

WSSR Workshop: Generalized Linear Regression Models for Social Scientists

Count Regression Models

JEFF GILL

Distinguished Professor

Departments of Government and Mathematics & Statistics

American University

The Poisson PMF

- ▶ The most basic and important probability mass function (PMF) for modeling counts of non-negative integer measured events is the Poisson.

- ▶ PMF

$$f(Y|\mu) = \frac{(\mu)^Y e^{-\mu}}{Y!}, \quad y = 0, 1, 2, \dots, \mu > 0$$

where μ is the intensity parameter.

- ▶ This is the probability that exactly Y arrivals occur in a given interval.

Poisson Technical Assumptions

- ▶ **Infinitesimal Interval.** The probability of an arrival in the interval: $(t : \delta t)$ equals $\mu\delta t + o(\delta t)$ where μ is the intensity parameter discussed above and $o(\delta t)$ is a time interval with the property: $\lim_{\delta t \rightarrow 0} \frac{o(\delta t)}{\delta t} = 0$. In other words, as the interval δt reduces in size towards zero, $o(\delta t)$ is negligible compared to δt . This assumption is required to establish that μ adequately describes the intensity or expectation of arrivals. Typically there is no problem meeting this assumption provided that the time measure is adequately granular with respect to arrival rates.
- ▶ **Non-Simultaneity of Events.** The probability of more than one arrival in the interval: $(t : \delta t)$ equals $o(\delta t)$. Since $o(\delta t)$ is negligible with respect to $\mu\delta t$ for sufficiently small $\mu\delta t$, the probability of simultaneous arrivals approaches zero in the limit.
- ▶ **I.I.D. Arrivals.** The number of arrivals in any two consecutive or non-consecutive intervals are independent and identically distributed. More specifically, $P(Y = y) \in (T_j : T_{j+1})$ does not depend on $P(Y = y) \in (T_k : T_{k+1})$ for any $j \neq k$.

Poisson Features

- ▶ The intensity parameter (μ) is both the mean and variance for a single Poisson distributed random variable.
- ▶ The intensity parameter is tied to a time interval, and rescaling time rescales the intensity parameter.
- ▶ Sums of independent Poisson random variables are themselves Poisson.
- ▶ We can also specifically model time by including it in the intensity parameter: $\mu^* = \mu t$.

Simple Example

- ▶ The data provide an index of social setting, an index of family planning effort, and the percent decline in the crude birth rate (CBR) 1965 to 1975 for 20 Latin American countries.
- ▶ See P.W. Mauldin and B. Berelson (1978). “Conditions of Fertility Decline in Developing countries, 1965-75.” *Studies in Family Planning*, 9:89-147.
- ▶ Load and look at the data:

```
effort <- read.table("https://jeffgill.org/files/jeffgill/files/program.effort.dat_.txt",
  header=TRUE)
effort
```

	setting	effort	change		setting	effort	change
Bolivia	46	0	1	Brazil	74	0	10
Chile	89	16	29	Colombia	77	16	25
CostaRica	84	21	29	Cuba	89	15	40
DominicanRep	68	14	21	Ecuador	70	6	0
ElSalvador	60	13	13	Guatemala	55	9	4
Haiti	35	3	0	Honduras	51	7	7
Jamaica	87	23	21	Mexico	83	4	9
Nicaragua	68	0	7	Panama	84	19	22
Paraguay	74	3	6	Peru	73	0	2
TrinidadTobago	84	15	29	Venezuela	91	7	11

Simple Example

- ▶ Run a Poisson GLM for percent decline in the CBR:

```
effort.out <- glm(change ~ setting + effort,data=effort,family=poisson)
summary(effort.out)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-4.0174	-1.2344	-0.4353	1.2384	2.7580

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.412505	0.459960	-0.897	0.37
setting	0.030470	0.006332	4.812	1.49e-06
effort	0.061287	0.009906	6.187	6.13e-10

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 206.93 on 19 degrees of freedom

Residual deviance: 61.96 on 17 degrees of freedom

AIC: 144.93

Number of Fisher Scoring iterations: 5

Relationships to Other Forms

- ▶ Poisson assumption is that there is no upper limit; if there is one use a binomial PMF.
- ▶ If $\mu = np$ as $n \rightarrow \infty$, then the Poisson is a good approximation for the binomial.
- ▶ If n is small, then $\text{logit}(p) \approx \log(p)$, so the logit model is close to the Poisson model.
- ▶ If counts are bins, then use the multinomial PMF (discussed later).

Derivation of the MLE

► PMF:

$$p(Y = y|\mu) = \frac{e^{-\mu} \mu^y}{y!}$$

► Likelihood function:

$$L(\mu|\mathbf{y}) = \prod_{i=1}^n \frac{e^{-\mu} \mu^{y_i}}{y_i!}$$

► Log-likelihood function:

$$\ell(\mu|\mathbf{y}) = -n\mu + \log(\mu) \sum_{i=1}^n y_i - \sum_{i=1}^n \log(y_i!)$$

► MLE:

$$\frac{d}{d\mu} \ell(\mu|\mathbf{y}) = -n + \frac{1}{\mu} \sum_{i=1}^n y_i \equiv 0 \Rightarrow n\mu = \sum_{i=1}^n y_i \Rightarrow \hat{\mu} = \bar{y}$$

Graphical View of the MLE

```
y.vals<-c(1,3,1,5,2,6,8,11,0,0)

# POISSON LIKELIHOOD AND LOG-LIKELIHOOD FUNCTION
llhfunc <- function(X,p,do.log=TRUE) {
  d <- rep(X,length(p))
  print(d) # PRINT THE DATA THE NUMBER OF TIMES TO BE RUN
  u.vec <- rep(p,each=length(X))
  print(u.vec) # PRINT THE TESTED PARAMETER THE NUMBER OF TIMES TO BE USED
  d.mat <- matrix(dpois(d,u.vec,log=do.log),ncol=length(p))
  print(d.mat) # PRINT THE INDIVIDUAL LIKELIHOOD CONTRIBUTIONS
  if (do.log==TRUE) apply(d.mat,2,sum)
  else apply(d.mat,2,prod)
}
```

Test the Function

```
llhfunc(y.vals,c(4,30))
```

```
[1]  1  3  1  5  2  6  8 11  0  0  1  3  1  5  2  6  8 11  0  0
```

```
[1]  4  4  4  4  4  4  4  4  4  4 30 30 30 30 30 30 30 30 30 30
```

```
      [,1]      [,2]
```

```
[1,] -2.6137 -26.599
```

```
[2,] -1.6329 -21.588
```

```
[3,] -2.6137 -26.599
```

```
[4,] -1.8560 -17.782
```

```
[5,] -1.9206 -23.891
```

```
[6,] -2.2615 -16.172
```

```
[7,] -3.5142 -13.395
```

```
[8,] -6.2531 -10.089
```

```
[9,] -4.0000 -30.000
```

```
[10,] -4.0000 -30.000
```

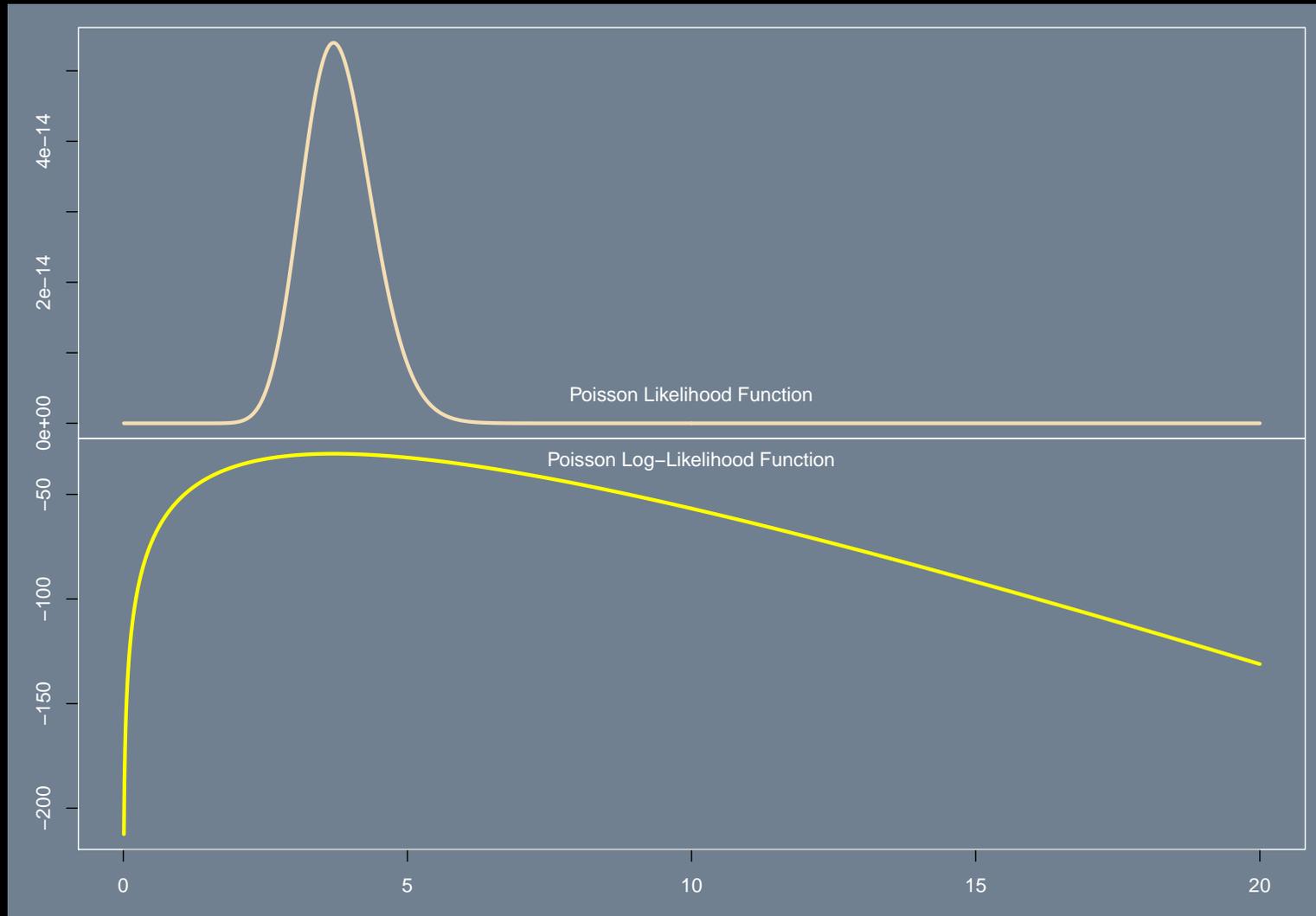
```
[1] -30.666 -216.114
```

Graphical View of the MLE

```
# THIS IS A VERSION OF THE mle CALL FROM, fnscale=-1 MAKES IT A MAXIMIZATION
mle <- optim(par=1,fn=llhfunc,X=y.vals,control=list(fnscale=-1),method="BFGS")

# MAKE A PRETTY GRAPH OF THE LOG AND NON-LOG VERSIONS
ruler <- seq(from=.01, to=20, by= .01)
poison.ll <- llhfunc(y.vals,ruler)
poison.l <- llhfunc(y.vals,ruler,do.log=FALSE)

postscript("poisson.like.ps")
par(oma=c(3,3,1,1),mar=c(0,0,0,0),mfrow=c(2,1),col.axis="white",
     col.lab="white",col.sub="white",col="white", bg="slategray")
plot(ruler,poison.l,col="wheat",type="l",xaxt="n",lwd=3)
text(mean(ruler),mean(poison.l),"Poisson Likelihood Function")
plot(ruler,poison.ll,col="yellow",type="l",lwd=3)
text(mean(ruler),mean(poison.ll)/2,"Poisson Log-Likelihood Function")
dev.off()
```



Derivation of the Variance

- ▶ Second derivative of the LL:

$$\frac{d^2}{d\mu^2}\ell(\mu|\mathbf{y}) = \frac{d}{d\mu} \left(-n + \frac{1}{\mu} \sum_{i=1}^n y_i \right) = -\mu^{-2} \sum_{i=1}^n y_i,$$

called the Hessian.

- ▶ Fisher Information:

$$FI = -E_{\mu} \left[\frac{d^2}{d\mu^2}\ell(\mu|\mathbf{y}) \right] = -E_{\mu} \left[-\mu^{-2} \sum_{i=1}^n y_i \right] = n\bar{y}E_{\mu} [\mu^{-2}] = \frac{n}{\bar{y}}$$

since $E\mu = \bar{y}$.

- ▶ Variance:

$$\text{Var}[\mu] = (FI)^{-1} = \bar{y}/n.$$

Link Function for Poisson Regression with Covariates

► Definition:

$$\log(\mu_i) = \eta_i \Rightarrow \mu_i = \exp(\eta_i) = \exp(\mathbf{X}_i\boldsymbol{\beta})$$

► Start with the substitution:

$$L(\boldsymbol{\beta}|\mathbf{y}) = \prod_{i=1}^n \frac{e^{-\mu} \mu^{y_i}}{y_i!} \Big|_{\mu_i=\exp(\mathbf{X}_i\boldsymbol{\beta})} = \prod_{i=1}^n e^{-\exp(\mathbf{X}_i\boldsymbol{\beta})} \exp(\mathbf{X}_i\boldsymbol{\beta})^{y_i} / y_i!$$

► Take the log:

$$\ell(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^n [-\exp(\mathbf{X}_i\boldsymbol{\beta}) + y_i(\mathbf{X}_i\boldsymbol{\beta}) - \log(y_i!)]$$

► Now take the first derivative:

$$\frac{d}{d\boldsymbol{\beta}} \ell(\boldsymbol{\beta}|\mathbf{y}) = \sum_{i=1}^n [\exp(\mathbf{X}_i\boldsymbol{\beta})\mathbf{X}_j + \mathbf{y}_i\mathbf{X}_j], \quad \forall j$$

► Or in full matrix terms: $\mathbf{X}'\mathbf{y} = \mathbf{X}'\hat{\boldsymbol{\mu}}$, where $\hat{\boldsymbol{\mu}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ (the normal equation for the Poisson model).

► Problem: there does not exist a closed form solution for $\hat{\boldsymbol{\beta}}$, so we use numerical methods.

Application: Poisson Model of Military Coups.

- ▶ Sub-Saharan Africa has experienced a disproportionately high proportion of regime changes due to the military takeover of government for a variety of reasons, including ethnic fragmentation, arbitrary borders, economic problems, outside intervention, and poorly developed governmental institutions.
- ▶ These data, selected from a larger set given by Bratton and Van De Walle (1994), look at potential causal factors for counts of military coups (ranging from 0 to 6 events) in 33 sub-Saharan countries over the period from each country's colonial independence to 1989.
- ▶ Seven explanatory variables are chosen here to model the count of military coups: **Military Oligarchy** (the number of years of this type of rule); **Political Liberalization** (0 for no observable civil rights for political expression, 1 for limited, and 2 for extensive); **Parties** (number of legally registered political parties); **Percent Legislative Voting**; **Percent Registered Voting**; **Size** (in one thousand square kilometer units); and **Population** (given in millions).
- ▶ Reading in the data:

```
africa <- read.table("https://jeffgill.org/files/jeffgill/files/africa.data_.txt",  
                    header=TRUE)
```

Application: Poisson Model of Military Coups.

- ▶ A generalized linear model for these data with the Poisson link function is specified as:

$$g^{-1}(\boldsymbol{\theta}) = g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \exp[\mathbf{X}\boldsymbol{\beta}] = \mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{Military Coups}].$$

- ▶ In this specification, the systematic component is $\mathbf{X}\boldsymbol{\beta}$, the stochastic component is $\mathbf{Y} = \mathbf{Military Coups}$, and the link function is $\boldsymbol{\theta} = \log(\boldsymbol{\mu})$.
- ▶ We can re-express this model by moving the link function to the left-hand side exposing the linear predictor: $g(\boldsymbol{\mu}) = \log(\mathbb{E}[\mathbf{Y}]) = \mathbf{X}\boldsymbol{\beta}$ (although this is now a less intuitive form for understanding the outcome variable).
- ▶ The R language GLM call for this model is:

```
africa.out <- glm(MILTCOUP ~ MILITARY+POLLIB+PARTY93+PCTVOTE+PCTTURN
                  +SIZE*POP+NUMREGIM*NUMELEC, family=poisson, data=africa).
```

- ▶ Recall that we use `family=poisson` where poisson is not capitalized.

Application: Poisson Model of Military Coups.

	Parameter Estimate	Standard Error	95% Confidence Interval
(Intercept)	2.9209	1.3368	[0.3008: 5.5410]
Military Oligarchy	0.1709	0.0509	[0.0711: 0.2706]
Political Liberalization	-0.4654	0.3319	[-1.1160: 0.1851]
Parties	0.0248	0.0109	[0.0035: 0.0460]
Percent Legislative Voting	0.0613	0.0218	[0.0187: 0.1040]
Percent Registered Voting	-0.0361	0.0137	[-0.0629:-0.0093]
Size	-0.0018	0.0007	[-0.0033:-0.0004]
Population	-0.1188	0.0397	[-0.1965:-0.0411]
Regimes	-0.8662	0.4571	[-1.7621: 0.0298]
Elections	-0.4859	0.2118	[-0.9010:-0.0709]
(Size)(Population)	0.0001	0.0001	[0.0001: 0.0002]
(Regimes)(Elections)	0.1810	0.0689	[0.0459: 0.3161]

Application: Poisson Model of Military Coups.

- ▶ Note that the two interaction terms are specified by using the multiplication character. The iteratively weighted least squares algorithm converged in only four iterations using Fisher scoring, and the results are provided in the table.
- ▶ The model appears to fit the data quite well:
 - ▷ an improvement from the null deviance of 62 on 32 degrees of freedom to a residual deviance of 7.5 on 21 degrees of freedom
 - ▷ evidence that the model does not fit would be supplied by a model deviance value in the tail of a χ_{n-k}^2 distribution
 - ▷ and nearly all the coefficients have 95% confidence intervals bounded away from zero and therefore appear reliable in the model.

Back to Residuals and Model Fit

- ▶ General Deviance Notation: $D = \sum_{i=1}^n d(\boldsymbol{\eta}, y_i)$, where the individual deviance function is defined as: $d(\boldsymbol{\eta}, y_i) = -2 [\ell(\hat{\boldsymbol{\eta}}, \psi|y_i) - \ell(\tilde{\boldsymbol{\eta}}, \psi|y_i)]$, where $\hat{\boldsymbol{\eta}}$ is the model estimate and $\tilde{\boldsymbol{\eta}}$ is the saturated estimate.
- ▶ Linear Model Residual Vector: $\mathbf{R}_{standard} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$.
- ▶ Response Residual Vector: $\mathbf{R}_{Response} = \mathbf{Y} - g^{-1}(\mathbf{X}\boldsymbol{\beta}) = \mathbf{Y} - \hat{\boldsymbol{\mu}}$.
- ▶ Pearson Residual Vector: $\mathbf{R}_{Pearson} = \frac{\mathbf{Y} - \hat{\boldsymbol{\mu}}}{\sqrt{VAR[\boldsymbol{\mu}]}}$ (the sum of the Pearson residuals for a Poisson generalized linear model is the Pearson χ^2 goodness-of-fit measure).
- ▶ Working Residual Vector: $\mathbf{R}_{Working} = (\mathbf{y} - \boldsymbol{\mu}) \frac{\partial}{\partial \boldsymbol{\eta}} \boldsymbol{\mu}$ (from the last step of Iteratively Reweighted Least Squares algorithm).

Deviance Summary

Table 1: DEVIANCE FUNCTIONS

Distribution	Canonical Parameter	Deviance Function
Poisson($\hat{\mu}$)	$\eta = \log(\hat{\mu})$	$2 \sum \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) - y_i + \hat{\mu}_i \right]$
Binomial(m, p)	$\eta = \log \left(\frac{\hat{\mu}}{1 - \hat{\mu}} \right)$	$2 \sum \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (m_i - y_i) \log \left(\frac{m_i - y_i}{m_i - \hat{\mu}_i} \right) \right]$
Normal($\hat{\mu}, \sigma$)	$\eta = \hat{\mu}$	$\sum [y_i - \hat{\mu}_i]^2$
Gamma($\hat{\mu}, \delta$)	$\eta = -\frac{1}{\hat{\mu}}$	$2 \sum \left[-\log \left(\frac{y_i}{\hat{\mu}_i} \right) \frac{y_i - \hat{\mu}_i}{\hat{\mu}_i} \right]$
Negative Binom($\hat{\mu}, p$)	$\eta = \log(1 - \hat{\mu})$	$2 \sum \left[y_i \log \left(\frac{y_i}{\hat{\mu}_i} \right) + (1 + y_i) \log \left(\frac{1 + \hat{\mu}_i}{1 + y_i} \right) \right]$

Focus on the Deviance for the Poisson Model

- ▶ The “G-statistic” (summed deviance) for this model is:

$$D_{\text{Poisson}} = 2 \sum_{i=1}^n (y_i \log(y_i/\hat{\mu}_i) - (y_i - \hat{\mu}_i)) \underset{\text{a}}{\sim} \chi_{n-p}^2,$$

where p is the number of explanatory variables including the constant, and $\hat{\mu}_i$ is the predicted outcome for the i th case.

- ▶ Individual Deviance Function:

$$R_{\text{Deviance}} = \frac{(y_i - \hat{\mu}_i)}{|y_i - \hat{\mu}_i|} \sqrt{|d(\boldsymbol{\eta}, y_i)|} \quad \text{where:} \quad d(\boldsymbol{\eta}, y_i) = -2 [\ell(\hat{\boldsymbol{\eta}}, \psi|y_i) - \ell(\tilde{\boldsymbol{\eta}}, \psi|y_i)].$$

- ▶ Recall also the Pearson’s statistic, which performs the same duty:

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i} \underset{\text{a}}{\sim} \chi_{n-p}^2.$$

- ▶ Generally the summed deviance is more robust than the Pearson’s statistic though.

Returning to the Effort Model

- ▶ We previously ran a model for the crude birth rate decline 1965 to 1975 for 20 Latin American countries.

- ▶ The deviance summary was:

```
Null deviance: 206.93 on 19 degrees of freedom  
Residual deviance: 61.96 on 17 degrees of freedom
```

- ▶ Showing a substantial improvement in summed deviance, which can be tested:

```
pchisq(206.93-61.96, df=19-17, lower.tail=FALSE)  
[1] 3.312566e-32
```

- ▶ Also (but not importantly): `pchisq(61.96-0, df=17-1, lower.tail=FALSE)` [1] 2.43867e-07

Poisson GLM of Capital Punishment Data, 1997

State	Executions	Median Income	Percent Poverty	Percent Black	Violent Crime/100K	South	Proportion w/Degrees
Texas	37	34453	16.7	12.2	644	1	0.16
Virginia	9	41534	12.5	20.0	351	1	0.27
Missouri	6	35802	10.6	11.2	591	0	0.21
Arkansas	4	26954	18.4	16.1	524	1	0.16
Alabama	3	31468	14.8	25.9	565	1	0.19
Arizona	2	32552	18.8	3.5	632	0	0.25
Illinois	2	40873	11.6	15.3	886	0	0.25
South Carolina	2	34861	13.1	30.1	997	1	0.21
Colorado	1	42562	9.4	4.3	405	0	0.31
Florida	1	31900	14.3	15.4	1051	1	0.24
Indiana	1	37421	8.2	8.2	537	0	0.19
Kentucky	1	33305	16.4	7.2	321	0	0.16
Louisiana	1	32108	18.4	32.1	929	1	0.18
Maryland	1	45844	9.3	27.4	931	0	0.29
Nebraska	1	34743	10.0	4.0	435	0	0.24
Oklahoma	1	29709	15.2	7.7	597	0	0.21
Oregon	1	36777	11.7	1.8	463	0	0.25
	EXE	INC	POV	BLK	CRI	SOU	DEG

Source: United States Census Bureau, United States Department of Justice.

Poisson GLM of Capital Punishment Data

The model is developed from the Poisson link function, $\boldsymbol{\eta} = \log(\boldsymbol{\mu})$, with the objective of finding the best $\boldsymbol{\beta}$ vector in:

$$\begin{aligned}
 \underbrace{g^{-1}(\boldsymbol{\eta})}_{17 \times 1} &= g^{-1}(\mathbf{X}\boldsymbol{\beta}) \\
 &= \exp[\mathbf{X}\boldsymbol{\beta}] \\
 &= \exp[\mathbf{1}\beta_0 + \mathbf{INC}\beta_1 + \mathbf{POV}\beta_2 + \mathbf{BLK}\beta_3 + \mathbf{CRI}\beta_4 + \mathbf{SOU}\beta_5 + \mathbf{DEG}\beta_6] \\
 &= \mathbb{E}[\mathbf{Y}] = \mathbb{E}[\mathbf{EXE}].
 \end{aligned}$$

```

dp.97 <- read.table("https://jeffgill.org/files/jeffgill/files/cpunish.dat_.txt",
  header=TRUE)
PROPDEGREE <- matrix(apply(dp.97[,12:14],1,sum)/apply(dp.97[8:14],1,sum),
  nrow(dp.97),1,dimnames=list(dimnames(dp.97)[[1]],"PROPDEGREE"))
dp.97 <- cbind(dp.97,PROPDEGREE)
dp.out <- glm(EXECUTIONS ~ INCOME + PERPOVERTY + PERBLACK + log(VC100k96) + SOUTH
  + PROPDEGREE, family=poisson, data=dp.97)

```

Poisson GLM of Capital Punishment Data

Table 2: MODELING CAPITAL PUNISHMENT IN THE UNITED STATES: 1997

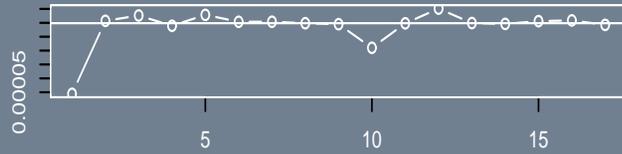
	Coefficient	Standard Error	95% Confidence Interval
(Intercept)	-6.30665	4.17678	[-14.49299: 1.87969]
Median Income	0.00027	0.00005	[0.00017: 0.00037]
Percent Poverty	0.06897	0.07979	[-0.08741: 0.22534]
Percent Black	-0.09500	0.02284	[-0.13978: -0.05023]
log(Violent Crime)	0.22124	0.44243	[-0.64591: 1.08838]
South	2.30988	0.42875	[1.46955: 3.15022]
Degree Proportion	-19.70241	4.46366	[-28.45102:-10.95380]
Null deviance: 136.573, $df = 16$			Maximized $\ell()$: -31.7375
Summed deviance: 18.212, $df = 11$			AIC: 77.475

Poisson GLM of Capital Punishment, Residuals

Table 3: RESIDUALS FROM POISSON MODEL OF CAPITAL PUNISHMENT

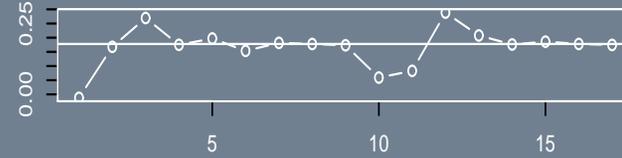
	Response	Pearson	Working	Deviance	Anscombe
Texas	1.70755431	0.28741478	0.04837752	0.28515874	0.28292493
Virginia	0.87407687	0.30671010	0.10762321	0.30136452	0.29629097
Missouri	4.59530299	3.86395636	3.24898061	2.86925916	2.27854829
Arkansas	0.26481208	0.13694108	0.07081505	0.13544624	0.13391171
Alabama	0.95958171	0.67097152	0.46916278	0.62736060	0.58874967
Arizona	0.95395198	0.93375106	0.91397549	0.82741022	0.74425671
Illinois	0.13924315	0.10197129	0.07467388	0.10084230	0.09963912
South Carolina	-0.38227185	-0.24752186	-0.16027167	-0.25478237	-0.26235519
Colorado	-0.95901329	-0.68428704	-0.48826435	-0.75706323	-0.84845827
Florida	-1.82216650	-1.08543456	-0.64657649	-1.25272634	-1.49557143
Indiana	-2.17726883	-1.21566195	-0.67880001	-1.42915840	-1.74185735
Kentucky	-2.31839936	-1.26926054	-0.69489994	-1.49593905	-1.83715998
Louisiana	-1.60160305	-0.99359914	-0.61640776	-1.13620002	-1.33738726
Maryland	0.10161119	0.10709684	0.11287657	0.10527242	0.10341466
Nebraska	0.07022962	0.07261924	0.07506941	0.07194451	0.07107841
Oklahoma	0.49917358	0.70406163	0.99304011	0.62019695	0.55401828
Oregon	-0.90510552	-0.65451282	-0.47330769	-0.72189767	-0.80517526

INCOME



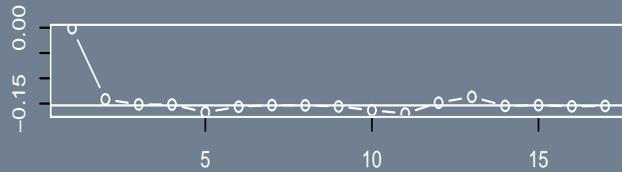
Index Number

PERPOVERTY



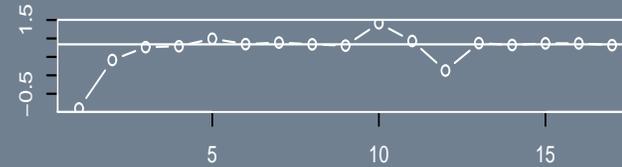
Index Number

PERBLACK



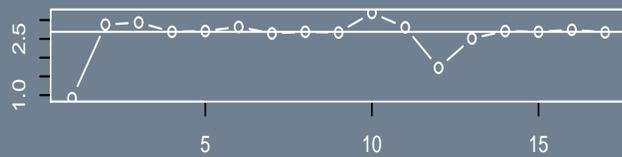
Index Number

log(VC100k96)



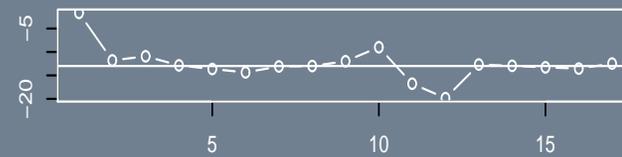
Index Number

SOUTH



Index Number

PROPDEGREE



Index Number

First Differences for Non-Linear Models

- ▶ We can no longer use “a one unit change in X gives a β change in Y .”
- ▶ Main idea:
 - ▷ pick one covariate of interest, \mathbf{X}_q
 - ▷ choose 2 levels of this variable, $\mathbf{X}_{1,q}$, $\mathbf{X}_{2,q}$
 - ▷ set all other covariates at their mean, $\bar{\mathbf{X}}_{-q}$
 - ▷ create two predictions by running these values through the link function:

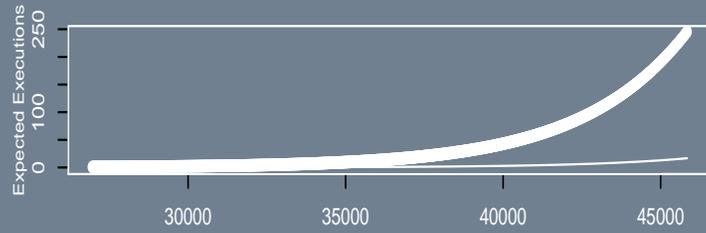
$$\hat{Y}_1 = g^{-1}(\bar{\mathbf{X}}_{-q}\hat{\boldsymbol{\beta}}_{-q} + \mathbf{X}_{1,q}\hat{\boldsymbol{\beta}}_q)$$

$$\hat{Y}_2 = g^{-1}(\bar{\mathbf{X}}_{-q}\hat{\boldsymbol{\beta}}_{-q} + \mathbf{X}_{2,q}\hat{\boldsymbol{\beta}}_q)$$

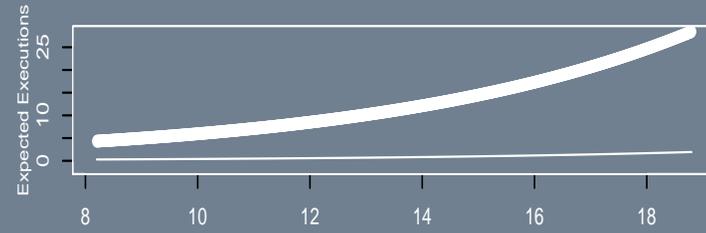
- ▷ Look at $\hat{Y}_1 - \hat{Y}_2$.

- ▶ For example:

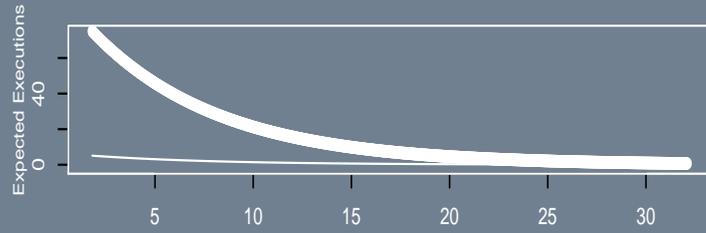
```
dp.1 <- dp.2 <- c(1, apply(dp.97[, c(3,4,5,6,7,15)], 2, mean))
dp.1[6] <- 0; dp.2[6] <- 1
y.1 <- exp(dp.1 %*% dp.out$coef); y.2 <- exp(dp.2 %*% dp.out$coef)
y.2 - y.1
```



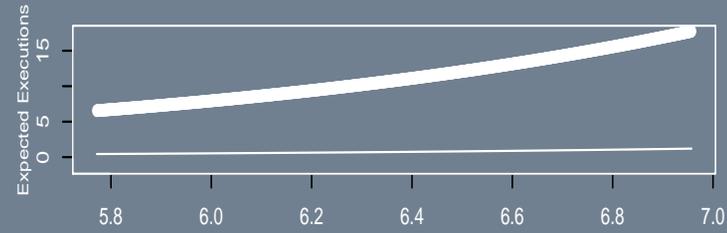
Levels of INCOME



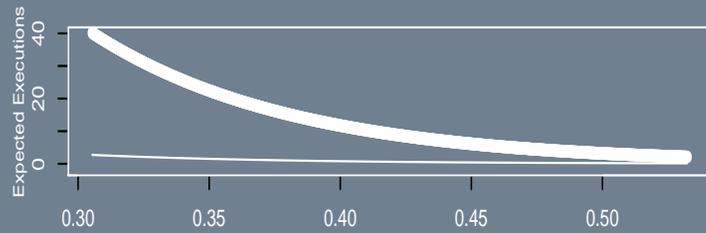
Levels of PERPOVERTY



Levels of PERBLACK



Levels of log(VC100k96)



Levels of PROPDEGREE

— South State
— Non-South State

Poisson GLM of Capital Punishment, First Difference Code

```
X <- cbind(rep(1,nrow(dp.97)), as.matrix(dp.97[,3:5]), as.matrix(log(dp.97[,6])),
          as.matrix(dp.97[,7]), as.matrix(dp.97[,15]))
X.0 <- cbind(X[,1:5],rep(0,length=nrow(X)),X[,7])
dimnames(X.0)[[2]] <- names(dp.out$coefficients)
X.1 <- cbind(X[,1:5],rep(1,length=nrow(X)),X[,7])
dimnames(X.1)[[2]] <- names(dp.out$coefficients)

postscript("glm.fig2.ps")
par(mfrow=c(3,2),mar=c(4,3,2,2),oma=c(3,1,1,1),col.axis="white",col.lab="white",
     col.sub="white",col="white",bg="slategray")
```

Poisson GLM of Capital Punishment, First Difference Code

```

for (i in 2:(ncol(X.0)-1)) {
  if (i==6) i <- i+1
  ruler <- seq(min(X.0[,i]),max(X.0[,i]),length=1000)
  xbeta0 <- exp(dp.out$coefficients[-i]*%apply(X.0[, -i], 2, mean)
              + dp.out$coefficients[i]*ruler)
  xbeta1 <- exp(dp.out$coefficients[-i]*%apply(X.1[, -i], 2, mean)
              + dp.out$coefficients[i]*ruler)
  plot(ruler,xbeta0,type="l",xlab="",ylab="",
       ylim=c(min(xbeta0,xbeta1)-2,max(xbeta0,xbeta1)) )
  lines(ruler,xbeta1,type="b")
  mtext(outer=F,side=1,paste("Levels of",dimnames(X.0)[[2]][i]),cex=0.8,line=3)
  mtext(outer=F,side=2,"Expected Executions",cex=0.6,line=2)
}
plot(ruler[100:200],rep(ruler[400],101),bty="n",xaxt="n",yaxt="n",xlab="",ylab="",
     type="l",xlim=range(ruler),ylim=range(ruler))
lines(ruler[100:200],rep(ruler[600],101),type="b")
text(ruler[445],ruler[400],"Non-South State",cex=1.4)
text(ruler[390],ruler[700],"South State",cex=1.4)
dev.off()

```

New and Old Ways to Look at Model Fit

- ▶ Approximation to Pearson's Statistic.

$$X^2 = \sum_{i=1}^n \mathbf{R}_{Pearson}^2 = \sum_{i=1}^n \left[\frac{\mathbf{Y} - \boldsymbol{\mu}}{\sqrt{VAR[\boldsymbol{\mu}]}} \right]^2.$$

- ▶ If the sample size is sufficiently large, then $\frac{X^2}{a(\psi)} \sim \chi_{n-p}^2$ where n is the sample size, p is the number of explanatory variables including the constant, and $a(\psi)$ is the scale function that we'll see in Chapter 6.
- ▶ For the summed deviance with sufficient sample size it is also true that $D(\boldsymbol{\eta}, \mathbf{y})/a(\psi) \sim \chi_{n-p}^2$.
- ▶ Recall that it is also common to contrast this with the *null deviance*: the deviance function calculated for a model with no covariates (mean function only).

New and Old Ways to Look at Model Fit

► Akaike Information Criterion.

minimizes the negative likelihood penalized by the number of parameters:

$$\text{AIC} = -2\ell(\hat{\boldsymbol{\beta}}|\mathbf{y}) + 2p$$

where $\ell(\hat{\boldsymbol{\beta}}|\mathbf{y})$ is the maximized model log likelihood value and p is the number of explanatory variables in the model (including the constant). (AIC has a bias towards models that overfit with extra parameters since the penalty component is obviously linear with increases in the number of explanatory variables, and the log likelihood often increases more rapidly.)

► Schwartz Criterion/Bayesian Information Criterion (BIC).

$$\text{BIC} = -2\ell(\hat{\boldsymbol{\beta}}|\mathbf{y}) + p\log(n)$$

where n is the sample size.

► There is also a Deviance Information Criterion (DIC) used in Bayesian MCMC estimation.

Covid Example Without Covariates

- ▶ These are Covid-19 cases count data from Washington State penal institutions through December 31, 2020.
- ▶ The single categorical definition is age group...

```
cases <- read.table("https://jeffgill.org/files/jeffgill/files/wash.prison.covid_.dat_.txt",row.names=1,  
  header=TRUE)  
cases
```

	count
Under-22	64
22-25	272
26-30	615
31-35	753
36-40	760
41-45	556
46-50	425
51-55	369
56-60	270
61-65	167
66-70	87
Over-70	69

Covid Example Without Covariates

- ▶ Here is something to worry about:

```
mean(cases[,1])  
[1] 367.25  
var(cases[,1])  
[1] 65589.48
```

Covid Example Without Covariates

- ▶ Now run a simple model:

```
cases.out <- glm(count ~ 1, data=cases)
summary(cases.out)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-303.25	-220.25	-46.75	203.50	392.75

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	367.25	73.93	4.967	0.000424

(Dispersion parameter for gaussian family taken to be 65589.48)

Null deviance: 721484 on 11 degrees of freedom
Residual deviance: 721484 on 11 degrees of freedom
AIC: 170.1

Number of Fisher Scoring iterations: 2

Over/Under Dispersion

- ▶ For Poisson models the mean and the variance of a single random variable are assumed to be the same.
- ▶ For the likelihood function as a statistic, the variance is scaled by n .
- ▶ Overdispersion, $\text{Var}(Y) > \mathbb{E}(Y)$, is relatively common, whereas underdispersion, $\text{Var}(Y) < \mathbb{E}(Y)$ is rare.
- ▶ Biggest effect is to make the standard errors wrong.
- ▶ One diagnostic: plot $\hat{\mu}$ versus $(y - \hat{\mu})^2$.
- ▶ Solution: make μ a random variable rather than a fixed constant to be estimated, with a gamma distribution: $G[\mu\alpha, \alpha]$. So

$$\mathbb{E}[Y] = \mu \qquad \text{Var}[Y] = \frac{\mu}{\phi}$$

- ▶ This is called the “Poisson-Gamma” model and it means that Y is distributed *negative binomial*.

Congressional Cosponsoring of Bills

- ▶ Fowler (2006) looks at patterns of sponsorship and cosponsorship in Congress from 1973 to 2004.

```
cosponsor <- read.table("https://jeffgill.org/files/jeffgill/files/fowler.dat_.txt"
  ,header=TRUE); head(cosponsor,4)
```

	Congress	Period	Total.Sponsors	Total.Bills
1	93rd	19731974	442	20994
2	94th	19751976	439	19275
3	95th	19771978	437	18578
4	96th	19791980	436	10478

	Mean.Bills.Per.Leg	Mean.Cos.Per.Leg	Mean.Cos.Per.Bill
1	48	129	3
2	44	151	3
3	42	170	4
4	24	187	8

	Cos.Per.Leg	Mean.Dist	Senate
1	70	1.95	0
2	79	1.89	0
3	93	1.83	0
4	111	1.76	0

Application to Congressional Cosponsoring of Bills

- ▶ Look at summary statistics:

```
mean(cosponsor$Mean.Bills.Per.Leg)
```

```
[1] 47.625
```

```
var(cosponsor$Mean.Bills.Per.Leg)
```

```
[1] 828.24
```

```
mean(cosponsor$Mean.Cos.Per.Leg)
```

```
[1] 247.5
```

```
var(cosponsor$Mean.Cos.Per.Leg)
```

```
[1] 6134.7
```

- ▶ This is also clear evidence of *overdispersion* in count data.

Negative Binomial

- ▶ Negative binomial distribution has the same sample space (i.e. on the counting measure) as the Poisson, but contains an additional parameter which can be thought of as gamma distributed and therefore used to model a variance function.
- ▶ Used by many to fit a count model with overdispersion.
- ▶ The binomial distribution measures the number of successes in a given number of fixed trials, whereas the negative binomial distribution measures *the number of failures, y before the r^{th} success*.
- ▶ An alternative but equivalent form,

$$f(y|r, p) = \binom{y-1}{r-1} p^r (1-p)^{y-r},$$

measures the number of trials necessary to get r successes.

- ▶ An important application of the negative binomial distribution is in survey research design. If the researcher knows the value of p from previous surveys, then the negative binomial can provide the number of subjects to contact in order to get the desired number of responses for analysis.

Negative Binomial

- ▶ The PMF is:

$$f(Y|k, p) = \binom{y-1}{k-1} p^k (1-p)^{y-k}, \quad y = 0, 1, 2, \dots, \quad 0 \leq p \leq 1.$$

- ▶ For this parameterization, we get:

$$\mathbb{E}[Y] = \mu, \quad \text{Var}[Y] = \frac{\mu(1 + \phi)}{\phi}.$$

- ▶ If ϕ (the dispersion parameter) is unknown, use the estimate:

$$\hat{\phi} = \frac{X^2}{n-p} = \frac{\sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i}}{n-p}.$$

- ▶ This gives an F-test for comparing models (big values implies a difference in models).

Negative Binomial

- ▶ There are two interpretations:
 - ▷ as a generalized Poisson,
 - ▷ with probability p , modeling the number of trials, Y , before the k th success (alternatively failure) where k is fixed in advance.
- ▶ For estimation, use `library(MASS)`, which has `glm.nb`.
- ▶ Note that there is also:

```
dnbinom(x, size, prob, mu, log = FALSE)
pnbinom(q, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
qnbinom(p, size, prob, mu, lower.tail = TRUE, log.p = FALSE)
rnbinom(n, size, prob, mu)
```

Returning to the Cosponsorship Data

- ▶ Run the negative binomial model:

```
library(MASS)
cosponsor.out <- glm.nb(Mean.Cos.Per.Bill ~ Total.Sponsors + Total.Bills
  + Senate + Mean.Dist, data=cosponsor)
summary(cosponsor.out)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-0.5059	-0.1050	0.0125	0.1023	0.3641

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.671e+00	1.297e+01	0.129	0.897438
Total.Sponsors	1.481e-02	2.994e-02	0.495	0.620777
Total.Bills	-8.401e-05	2.721e-05	-3.087	0.002019
Senate	2.421e+00	1.002e+01	0.242	0.809158
Mean.Dist	-2.859e+00	8.129e-01	-3.517	0.000437

Returning to the Cosponsorship Data

- ▶ Which also has a huge improvement in summed deviance:

(Dispersion parameter for Negative Binomial(1631453) family taken to be 1)

```
Null deviance: 133.4743 on 31 degrees of freedom
Residual deviance: 1.2843 on 27 degrees of freedom
AIC: 130.56
```

```
Number of Fisher Scoring iterations: 1
```

Negative Binomial GLM, Congressional Activity: 1995

- ▶ Compare the number of bills assigned to committee in the first 100 days of the 103rd and 104th Houses as a function of the number of members on the committee, the number of subcommittees, the number of staff assigned to the committee, and a dummy variable indicating whether or not it is a high prestige committee.
- ▶ The model is developed with the link function:

$$\eta = g(\mu) = \log\left(\frac{\mu}{\mu + \frac{1}{k}}\right) \quad \longrightarrow \quad \mu = g^{-1}(\eta) = \frac{\exp(\eta)}{k(1 - \exp(\eta))},$$

where $\eta = \mathbf{X}\boldsymbol{\beta}$, and $k \geq 1$ is the overdispersion term.

Negative Binomial GLM, Bills Assigned to Committed, First 100 Days

Committee	Size	Subcommittees	Staff	Prestige	Bills-103 rd	Bills-104 th
Appropriations	58	13	109	1	9	6
Budget	42	0	39	1	101	23
Rules	13	2	25	1	54	44
Ways and Means	39	5	23	1	542	355
Banking	51	5	61	0	101	125
Economic/Educ. Opportunities	43	5	69	0	158	131
Commerce	49	4	79	0	196	271
International Relations	44	3	68	0	40	63
Government Reform	51	7	99	0	72	149
Judiciary	35	5	56	0	168	253
Agriculture	49	5	46	0	60	81
National Security	55	7	48	0	75	89
Resources	44	5	58	0	98	142
Transport./Infrastructure	61	6	74	0	69	155
Science	50	4	58	0	25	27
Small Business	43	4	29	0	9	8
Veterans Affairs	33	3	36	0	41	28
House Oversight	12	0	24	0	233	68
Standards of Conduct	10	0	9	0	0	1
Intelligence	16	2	24	0	2	4

Model Code

```
committee.dat <-  
  read.table("https://jeffgill.org/files/jeffgill/files/committee.dat_.txt",  
            header=TRUE)  
  
committee.poisson <- glm(BILLS104 ~ SIZE + SUBS * (log(STAFF)) + PRESTIGE +  
  BILLS103, family=poisson, data=committee.dat)  
1 - pchisq(summary(committee.poisson)$deviance,  
          summary(committee.poisson)$df.residual)  
[1] 0 # IN THE TAIL INDICATES OVERDISPERSION  
  
committee.out <- glm.nb(BILLS104 ~ SIZE + SUBS * (log(STAFF)) + PRESTIGE +  
  BILLS103, data=committee.dat)  
  
resp <- resid(committee.out,type="response")  
pears <- resid(committee.out,type="pearson")  
working <- resid(committee.out,type="working")  
devs <- resid(committee.out,type="deviance")  
cbind(resp,pears,working,devs)
```

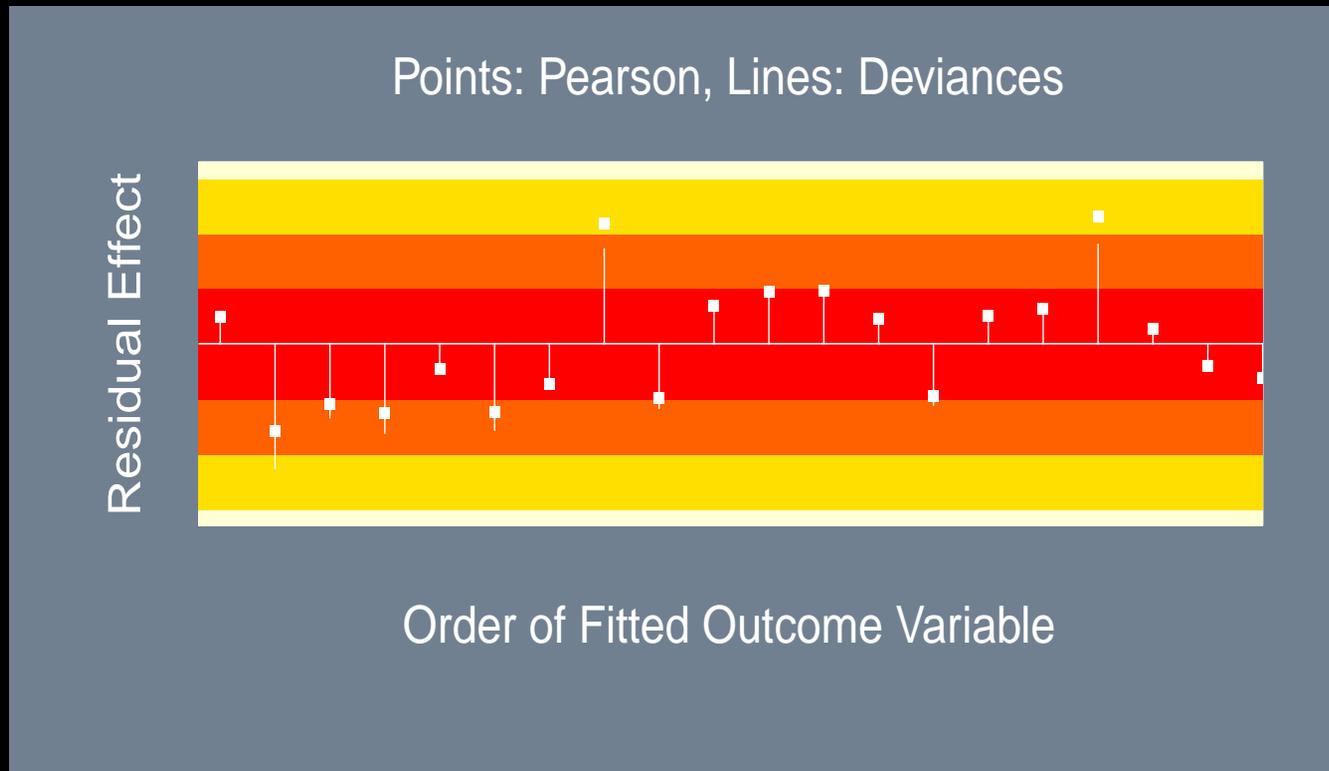
Negative Binomial GLM, Congressional Activity: 1995

	resp	pears	working	devs
Appropriations	-7.38308	-0.99451	-0.55167	-1.22671
Budget	-6.17325	-0.40931	-0.21161	-0.43997
Rules	22.54158	1.98665	1.05048	1.56745
Ways_and_Means	-135.06135	-0.56848	-0.27560	-0.63081
Banking	21.00117	0.40998	0.20194	0.38568
Economic_Educ_Oppor	-93.92104	-0.85695	-0.41757	-1.01572
Commerce	-58.03818	-0.36306	-0.17639	-0.38675
International_Relations	-49.33480	-0.89295	-0.43918	-1.06810
Government_Reform	32.60986	0.57003	0.28018	0.52480
Judiciary	27.80878	0.25343	0.12349	0.24378
Agriculture	24.21181	0.85168	0.42635	0.75680
National_Security	27.14348	0.87911	0.43881	0.77861
Resources	26.13708	0.45893	0.22559	0.42884
TransInfrastructure	79.10378	2.10068	1.04226	1.64133
Science	-34.35454	-1.12146	-0.55993	-1.43001
Small_Business	-12.50419	-1.14887	-0.60984	-1.48074
Veterans_Affairs	-14.18802	-0.66378	-0.33630	-0.75200
House_Oversight	16.14917	0.62009	0.31145	0.56716
Stds_of_Conduct	0.37836	0.44850	0.60864	0.40700
Intelligence	-13.58498	-1.43490	-0.77253	-2.05981

Modeling Bill Assignment – 104th House, Results

	Coefficient	Standard Error	95% Confidence Interval
(Intercept)	-6.80543	2.54651	[-12.30683:-1.30402]
Size	-0.02825	0.02093	[-0.07345: 0.01696]
Subcommittees	1.30159	0.54370	[0.12701: 2.47619]
log(Staff)	3.00971	0.79450	[1.29329: 4.72613]
Prestige	-0.32367	0.44102	[-1.27644: 0.62911]
Bills in 103 rd	0.00656	0.00139	[0.00355: 0.00957]
Subcommittees:log(STAFF)	-0.32364	0.12489	[-0.59345:-0.05384]
Null deviance: 107.314, $df = 19$			Maximized $\ell()$: 10559
Summed deviance: 20.948, $df = 13$			AIC: 121130

Modeling Bill Assignment – 104th House, Residuals Diagnostics



Rate Models

- ▶ Accounts for occurrences, maximum possible events, time.
- ▶ Note that the binomial does not account for repeat events on the same unit.
- ▶ A key problem is that units may differ in size: crime events are higher in bigger cities.
- ▶ Focus on rate:

$$\text{Rate} = \frac{\text{\#events}}{\text{unit}} = \frac{\text{occurrences}}{\text{possibilities}}$$

- ▶ Example from Faraway:
 - ▷ gamma radiation leads to cell abnormalities,
 - ▷ **ca** is the count of chromosomal abnormalities,
 - ▷ **cells** is the number (in hundreds) of exposed cells,
 - ▷ **doseamt** = dose amount,
 - ▷ **doserate** = rate of application.

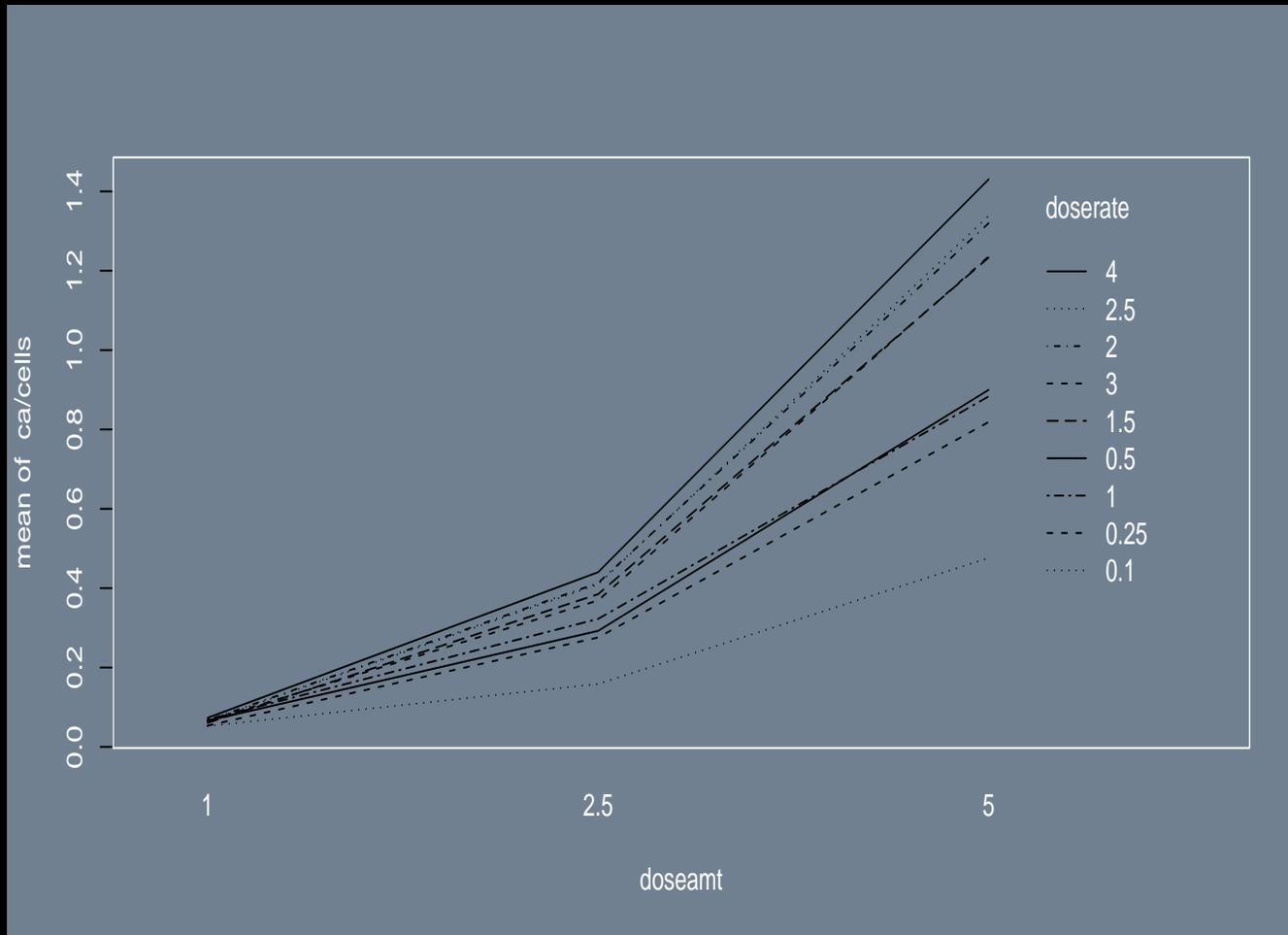
Rate Models

```
library(faraway)
data(dicentric)
round(xtabs(ca/cells ~ doseamt + doserate, dicentric),2)
```

	doserate									
doseamt	0.1	0.25	0.5	1	1.5	2	2.5	3	4	5
1	0.05	0.05	0.07	0.07	0.06	0.07	0.07	0.07	0.07	0.07
2.5	0.16	0.28	0.29	0.32	0.38	0.41	0.41	0.37	0.44	0.44
5	0.48	0.82	0.90	0.88	1.23	1.32	1.34	1.24	1.43	1.43

```
postscript("dicentric.ps")
par(mfrow=c(1,1),col.axis="white",col.lab="white",col.sub="white",col="white",
     bg="slategray")
with(dicentric,interaction.plot(doseamt,doserate,ca/cells))
dev.off()
```

Rate Models



Rate Models

- ▶ MODEL 1: Linearly modeling the ratio directly seems to fit well, but there is overdispersion.

```
lmod <- lm(ca/cells ~ log(doserate)*factor(doseamt), dicentric); summary(lmod)
```

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	0.06349	0.01953	3.25	0.0038
log(doserate)	0.00457	0.01669	0.27	0.7868
factor(doseamt)2.5	0.27631	0.02762	10.01	1.9e-09
factor(doseamt)5	1.00412	0.02762	36.36	< 2e-16
log(doserate):factor(doseamt)2.5	0.06393	0.02361	2.71	0.0132
log(doserate):factor(doseamt)5	0.23913	0.02361	10.13	1.5e-09

Residual standard error: 0.0586 on 21 degrees of freedom

Multiple R-squared: 0.987, Adjusted R-squared: 0.984

F-statistic: 330 on 5 and 21 DF, p-value: <2e-16

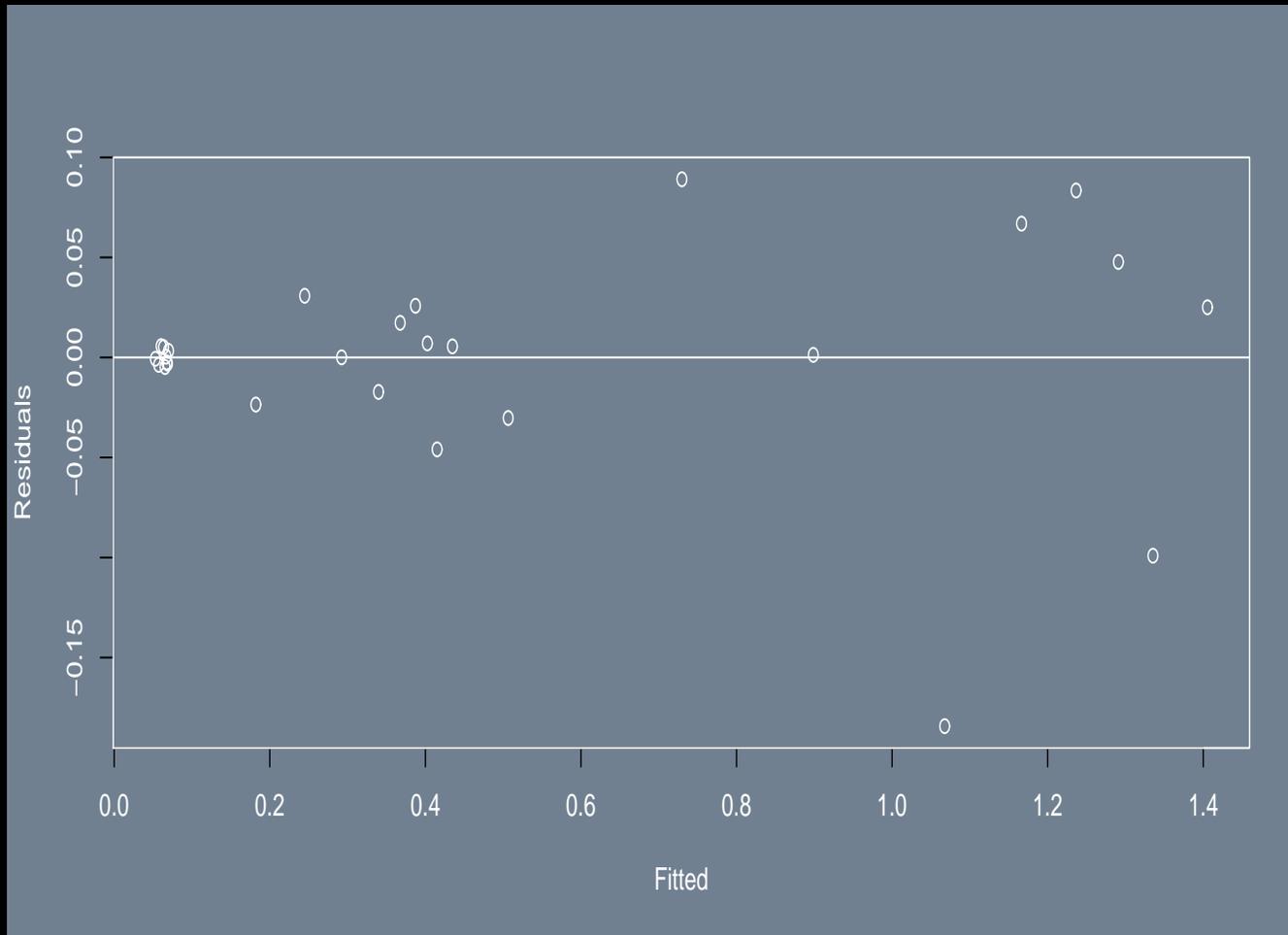
[1]

Rate Models

```
pchisq(sum(lmod$residual), lmod$df.residual)
[1]

postscript("rate.diag.ps")
par(mfrow=c(1,1),col.axis="white",col.lab="white",col.sub="white",
     col="white",bg="slategray")
plot(residuals(lmod) ~ fitted(lmod),xlab="Fitted",ylab="Residuals")
abline(h=0)
dev.off()
```

Rate Models



Rate Models

- ▶ MODEL 2: Poisson modeling directly the counts, starting with logging the number of cells since it has a multiplicative effect on the outcome, and make `doseamt` a factor:

```
dicentric$dosef <- factor(dicentric$doseamt)
pmod <- glm(ca ~ log(cells)+log(doserate)*dosef,family=poisson,dicentric)
summary(pmod)
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.7653	0.3812	-7.25	4e-13
log(cells)	1.0025	0.0514	19.52	< 2e-16
log(doserate)	0.0720	0.0355	2.03	0.04240
dosef2.5	1.6298	0.1027	15.87	< 2e-16
dosef5	2.7667	0.1229	22.52	< 2e-16
log(doserate):dosef2.5	0.1611	0.0484	3.33	0.00087
log(doserate):dosef5	0.1932	0.0430	4.49	7e-06

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 916.127 on 26 degrees of freedom
Residual deviance: 21.748 on 20 degrees of freedom
AIC: 211.2
```

Using an Offset

- ▶ We just modeled these as counts independent of the amount of exposure.
- ▶ But the deaths are actually out of a number of cases exposed.
- ▶ This is called a rate model in the count literature: events per unit of exposed.
- ▶ Thus we want to put exposure on the RHS of the model, being careful about logs:

$$\log\left(\frac{\mathbb{E}[Y|\boldsymbol{\beta}, \mathbf{X}]}{\text{exposure}}\right) = \mathbf{X}\boldsymbol{\beta}$$

$$\log(\mathbb{E}[Y|\boldsymbol{\beta}, \mathbf{X}]) - \log(\text{exposure}) = \mathbf{X}\boldsymbol{\beta}$$

$$\log(\mathbb{E}[Y|\boldsymbol{\beta}, \mathbf{X}]) = \mathbf{X}\boldsymbol{\beta} + \log(\text{exposure})$$

$$\mathbb{E}[Y|\boldsymbol{\beta}, \mathbf{X}] = \exp[\mathbf{X}\boldsymbol{\beta} + \log(\text{exposure})]$$

which justifies putting a log-constant on the RHS to reflect the number exposed in each case.

- ▶ In R this is done with the `offset()` specification, for example:

```
glm(Y ~ X1 + X2 + offset(X3), family=poisson, data=swe07)
```

Rate Models

- ▶ MODEL 3: make this intuitive like a standard Poisson model:

$$\log\left(\frac{\text{ca}}{\text{cells}}\right) = \mathbf{X}\boldsymbol{\beta} \quad \implies \quad \log(\text{ca}) = \log(\text{cells}) + \mathbf{X}\boldsymbol{\beta}.$$

- ▶ Note also the estimate `log(cells) 1.0025` in the previous model, which suggests that this parameter is really just 1, so fix it at one using an offset:

```
rmod <- glm(ca ~ offset(log(cells))+log(doserate)*dosef, family=poisson,dicentric);
```

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.7467	0.0343	-80.16	< 2e-16
log(doserate)	0.0718	0.0352	2.04	0.04130
dosef2.5	1.6254	0.0495	32.86	< 2e-16
dosef5	2.7611	0.0435	63.49	< 2e-16
log(doserate):dosef2.5	0.1612	0.0483	3.34	0.00084
log(doserate):dosef5	0.1935	0.0424	4.56	5.1e-06

(Dispersion parameter for poisson family taken to be 1)

```
Null deviance: 4753.00 on 26 degrees of freedom
Residual deviance: 21.75 on 21 degrees of freedom
AIC: 209.2
```

Zero-Inflated Poisson Model

- ▶ Zero-inflated Poisson (ZIP) regression is first introduced Lambert (1992) although the ZIP distribution, without covariates, has been discussed early in literatures (Cohen 1963, Yip 1988).
- ▶ The main advantage of this model is to deal with so called “structural” zeros in modeling count data.
- ▶ The ZIP regression model assumes that zeros are observed with probability π , and the rest of observations come from a **Poisson**(λ) with probability $1 - \pi$.

Zero-Inflated Poisson Model

► Let Y_1, \dots, Y_N be a sample of size N independently drawn from

$$Y_i \sim \begin{cases} 0 & \text{with probability } \pi_i \\ \text{Poisson}(\lambda_i) & \text{with probability } 1 - \pi_i \end{cases}$$

► So the probability mass function is given by

$$P(Y_i = h) = \begin{cases} \pi_i + (1 - \pi_i)e^{-\lambda_i} & \text{for } h = 0 \\ (1 - \pi_i)e^{-\lambda_i}\lambda_i^h/h! & \text{for } h = 1, 2, \dots \end{cases}$$

Zero-Inflated Poisson Model

- ▶ The regression model with this zero-inflated Poisson distribution now consists of two generalized linear models.
- ▶ The first part is a logistic regression, specified by $\text{logit}(\pi_i) = \mathbf{u}_i^T \boldsymbol{\gamma}$, where the response variable states zero or nonzero status and $\boldsymbol{\gamma}$ is a regression coefficient vector for covariates \mathbf{u}_i^T .
- ▶ The second part is a poisson regression, specified by $\log(\lambda_i) = \mathbf{x}_i^T \boldsymbol{\beta}$, where the response variable is a non-negative count from a $\text{Poisson}(\lambda_i)$ and $\boldsymbol{\beta}$ is a regression coefficient vector for covariates \mathbf{x}_i^T .
- ▶ This separation allows the predictors in each model to perform different roles; for example, what causes exact zeros (no-movement) is different from what causes vigorous activities.

Hurdle Model

- ▶ A similar approach to handle zero-inflated count data is also introduced in Mullahy (1986) referred as a hurdle model.
- ▶ This model utilizes a zero-truncated Poisson distribution:

$$P(Y_i = h | Y_i > 0) = \lambda_i^h / \{(e^{\lambda_i} - 1)h!\}$$

- ▶ The probability mass function in the ZIP model is modified to

$$P(Y_i = h) = \begin{cases} \pi_i & \text{for } h = 0 \\ (1 - \pi_i)\lambda_i^h / \{(e^{\lambda_i} - 1)h!\} & \text{for } h = 1, 2, \dots \end{cases}$$

- ▶ The hurdle model has the advantage of handling both zero-inflated and zero-deflated count data.

Congress and the Supreme Court

- ▶ Zorn (1996) observes...

Whether due to institutional deference, agreement with case outcomes, or simple inattention, the typical Supreme Court decision is final: Congress rarely intervenes to modify or overturn the high Courts ruling. As a result, the vast majority of Supreme Court cases are never addressed by the Congress.

- ▶ So this is a perfect application for ZIP and hurdle models.

Descriptive Statistics for Dependent and Independent Variables				
Variables	Mean	Std. Dev.	Min.	Max.
Number of Actions Taken	0.11	0.64	0	11
ln(Exposure)	2.04	0.55	0	2.30
Year of Decision	1972.4	9.85	1953	1988
Liberal Decision	0.52	0.50	0	1
Lower Court Disagreement	0.23	0.42	0	1
Alteration of Precedent	0.02	0.15	0	1
Declaration of Unconstitutionality	0.08	0.27	0	1
Unanimous Vote	0.34	0.47	0	1

Note: N = 4052. Data are all Supreme Court decisions handed down during the 1953-1987 terms and which fall under the jurisdiction of House and Senate Judiciary committees. See Zorn and Caldeira (1995) and Eskridge (1991) for a fuller description of how the cases were selected and coded for analysis.

Congress and the Supreme Court

- ▶ The vast majority of decisions received no Congressional scrutiny.
- ▶ Of those that did, the total number of such actions ranged from one to eleven, with a mean of 2.6.
- ▶ The data contain significantly more zeros than would be predicted by a Poisson with a mean of 0.11.
- ▶ In nearly 96 percent of all cases analyzed here no Congressional response occurred during the 1979-1988 period.

Frequencies: Numbers of House and Senate Actions Taken in Response to Supreme Court Decisions, 1979-1988

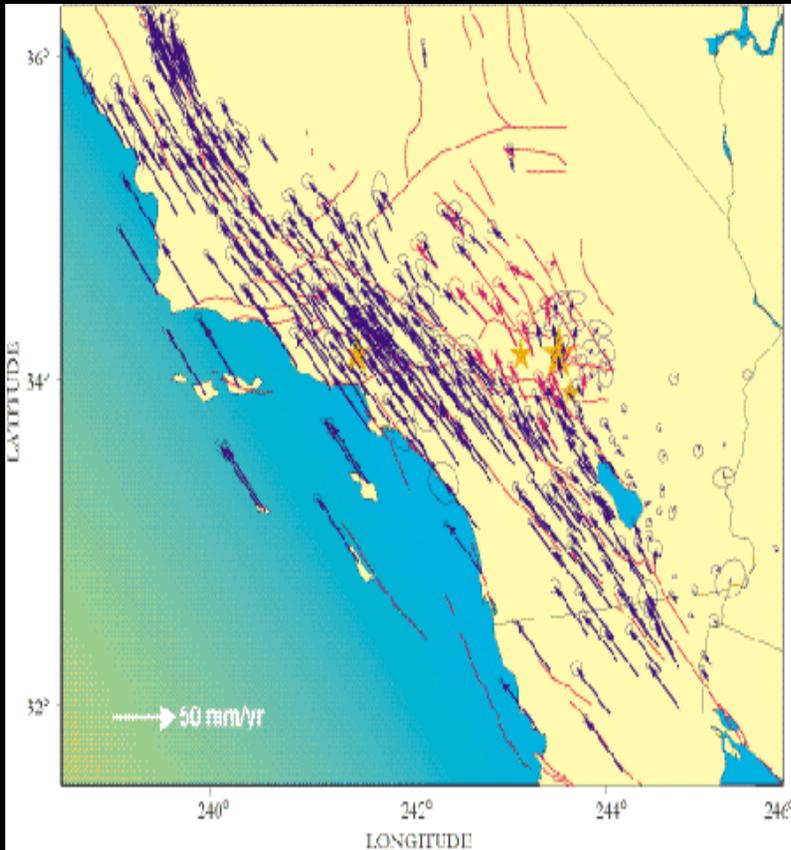
Number of Actions	Frequency	Percentage
0	3882	95.80
1	63	1.55
2	38	0.94
3	32	0.79
4	8	0.20
5	12	0.30
6	12	0.30
7	3	0.07
10	1	0.02
11	1	0.02
Total	4052	100.0

Model Results (Numbers in parentheses are t-ratios)

Variables	Poisson	Negative Binomial
(Constant)	-160.125 (-9.91)	-134.411 (-4.93)
log(Exposure)	0.544 (4.77)	0.178 (0.67)
Year of Decision	0.079 (9.82)	0.067 (4.89)
Liberal Decision	0.296 (3.02)	0.099 (0.45)
Lower Court Disagreement	-0.212 (-1.79)	-0.321 (-1.22)
Alteration of Precedent	-0.254 (-0.67)	-0.102 (-0.13)
Declaration of Unconstitutionality	-1.838 (-4.78)	-1.538 (-2.89)
Unanimous Decision	-0.407 (-3.74)	-0.297 (-1.28)
(σ)	-	32.233 (30.96)
Log-Likelihood	-1636.308	-989.542

Variables	Zero-Inflated Poisson		Hurdle Poisson	
	Prob(Y=0)	E(Y)	Prob(Y>0)	E(Y)
(Constant)	153.580 (6.35)	-8.793 (-0.63)	-153.217 (-5.86)	-9.967 (-0.60)
log(Exposure)	-0.487 (-2.64)	0.089 (0.65)	0.510 (2.76)	0.079 (0.62)
Year of Decision	-0.076 (-6.24)	0.005 (0.68)	0.076 (5.77)	0.005 (0.64)
Liberal Decision	-0.091 (-0.54)	0.190 (2.08)	0.139 (0.87)	0.192 (1.70)
Lower Court Disagreement	0.043 (0.22)	-0.138 (-1.30)	-0.079 (-0.43)	-0.147 (-1.01)
Alteration of Precedent	-0.401 (-0.65)	-0.582 (-1.08)	0.171 (0.34)	-0.601 (-1.11)
Declaration of Unconstitutionality	1.590 (2.42)	-0.421 (-0.69)	-1.696 (-2.88)	-0.367 (-0.78)
Unanimous Decision	0.499 (2.58)	0.098 (0.96)	-0.460 (-2.59)	0.088 (0.70)
Log-Likelihood	-979.483		-671.428	

Multivariate Application: the Predicting Earthquake Aftershocks



- ▶ Topical.
- ▶ Immediately after a powerful earthquake in a high population density area decisions must be made about operating powerplants, schools, and transportation facilities.
- ▶ A series of aftershocks can be equally deadly and destructive as a mainshock.
- ▶ Predicting aftershocks based on empirical evidence is far reliable than predicting mainshocks.

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Why is this relevant?
- ▶ Some geopolitical events are very hard to predict, but their after-effects may be much more reliably anticipated.
- ▶ Examples: terrorist attacks, unannounced nuclear tests, civil wars, coups.
- ▶ Bayesian learning may (over time) increase our knowledge.
- ▶ The need for real-time analysis parallels necessary government reactions after such events.

How Aftershocks Are Described

Infographic



Bay Area's 6.0 quake and aftershocks

[READ THE STORY >](#)

A little more than two hours after the quake, a shallow magnitude 3.6 tremor was reported by the USGS. The aftershock occurred at 5:47 a.m. at a depth of five miles. The National California Seismic System **put the chance of a strong aftershock** in the next week at 54%. Scientists at UC Berkely released a video showing an early-warning system that **sent an alert 10 seconds** before the earthquake.

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Model aftershocks as a *non-homogeneous* Poisson process with the intensity parameter:

$$N(t) \propto \frac{1}{(t + c)^p}.$$

This is actually called “Omori’s Law” where t is time, and the rest are constants: c is a time offset, p is a rate of decay.

- ▶ So the probability of n aftershocks at time period t is:

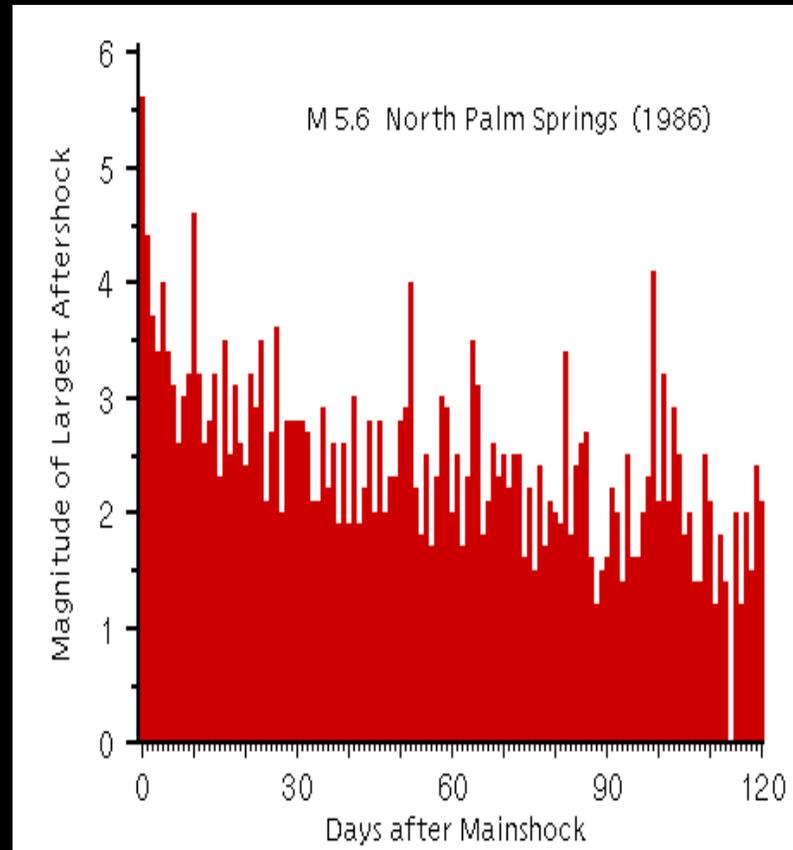
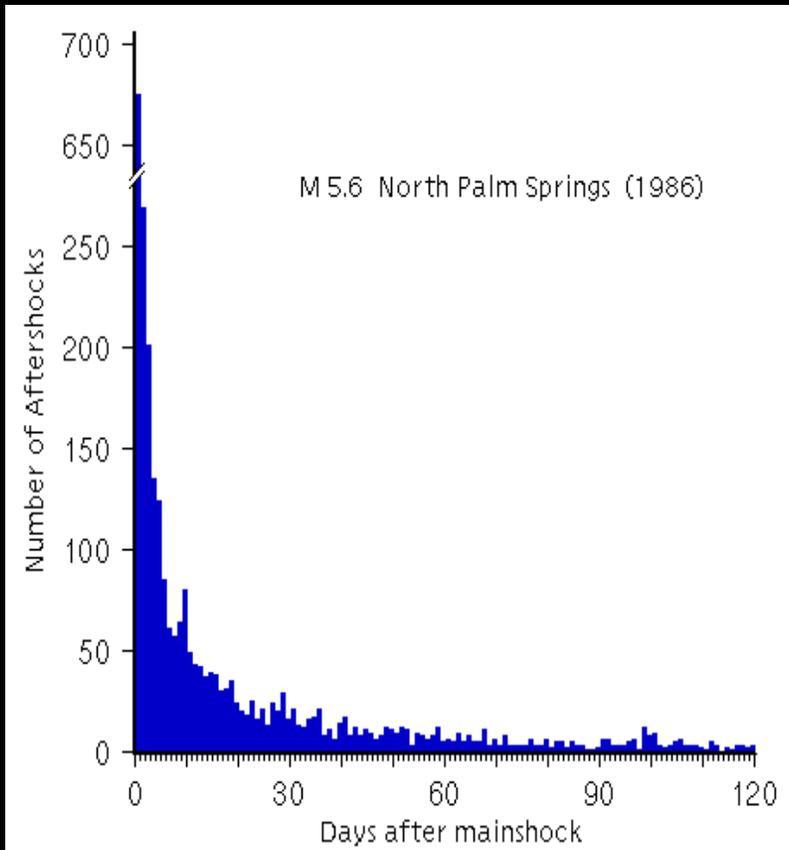
$$P(n|t) = \frac{N(t)^n e^{-N(t)}}{n!}.$$

- ▶ Use the Gutenberg-Richter relation (an empirical law), aftershock version:

$$\log_{10} N(M) = a + b(M_{\text{mainshock}} - M_{\text{aftershock}})$$

where $N(M)$ is the number per year of aftershocks of magnitude greater than $M_{\text{aftershock}}$ following a mainshock of magnitude $M_{\text{mainshock}}$, a and b are constants.

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)



Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Putting these two principles together gives the **rate** of aftershocks of magnitude $M_{\text{aftershock}}$ or larger at time t following a mainshock:

$$\lambda(t, M) = 10^{a+b(M_{\text{mainshock}} - M_{\text{aftershock}})}(t + c)^{-p}$$

- ▶ More usefully, the **probability** of an aftershock between M_1 and M_2 , both less than $M_{\text{mainshock}}$, and between time t_1 and t_2 after the mainshock:

$$p(t, M) = 1 - \exp \left[- \int_{M_1}^{M_2} \int_{t_1}^{t_2} \lambda(t, M) dt dM \right]$$

under the assumption that the joint instantaneous rate is distributed exponential (see the figure!)

- ▶ What we need now is a **posterior distribution** for $\boldsymbol{\mu} = (a, b, p, c)$ conditional on the mainshock.

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Start with some (regionalized) data, calculate posteriors with a Bayesian gaussian model and update as new data (earthquakes) occur.
- ▶ Multivariate priors:

$$\boldsymbol{\mu}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\mathbf{m}, \frac{\boldsymbol{\Sigma}}{n_0}\right), \quad \boldsymbol{\Sigma}^{-1} \sim \mathcal{W}(\alpha, \boldsymbol{\beta}),$$

where n_0/n measures our belief in the representativeness prior data.

- ▶ This produces posteriors:

$$\hat{\boldsymbol{\mu}}|\boldsymbol{\Sigma} \sim \mathcal{N}_k\left(\frac{n_0\mathbf{m} + n\bar{\mathbf{x}}}{n_0 + n}, \frac{\boldsymbol{\Sigma}}{n_0 + n}\right)$$

$$\widehat{\boldsymbol{\Sigma}}^{-1} \sim \mathcal{W}_k\left(\alpha + n, \boldsymbol{\beta}^{-1} + S^2 + \frac{n_0n}{n_0 + n}(\bar{\mathbf{x}} - \mathbf{m})(\bar{\mathbf{x}} - \mathbf{m})'\right).$$

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Some information to build “Generic California” priors:
 - ▷ 62 aftershock sequences with $M_{\text{mainshock}} \geq 5$, occurring from 1933 to 1987 in California (exclusive of two unusual events),
 - ▷ Omori’s Law parameters (a, p) from $M_{\text{mainshock}} - M_{\text{aftershock}} \geq 3$,
 - ▷ b from $M_{\text{mainshock}} - M_{\text{aftershock}} \geq 2$,
 - ▷ c picked to get maximum distinction between mainshock “coda” and aftershocks using post-1970 data.
- ▶ Reasenber and Jones (1989) assume Σ^{-1} is diagonal and produce normal priors with means:

$$\bar{a} = -1.67, \quad \bar{b} = 0.91, \quad \bar{p} = 1.08, \quad c = 0.05$$

($\sigma_a = 0.0.7$, $\sigma_b = 0.02$, $\sigma_p = 0.03$, c deterministic).

Multivariate Application: the Predicting Earthquake Aftershocks (cont.)

- ▶ Data taken from real-time sequence of aftershocks for two excluded events:
 - ▷ Coalinga (1983), $M_{\text{mainshock}} = 6.5$
 - ▷ Whittier-Narrows (1987), $M_{\text{mainshock}} = 5.9$and updated *during* aftershock times.
- ▶ Thus probabilities are Bayesianly improved during risk period for an event greater than the mainshock.
- ▶ Updating the “Generic California” priors with the conjugate-normal Bayesian model gives any desired set of probabilities over a period of time after the mainshock by integrating some region of the posterior.

Probability of $M_{\text{aftershock}} > M_{\text{mainshock}} - 1$

Within $t_2 - t_1$	Time After Mainshock, Coalinga								
	15 min.	6 hrs.	12 hrs.	1 day	3 days	7 days	15 days	30 days	60 days
1 Day	0.330	0.176	0.125	0.081	0.033	0.015	0.007	0.003	0.002
3 Days	0.413	0.265	0.209	0.153	0.077	0.039	0.020	0.010	0.005
7 Days	0.467	0.330	0.276	0.218	0.129	0.074	0.040	0.022	0.011
30 Days	0.545	0.427	0.378	0.324	0.234	0.165	0.109	0.069	0.039
60 Days	0.577	0.466	0.420	0.370	0.283	0.214	0.154	0.105	0.066

Within $t_2 - t_1$	Time After Mainshock, Whittier-Narrows								
	15 min.	6 hrs.	12 hrs.	1 day	3 days	7 days	15 days	30 days	60 days
1 Day	0.393	0.141	0.084	0.044	0.012	0.004	0.001	0.000	0.000
3 Days	0.431	0.185	0.123	0.074	0.026	0.010	0.004	0.001	0.000
7 Days	0.488	0.208	0.146	0.095	0.040	0.017	0.007	0.003	0.001
30 Days	0.465	0.232	0.171	0.120	0.062	0.034	0.017	0.009	0.004
60 Days	0.470	0.238	0.178	0.127	0.069	0.040	0.023	0.012	0.006

Example for Whittier-Narrows, if the main shock happened within the last 15 minutes then the probability of a serious aftershock in the next 24 hours is 0.393.