		.381	.266	.286	.272
		(n=15)	(n=11)	(n=9)	(n=12)
rate ratio		1.60	0.96	1.95	1.18

Estimated means $= \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}_v^2/2)$

drug	baseline	period 1	period 2	period 3	period 4
placebo	.228	.252	.249	.185	.211
aspartame		.334	.330	.244	.279
rate ratio		1.32	1.32	1.32	1.32

SAS code: `aspart.sas` (continued)

```
/* random intercept Poisson model with offset */
PROC GLIMMIX METHOD=QUAD;
CLASS id;
logNTDays = log(NTDays);
MODEL Headache = Period1 Period2 Period3 Period4 DrugAsp
    / LINK=LOG S DIST=POISSON OFFSET=logNTDays;
RANDOM INT / SUBJECT=id;
COVTEST 'test of random intercept var' GLM;
RUN;
```

- User must take log of offset variable:
  `logNTDays = log(NTDays);`

- Include MODEL option: `OFFSET=logNTDays`

```
                          The GLIMMIX Procedure

                          Model Information

            Data Set                    WORK.ONE
            Response Variable           Headache
            Response Distribution       Poisson
            Link Function               Log
            Variance Function           Default
            Offset Variable             logNTDays = log(NTDays);
            Variance Matrix Blocked By  ID
            Estimation Technique        Maximum Likelihood
            Likelihood Approximation    Gauss-Hermite Quadrature
            Degrees of Freedom Method   Containment




                          Fit Statistics

                  -2 Log Likelihood               404.87
                  AIC  (smaller is better)        418.87
                  AICC (smaller is better)        419.86
                  BIC  (smaller is better)        427.94
                  CAIC (smaller is better)        434.94
                  HQIC (smaller is better)        421.57
```

```
                     Covariance Parameter Estimates

                                               Standard
                   Cov Parm    Subject    Estimate      Error

                   Intercept   ID           0.4770     0.1921


                     Solutions for Fixed Effects

                               Standard
            Effect      Estimate      Error      DF    t Value    Pr > |t|

            Intercept    -1.7126     0.2104      26      -8.14     <.0001
            Period1       0.1001     0.2357      90       0.42     0.6721
            Period2       0.08792    0.2250      90       0.39     0.6969
            Period3      -0.2116     0.2567      90      -0.82     0.4119
            Period4      -0.07871    0.2545      90      -0.31     0.7578
            DrugAsp       0.2797     0.1641      90       1.70     0.0918


                     Tests of Covariance Parameters
                        Based on the Likelihood

       Label                       DF    -2 Log Like    ChiSq    Pr > ChiSq    Note

       test of random intercept var  1       462.78     57.91       <.0001      MI

    MI: P-value based on a mixture of chi-squares.
```

Empirical Bayes estimates of random effects

$$\log(\mu_{ij}) = \log(t_{ij}) + \boldsymbol{x}'_{ij}\boldsymbol{\beta} + v_i \qquad \text{where } v_i \sim N(0, \sigma_v^2)$$

- Random effects v_i are also estimated
- can be of interest to indicate how particular subjects are doing
- can be used to rank, compare, or indicate unusual subjects
- From `Aspart.do` :

```
* get random effect estimates (and std errs)
predict u0, reffects reses(u0se)
* assign a value of 1 for one obs of each subject
* and rank the random effects
egen pick1obs = tag(ID)
egen u0rank = rank(u0) if pick1obs==1
list ID u0 u0se u0rank if pick1obs==1
* standard error bar chart of random effects
serrbar u0 u0se u0rank if pick1obs==1, scale(1.96) yline(0)
```

```
. list ID u0 u0se u0rank if pick1obs==1


        +-------------------------------------+
        | ID          u0        u0se   u0rank |
        |-------------------------------------|
  1.    |  2     .2936098   .3489556       18 |
  4.    |  5    -.2885217   .3663934        9 |
  9.    | 13     1.469566   .1751805       27 |
 14.    | 16     .0260536   .3832188       17 |
 17.    | 19     -.585127    .404517        4 |
        |-------------------------------------|
 22.    | 23     .7595504   .2416335       26 |
 27.    | 25     .6363964   .2548456       23 |
 32.    |  1    -.0479451   .3349817       14 |
 37.    |  3     .3374991   .2880078       19 |
 42.    |  6    -.1651728   .3497926       12 |
        |-------------------------------------|
 47.    |  9    -.0790354   .3773501       13 |
 51.    | 17     .7314795    .430931       25 |
 53.    | 18    -.5895116   .4042242        3 |
 58.    | 21     .6306354   .2547467       22 |
 63.    | 22    -.4340891   .3842416        6 |
        |-------------------------------------|
 68.    |  7     .3381366   .2880236       20 |
```

```
 73. |  10   -1.168355    .4763907          1 |
 78. |  11   -.1970716     .371061         11 |
 83. |  14   -.0473549    .3350045         15 |
 88. |  24   -.4335562    .3842723          7 |
      |-----------------------------------|
 93. |  27    .4194811    .2900531         21 |
 98. |   4   -1.164856    .4767669          2 |
103. |   8   -.3419288    .3896556          8 |
108. |  12   -.2879634    .3664214         10 |
113. |  15   -.0422512    .3352014         16 |
      |-----------------------------------|
118. |  20    .6717975    .3446973         24 |
121. |  26   -.4439119    .5051993          5 |
      +-----------------------------------+
```

. * standard error bar chart of random effects
. serrbar u0 u0se u0rank if pick1obs==1, scale(1.96) yline(0)
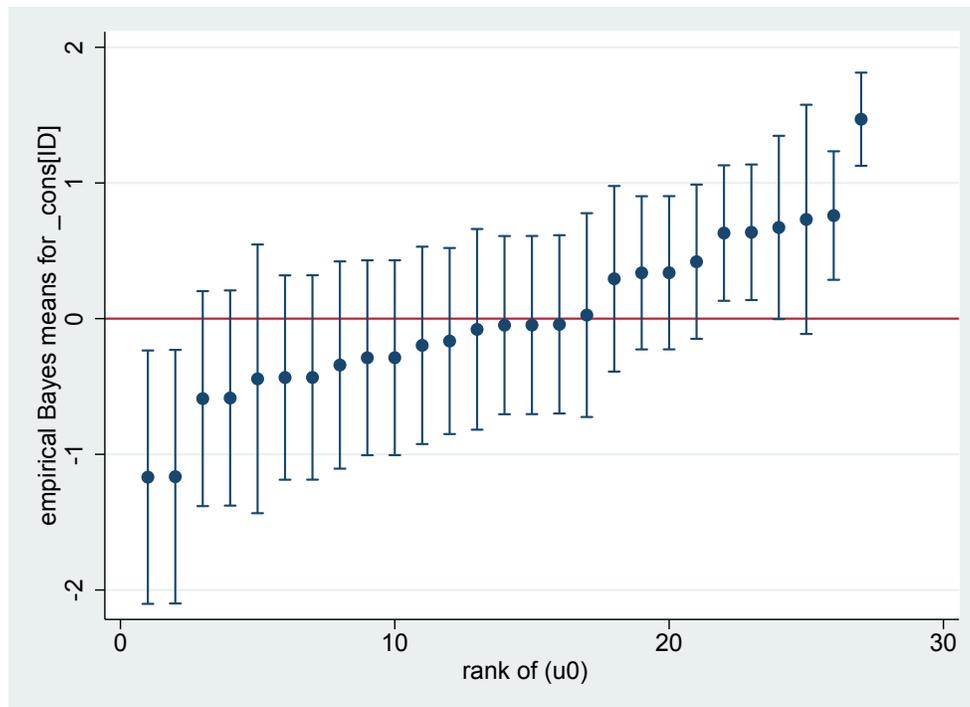
From `aspart.sas`: include `SOLUTION` or `S` on `RANDOM`
statement and use `ODS` to direct these estimates to a new dataset
named `ebest`

```
PROC GLIMMIX METHOD=QUAD;
CLASS id;
logNTDays = log(NTDays);
MODEL Headache = Period1 Period2 Period3 Period4 DrugAsp
     / LINK=LOG S DIST=POISSON OFFSET=logNTDays;
RANDOM INT / SUBJECT=id S;
ODS OUTPUT SOLUTIONR=ebest;
COVTEST 'test of random intercept var' GLM;
RUN;

/* print out the estimated random effects */
PROC PRINT DATA=ebest;
RUN;
```

| Obs	Effect	Subject	Estimate	StdErr Pred	DF	tValue	Probt
1	Intercept	ID 1	-0.00585	0.3536	90	-0.02	0.9868
2	Intercept	ID 2	0.3379	0.3712	90	0.91	0.3651
3	Intercept	ID 3	0.3711	0.3179	90	1.17	0.2462
4	Intercept	ID 4	-1.1056	0.5204	90	-2.12	0.0364
5	Intercept	ID 5	-0.2411	0.3800	90	-0.63	0.5274
6	Intercept	ID 6	-0.1205	0.3657	90	-0.33	0.7426
7	Intercept	ID 7	0.3717	0.3180	90	1.17	0.2456
8	Intercept	ID 8	-0.2910	0.4015	90	-0.72	0.4704
9	Intercept	ID 9	-0.03007	0.3909	90	-0.08	0.9389
10	Intercept	ID 10	-1.1091	0.5202	90	-2.13	0.0357
11	Intercept	ID 11	-0.1490	0.3843	90	-0.39	0.6992
12	Intercept	ID 12	-0.2405	0.3801	90	-0.63	0.5285
13	Intercept	ID 13	1.4836	0.2383	90	6.23	<.0001
14	Intercept	ID 14	-0.00525	0.3537	90	-0.01	0.9882
15	Intercept	ID 15	-0.00012	0.3539	90	-0.00	0.9997
16	Intercept	ID 16	0.07582	0.3977	90	0.19	0.8492
17	Intercept	ID 17	0.7864	0.4676	90	1.68	0.0961
18	Intercept	ID 18	-0.5365	0.4172	90	-1.29	0.2018
19	Intercept	ID 19	-0.5321	0.4174	90	-1.27	0.2056
20	Intercept	ID 20	0.7152	0.3772	90	1.90	0.0611
21	Intercept	ID 21	0.6581	0.2939	90	2.24	0.0276
22	Intercept	ID 22	-0.3839	0.3967	90	-0.97	0.3358
23	Intercept	ID 23	0.7846	0.2845	90	2.76	0.0071
24	Intercept	ID 24	-0.3834	0.3968	90	-0.97	0.3366
25	Intercept	ID 25	0.6639	0.2939	90	2.26	0.0263
26	Intercept	ID 26	-0.3855	0.5172	90	-0.75	0.4580
27	Intercept	ID 27	0.4534	0.3205	90	1.41	0.1607

SAS code: `aspart.sas` (continued)

```
DATA ebestci; SET ebest;
lo = Estimate-PROBIT(.975)*StdErrPred;
hi = Estimate+PROBIT(.975)*StdErrPred;
RUN;

/* standard error plot with estimated random effects */
PROC SORT; BY Estimate;
PROC SGPLOT DATA=ebestci;
SCATTER X=Subject Y=Estimate / YERRORLOWER=lo YERRORUPPER=hi;
RUN;
```
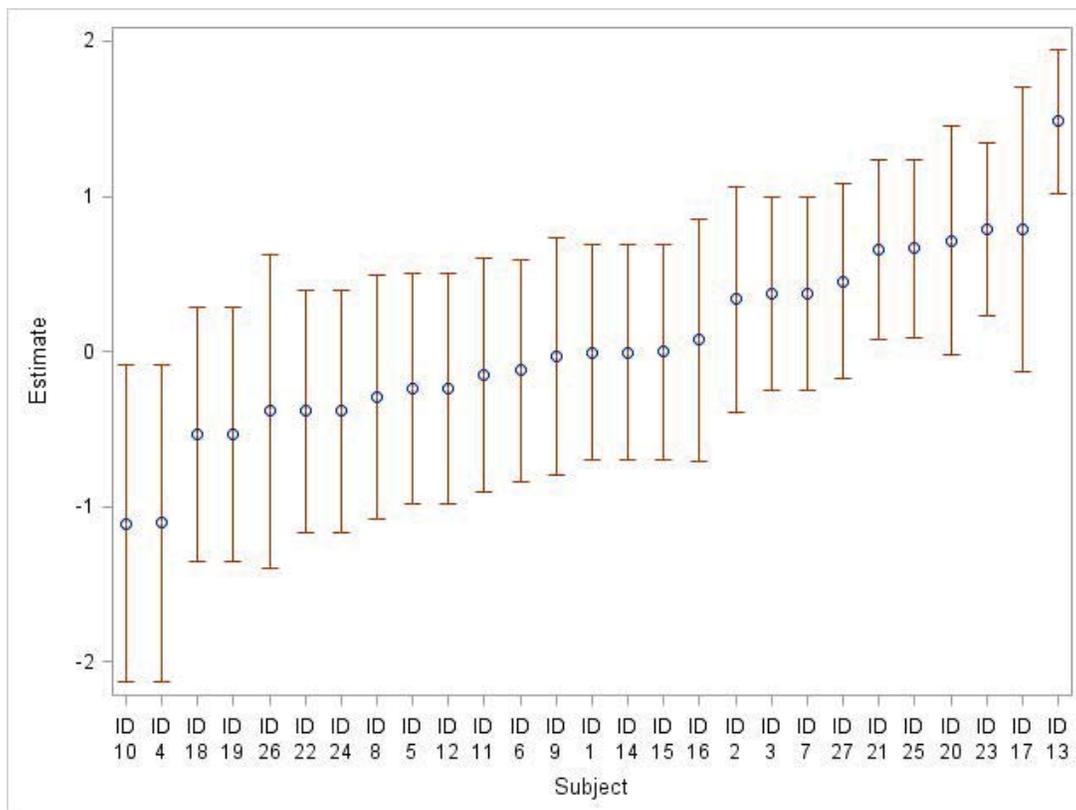
Large empirical Bayes estimates of random effects

- ID 13 has large positive value (1.470) and has observed data: 7, 7, 7, 6, 7 headaches

  note, also that this subject has the smallest standard error (0.175) due to number and consistency of responses

  95% C.I. $= 1.470 \pm 1.96 \times 0.175 = (1.127, 1.813)$

- IDs 4 and 10 have large negative values ($<$ -1) and have observed headaches of 0 for all periods

Random drug effect?

- In many studies, drug or treatment is a subject-level variable and doesn't vary across time

- In a crossover study, however, drug is a time-varying variable and DOES vary across time

- A time-varying variable can be considered as random at the subject level

  - Does the drug effect vary across subjects?
  - Is there subject heterogeneity in the number of headaches for aspartame relative to placebo?

Model in multilevel representation

$i = 1, \ldots, 27$ subjects $\qquad j = 1, \ldots, n_i$ periods (max = 5)

<u>Level-1 model</u> (within-subjects)

$$\log(\mu_{ij}) \; = \; \log(t_{ij}) + b_{0i} + b_{1i}P1_j + b_{2i}P2_j + b_{3i}P3_j + b_{4i}P4_j + b_{5i}Drug_{ij}$$

<u>Level-2 model</u> (between-subjects)

$$
\begin{aligned}
b_{0i} &= \beta_0 + \upsilon_{0i} \\
b_{1i} &= \beta_1 \\
b_{2i} &= \beta_2 \\
b_{3i} &= \beta_3 \\
b_{4i} &= \beta_4 \\
b_{5i} &= \beta_5 + \upsilon_{5i}
\end{aligned}
$$

Does the effect of aspartame on headaches vary across subjects?
(is υ_{5i} necessary?)

From `Aspart.do` :

```
* random int & Drug Poisson regression model with varying exposure
mepoisson Headache Period1 Period2 Period3 Period4 DrugAsp, exposure(NTDays) ///
 || ID:DrugAsp, covariance(unstructured)
scalar m2=e(ll)
```

```
Mixed-effects Poisson regression              Number of obs     =         122
Group variable:               ID              Number of groups  =          27

                                              Obs per group:
                                                            min =           2
                                                            avg =         4.5
                                                            max =           5

Integration method: mvaghermite               Integration pts.  =           7

                                              Wald chi2(5)      =        3.71
Log likelihood = -200.08337                   Prob > chi2       =      0.5920
--------------------------------------------------------------------------------
      Headache |     Coef.   Std. Err.      z    P>|z|    [95% Conf. Interval]
---------------+----------------------------------------------------------------
       Period1 |   .1235719   .2479087     0.50   0.618   -.3623203     .609464
       Period2 |   .0567582   .2318316     0.24   0.807   -.3976233    .5111397
       Period3 |  -.2126376   .2655232    -0.80   0.423   -.7330535    .3077782
       Period4 |  -.0662054   .2621424    -0.25   0.801   -.5799951    .4475843
```

```
        DrugAsp |    .2468226    .2363062      1.04   0.296     -.216329    .7099742
          _cons |   -1.742829    .2224108     -7.84   0.000    -2.178746   -1.306912
      ln(NTDays) |           1   (exposure)
-----------------+----------------------------------------------------------------
ID               |
      var(DrugAsp)|    .3938288    .2880716                        .093904    1.651698
       var(_cons)|    .5281057    .2379992                         .21833    1.277404
-----------------+----------------------------------------------------------------
ID               |
cov(_cons,DrugAsp)|   -.144784    .2073222     -0.70   0.485     -.551128      .26156
----------------------------------------------------------------------------------
LR test vs. Poisson model: chi2(3) = 62.62              Prob > chi2 = 0.0000

Note: LR test is conservative and provided only for reference.
```

Likelihood ratio test: comparing random DrugAsp to random intercept model

$$H_0 : \sigma^2_{v_5} = \sigma_{v_{05}} = 0$$

need `chibar2(12)`, 50:50 mixture of a χ^2_1 and a χ^2_2 distribution; p-value is obtained from the average of χ^2_1 and χ^2_2 (i.e., q and $q-1$, where q is the number of (co)variance parameters in the null)

From `Aspart.do` :

```
. * get LR test for comparing models
. display "chibar2(12) = " 2*(m2-m1)
chibar2(12) = 4.7010964

. display "Prob > chibar2(12) = " .5*chi2tail(1, 2*(m2-m1)) + ///
>   .5*chi2tail(2, 2*(m2-m1))
Prob > chibar2(12) = .06273014
```

Observed means: Headaches/Day across time by drug

drug	baseline	period 1	period 2	period 3	period 4
placebo	.228	.238	.276	.147	.230
	(n=27)	(n=12)	(n=14)	(n=13)	(n=9)
aspartame		.381	.266	.286	.272
		(n=15)	(n=11)	(n=9)	(n=12)
	rate ratio	1.60	0.96	1.95	1.18

Estimated means $= \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}} + \hat{\sigma}^2_{v_0}/2)$ *for placebo*
$= \exp(\boldsymbol{x}'\hat{\boldsymbol{\beta}} + 1/2(\hat{\sigma}^2_{v_0} + \hat{\sigma}^2_{v_5} + 2\hat{\sigma}_{v_{05}}))$ *for aspartame*

drug	baseline	period 1	period 2	period 3	period 4
placebo	.228	.258	.242	.184	.217
aspartame		.348	.326	.249	.288
	rate ratio	1.35	1.35	1.35	1.35

From `aspart.sas`

```
/* random intercept & DrugAsp Poisson model with offset */
PROC GLIMMIX METHOD=QUAD DATA=one;
CLASS id;
logNTDays = log(NTDays);
MODEL Headache = Period1 Period2 Period3 Period4 DrugAsp
     / LINK=LOG S DIST=POISSON OFFSET=logNTDays;
RANDOM INT DrugAsp / SUBJECT=id TYPE=UN G GCORR S;
ODS OUTPUT SOLUTIONR=ebest2;
COVTEST 'test of random effects' GLM;
COVTEST 'test of random slope' . 0 0;
RUN;
```

- `COVTEST 'test of random effects' GLM;`
  will compare this model to model without random effects using LR test:
  $H_0 : \sigma^2_{v_0} = \sigma^2_{v_5} = \sigma_{v_{05}} = 0$

- `COVTEST 'test of random slope' .  0 0;`
  will compare this model to model without random slope and intercept
  slope covariance using LR test: $H_0 : \sigma^2_{v_5} = \sigma_{v_{05}} = 0$

- `SOLUTION` or `S` on `RANDOM` statement will list the estimates of the random
  effects; `ODS` statement saves these to dataset `ebest2`

The GLIMMIX Procedure

Model Information

                 Data Set   WORK.ONE
                 Response Variable   Headache
                 Response Distribution   Poisson
                 Link Function   Log
                 Variance Function   Default
                 Offset Variable   logNTDays = log(NTDays);
                 Variance Matrix Blocked By   ID
                 Estimation Technique   Maximum Likelihood
                 Likelihood Approximation   Gauss-Hermite Quadrature
                 Degrees of Freedom Method   Containment

Class Level Information

      Class   Levels   Values

      ID   27   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
                         19 20 21 22 23 24 25 26 27

                 Number of Observations Read   122
                 Number of Observations Used   122

Dimensions

                 G-side Cov. Parameters   3
                 Columns in X   6
                 Columns in Z per Subject   2
                 Subjects (Blocks in V)   27
                 Max Obs per Subject   5

Fit Statistics

                 -2 Log Likelihood   400.17
                 AIC (smaller is better)   418.17
                 AICC (smaller is better)   419.78
                 BIC (smaller is better)   429.83
                 CAIC (smaller is better)   438.83
                 HQIC (smaller is better)   421.64

Estimated G Matrix

            Effect   Row   Col1   Col2

            Intercept   1   0.5271   -0.1446
            DrugAsp   2   -0.1446   0.3934

```
                    Estimated G Correlation Matrix

                 Effect      Row       Col1        Col2

                 Intercept     1      1.0000     -0.3175
                 DrugAsp       2     -0.3175      1.0000


                    Covariance Parameter Estimates

            Cov                                  Standard
            Parm       Subject     Estimate       Error

            UN(1,1)    ID            0.5271       0.2373
            UN(2,1)    ID           -0.1446       0.2075
            UN(2,2)    ID            0.3934       0.2876


                      Solutions for Fixed Effects

                              Standard
           Effect      Estimate    Error      DF    t Value    Pr > |t|

           Intercept   -1.7428    0.2224      26     -7.83     <.0001
           Period1      0.1235    0.2479      65      0.50     0.6200
           Period2      0.05675   0.2318      65      0.24     0.8074
           Period3     -0.2127    0.2655      65     -0.80     0.4261
           Period4     -0.06624   0.2621      65     -0.25     0.8013
           DrugAsp      0.2468    0.2364      25      1.04     0.3063
```

```
                    Tests of Covariance Parameters
                       Based on the Likelihood

        Label                    DF    -2 Log Like    ChiSq    Pr > ChiSq    Note

        test of random effects    3       462.78      62.61      <.0001       --
        test of random slope      2       404.87       4.70      0.0627       MI
```

MI: P-value based on a mixture of chi-squares.
--: Standard test with unadjusted p-values.

- **test of random effects** compares random trend model to ordinary Poisson regression using ordinary LR test (too conservative)

- **test of random slope** compares random trend to random intercept model

$H_0 : \sigma^2_{v_5} = \sigma_{v_{05}} = 0$

uses `chibar2(12)`, 50:50 mixture of a χ^2_1 and a χ^2_2 distribution; p-value is obtained from the average of χ^2_1 and χ^2_2 (i.e., q and $q-1$, where q is the number of (co)variance parameters in the null)

Graph of Random Drug Effects

From `Aspart.do` :

```
 * get random effect estimates (and std errs)
predict u2 u1, reffects reses(u2se u1se)

* rank the random DrugAsp effects
egen u2rank = rank(u2) if pick1obs==1
list ID u2 u2se u2rank if pick1obs==1

* standard error bar chart of random effects
serrbar u2 u2se u2rank if pick1obs==1, scale(1.96) yline(0)
```
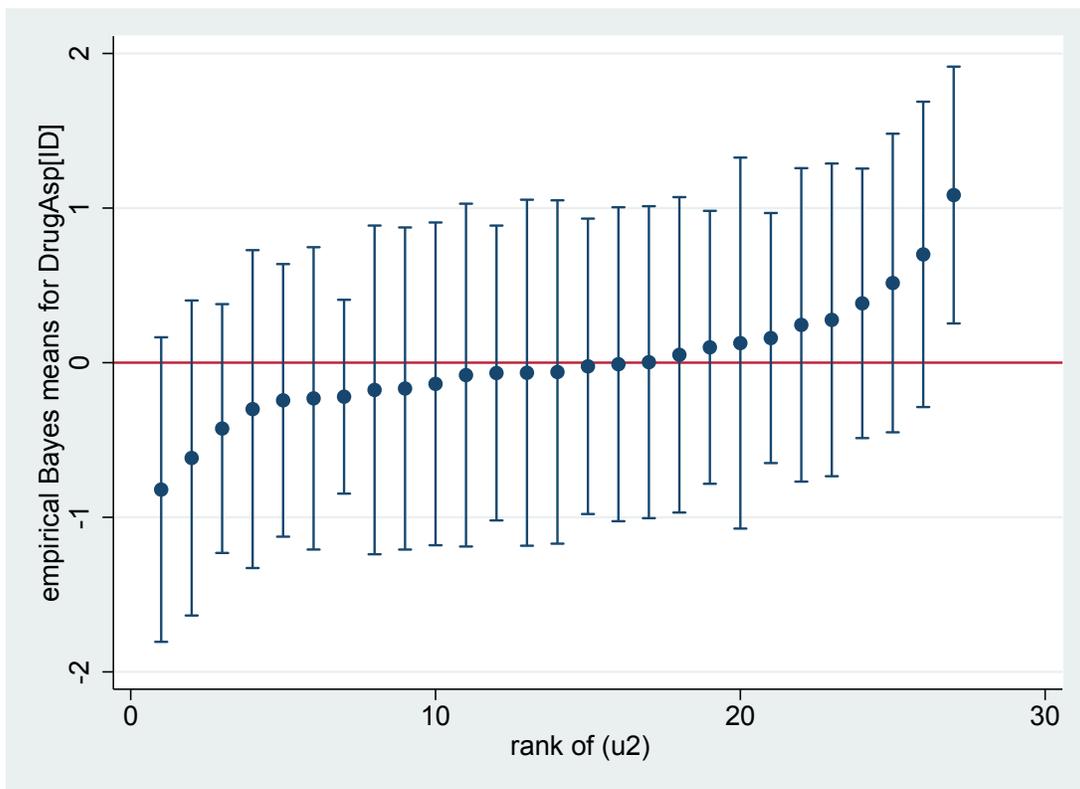
From `aspart.sas`

```
/* print out the estimated random effects */
PROC PRINT DATA=ebest2;
RUN;

/* select only the random slopes */
DATA ebest2ci; SET ebest2; if Effect = "DrugAsp";
lo = Estimate-PROBIT(.975)*StdErrPred;
hi = Estimate+PROBIT(.975)*StdErrPred;
RUN;

/* standard error plot with estimated random effects */
PROC SORT; BY Estimate;
PROC SGPLOT DATA=ebest2ci;
SCATTER X=Subject Y=Estimate / YERRORLOWER=lo YERRORUPPER=hi;
REFLINE 0 /AXIS=Y;
RUN;
```
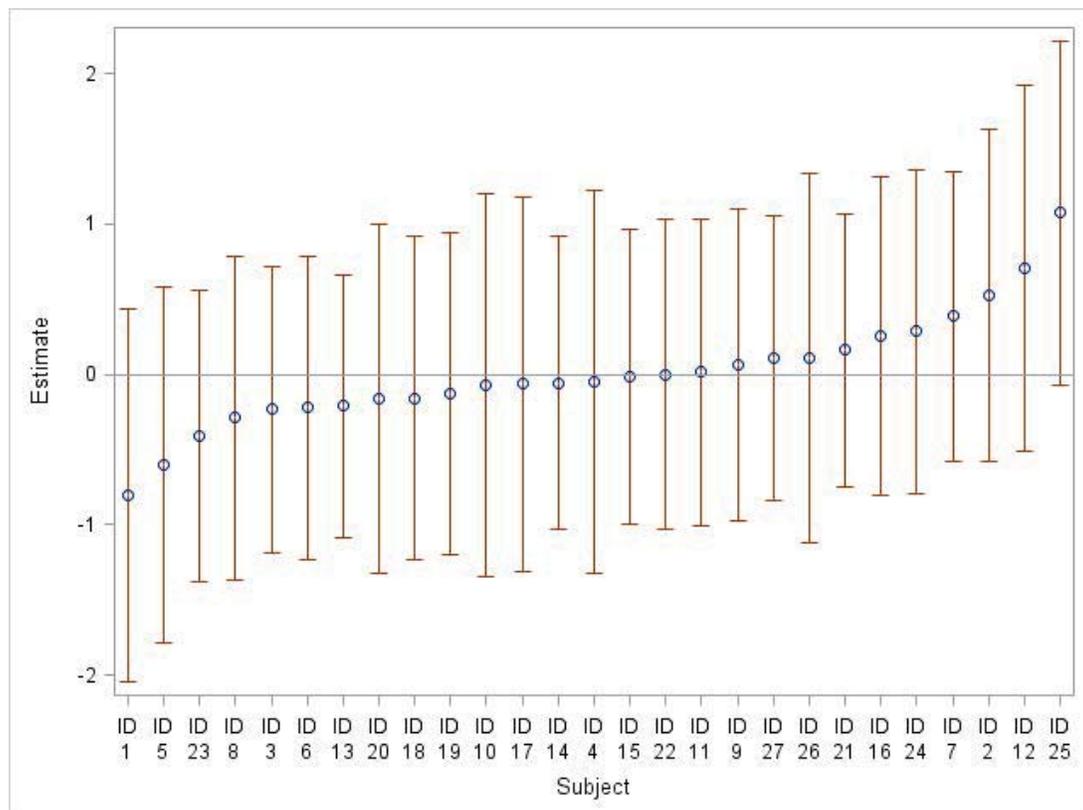
Very interesting subject

Empirical Bayes estimate of drug effect is very large (≈ 1) for
ID=25

observed number of headaches:
1 (placebo), 6 (drug), 1 (placebo), 7 (drug), 0 (placebo)

Drug Effect Estimates

	model	estimate	std error	p-value
with ID=25	rand int	0.2797	0.1641	0.088
	rand drug	0.2468	0.2363	0.296
without ID=25	rand int	0.1384	0.1698	0.415
	rand drug	0.1462	0.1955	0.455

also, from random drug models, estimate of drug variance goes
from 0.3938 (se = 0.288) to 0.0030 (se = 0.033) when this
subject is removed

Overdispersion

- Poisson model assumes that the mean equals the variance

- overdispersion occurs when the variance exceeds the mean

  - often present in real data, can change model estimates
  - inclusion of random effects, by accounting for individual
    differences, may decrease possibility of overdispersion

- Negative Binomial model relaxes this assumption by including
  an overdispersion parameter

  - Poisson model is a special case of Negative Binomial when
    this overdispersion parameter equals 0

- in `Stata` use `menbreg`

- in `SAS` use `MODEL` option `DIST=NEGBINOMIAL`

From `Aspart.do` :

```
* random intercept Negative Binomial model with varying exposure
menbreg Headache Period1 Period2 Period3 Period4 DrugAsp, exposure(NTDays) ///
 || ID:
scalar m3=e(ll)

* get LR test for comparing models
display "chibar2(01) = " 2*(m3-m1)
display "Prob > chibar2(01) = " .5*chi2tail(1, 2*(m3-m1))
```

```
. menbreg Headache Period1 Period2 Period3 Period4 DrugAsp, exposure(NTDays) ///
>  || ID:


Mixed-effects nbinomial regression            Number of obs      =         122
Overdispersion:              mean
Group variable:                ID             Number of groups   =          27

                                              Obs per group:
                                                            min =           2
                                                            avg =         4.5
                                                            max =           5

Integration method: mvaghermite               Integration pts.   =           7

                                              Wald chi2(5)       =        5.76
Log likelihood =  -202.1028                   Prob > chi2        =      0.3298
------------------------------------------------------------------------------
    Headache |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     Period1 |   .1117396   .2539913     0.44   0.660    -.3860743    .6095534
     Period2 |   .0869462   .2445365     0.36   0.722    -.3923365    .5662288
```

```
    Period3 |  -.2260488    .2760435    -0.82   0.413    -.767084     .3149865
    Period4 |  -.0934193    .2769078    -0.34   0.736    -.6361486      .44931
    DrugAsp |   .2860314    .1773072     1.61   0.107    -.0614843     .6335472
      _cons |  -1.706765    .2180453    -7.83   0.000    -2.134126    -1.279404
  ln(NTDays) |          1   (exposure)
------------+----------------------------------------------------------------
   /lnalpha |  -2.601152    1.389024                     -5.323588     .1212846
------------+----------------------------------------------------------------
ID          |
  var(_cons)|   .4614995     .193958                      .2025023      1.05175
-----------------------------------------------------------------------------
LR test vs. nbinomial model: chibar2(01) = 23.17     Prob >= chibar2 = 0.0000

. scalar m3=e(ll)

. * get LR test for comparing models

. display "chibar2(01) = " 2*(m3-m1)
chibar2(01) = .66222995

. display "Prob > chibar2(01) = " .5*chi2tail(1, 2*(m3-m1))
Prob > chibar2(01) = .20788689
```

From `aspart.sas`

```
/* random intercept Negative Binomial model with offset */
PROC GLIMMIX METHOD=QUAD;
CLASS id;
logNTDays = log(NTDays);
MODEL Headache = Period1 Period2 Period3 Period4 DrugAsp
    / LINK=LOG S DIST=NEGBINOMIAL OFFSET=logNTDays;
RANDOM INT / SUBJECT=id;
COVTEST 'test of random intercept var' GLM;
RUN;
```

The GLIMMIX Procedure

Model Information

            Data Set   WORK.ONE
            Response Variable   Headache
            Response Distribution   Negative Binomial
            Link Function   Log
            Variance Function   Default
            Offset Variable   logNTDays = log(NTDays);
            Variance Matrix Blocked By   ID
            Estimation Technique   Maximum Likelihood
            Likelihood Approximation   Gauss-Hermite Quadrature
            Degrees of Freedom Method   Containment

Class Level Information

    Class   Levels   Values

    ID   27   1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18
                       19 20 21 22 23 24 25 26 27

            Number of Observations Read   122
            Number of Observations Used   122

Dimensions

            G-side Cov. Parameters   1
            R-side Cov. Parameters   1
            Columns in X   6
            Columns in Z per Subject   1
            Subjects (Blocks in V)   27
            Max Obs per Subject   5

Fit Statistics

            -2 Log Likelihood   404.37
            AIC (smaller is better)   420.37
            AICC (smaller is better)   421.65
            BIC (smaller is better)   430.74
            CAIC (smaller is better)   438.74
            HQIC (smaller is better)   423.45

Covariance Parameter Estimates

                                        Standard
            Cov Parm   Subject   Estimate   Error

            Intercept   ID   0.4541   0.1904
            Scale   0.07969   0.1036

```
                    Solutions for Fixed Effects

                               Standard
            Effect      Estimate      Error      DF     t Value    Pr > |t|

            Intercept    -1.7047      0.2180      26      -7.82     <.0001
            Period1       0.1123      0.2552      90       0.44     0.6610
            Period2       0.08660     0.2458      90       0.35     0.7255
            Period3      -0.2273      0.2773      90      -0.82     0.4145
            Period4      -0.09455     0.2783      90      -0.34     0.7349
            DrugAsp       0.2863      0.1782      90       1.61     0.1116


                    Tests of Covariance Parameters
                       Based on the Likelihood

      Label                       DF     -2 Log Like     ChiSq    Pr > ChiSq    Note

      test of random intercept var  1        427.37      23.00       <.0001      MI

   MI: P-value based on a mixture of chi-squares.
```

Conclusions

- Van Den Eeden *et al.*, (1994), Aspartame ingestion and headaches: A randomized crossover trial, *Neurology, 44*, 1787-1793.

  "the proportion of days subjects reported having a headache was higher during aspartame treatment compared with placebo treatment (aspartame = .33, placebo = .24, $p = .04$)"

- Levy, Hedeker, & Sanders (1995) To the editor: Aspartame and headache, *Neurology, 45(8)*:1631-2; author reply 1632-3.

  Increase of headaches by aspartame only for 1-tailed test in random intercept model; random drug model and model without subject 25 (very influential subject) shows no drug effect.

- Butchkoa & Stargelb (2001), Aspartame: Scientific evaluation in the postmarketing period, *Regulatory Toxicology and Pharmacology, 34*, 221-233.

  "Evaluation of the anecdotal reports of adverse health effects, the first such system for a food additive, revealed that the reported effects were generally mild and also common in the general population and that there was no consistent or unique pattern of symptoms that could be causally linked to consumption of aspartame. Finally, the results of the extensive scientific research done to evaluate these allegations did not show a causal relationship between aspartame and adverse effects. Thus, the weight of scientific evidence confirms that, even in amounts many times what people typically consume, aspartame is safe for its intended uses as a sweetener and flavor enhancer."

Summary

- Poisson model useful for modeling counts; $\exp\beta$ = rate ratio
  - Poisson model assumes mean = variance
- Poisson overdispersion can be handled by
  - random effects
  - inclusion of overdispersion parameter (Negative Binomial regression)
  - random effects and overdispersion (mixed Negative Binomial regression)
- Zero-inflated models (ZIP, ZINB) are available in Stata and SAS without random effects; perhaps with random effects in the future (can use SAS PROC NLMIXED, but involves some programming)