Winter Institute in Data Science and Big Data

The Generalized Linear Regression Model

JEFF GILL
Distinguished Professor
Departments of Government and Mathematics & Statistics
American University

Overview

- We will create a regression model for dichotomous outcome variables: vote/not-vote, war/no-war, pass/fail, etc.

- Note that this is different than having dichotomous explanatory variables.

- Remember that regression is really conditional average, $\mathbb{E}[\mathbf{Y}|\mathbf{X}]$, which does not have the same implications for $0/1$ outcomes on the LHS.

- Consider the probability that a single case has a 0 or a 1 as the outcome:

$$\pi_i = p(Y_i) = p(Y = 1|\mathbf{X} = \mathbf{x}_i), \quad \text{where} \quad \pi \in [0{:}1].$$

- So:

$$\mathbb{E}(Y_i|\mathbf{x}_i) = (\pi_1)(1) + (1 - \pi_i)(0) = \pi_i.$$

  (recall that for discrete RV $\mathbb{E}(A) = \sum_{\text{over events}} P(\text{A}) \times A$)

- This means that we are *estimating* an underlying probability value for given levels of a vector of explanatory variable values.

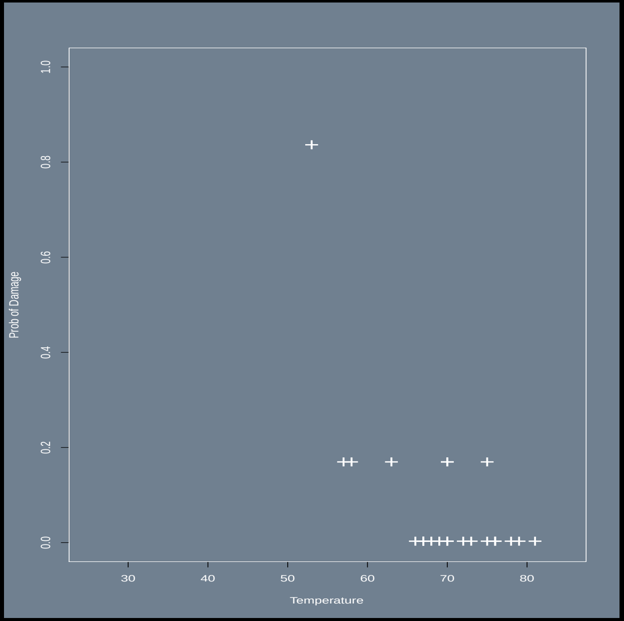
Challenger Disaster Example

- The Challenger Shuttle exploded shortly after takeoff in January 1986 due to a failure of one of it's six O-rings.

- We will model the probability of failure based on past year's temperature and o-ring damage data.

- R code:

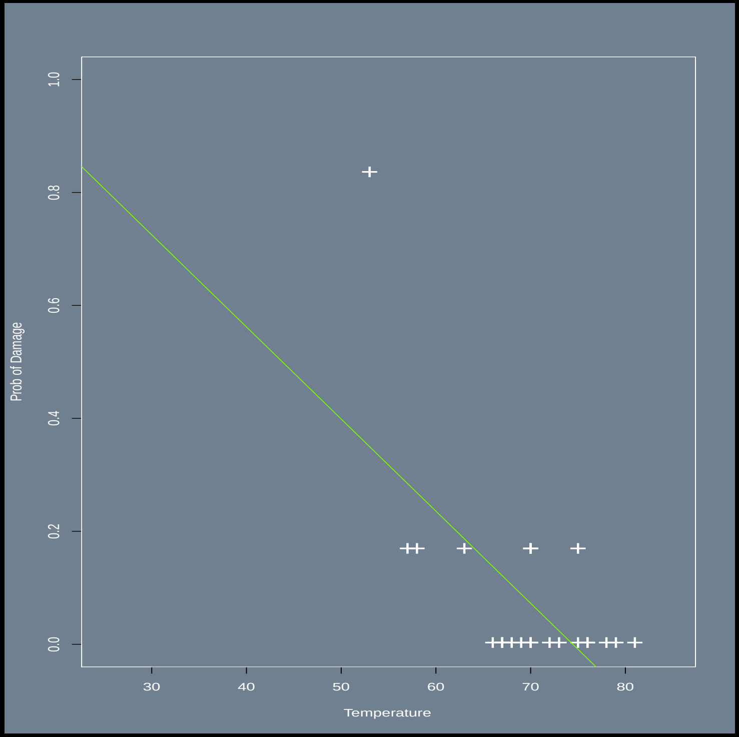
```
library(faraway)
data(orings)
t(orings)

        1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23
temp   53 57 58 63 66 67 67 67 68 69 70 70 70 70 72 73 75 75 76 76 78 79 81
damage  5  1  1  1  0  0  0  0  0  0  1  0  1  0  0  0  0  1  0  0  0  0  0

postscript("faraway.ch2.fig1.ps")
par(col.axis="white",col.lab="white",col.sub="white",col="white",
    bg="slategray")
plot(damage/6 ~ temp, orings, pch="+", xlim=c(25,85), ylim = c(0,1),
     xlab="Temperature", ylab="Prob of Damage", cex=2)
dev.off()
```



Challenger Disaster Example

- Naïve linear-probability model:

```
lmod <- lm(damage/6 ~ temp, orings)
summary(lmod)

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.21429    0.29993    4.05  0.00058
temp        -0.01631    0.00429   -3.80  0.00104

Residual standard error: 0.142 on 21 degrees of freedom
Multiple R-squared: 0.408,        Adjusted R-squared: 0.379
F-statistic: 14.5 on 1 and 21 DF,  p-value: 0.00104

postscript("faraway.ch2.fig2.ps")
par(col.axis="white",col.lab="white",col.sub="white",col="white",
    bg="slategray")
plot(damage/6 ~ temp, orings, pch="+", xlim=c(25,85), ylim = c(0,1),
     xlab="Temperature", ylab="Prob of Damage", cex=2)
abline(lmod, col="lawngreen")
dev.off()
```



Problems with the Linear-Probability Approach

- Allows predictions outside of $[0:1]$.

- Deceptive sense of fit:

```
Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.21429    0.29993    4.05  0.00058
temp        -0.01631    0.00429   -3.80  0.00104
```

- Wrong distributional implications:

$$Y_i = \alpha + \beta x_i + \epsilon \;\Rightarrow\; \pi_i = \alpha + \beta x_i,$$

but since $Y_i \in \{0, 1\}$, then ϵ_i is dichotomous not normally distributed:

$$\begin{aligned} \epsilon_i &= 1 - \mathbb{E}[Y_i] = 1 - (\alpha + \beta x_i) = 1 - \pi_i \\ &\text{or} \\ \epsilon_i &= 0 - \mathbb{E}[Y_i] = 0 - (\alpha + \beta x_i) = -\pi_i \end{aligned}$$

Problems with the Linear-Probability Approach

- The expectation is okay:

$$\mathbb{E}[\epsilon_i] = \mathbb{E}[Y_i - \alpha - \beta x_i] = \pi_i - \pi_i = 0,$$

but the variance is wrong:

$$\text{Var}[\epsilon_i] = \mathbb{E}[\epsilon_i^2] - (\mathbb{E}[\epsilon_i])^2 = \mathbb{E}[\epsilon_i^2],$$

which turns out to be:

$$\begin{aligned}
\mathbb{E}[\epsilon_i^2] &= \sum_{i=0}^{1} \epsilon_i^2 p(\epsilon_i) \\
&= (1-\pi_i)^2(\pi_i) + (-\pi_i)^2(1-\pi_i) \\
&= (1-\pi_i)[(1-\pi_i)(\pi_i) + \pi_i^2] \\
&= (1-\pi_i)\pi_i \\
&= \pi_i - \pi_i^2.
\end{aligned}$$

This is a quadratic form and is therefore heteroscedastic, especially near zero and one.

Ad Hoc "Fix:" Constrained Linear-Probability Model

- Fix π artificially:

$$\pi = \begin{cases} 0 & 0 > \alpha + \beta x \\ \alpha + \beta x & 0 \leq \alpha + \beta x \leq 1 \\ 1 & \alpha + \beta x > 1 \end{cases}$$

- The hardest part is finding a criterion for 0 and 1 on the x-axis.
- The effect of x is difficult to interpret.
- This is a difficult estimation problem.
- The abrupt changes are substantively unreasonable (do not have derivatives).

Implementing the Constrained Linear-Probability Model

- We need "corner points" so I'm going to cheat and run a logit model and look at the expected values for good candidate values.

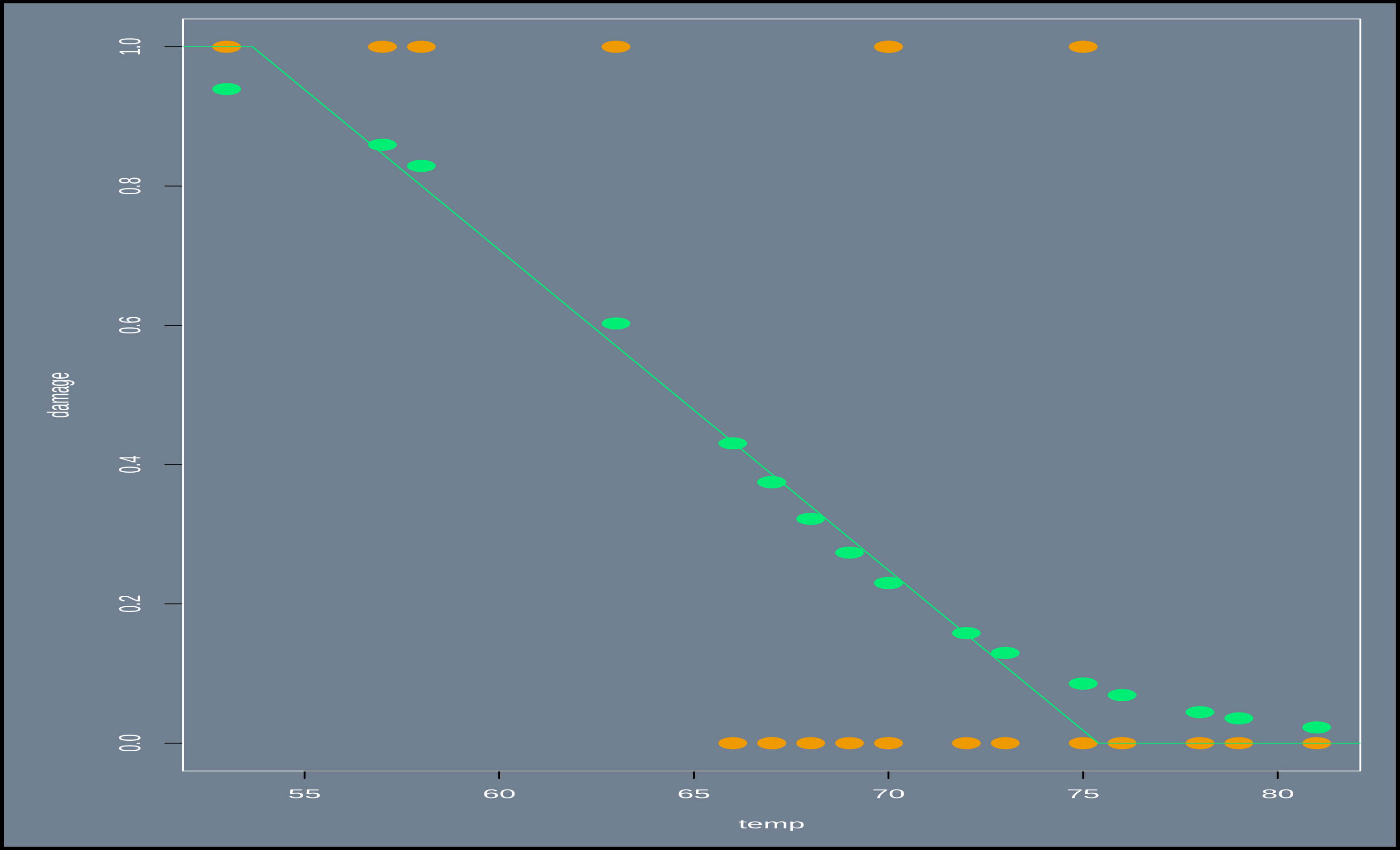
- The first row of the data has the explosion value of 5, which does not work for this example.

```
orings2 <- orings; orings2[1,2] <- 1
logitmod2 <- glm(damage ~ temp, family=binomial(link=logit), orings2)
logitmod2$fitted.values
sort(logitmod2$fitted.values)
      23       22       21       19       20       17       18       16       15
0.022703 0.035641 0.044541 0.069044 0.069044 0.085544 0.085544 0.129546 0.158049
      11       12       13       14       10        9        6        7        8
0.229968 0.229968 0.229968 0.229968 0.273621 0.322094 0.374724 0.374724 0.374724
       3        2        1
0.828845 0.859317 0.939248
```

Implementing the Constrained Linear-Probability Model

```
# SETUP PLOT WINDOW, PLOT DATA POINTS
postscript("faraway.ch2.fig6.ps")
par(col.axis="white",col.lab="white",col.sub="white",col="white",
    col.main="white",bg="slategray",mar=c(5,5,1,1))
plot(orings2,pch=19,cex=2,col="orange2")
points(orings2$temp,logitmod2$fitted.values,pch=19,cex=2,col="springgreen2")

# ARBITRARILY CHOOSE CORNER POINTS FROM LINEAR MODEL FITTED VALUES
# BY EXCLUDING VALUES ON EITHER END OF THE DATA
partial.lm <- lm(logitmod2$fitted.values[-c(1,2,18:23)] ~ orings2$temp[-c(1,2,18:23)])
( y0.xval <-  abs(coef(partial.lm)[1]/coef(partial.lm)[2]) )
[1] 75.397
( y1.xval <-  (1-coef(partial.lm)[1])/coef(partial.lm)[2] )
[1] 53.663

segments(y0.xval,0,y1.xval,1,cex=1.5,col="springgreen2")
segments(y0.xval,0,90,0,cex=1.5,col="springgreen2")
segments(y1.xval,1,40,1,cex=1.5,col="springgreen2")
dev.off()
```

Implementing the Constrained Linear-Probability Model



New Conceptual Model

- Start with the linear predictor $\boldsymbol{\eta} = \alpha + \beta \mathbf{x}$.
- Now let's specify a link function that relates the linear additive RHS component to the expected value of the nonlinear LHS component:

$$\pi_i = g^{-1}(\eta_i) = p(\alpha_i + \beta_i x) \;\Rightarrow\; g(\pi_i) = \eta_i = \alpha_i + \beta_i x.$$

- Objectives for $g^{-1}()$:
  - ▷ smooth on $[0{:}1]$
  - ▷ For a positive effect of $\mathbf{x}_i$ on π_i:
    - $g^{-1} \;\rightarrow\; 0 \text{ as } x_i \;\rightarrow, -\infty$
    - $g^{-1} \;\rightarrow\; 1 \text{ as } x_i \;\rightarrow, +\infty.$
  - ▷ For a negative effect of $\mathbf{x}_i$ on π_i:
    - $g^{-1} \;\rightarrow\; 1 \text{ as } x_i \;\rightarrow, -\infty$
    - $g^{-1} \;\rightarrow\; 0 \text{ as } x_i \;\rightarrow, +\infty.$

New Conceptual Model

- There are two common solutions for $g^{-1}()$.
- Logit:

$$\Lambda(\eta_i) = [1 + \exp(-\eta_i)]^{-1}$$

- Probit:

$$\Phi(\eta_i) = (2\pi)^{-\frac{1}{2}} \int_{-\infty}^{\eta_i} \exp[-\frac{1}{2}\eta_i^2] d\eta_i$$

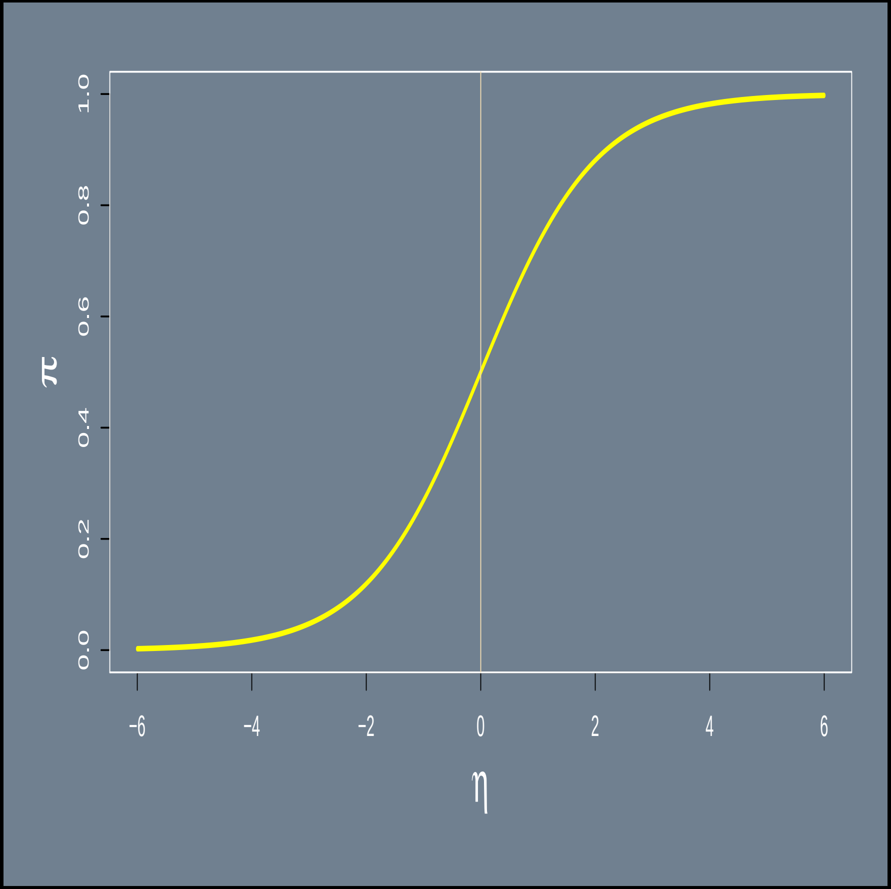
- These are sometimes given in $g()$ form: $\Phi^{-1}(\pi_i)$ and $\Lambda^{-1}(\eta_i) = \text{logit}(\pi_i) = \log\left(\frac{p_i}{1-p_i}\right)$.
- Less common is the cloglog function:

$$g(\mu) = -\log\left(-\log(1-\mu)\right) \qquad\qquad g^{-1}(\eta) = 1 - \exp\left(-\exp(\eta)\right)$$

Latent Variable Justification

- Humans make dichotomous decisions from smooth preference structures, but we only see discrete choices in the data.

- The Index Function (Utility) model states that if *benefits* - *costs* = U is greater than zero then the choice should be a one, and vice-versa.



Latent Variable Justification

- Utility model states: $U_i = \mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i$ (subsume the constant into the vector), and $p(U_i > 0) = p(\mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\epsilon}_i > 0) = p(\boldsymbol{\epsilon}_i > -\mathbf{x}_i\boldsymbol{\beta})$.
- Political Example:
  - U^R, the utility of voting for the Republican candidate
  - U^D, the utility of voting for the Democratic candidate
  - direction is arbitrary, so pick $Y = 1$ the decision to vote for the Republican candidate
  - Define the two utility functions in regression terms:

$$U_i^R = \mathbf{x}_i\boldsymbol{\beta}_R + \boldsymbol{\epsilon}_{iR} \qquad\qquad U_i^D = \mathbf{x}_i\boldsymbol{\beta}_D + \boldsymbol{\epsilon}_{iD}$$

  - So now:

$$\begin{aligned} p(Y_i = 1|\mathbf{x}_i) &= p(U_i^R > U_i^D) \\ &= p(\mathbf{x}_i\boldsymbol{\beta}_R + \boldsymbol{\epsilon}_{iR} > \mathbf{x}_i\boldsymbol{\beta}_D + \boldsymbol{\epsilon}_{iD}|\mathbf{x}_i) \\ &= p(\mathbf{x}_i[\boldsymbol{\beta}_R - \boldsymbol{\beta}_D] + \boldsymbol{\epsilon}_{iR} - \boldsymbol{\epsilon}_{iD} > 0) \\ &= p(\mathbf{x}_i\boldsymbol{\beta} + \boldsymbol{\epsilon} > 0) \end{aligned}$$

  which is just 1-CDF.

Binomial Regression Model

- What we are really doing here is a binomial since there are $i = 1, 2, \ldots, n$ "attempts" and some of these will be successful, 1, and some will not, 0.

- If Y_i for $i = 1, \ldots, n$ is iid binomial $B(n_i, p_i)$, then:

$$p(Y_i = y_i) = \binom{n_i}{y_i} p_i^{y_i} (1 - p_i)^{n_i - y_i}$$

- Further suppose that these are affected by the same k predictors (covariates, explanatory variables), $x_{i1}, \ldots, x_{ik}$.

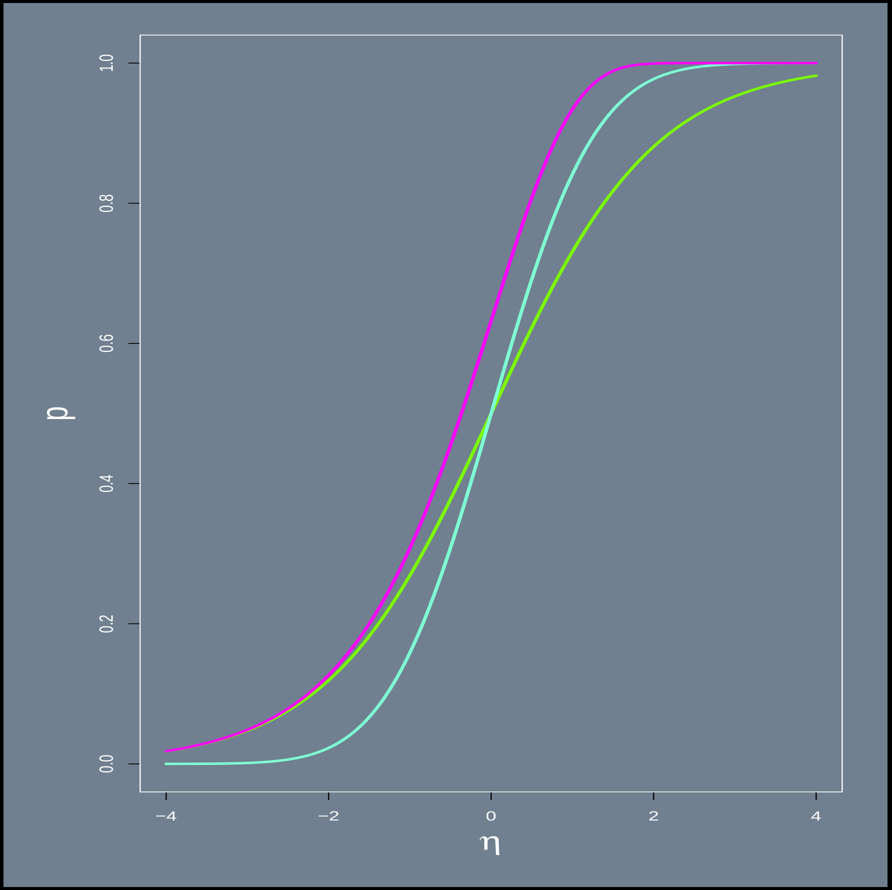
- The tool that connects these predictors to p is the linear predictor:

$$\eta_i = \beta_0 + \beta_1 x_{i1} + \ldots + \beta_k x_{ik}.$$

- We still need a link function, $\eta_i = g(p_i)$, that is not an identity $(\eta_i = p_i)$ since we need $0 \le p_i \le 1$.

Binomial Link Functions From Before

- Logit (logistic): $\eta = \log\left(\frac{p}{1-p}\right)$, $p = \frac{\exp(\eta)}{1+\exp(\eta)} [1 + \exp(-\eta)]$.
- Probit: $\eta = \Phi^{-1}(p)$, $p = \Phi(\eta)$.
- Complementary log-log:
  $\eta = \log(-\log(1-p))$,
  $p = 1 - \exp(-\exp(\eta))$.



```
ruler <- seq(-4,4,length=200)
postscript("faraway.ch2.fig3.ps")
par(col.axis="white",col.lab="white",col.sub="white",
    col="white", bg="slategray",cex.lab=2,mar=c(6,6,2,2))
plot(ruler,exp(ruler)/(1+exp(ruler)),type="l",lwd=3,
     col="lawngreen",ylim=c(0,1),
     xlab=expression(eta),ylab="p")
lines(ruler,pnorm(ruler),lwd=3,col="aquamarine")
lines(ruler,1-exp(-exp(ruler)),lwd=3,col="magenta")
dev.off()
```

Binomial Treatment of the Challenger Data

- Consider the Challenger data as a binomial experiment:

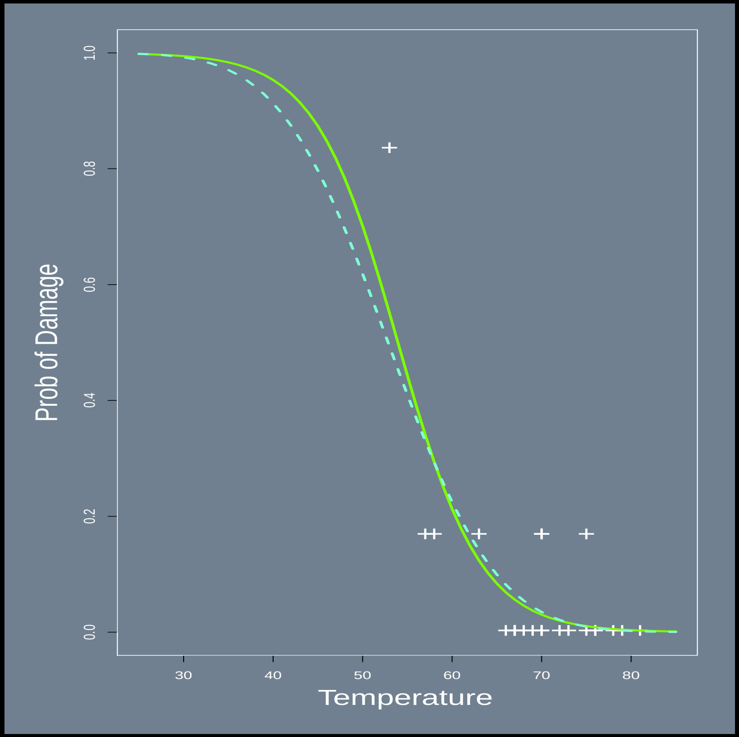
```
t(cbind(orings$damage,6-orings$damage))
     [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9] [,10] [,11] [,12] [,13] [,14]
[1,]    5    1    1    1    0    0    0    0    0     0     1     0     1     0
[2,]    1    5    5    5    6    6    6    6    6     6     5     6     5     6
     [,15] [,16] [,17] [,18] [,19] [,20] [,21] [,22] [,23]
[1,]     0     0     0     1     0     0     0     0     0
[2,]     6     6     6     5     6     6     6     6     6
```

- We will first model these as probabilities (something over six).

Challenger Disaster Example

- Now that we have a reasonable set of assumptions, the regression model is estimated with *maximum likelihood* (note the binomial treatment of the data):

```
logitmod <- glm(cbind(damage,6-damage) ~ temp, family=binomial, orings)
summary(logitmod)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)  11.6630     3.2963    3.54    4e-04
temp         -0.2162     0.0532   -4.07  4.8e-05

probitmod <- glm(cbind(damage,6-damage) ~ temp,
                 family=binomial(link=probit), data=orings)
summary(probitmod)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)   5.5915     1.7105    3.27   0.0011
temp         -0.1058     0.0266   -3.98  6.8e-05

postscript("faraway.ch2.fig4.ps")
par(col.axis="white",col.lab="white",col.sub="white",
    col="white", bg="slategray",cex.lab=2,mar=c(6,6,2,2))
plot(damage/6 ~ temp, orings, pch="+", xlim=c(25,85), ylim = c(0,1),
     xlab="Temperature", ylab="Prob of Damage", cex=2)
x <- seq(25,85,1)
lines(x,ilogit(11.6630-0.2162*x),col="lawngreen",lwd=3)
lines(x,pnorm(5.5915-0.1058*x),lty=2,lwd=3,col="aquamarine")
dev.off()
ilogit(11.6630-0.2162*31); pnorm(5.5915-0.1058*31)
```



Bernoulli Treatment of the Same Data

- Now modify the problem slightly where the criteria is “any damage at all, yes or no.”

```
orings2 <- orings; orings2[1,2] <- 1  # MOVE THE 5 VALUE TO 1
logitmod2 <- glm(damage ~ temp, family=binomial(link=logit), orings2)
summary(logitmod2)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)   15.043      7.379    2.04    0.041
temp          -0.232      0.108   -2.14    0.032

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 28.267  on 22  degrees of freedom
Residual deviance: 20.315  on 21  degrees of freedom
AIC: 24.32
```

- We could have also used:

```
family=binomial(link=probit)
family=binomial(link=cloglog)
```

Binomial Model Estimation

- Define a likelihood function for observed iid y_i, where $i = 1, \ldots, n$ from $f(y|p)$.

- Then the *joint distribution* of these observed data is:

$$p(y_1, y_2, \ldots, y_n) = p(y_1|\boldsymbol{\beta}, \mathbf{x}_1) f(y_2|\boldsymbol{\beta}, \mathbf{x}_2) \cdots f(y_n|\boldsymbol{\beta}, \mathbf{x}_n) = \prod_{i=1}^{n} f(y_i|\boldsymbol{\beta}, \mathbf{x}_i).$$

- If we consider that p is really the unknown and the y_i are known, then it makes sense to think of this joint function as a function that reveals something about $\boldsymbol{\beta}$.

- Denote it $L(\boldsymbol{\beta}|\mathbf{x}, \mathbf{y})$, which is called a likelihood function.

Binomial Model Estimation

- More precisely, we can incorporate the information that Y can only be 0 or 1:

$$
\begin{aligned}
L(\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}) &= \prod_{y_i=0}[1-F(\mathbf{X}_i\boldsymbol{\beta})]\prod_{y_i=1}[F(\mathbf{X}_i\boldsymbol{\beta})] \\
&= \prod_{i=1}^{n}[1-F(\mathbf{X}_i\boldsymbol{\beta})]^{1-y_i}[F(\mathbf{X}_i\boldsymbol{\beta})]^{y_i} \\
\ell(\boldsymbol{\beta}|\mathbf{X},\mathbf{Y}) &= \sum_{i=1}^{n}[(1-y_i)\log(1-F(\mathbf{X}_i\boldsymbol{\beta}))+y_i\log(F(\mathbf{X}_i\boldsymbol{\beta}))]
\end{aligned}
$$

- The log-likelihood is concave to the x-axis for common choices of $F()$, and produces coefficient estimates that are distributed student's-t.

- Generally with the binomial setup it is easier to think in terms of the CDF, $F()$, rather than the PDF, $f()$, since the former directly describes the S-curve of theoretical interest.

Binomial Model MLE

- The gradient is given by:

$$G = \frac{\partial}{\partial \boldsymbol{\beta}} \ell(\boldsymbol{\beta} | \mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{n} \left[\frac{y_i f_i}{F_i} + (1 - y_i) \frac{-f + i}{1 - F_i} \right] \mathbf{x}_i$$

- The Hessian is given by:

$$H = \frac{\partial^2}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'} \ell(\boldsymbol{\beta} | \mathbf{X}, \mathbf{Y}) = \sum_{i=1}^{n} \frac{f_i^2}{F_i(1 - F_i)} \mathbf{x}_i \mathbf{x}_i'$$

- The Variance-Covariance Matrix is calculated as:

$$VC_{\boldsymbol{\beta}} = E\left[-H^{-1} \right]$$

Common Forms

- Probit, where $\phi_i = \phi_i(\mathbf{x}_i\boldsymbol{\beta})$ and $\Phi_i = \Phi_i(\mathbf{x}_i\boldsymbol{\beta})$:

$$
\begin{aligned}
G &= \sum_{y=0} \frac{-\phi_i}{1-\Phi_i}\boldsymbol{\beta}\mathbf{x}_i + \sum_{y_i=1} \frac{\phi_i}{\Phi_i}\boldsymbol{\beta}\mathbf{x}_i \\
H &= \left\{ \sum_{i=0} \left[-\frac{-\phi_i^2}{(1-\Phi_i)^2} + \frac{\mathbf{x}_i\boldsymbol{\beta}\phi_i}{1-\Phi_i} \right] + \sum_{i=1} \left[-\frac{\mathbf{x}_i\boldsymbol{\beta}\phi_i}{\Phi_i} - \phi_i^2 \right] \right\} \mathbf{x}_i\mathbf{x}_i' \\
VC_{\boldsymbol{\beta}} &= \sum_{i=1}^{n} \frac{\phi_i^2}{\Phi_i(1-\Phi_1)}\mathbf{x}_i\mathbf{x}_i'
\end{aligned}
$$

- Logit, where $\Lambda_i = 1/[1+\exp(\mathbf{X}_i\boldsymbol{\beta})]$:

$$
G = \sum_{i=1}^{n} (y_i - \Lambda_i)\mathbf{x}_i \qquad\qquad H = \sum_{i=1}^{n} \left\{ -\Lambda_i(1-\Lambda_i) \right\} \mathbf{x}_i\mathbf{x}_i'
$$

$$
VC_{\boldsymbol{\beta}} = \left[\sum_{i=1}^{n} \left\{ \Lambda_i(1-\Lambda_i) \right\} \mathbf{x}_i\mathbf{x}_i' \right]^{-1}
$$

Interpretation of Individual Binomial $\boldsymbol{\beta}$ Results

- sign of the parameter estimate
- predicted/fitted values
- marginal effects, including first differences
- derivative methods
- Note $\text{logit}(\boldsymbol{\beta}) \approx \frac{\pi}{\sqrt{3}}\text{probit}(\boldsymbol{\beta}) \approx 1.8138$ (although some books say 1.6.
- Wald (t-tests) for significance:

$$W = (R\hat{\boldsymbol{\beta}} - q)\left[R(VC_{\hat{\boldsymbol{\beta}}})R'\right]^{-1}(R\hat{\boldsymbol{\beta}} - q)$$

  for $H_0\!: R\hat{\boldsymbol{\beta}} = q$ (commonly $R = 1, q = 0$, so that $W \sim F_{df=J,n-K}$. (where J is the number of restrictions stipulated in R). For individual coefficients, this reduces to:

$$W_k = (\hat{\boldsymbol{\beta}}'_k\hat{\boldsymbol{\beta}}_k/VC_{\hat{\boldsymbol{\beta}}}[k,k])^{\frac{1}{2}} \sim t_{df=n-k}$$

  (where $n \times k$ is the dimension of the $\mathbf{X}$ matrix).

- Note that the F-test is more robust than the t-test (Hauck-Donner effect, JASA 1977).

Example: Model of Vote Choice 1994 American National Election Study

	Parameter Estimate	Standard Error	z-statistic	p-value
Choice Parameters				
Intercept	-1.116	0.387	-2.882	0.004
Democratic Support for Clinton	-0.015	0.008	-1.943	0.052
Republican Support for Clinton	0.030	0.011	2.701	0.007
Democratic Crime Concern	0.044	0.009	4.960	0.000
Republican Crime Concern	0.007	0.009	0.699	0.485
Democratic Gvt. Help Disadv.	0.029	0.011	2.698	0.007
Republican Gvt. Help Disadv.	-0.006	0.013	-0.438	0.661
Democratic Gvt. Spending	0.114	0.025	4.633	0.000
Republican Gvt. Spending	-0.100	0.025	-4.030	0.000
Democratic Federal Healthcare	0.031	0.008	3.670	0.000
Republican Federal Healthcare	-0.017	0.010	-1.691	0.091
Democratic Ideology Entropy	0.104	0.131	0.794	0.427
Republican Ideology Entropy	0.303	0.068	4.437	0.000
Party Identification Scale	0.368	0.028	13.158	0.000

Goodness of Fit Test: $LRT = 359.3869, p < 0.0001$ for $\chi^2_{df=19}$
Percent Correctly Classified: 78.66% (using the "naive criteria")

- The outcome variable is candidate choice in the 1994 House election by party, Democrat 0, Republican 1.

Percent Predicted Correctly

- It would be nice to compare actual against predicted values in a 2-by-2 table:

		Prediction	
		0	1
Data	0	correct	incorrect
	1	incorrect	correct

- But wait! These models do not produce predicted 0/1 values, for instance from the 0/1 O-rings model:

```
round(logitmod2$fitted.values,3)
    1     2     3     4     5     6     7     8     9    10    11    12    13
0.939 0.859 0.829 0.603 0.430 0.375 0.375 0.375 0.322 0.274 0.230 0.230 0.230
   14    15    16    17    18    19    20    21    22    23
0.230 0.158 0.130 0.086 0.086 0.069 0.069 0.045 0.036 0.023
```

from the Bernoulli treatment.

Percent Predicted Correctly

- The naïve criteria:

$$p_i = 1 \text{ if, } F(\mathbf{x}_i\boldsymbol{\beta}) > 0.5 \qquad\qquad p_i = 0 \text{ if, } F(\mathbf{x}_i\boldsymbol{\beta}) < 0.5$$

- Create the table:

```
ppc <- cbind(orings2$damage, round(logitmod2$fitted.values,3))
( naive <- matrix(c(
    nrow(ppc[(ppc[,1] == 0) & (ppc[,2] < 0.5),])/nrow(ppc),
    nrow(ppc[(ppc[,1] == 0) & (ppc[,2] > 0.5),])/nrow(ppc),
    nrow(ppc[(ppc[,1] == 1) & (ppc[,2] < 0.5),])/nrow(ppc),
    nrow(ppc[(ppc[,1] == 1) & (ppc[,2] > 0.5),])/nrow(ppc)),
    byrow=TRUE,ncol=2) )

        [,1]    [,2]
[1,] 0.69565 0.00000
[2,] 0.13043 0.17391
```

- Better criteria: mean of $\hat{y}_i$, substantive/theoretical point.

Binomial Model Comparison

- Compare two models, one with ℓ parameters and one with s parameters such that $\ell > s$ and every parameter in the s set is also in the ℓ set: nesting.

- Denote the first as $L(p|\mathbf{y}, \mathbf{X}_L) = L_L$ and the second as $L(p|\mathbf{y}, \mathbf{X}_S) = L_S$.

- A tool for comparing these models is the likelihood ratio statistic:

$$LRT = 2\log\frac{L_L}{L_S} = 2(\log(L_L) - \log(L_S)) = -2\log\frac{L_S}{L_L} = -2(\log(L_S) - \log(L_L)).$$

- This test statistic is distributed asymptotically χ^2 with degrees of freedom the difference between the number of parameters in the two models.

- Tail values support the nesting model, meaning that the restricted in the nested model are not supported.

Binomial Model Comparison

- The most extreme case of L_L fits a “covariate” to every datapoint as an indicator function, and is thus a regression model where every datapoint is a separate inference.

- This is called the saturated model and provides no data-reduction and no modeling value, but serves as a reference point.

- For the binomial model, the saturated model can be described by $\hat{p}_i = y_i/n_i$, which is the number of success over the number of trials for the i th case (frequently $n_i = 1$).

- Another reference point is a model that uses β_0 only and is called a *mean model*.

- Thus any model we specify “lives” between these two extremes of model fit.

- Residuals in the nonlinear regression sense are called deviances to distinguish them from the assumptions in linear models.

- So a more general and more robust approach is to compare the summed deviances for each model of interest.

Binomial Model Comparison

- If we consider/run a single GLM model, then:

$$\sum D_{\text{saturated model}} < \sum D_{\text{our specified model}} < \sum D_{\text{mean model}}$$

- For the binomial model, the LRT reduces to a ratio of the saturated model to the specified model, given by:

$$D = 2\sum_{i=1}^{n}\{y_i \log(y_i/\hat{y}_i) + (n_i - y_i)\log((n_i - y_i)/(n_i - \hat{y}_i))\},$$

  where $\hat{y}_i$ are the fitted values from the smaller (specified) model.

- The mean model provides the largest value of D called the *null deviance*.

- D for assessing a model with p covariates is asymptotically distributed χ^2_{n-p}, where $n - p$ is the degrees of freedom.

- Returning the Challenger example ($n = 23$), I left off the following information before:

```
summary(logitmod)
    :
    Null deviance: 38.898  on 22  degrees of freedom
Residual deviance: 16.912  on 21  degrees of freedom
```

Binomial Model Comparison

- Formal tests:
  - Specified model versus saturated model:

    ```
    pchisq(deviance(logitmod),df.residual(logitmod),lower=FALSE)
    0.71641
    ```

    which is not in the χ^2_{21} tail, so it is statistically "close" to the saturated model and therefore a good fit.

  - Mean model versus saturated model:

    ```
    pchisq(38.9,22,lower=FALSE)
    0.014489
    ```

    which is in the χ^2_{22} tail, so it is statistically "far" from the saturated model and therefore not a good fit.

  - Specified model (with temperature) versus mean model ($D_S - D_L$):

    ```
    pchisq(38.9-16.9,1,lower=FALSE)
    2.7265e-06
    ```

    which is in the χ^2_{22} tail, so L_S is statistically "far" from L_L.

Binomial Model Comparison

- Cautions:
    - The approximation of D to a χ^2 distributed statistic is poor for small n_i and "lumpy" distribution of n_i as well.
    - Most texts recommend $n_i \geq 5, \ \forall i$, but this is just an arbitrary choice.
    - We could also have done a Wald test on temperature:

      ```
                   Estimate Std. Error z value Pr(>|z|)
      temp          -0.2162     0.0532   -4.07  4.8e-05
      ```

      but differences of deviances are usually more accurate than tests on a single deviance.
    - It is possible that a Wald test provides significant results but a deviance comparison doesn't (the Hauck-Donner effect).

Binomial Model Comparison

- Confidence interval for the j th coefficient: $\hat{\beta}_j \pm z^{\alpha/2} se(\hat{\beta}_j)$.

- Low-tech method (for just the covariate):

```
c(-0.2162-1.96*0.0532,-0.2162+1.96*0.0532)
-0.32047 -0.11193
```

- Hi-tech method:

```
summary(logitmod)$coefficients[,1]
        + qnorm(0.975) * t(c(-1,1) %o% summary(logitmod)$coefficients[,2])
(Intercept)  5.20243 18.12355
temp        -0.32046 -0.11201
```

- Profile likelihood version (accounts for covariance):

```
library(MASS)
confint(logitmod)
Waiting for profiling to be done...
               2.5 %   97.5 %
(Intercept)  5.57520 18.73760
temp        -0.33266 -0.12018
```

Example: Anaemia

- Consider again a study of anaemia in women in a given clinic where 20 cases are chosen at random from the full study to get the data here.

- From a blood sample we get:

  - haemoglobin level (Hb) in grams per deciliter (12–15 g/dl is normal in adult females)

  - packed cell volume (PCV) in percent of blood volume that is occupied by red blood cells (also called hematocrit, Ht or HCT, or erythrocyte volume fraction, EVF). 38% to 46% is normal in adult females.

- We also have:

  - age in years

  - menopausal (0=no, 1=yes)

- There is an obvious endogeneity problem in modeling Hb(g/dl) versus PCV(%).

Anaemia Data

- First some house-keeping: note that the logit and inverse logit function are not in the `R` base package, although many package writers include them.

- So you might want to have these handy:

```
ilogit <- inv.logit <- function(Xb)  1/(1+exp(-Xb))
logit <- function(mu)  log(mu/(1-mu))
```

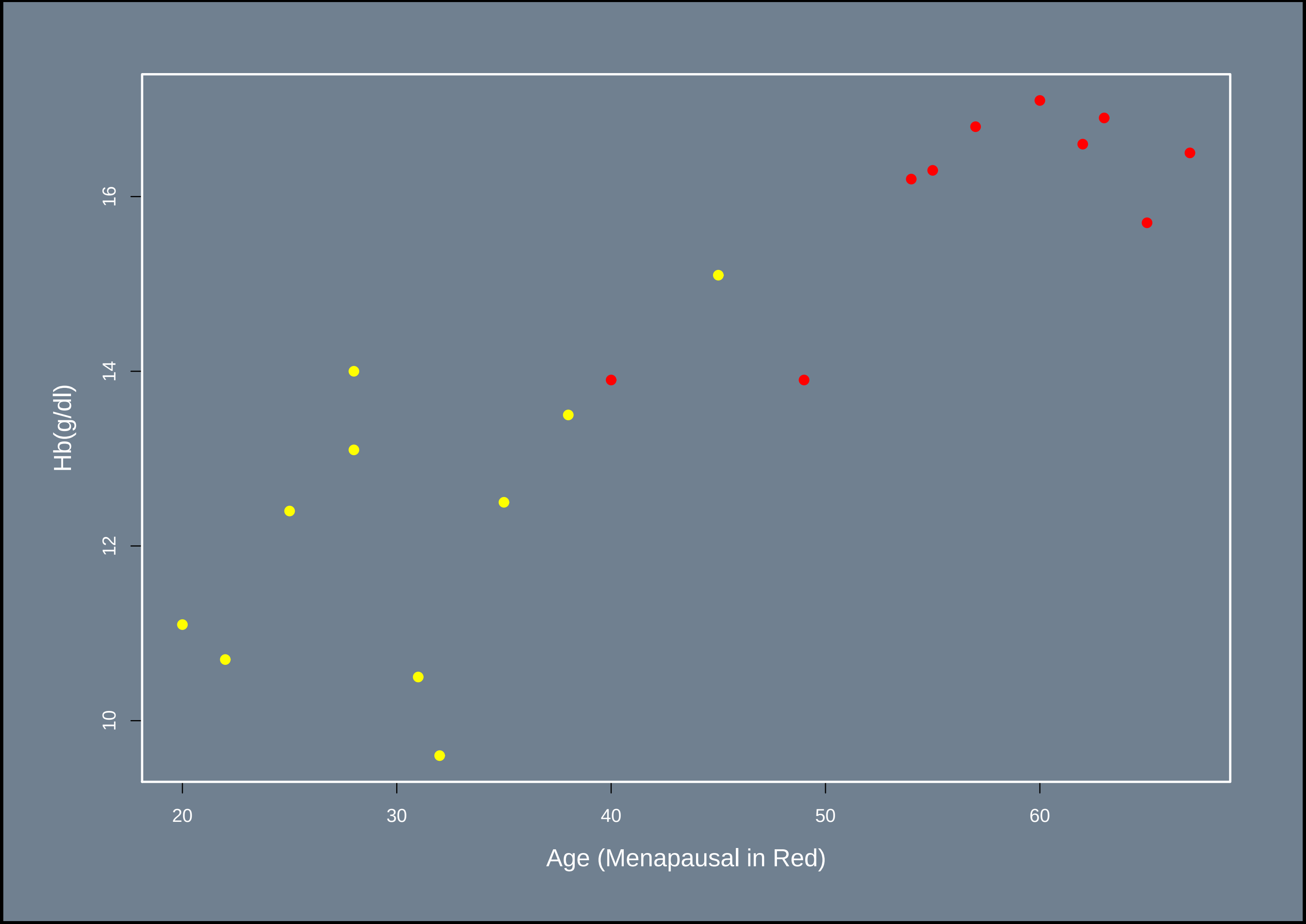
- Load the anaemia data:

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

Anaemia Data

Subject	Hb(g/dl)	PCV(%)	Age	Menopausal
1	11.1	35	20	0
2	10.7	45	22	0
3	12.4	47	25	0
4	14.0	50	28	0
5	13.1	31	28	0
6	10.5	30	31	0
7	9.6	25	32	0
8	12.5	33	35	0
9	13.5	35	38	0
10	13.9	40	40	1
11	15.1	45	45	0
12	13.9	47	49	1
13	16.2	49	54	1
14	16.3	42	55	1
15	16.8	40	57	1
16	17.1	50	60	1
17	16.6	46	62	1
18	16.9	55	63	1
19	15.7	42	65	1
20	16.5	46	67	1

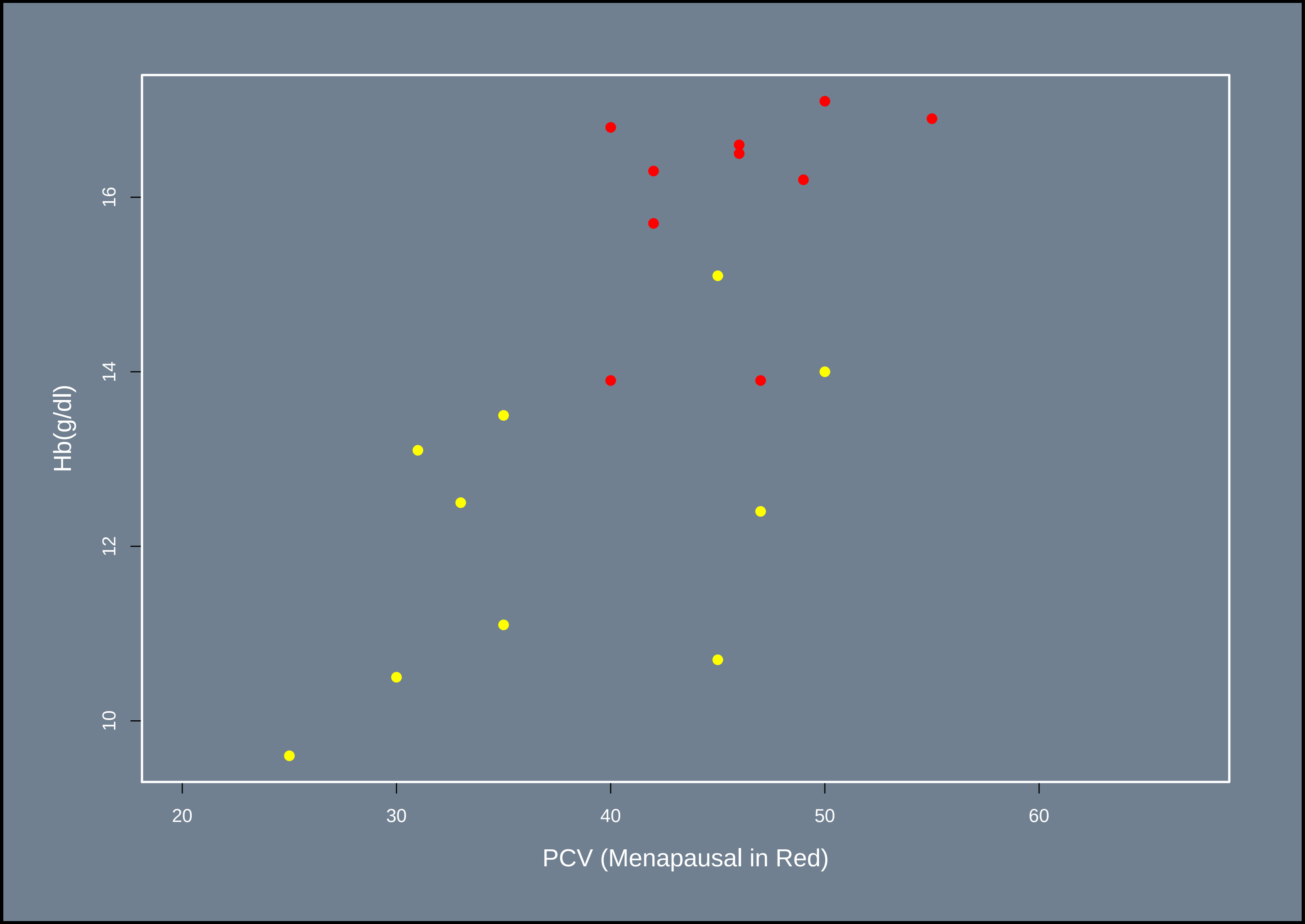
Scatterplot of the Anaemia Data



Scatterplot of the Anaemia Data

```
postscript("anaemia1.fig.ps")
par(mfrow=c(1,1),mar=c(5,5,2,2),lwd=2,col.axis="white",col.lab="white",
        col.sub="white", col="white",bg="slategray", cex.lab=1.3)
plot(anaemia$Age[anaemia$Menapause==0],anaemia$Hb[anaemia$Menapause==0],
        pch=19,col="yellow",
        xlim=range(anaemia$Age),ylim=range(anaemia$Hb),
        xlab="Age (Menapausal in Red)",ylab="Hb(g/dl)")
points(anaemia$Age[anaemia$Menapause==1],anaemia$Hb[anaemia$Menapause==1],
        pch=19,col="red")
dev.off()
```

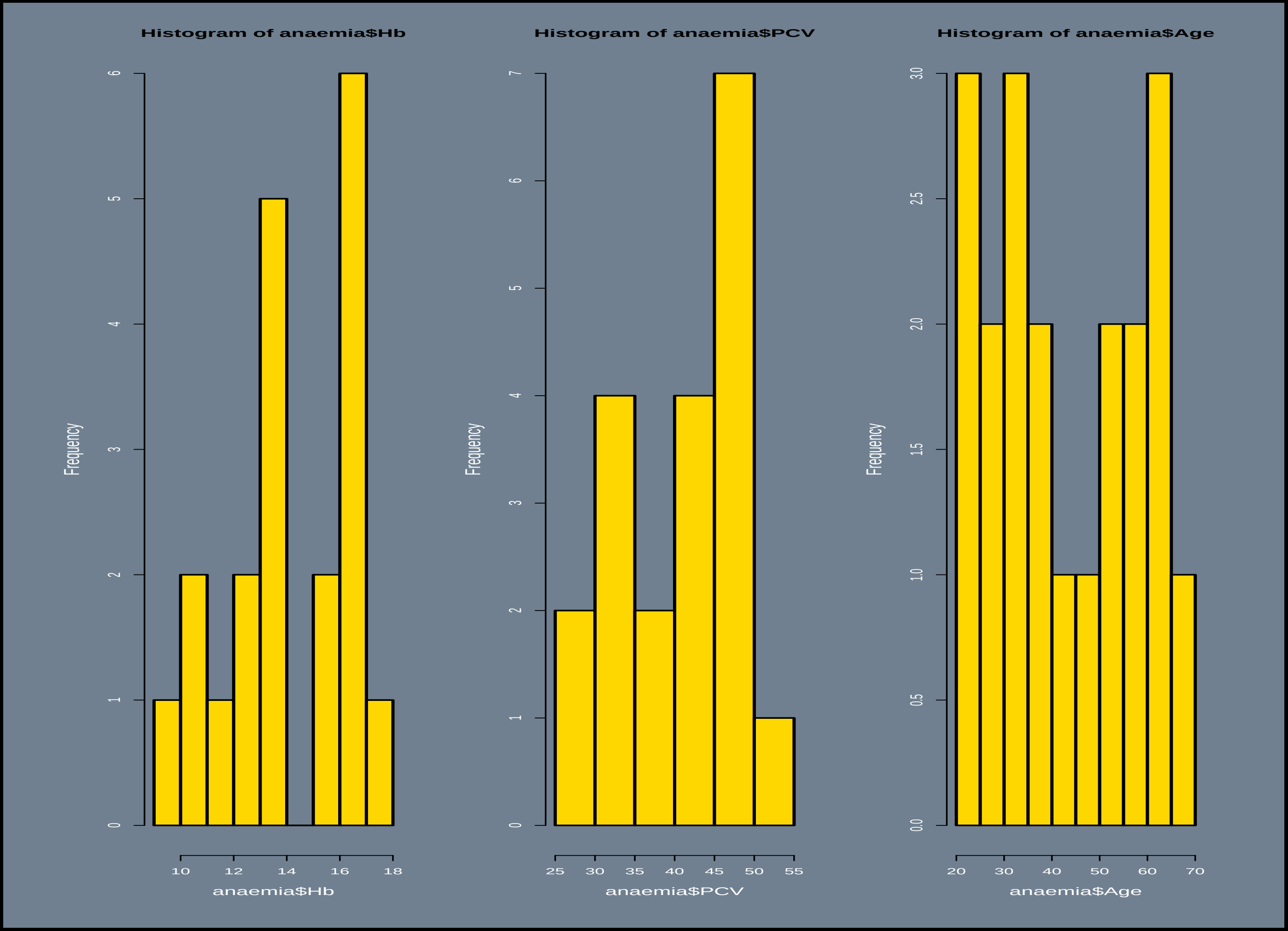
Scatterplot of the Anaemia Data



Scatterplot of the Anaemia Data

```
postscript("anaemia2.fig.ps")
par(mfrow=c(1,1),mar=c(5,5,2,2),lwd=2,col.axis="white",col.lab="white",
        col.sub="white",col="white",bg="slategray", cex.lab=1.3)
plot(anaemia$PCV[anaemia$Menapause==0],anaemia$Hb[anaemia$Menapause==0],
        pch=19,col="yellow",
        xlim=range(anaemia$Age),ylim=range(anaemia$Hb),
        xlab="PCV (Menapausal in Red)",ylab="Hb(g/dl)")
points(anaemia$PCV[anaemia$Menapause==1],anaemia$Hb[anaemia$Menapause==1],
        pch=19,col="red")
dev.off()
```

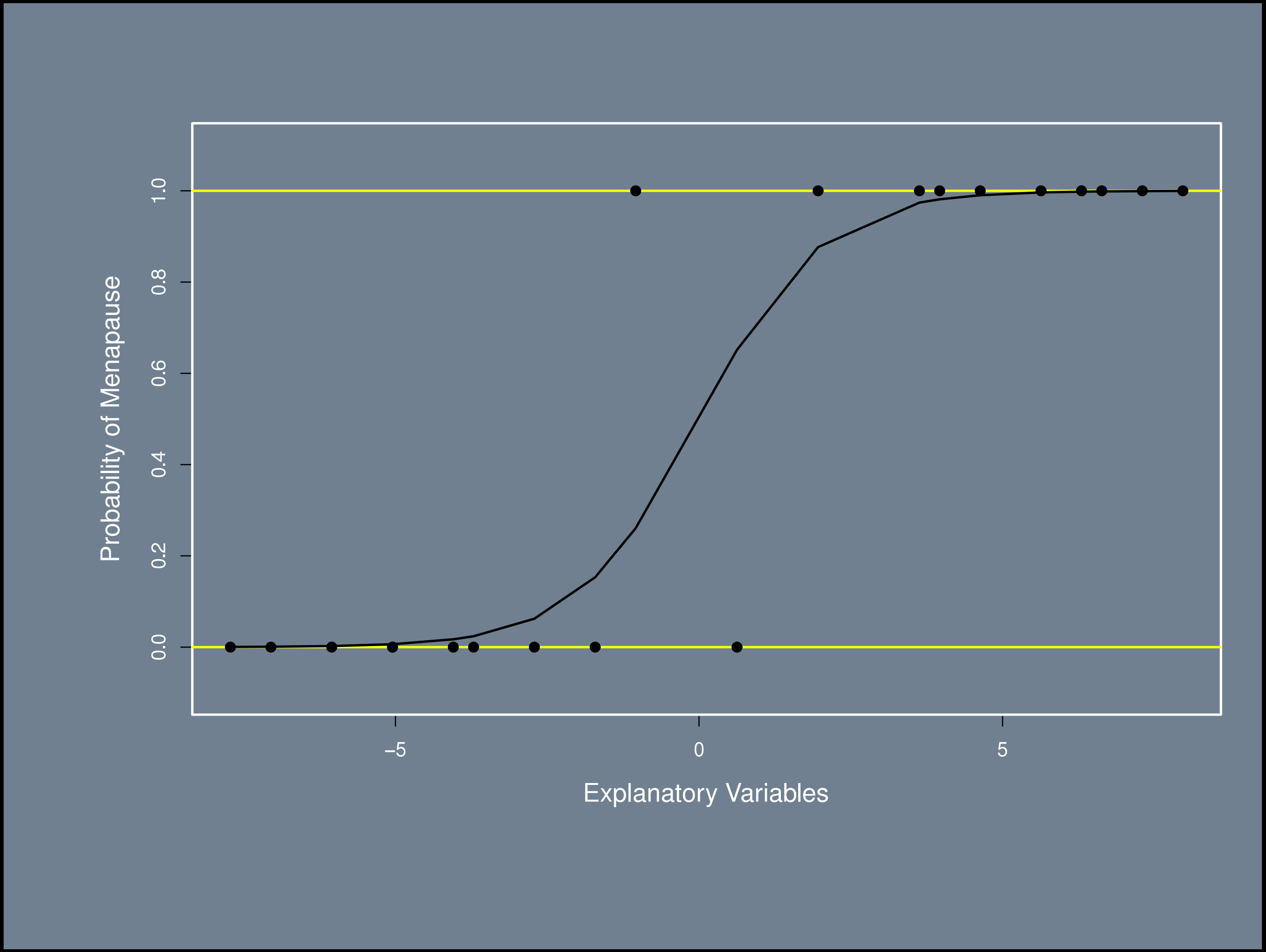
Distribution of the Anaemia Data?



Logistic Regression: Anaemia Example

```
summary( glm(Menapause~Age, data=anaemia, family=binomial(link=logit)) )
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.45227  -0.13139  -0.00176   0.09818   1.63990

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -14.395      7.462   -1.93    0.054
Age            0.334      0.174    1.92    0.055

    Null deviance: 27.7259  on 19  degrees of freedom
Residual deviance:  5.7632  on 18  degrees of freedom
```

Logistic Empirical Illustration



Logistic Empirical Illustration

```
ana.logit <- glm(Menapause ~ Age, data=anaemia, family=binomial(link=logit))
postscript("/Users/jgill/CLASSES/Class.MLE/Images/logit.anaemia1.fig.ps")
par(mfrow=c(1,1),mar=c(5,5,2,2),lwd=2,col.axis="white",col.lab="white",
    col.sub="white",col="white",bg="slategray", cex.lab=1.3,oma=c(4,2,2,2))
xbeta <- as.matrix(cbind(rep(1,length=nrow(anaemia)),anaemia$Age)) %*% coef(ana.logit)
plot(range(xbeta),c(-0.1,1.1),type="n",xlab="Explanatory Variables",
    ylab="Probability of Menapause")
abline(h=1,col="red"); abline(h=0,col="yellow")
x <- seq(from=min(xbeta),to=max(xbeta),length=100)
points(xbeta,anaemia$Menapause,col="black",pch=19)
lines(xbeta,ilogit(xbeta),col="black")
dev.off()
```

Logit Model for Survey Responses in Scotland

- These data come from the British General Election Study, Scottish Election Survey, 1997 (ICPSR Study Number 2617).

- These data contain 880 valid cases, each from an interview with a Scottish national after the election.

- Our outcome variable of interest is their party choice in the UK general election for Parliament where we collapse all non-Conservative party choices (abstention, Labour, Liberal Democrat, Scottish National, Plaid Cymru, Green, Other, Referendum) to one category, which produces 104 Conservative votes.

- For probit, $\sigma^2 = 1$ to establish the scale and provide an intuitive (standard) probit metric.

Logit Model for Survey Responses in Scotland, Explanatory Variables

- `POLITICS`, which asks how much interest the respondent has in political events (increasing scale: none at all, not very much, some, quite a lot, a great deal).
- `READPAP`, which asks about daily morning reading of the newspapers (yes=1 or no=0).
- `PTYTHNK`, how strong that party affiliation is for the respondent (categorical by party name).
- `IDSTRNG` (increasing scale: not very strong, fairly strong, very strong).
- `TAXLESS` asks if “it would be better if everyone paid less tax and had to pay more towards their own healthcare, schools and the like” (measured on a five point increasing Likert scale).
- `DEATHPEN` asks whether the UK should bring back the death penalty ((measured on a five point increasing Likert scale).
- `LORDS` queries whether the House of Lords should be reformed (asked as *remain as is* coded as zero and *change is needed* coded as one).
- `SCENGBEN` asks how economic benefits are distributed between England and Scotland with the choices: England benefits more $= -1$, neither/both lose $= 0$, Scotland benefits more $= 1$.

Logit Model for Survey Responses in Scotland, Explanatory Variables

- INDPAR asks which of the following represents the respondent's view on the role of the Scottish government in light of the new parliament: (1) Scotland should become independent, separate from the UK and the European Union, (2) Scotland should become independent, separate from the UK but part of the European Union, (3) Scotland should remain part of the UK, with its own elected parliament which has some taxation powers, (4) Scotland should remain part of the UK, with its own elected parliament which has no taxation powers, and (5) Scotland should remain part of the UK without an elected parliament.

- SCOTPREF1 asks "should there be a Scottish parliament within the UK? (yes=1, no=0).

- RSEX, the respondent's sex.

- RAGE, the respondent's age.

- RSOCCLA2, the respondents social class (7 category ascending scale).

- TENURE1, whether the respondent rents (0) or owns (1) their household.

- PRESBm a categorical variable for church affiliation, measurement of religion is collapsed down to one for the dominant historical religion of Scotland (Church of Scotland/Presbyterian) and zero otherwise and designated

Logit Model for Survey Responses in Scotland

- Run a probit model for the conservative/not-conservative outcome with these covariates:
- Results give across two slides...

```
scot.mat <- read.table("https://jeffgill.org/files/jeffgill/files/scotland.dat_.txt",
            sep=",",header=TRUE)
Y        <- as.numeric(scot.mat[,1])
X        <- as.matrix(scot.mat[,2:ncol(scot.mat)])
glm.out  <- glm(Y ~ X, family=binomial(link=probit))
```

Logit Model for Survey Responses in Scotland, Results (not in order)

```
summary(glm.out)

Call:
glm(formula = Y ~ X[, -1], family = binomial(link = probit))

Deviance Residuals:
   Min       1Q  Median      3Q     Max
-2.223  -0.287  -0.120  -0.022   3.598

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 639.38  on 879  degrees of freedom
Residual deviance: 338.98  on 864  degrees of freedom
AIC: 371

Number of Fisher Scoring iterations: 8
```

Logit Model for Survey Responses in Scotland, Results (not in order)

```
Coefficients:
                Estimate Std. Error z value Pr(>|z|)
(Intercept)      -0.8032     0.5655   -1.42   0.1555
X[, -1]POLITICS   0.1999     0.0777    2.57   0.0101
X[, -1]READPAP    0.2626     0.1840    1.43   0.1536
X[, -1]PTYTHNK   -0.5765     0.0928   -6.21  5.3e-10
X[, -1]IDSTRNG    0.2114     0.0775    2.73   0.0064
X[, -1]TAXLESS    0.1059     0.0736    1.44   0.1501
X[, -1]DEATHPEN   0.0817     0.0578    1.41   0.1573
X[, -1]LORDS     -0.4267     0.1597   -2.67   0.0075
X[, -1]SCENGBEN   0.3279     0.1107    2.96   0.0031
X[, -1]SCOPREF1  -0.9728     0.1889   -5.15  2.6e-07
X[, -1]RSEX       0.3785     0.1712    2.21   0.0270
X[, -1]RAGE       0.0118     0.0043    2.74   0.0062
X[, -1]RSOCCLA2  -0.1218     0.0582   -2.09   0.0363
X[, -1]TENURE1    0.4634     0.1808    2.56   0.0104
X[, -1]PRESB     -0.1417     0.1675   -0.85   0.3975
X[, -1]IND.PAR    0.2500     0.1925    1.30   0.1940
```

Percent Predicted Correctly

```
scot.pred <- scot.out$fitted.values
scot.pred[scot.pred < 0.5] <- 0
scot.pred[scot.pred > 0.5] <- 1
table(scot.pred,scot.mat$VOTE)

scot.pred   0   1
        0 750  50
        1  26  54

sum(diag(table(scot.pred,scot.mat$VOTE)))/nrow(scot.mat)
[1] 0.91364
```

Percent Predicted Correctly

```
mean(scot.pred)
[1] 0.09091
scot.pred <- scot.out$fitted.values
scot.pred[scot.pred < mean(scot.pred)] <- 0
scot.pred[scot.pred > mean(scot.pred)] <- 1
table(scot.pred,scot.mat$VOTE)

scot.pred   0   1
        0 663  11
        1 113  93

sum(diag(table(scot.pred,scot.mat$VOTE)))/nrow(scot.mat)
[1] 0.85909
```

Model Using 2012 ANES Data

- The dichotomous outcome is voting for Mitt Romney (1) rather than Barrack Obama (0).
- The selected explanatory variables are the typical demographic questions along with some issues topical to this particular presidential election.
- `gender_respondent_x`, with 0 for male and 1 for female.
- `dem_birthyr`, the respondent's year of birth.
- `dem_racecps_black`, respondent self-identifies as black.
- `dem_hisp`, respondent self-identifies as Hispanic
- `pid_x`, a seven-point party identification variable going from Democrat to Republican
- `libcpre_self`, an ideology scale going from liberal to conservative,
- `dem_edugroup_x`, a dichotomous indicate of high school graduate (1) or not (0).
- `candrel_dpc`, whether the respondent thinks that Obama is Muslim (20% said yes).
- `hlthlaw_qual`, for evaluation of the Affordable Care Act.

Model Using 2012 ANES Data

- The variable `libcpre_self` is recoded as a factor:

```
libcpre_self <- factor(libcpre_self)
contrasts(libcpre_self) <- contr.treatment(n=7,base=4)
```

- The variable `libcpre_self`, which also has seven categories, is estimated as a non-factor and this imposes the assumption that the categories are actually even-spaced on the underlying latent metric.

- The model statement is:

```
prez.out.anes <- glm(presvote2012_x ~ gender_respondent_x + dem_birthyr
    + dem_racecps_black + dem_hisp + pid_x + libcpre_self + dem_edugroup_x
    + candrel_dpc + hlthlaw_qual,
    family=binomial(link=probit), data=current.anes.df)
```

Model Using 2012 ANES Data, Results

```
summary(prez.out.anes)
                     Estimate Std. Error  t value Pr(>|t|) exp(beta) 95% Lower CI 95% Upper CI
(Intercept)           -2.5766     0.4009  -6.4267   0.0000    0.0760       0.0393       0.1470
gender_respondent_x    0.0738     0.1319   0.5593   0.2934    1.0765       0.8666       1.3373
dem_birthyr            0.0001     0.0003   0.4841   0.3195    1.0001       0.9997       1.0006
dem_racecps_black     -1.3708     0.2422  -5.6590   0.0000    0.2539       0.1705       0.3782
dem_hisp              -0.3071     0.0857  -3.5821   0.0019    0.7356       0.6388       0.8470
pid_x1                -3.1297     0.2966 -10.5530   0.0000    0.0437       0.0268       0.0712
pid_x2                -1.4570     0.1808  -8.0580   0.0000    0.2329       0.1730       0.3136
pid_x3                -1.3895     0.1924  -7.2217   0.0000    0.2492       0.1816       0.3420
pid_x5                 1.2741     0.2321   5.4886   0.0001    3.5756       2.4407       5.2384
pid_x6                 1.6826     0.2369   7.1030   0.0000    5.3796       3.6435       7.9431
pid_x7                 2.7143     0.2600  10.4397   0.0000   15.0942       9.8415      23.1504
libcpre_self           0.4216     0.0623   6.7651   0.0000    1.5244       1.3759       1.6890
dem_edugroup_x         0.4315     0.2052   2.1029   0.0291    1.5396       1.0985       2.1578
candrel_dpc            1.0963     0.1723   6.3631   0.0000    2.9932       2.2544       3.9740
hlthlaw_qualImproved  -1.2624     0.2026  -6.2297   0.0000    0.2830       0.2028       0.3949
hlthlaw_qualWorsened   0.9918     0.1366   7.2600   0.0000    2.6959       2.1534       3.3753

Sample size: 5914
    Null deviance: 7920.18 on 5913 degrees of freedom
Residual deviance: 2907.18 on 5898 degrees of freedom
AIC: 2939.176
```

Model Using 2012 ANES Data, Results

- All but two of the coefficient estimates are statistically reliable at the conventional 0.05 level.
- Signs and magnitudes are interesting.
- This summary also includes the odds interpretation of the coefficients, which is calculated as:

$$\exp(\beta_k) = \mathbb{E}\left[\Delta \frac{\frac{\pi_1}{1-\pi_1}}{\frac{\pi_0}{1-\pi_0}}\right]$$

  which is the expected change in odds of success relative to odds of failure (odds ratio) when we increase X_k by one unit.
- Finally, notice the dramatic decrease in the deviance from the null model to the specified model. This also is a result of the sample size. The AIC (Akaike Information Criterion) is explained below.
- The sample size is 5,913.
- This leads to a discussion of odds. . .

Tabular Analysis of Binary Outcomes

- Binary outcomes are often called *events*, meaning they either happened or didn't.
- Usually these are labeled 0 and 1, where the one denotes "happened."
- Sometimes the 1 is called a "success."
- These are only labels and switching the assignment never changes the construction or reliability of the statistical model.
- Tables of events have a very specific construction:

2 × 2 Contingency Table

	Experimental-Manipulation		
Outcome	Treatment	Control	Row Total
Positive	a	b	$a+b$
Negative	c	d	$c+d$
Column Total	$a+c$	$b+d$	

- Hypothesized relationships are usually down the primary diagonal of the table.

Odds and Odds Ratios

- Odds of an event is the ratio of the probability of an event *happening* to the probability of the event *not happening*:

$$Odds = \frac{p}{1-p},$$

  where p is the probability of the event.

- Odds Ratio compares the odds of an event under treatment to odds under control:

$$OR = \frac{\left(\frac{p_T}{1-P_T}\right)}{\left(\frac{P_C}{1-P_C}\right)} = \frac{\frac{\frac{a}{a+c}}{1-\frac{a}{a+c}}}{\frac{\frac{b}{b+d}}{1-\frac{b}{b+d}}} = \frac{\frac{\frac{a}{a+c}}{\frac{a+c}{a+c}-\frac{a}{a+c}}}{\frac{\frac{b}{b+d}}{\frac{b+d}{b+d}-\frac{b}{b+d}}} = \frac{\left(\frac{a}{c}\right)}{\left(\frac{b}{d}\right)} = \frac{ad}{bc}.$$

- For rare events, the odds and probability are close since $a \ll c$, so $a/c \approx a/(a+c)$, and the OR is close to the RR ($RR \approx \frac{p_T}{p_C}$).

- Nicely, the OR for failure is just the inverse of the OR for success (symmetry).

Interpreting Odds

- Some people prefer to think in terms of *odds* rather than probability:

  $$o = \frac{p}{1-p} = \frac{p(y=1)}{p(y=0)} \qquad\qquad p = \frac{o}{1+o}$$

  where 0 is obviously on the support $(0:\infty)$.

- This is essentially how logit works since:

  $$\log(\text{odds}) = \log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2.$$

- So if x_2 is held constant, then a one-unit change in x_1 gives a β_1 change in the log-odds of success (or a $\exp(\beta_1)$ change in the odds).

- Relatedly, if p_1 is the probability of success under condition 1 and p_2 is the probability of success under condition 2, then the relative risk is simply:

  $$RR = \frac{p_1}{p_2}$$

Example: Cohort Study of Adolescents

- A random sample of size 2437, asking about cannabis and psychotic symptoms up to 4 years later(!).

- Summary table (Henquet, et al. 2005):

Cannabis Use and Psychosis

	Cannabis	*No Cannabis*	Total
Event	82	342	424
No Event	238	1775	2013
Total	320	2117	2437

- Thus the odds ratio for psychosis is:

$$OR = \frac{ad}{bc} = \frac{82 \times 1775}{342 \times 238} = 1.79.$$

- Since psychosis is a relatively rare event, this close to the relative risk:

$$RR = \frac{p_T}{p_C} = \frac{\left(\frac{82}{320}\right)}{\left(\frac{342}{2117}\right)} = 1.59.$$

Interpreting Odds, Respiratory Disease

- Respiratory Disease in < 1 year-olds:

```
library(MASS); data(babyfood)
xtabs(disease/(disease+nondisease)~sex+food,babyfood)

       Bottle   Breast    Suppl
Boy  0.168122 0.095142 0.129252
Girl 0.125000 0.066810 0.125984

mdl <- glm(cbind(disease,nondisease) ~ sex + food, family=binomial,babyfood)
summary(mdl)

             Estimate Std. Error z value Pr(>|z|)
(Intercept)   -1.613      0.112  -14.35  < 2e-16
sexGirl       -0.313      0.141   -2.22    0.027
foodBreast    -0.669      0.153   -4.37  1.2e-05
foodSuppl     -0.173      0.206   -0.84    0.401

Null deviance: 26.37529  on 5  degrees of freedom
Residual deviance:  0.72192  on 2  degrees of freedom
```

Interpreting Odds, Respiratory Disease

- The interaction model is the saturated model for these data since $k - 1$ degrees of freedom gets consumed by 1 sex and 2 food categories.

- A deviance (not Wald) test for each of the main effects relative to the full is done with:

  ```
  drop1(mdl,test="Chi")

  Single term deletions
  Model:
  cbind(disease, nondisease) ~ sex + food
         Df Deviance  AIC  LRT Pr(Chi)
  <none>          0.7 40.2
  sex     1       5.7 43.2  5.0   0.026
  food    2      20.9 56.4 20.2 4.2e-05
  ```

  where the LRTs show strong evidence for inclusion.

Interpreting Odds, Respiratory Disease

- Coefficient interpretations:
  - `foodBreast -0.669`, so $\exp(-0.669) = 0.51222$, meaning that breast feeding reduces the odds of respiratory disease to 51% of bottle only feeding (the reference).
  - Computing a confidence interval on the log-odds scale (better coverage properties for categorical variables):

    ```
    exp(c(-0.669-1.96*0.153,-0.669+1.96*0.153))
    0.37951 0.69134
    ```

    or:

    ```
    library(MASS); exp(confint(mdl))
    Waiting for profiling to be done...
                  2.5 %  97.5 %
    (Intercept) 0.15920 0.24743
    sexGirl     0.55362 0.96292
    foodBreast  0.37819 0.68952
    foodSuppl   0.55554 1.24643
    ```

More Measures of Goodness of Fit

- Pearson's X^2 is intended to look like the Sum of Squares Error:

$$X^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

  and for successes: $E_i = n_i\hat{p}_i \; O_i = y_i$, but for failures $E_i = n_i(1 - \hat{p}_i) \; O_i = n_i - y_i$, giving:

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

  courtesy of some unpleasant algebra.

- We can also define an individual Pearson residual:

$$r_i^p = (y_i - n_i\hat{p}_i)/\sqrt{\text{var}(y_i)}$$

  which means that:

$$X^2 = \sum_{i=1}^{n} (r_i^p)^2$$

More Measures of Goodness of Fit

- New dataset, insects dying at differing levels of insecticide concentration.

```
data(bliss)
bliss

  dead alive conc
1    2    28    0
2    8    22    1
3   15    15    2
4   23     7    3
5   27     3    4

bliss.out <- glm(cbind(dead,alive) ~ conc, family=binomial, data=bliss)
sum(residuals(bliss.out,type="pearson")^2)

0.36727

deviance(bliss.out)

0.37875
```

Proportion of Deviance Explained

- Meant to be like the R^2 measure for linear models (Nagelkerke 1991).
- Definition:

$$R^2 = \frac{1-(\hat{L}_0/\hat{L})^{2/n}}{1-\hat{L}_0} = \frac{1-\exp((D-D_{null})/n)}{1-\exp(D_{null}/n)}$$

  where $\hat{L}_0$ is the maximized likelihood under the null.

- Implementation:

```
(1-exp((modl$dev-bliss.out$null)/150))/(1-exp(-bliss.out$null/150))

0.99532
```

Prediction and Effective Doses

- We want to predict an outcome, in this case the probability of success, for levels of the explanatory variables: $g^{-1}(\hat{\eta}) = g^{-1}(x_0\hat{\beta})$.

- Returning to the insect data, predict the response at a dose of 2.5:

```
bliss.out <- glm(cbind(dead,alive) ~ conc, family=binomial,data=bliss)
lmodsum <- summary(bliss.out)
x0 <- c(1,2.5)
( eta0 <- sum(x0*coef(bliss.out)) )

0.58095

ilogit(eta0)

0.64129
```

meaning that 64% are predicted die at this level.

Prediction and Effective Doses

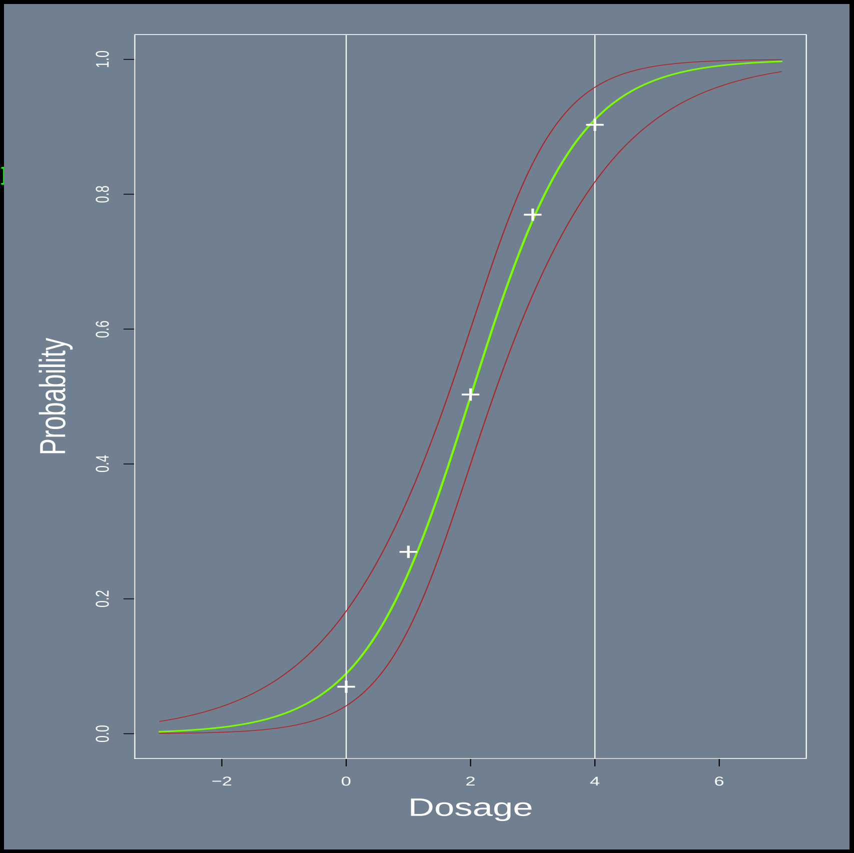
- We also want a measure of uncertainty around this prediction in the form of a 95% CI:

```
( cm <- lmodsum$cov.unscaled )

            (Intercept)      conc
(Intercept)    0.174630 -0.065823
conc          -0.065823  0.032912

( se <- sqrt( t(x0) %*% cm %*% x0) )

0.2263

ilogit(c(eta0-1.96*se,eta0+1.96*se))

0.53430 0.7358
```

Confidence Bands for Effective Doses

- Here is a better tool. . .

```
ruler <- seq(-3,7,length=200)
predicts <- predict(bliss.out,newdata=data.frame(conc=ruler),se=TRU

ci <- cbind( ilogit(predicts$fit-qnorm(0.975)*predicts$se.fit),
             ilogit(predicts$fit+qnorm(0.975)*predicts$se.fit) )

postscript("faraway.ch2.fig5.ps")
par(col.axis="white",col.lab="white",col.sub="white",
    col="white", bg="slategray",cex.lab=2,mar=c(6,6,2,2))
plot(ruler,ilogit(predicts$fit),xlab="Dosage",ylab="Probability",
     type="l",lwd=2, col="lawngreen")
abline(v=c(0,4),col="white")
lines(ruler,ci[,1],col="firebrick")
lines(ruler,ci[,2],col="firebrick")
points(0:4,bliss$dead/(bliss$alive+bliss$dead),pch="+",cex=2)
dev.off()
```



- Where the indicated points are from the original data values (5 concentrations).

LD50 Calculations

- Sometimes we would like to go backwards: what levels of x product a certain probability?

- One common question: what is the effective dose required to get a prediction of 50% killed (LD50)?

- For a logit link this is just:

$$\begin{aligned} p(y=1|x) = \frac{1}{2} &= [1+\exp(-\beta_0 - \beta_1 x)]^{-1} \\ 2 &= 1+\exp(-\beta_0 - \beta_1 x) \\ 0 &= -\beta_0 - \beta_1 x \\ \widehat{\text{LD50}} &= -\hat{\beta}_0/\hat{\beta}_1 \end{aligned}$$

- Returning to the Bliss data:

```
(ld50 <- -bliss.out$coef[1]/bliss.out$coef[2])
(Intercept)
          2
```

LD50 For Calculations

- The variance of a function of a random variable $\hat{\theta}$ can often be obtained by the *delta method*:

$$\mathrm{Var} g(\hat{\theta}) \approxeq g'(\hat{\theta})\mathrm{Var}(\hat{\theta})g'(\hat{\theta})$$

- Here:

$$\mathrm{ld50} = \hat{\theta}.$$

- So:

$$\frac{d}{d\hat{\beta}_1} g(\hat{\theta}) = \frac{d}{d\hat{\beta}_1}(-\hat{\beta}_0/\hat{\beta}_1) = -1/\hat{\beta}_2$$

  and:

$$\frac{d}{d\hat{\beta}_2} g(\hat{\theta}) = \frac{d}{d\hat{\beta}_2}(-\hat{\beta}_0/\hat{\beta}_1) = \hat{\beta}_0/\hat{\beta}_2$$

- Executing:

```
dr <- c(-1/bliss.out$coef[2],bliss.out$coef[1]/bliss.out$coef[2]^2)
( sqrt(dr %*% lmodsum$cov.unscaled %*% dr)[,] )
[1] 0.17844
```

Overdispersion in Dichotomous Choice Models

- If we meet the described assumptions, then the two times the residual (summed) deviance is approximately χ^2 with $n-p$ degrees of freedom.
- However, sometimes we are in the tail of this distribution not because we have chosen the wrong explanatory variables, but because of:
  - outliers,
  - sparse data,
  - overdispersion: $\text{Var}(Y) \gg mp(1-p)$, where m is the size of the binomial trial group (often denoted n_i when there are differences).
- Underdispersion is rare.
- Typical causes of overdispersion:
  - variation in p across binomial trials (violates iid assumption),
  - unmeasured clustering in the data,
  - dependence between trials (which can come from clustering).
- One diagnostic: plot $\hat{\mu}$ versus $(y-\hat{\mu})^2$.

Overdispersion in Dichotomous Choice Models

- In regular models $\sigma^2 = \phi = 1$, and `R` even reminds us of this assumption.
- A test for $\phi > 1$ can be constructed by modifying the Pearson statistic according to:

$$\hat{\sigma}^2 = X^2/(n-k) = \frac{1}{n-k}\sum_{i=1}^{n}\frac{(y_i - n_i\hat{p}_i)^2}{n_i\hat{p}_i(1-\hat{p}_i)}.$$

- Then the variance of the coefficient variance is adjusted with:

$$\widehat{\text{Var}}\hat{\boldsymbol{\beta}} = \hat{\sigma}^2(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1},$$

  where $\mathbf{W} = \text{diag}(mp(1-p))$ (the coefficient estimate is still unbiased).

- This added uncertainty replaces the chi-square model comparison with an approximate F-test:

$$F \approx \frac{D_{small} - D_{large}}{\widehat{\text{Var}}\hat{\boldsymbol{\beta}}(df_{small} - df_{large})}.$$

Overdispersion Example: Teenage Conformance and Survival in a Social Group

```
data(tg)
ftable(xtabs(cbind(survive,total) ~ location+period, tg))
location period survive total   period survive total
1        4       89       94     7      94      98
         8       77       86    11     141     155
2        4      106      108     7      91     106
         8       87       96    11     104     122
3        4      119      123     7     100     130
         8       88      119    11      91     125
4        4      104      104     7      80      97
         8       67       99    11     111     132
5        4       49       93     7      11     113
         8       18       88    11       0     138
```

Overdispersion Example: Teenage Conformance and Survival in a Social Group

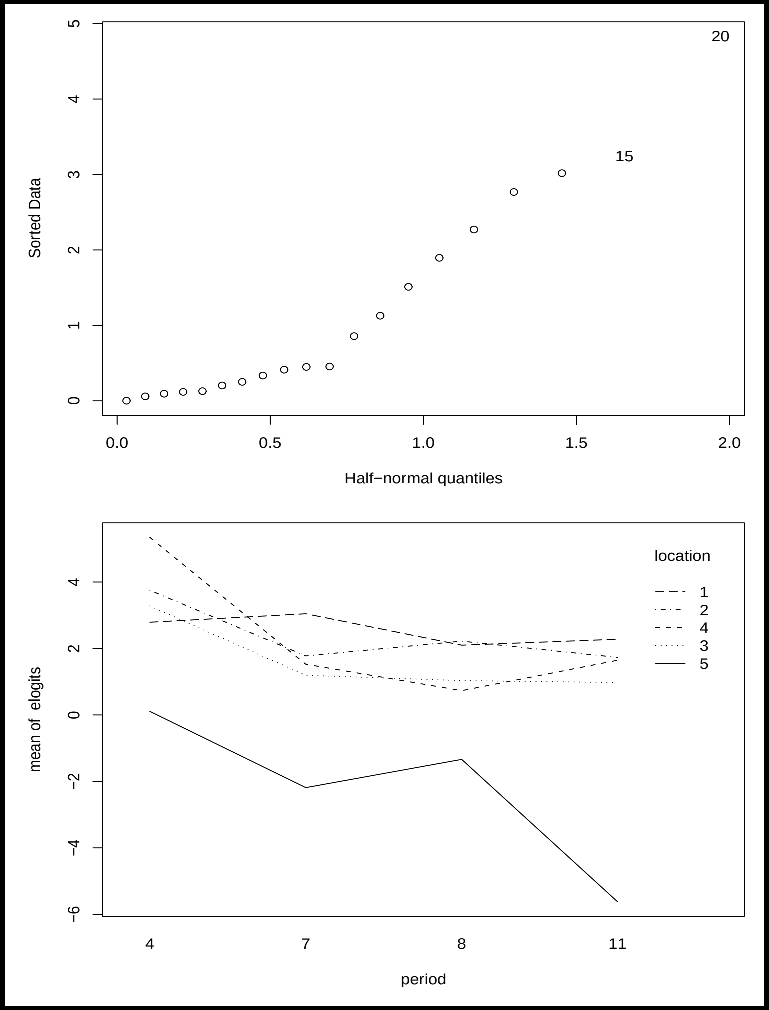
```
teen.out <- glm(cbind(survive,total-survive) ~ location+period, family=binomial,
    data=tg)
teen.out

Coefficients:
(Intercept)    location2    location3    location4    location5      period7
      4.636       -0.417       -1.242       -0.951       -4.614       -2.170
   period11      period8
     -2.450       -2.326

Degrees of Freedom: 19 Total (i.e. Null);  12 Residual
Null Deviance:          1020
Residual Deviance: 64.5         AIC: 157
```

- Since 64.5 is way into the tail of a χ^2_{12} distribution, we know to be worried.

Overdispersion Example: Teenage Conformance and Survival in a Social Group

- Sparseness? No, `min(tg$total)` returns 86.
- Outliers? No, `halfnorm(residuals(teen.out))` shows no problems.
- Specification error? No, an interaction plot of the *empirical logits* ($\log(y+0.5)-\log(m-y+0.5)$) shows no major relationships.

  ```
  elogits <- log((tg$survive+0.5)/
  (tg$total-tg$survive+0.5))
  with(tg,interaction.plot(period,
  location,elogits))
  ```



Overdispersion Example: Teenage Conformance and Survival in a Social Group

- Estimating $\hat{\sigma}^2$ shows it to be much larger than 1:

```
(sigma2 <- sum(residuals(teen.out,type="pearson")^2)/12)
5.3303
```

- And summarize the new results using the new value of $\hat{\sigma}^2$:

```
summary(teen.out,dispersion=sigma2)

            Estimate Std. Error z value Pr(>|z|)
(Intercept)    4.636      0.649    7.14  9.5e-13
location2     -0.417      0.568   -0.73    0.463
location3     -1.242      0.507   -2.45    0.014
location4     -0.951      0.528   -1.80    0.072
location5     -4.614      0.578   -7.99  1.4e-15
period7       -2.170      0.550   -3.94  8.1e-05
period8       -2.326      0.561   -4.15  3.4e-05
period11      -2.450      0.540   -4.53  5.8e-06

(Dispersion parameter for binomial family taken to be 5.3303)
```

Overdispersion Example: Teenage Conformance and Survival in a Social Group

- Another strategy for dealing with overdispersion in dichotomous outcome models is using Quasi-likelihood.

- This relaxes the form of the relevant likelihood function such that it relies on moments rather than full contributions.

```
teen.out.q <- glm(cbind(survive,total-survive) ~ location+period,
    family = quasibinomial(logit),data=tg)
```

- Note the modification to `family`.

- The results are the same, subject to algorithmic rounding from:

```
summary(teen.out.q)
```

Overdispersion Example: Teenage Conformance and Survival in a Social Group

```
Deviance Residuals:
    Min        1Q    Median       3Q       Max
-4.8305  -0.3650  -0.0303   0.6191   3.2434

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)   4.6358     0.6495   7.138 1.18e-05
location2    -0.4168     0.5682  -0.734 0.477315
location3    -1.2421     0.5066  -2.452 0.030501
location4    -0.9509     0.5281  -1.800 0.096970
location5    -4.6138     0.5777  -7.987 3.82e-06
period7      -2.1702     0.5504  -3.943 0.001953
period8      -2.3256     0.5609  -4.146 0.001356
period11     -2.4500     0.5405  -4.533 0.000686

(Dispersion parameter for quasibinomial family taken to be 5.330358)
    Null deviance: 1021.469  on 19  degrees of freedom
Residual deviance:   64.495  on 12  degrees of freedom
AIC: NA                       Number of Fisher Scoring iterations: 5
```

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
$$f(Y|\mu) = \frac{(\mu)^Y e^{-\mu}}{Y!}, \qquad\qquad y = 0, 1, 2, \ldots, \ \mu > 0$$
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 $\mu \delta t + \circ(\delta t)$ where μ is the intensity parameter discussed above and $\circ(\delta t)$ is a time interval with the property: $\lim_{\delta t \to 0} \frac{\circ(\delta t)}{\delta t} = 0$. In other words, as the interval δt reduces in size towards zero, $\circ(\delta t)$ 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 $\circ(\delta t)$. Since $\circ(\delta t)$ 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

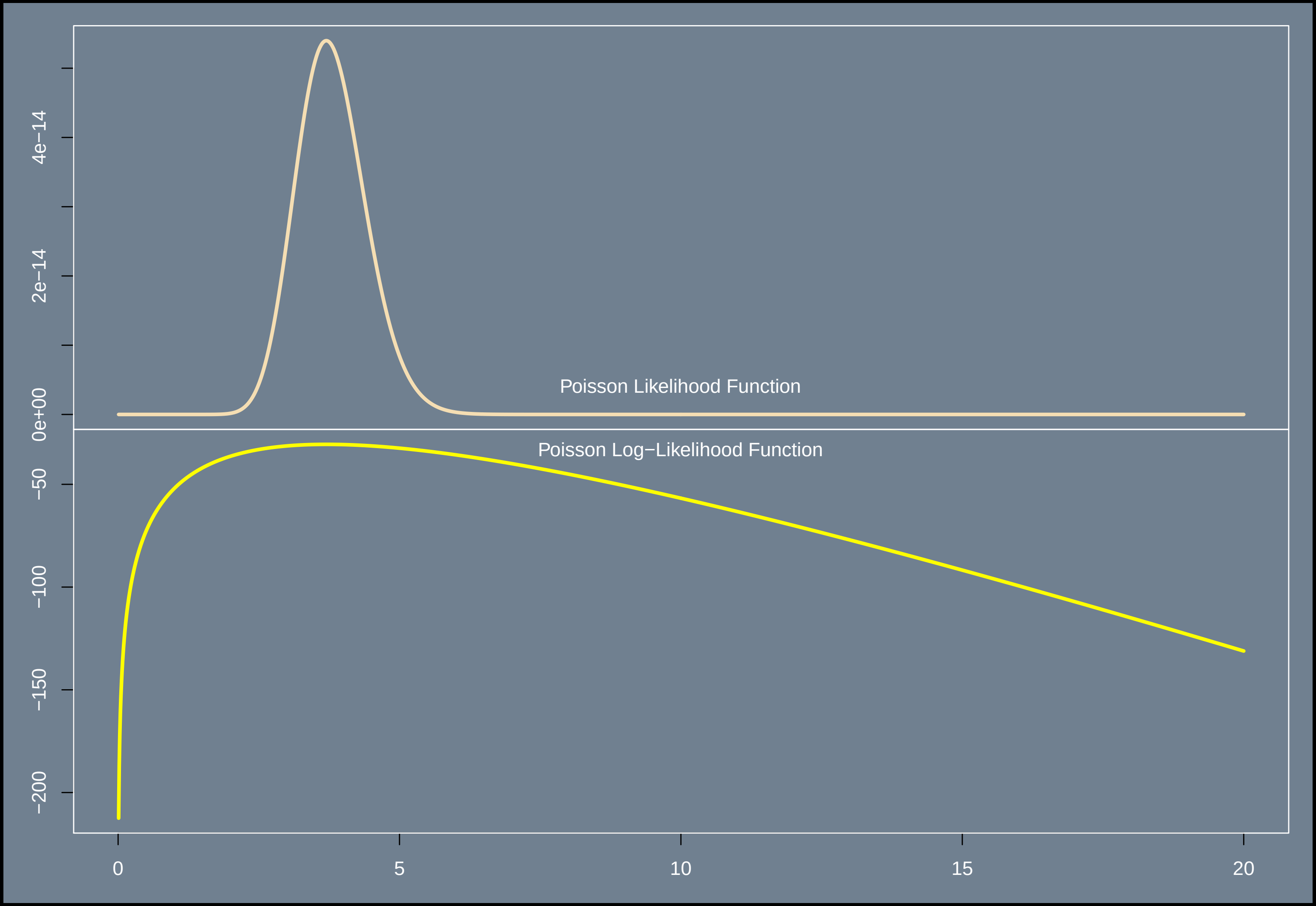
```
y.vals<-c(1,3,1,5,2,6,8,11,0,0)

# 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()
```


Poisson Likelihood Function
Poisson Log−Likelihood Function
4e−14
2e−14
0e+00
−50
−100
−150
−200
0
5
10
15
20

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:

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

  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:

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

- Now take the first deriviative:

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

- Or in full matrix terms: $\mathbf{X}'\mathbf{y} = \mathbf{X}'\hat{\mu}$, where $\hat{\mu} = \mathbf{X}\hat{\boldsymbol{\beta}}$ (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:

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

- In this specification, the systematic component is $\mathbf{X}\boldsymbol{\beta}$, the stochastic component is $\mathbf{Y} = \textbf{Military Coups}$, 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^2_{n-k} 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: $d(\boldsymbol{\eta}, y_i) = -2\left[\ell(\hat{\boldsymbol{\eta}}, \psi|y_i) - \ell(\tilde{\boldsymbol{\eta}}, \psi|y_i)\right]$, where $\hat{\boldsymbol{\eta}}$ is the model estimate and $\tilde{\boldsymbol{\eta}}$ is the saturated estimate.

- Linear Model Residual Vector: $\mathbf{R}_{standard} = \mathbf{Y} - \mathbf{X}\beta$.

- Response Residual Vector: $\mathbf{R}_{Response} = \mathbf{Y} - g^{-1}(\boldsymbol{X}\boldsymbol{\beta}) = \mathbf{Y} - \hat{\boldsymbol{\mu}}$.

- 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: $\mathbf{R}_{Working} = (\mathbf{y} - \boldsymbol{\mu})\frac{\partial}{\partial\eta}\boldsymbol{\mu}$ (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($\hat{\mu}, \sigma$)	$\eta = \hat{\mu}$	$\sum\left[y_i - \hat{\mu}_i\right]^2$
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:

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

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

- Individual Deviance Function:

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

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

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

- Generally the summed deviance is more robust that 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

		Median	Percent	Percent	Violent		Proportion
State	Executions	Income	Poverty	Black	Crime/100K	South	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:

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

```
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

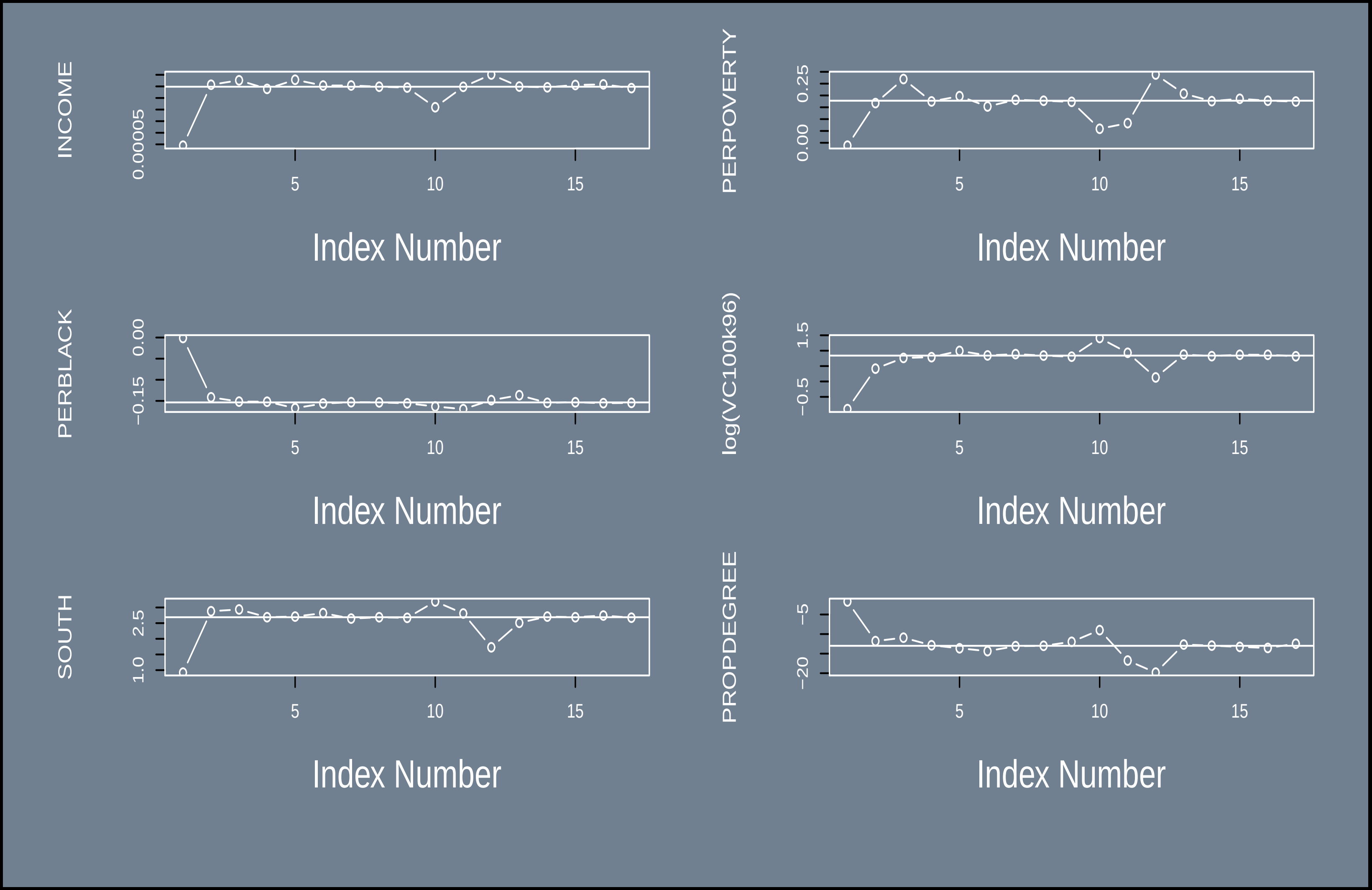
Table 2: Modeling Capital Punishment in the United States: 1997

	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
0.00005
5
10
15
Index Number
PERPOVERTY
0.00
0.25
5
10
15
Index Number
PERBLACK
−0.15
0.00
5
10
15
Index Number
log(VC100k96)
−0.5
1.5
5
10
15
Index Number
SOUTH
1.0
2.5
5
10
15
Index Number
PROPDEGREE
−20
−5
5
10
15
Index Number

First Differences for Non-Linear Models

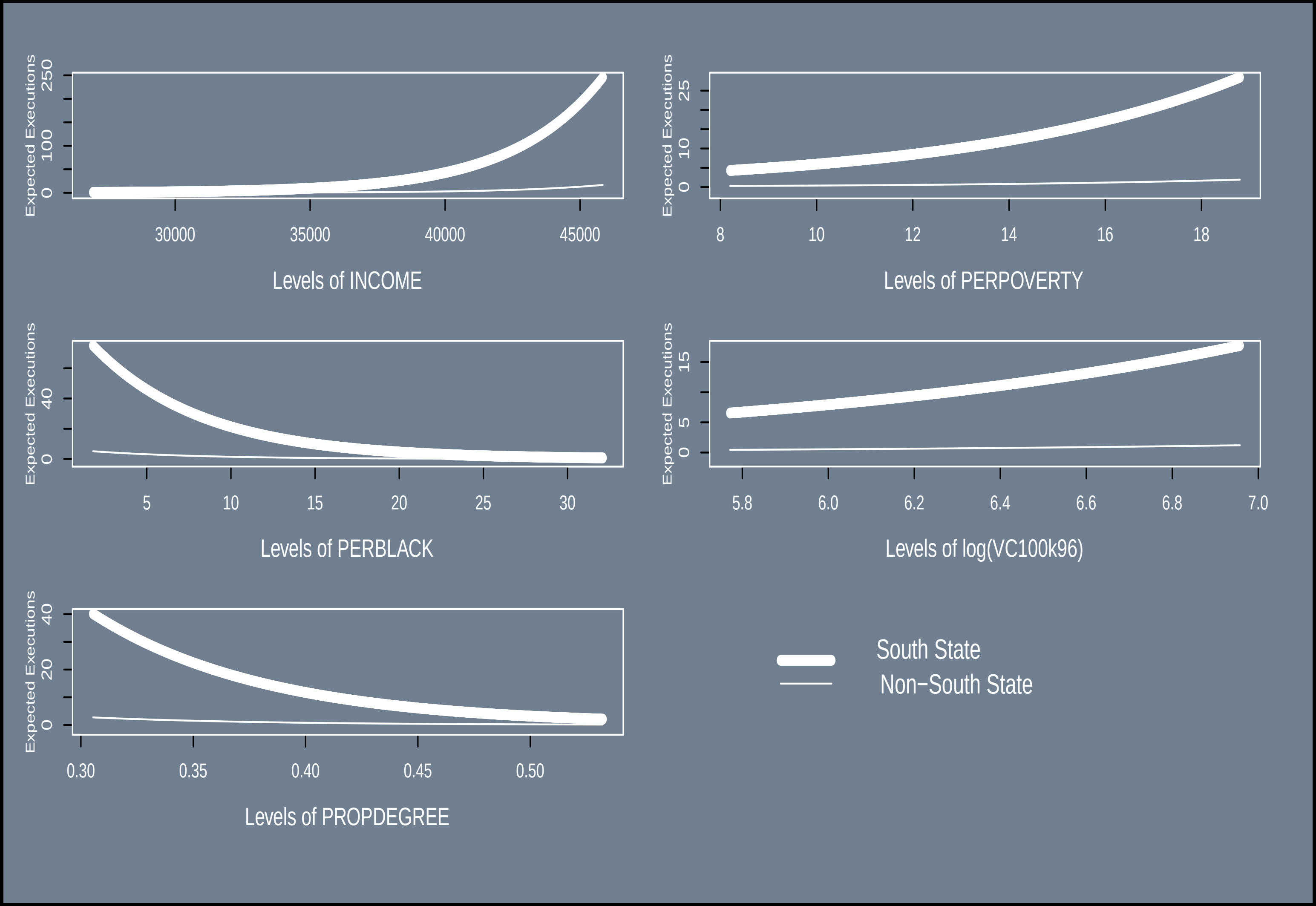
- We can no longer use "a one unit change in X gives a β change in Y."
- Main idea:
  - ▷ pick one covariate of interest, $\mathbf{X}_q$
  - ▷ choose 2 levels of this variable, $\mathbf{X}_{1,q}$, $\mathbf{X}_{2,q}$
  - ▷ set all other covarates at their mean, $\bar{\mathbf{X}}_{-q}$
  - ▷ create two 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
```


Expected Executions
0
100
250
30000
35000
40000
45000
Levels of INCOME
Expected Executions
0
10
25
8
10
12
14
16
18
Levels of PERPOVERTY
Expected Executions
0
40
5
10
15
20
25
30
Levels of PERBLACK
Expected Executions
0
5
15
5.8
6.0
6.2
6.4
6.6
6.8
7.0
Levels of log(VC100k96)
Expected Executions
0
20
40
0.30
0.35
0.40
0.45
0.50
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.

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

- If the sample size is sufficiently large, then $\frac{X^2}{a(\psi)} \sim \chi^2_{n-p}$ 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 $D(\boldsymbol{\eta}, \mathbf{y})/a(\psi) \sim \chi^2_{n-p}$.

- 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).

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

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

- 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:

$$f(Y|k,p) = \binom{y-1}{k-1} p^k (1-p)^{y-k}, \qquad y = 0,1,2,\ldots, \qquad 0 \le p \le 1.$$

- For this parameterization, we get:

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

- 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 kth 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–103rd	Bills–104th
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)
```

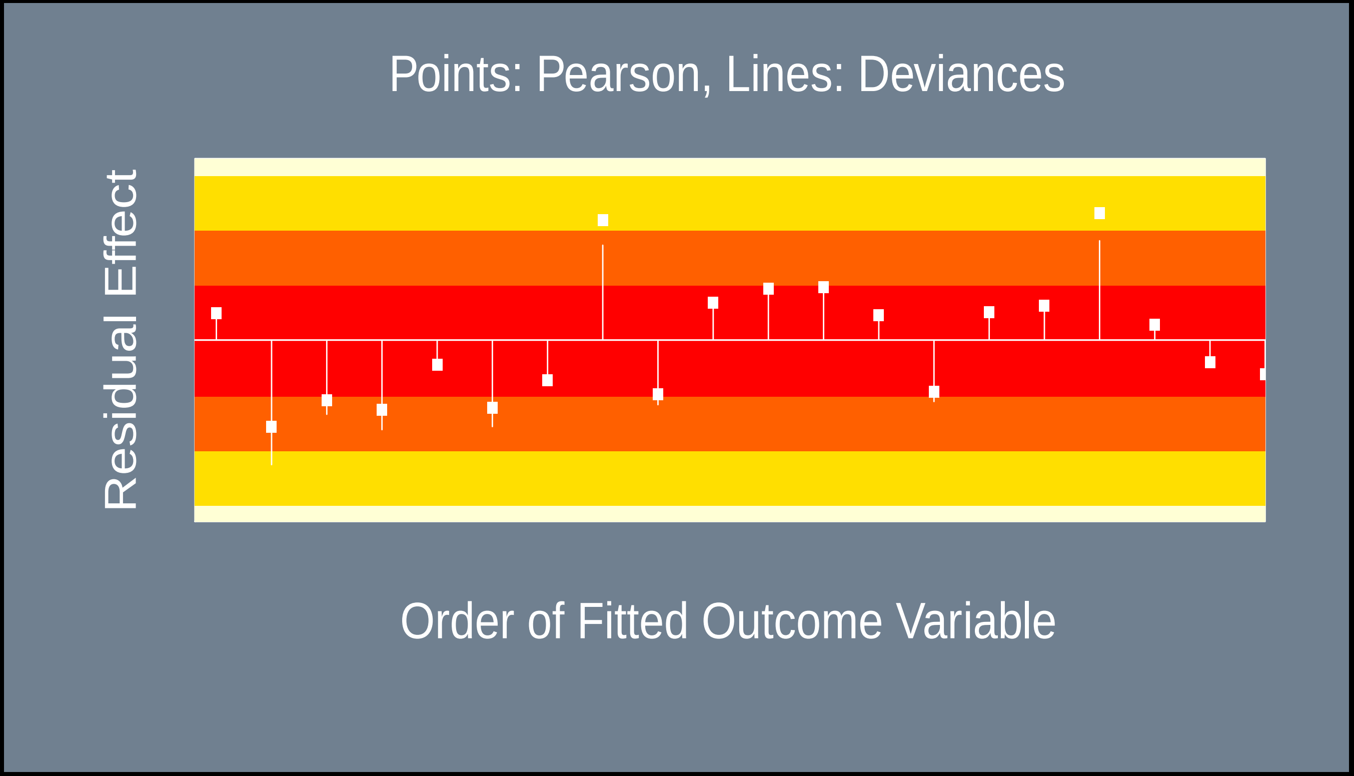
Negative Binomial GLM, Congressional Activity: 1995

```
                              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 103rd	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 occurrances, 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:

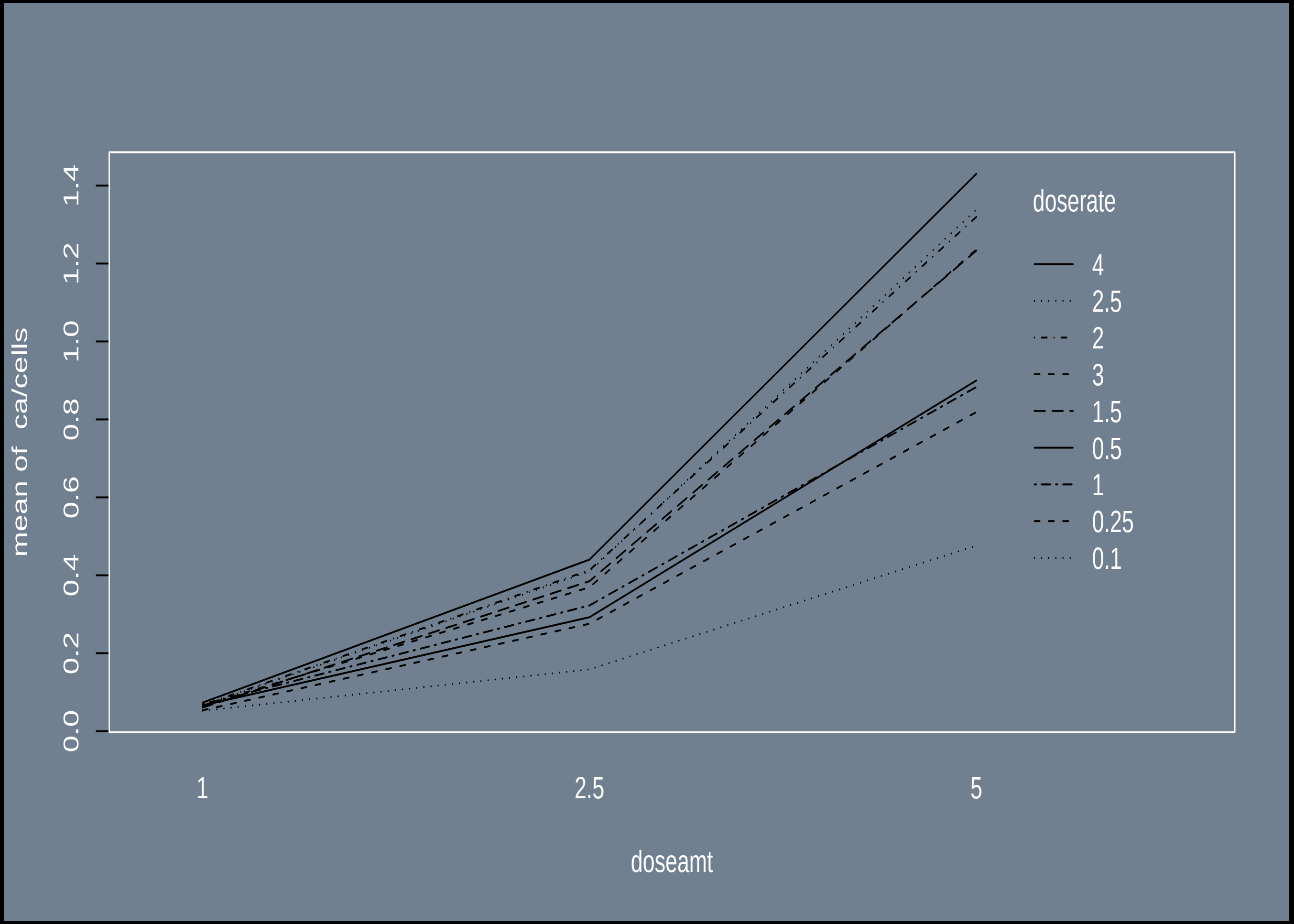
$$\text{Rate} = \frac{\#\text{events}}{\text{unit}} = \frac{\text{occurances}}{\text{possibilities}}$$

- Example from Faraway:
  - gamma radiation leads to cell abnormalities,
  - `ca` is the count of chromosonal 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
    1   0.05 0.05 0.07 0.07 0.06 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
    5   0.48 0.82 0.90 0.88 1.23 1.32 1.34 1.24 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


mean of ca/cells
0.0
0.2
0.4
0.6
0.8
1.0
1.2
1.4
1
2.5
5
doseamt
doserate
4
2.5
2
3
1.5
0.5
1
0.25
0.1

Rate Models

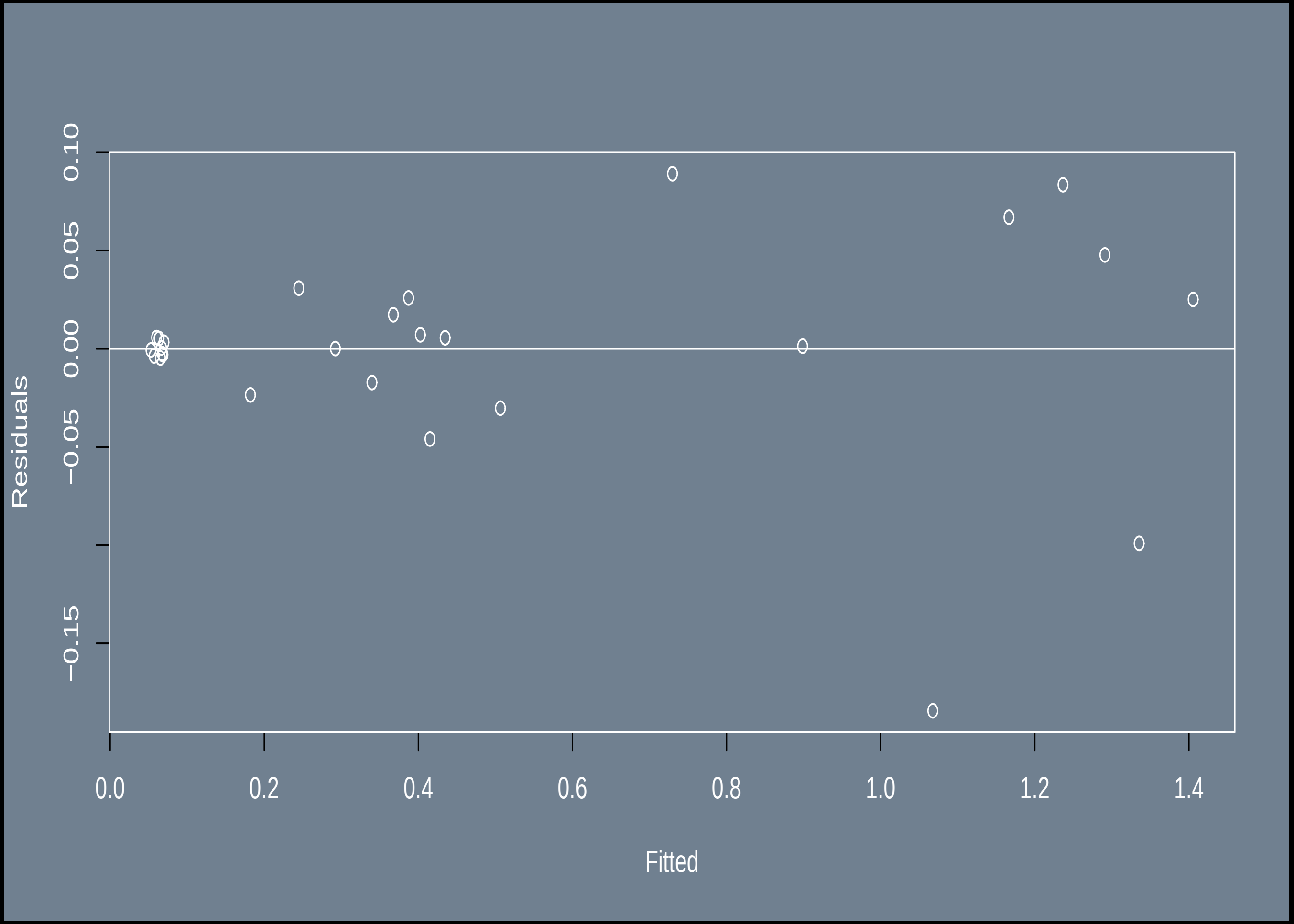
- MODEL 1: Linearly modeling the ratio directly seems to fit well, but there is overdispersion.

```
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:

$$\begin{aligned}
\log\left(\frac{\mathbb{E}[Y|\boldsymbol{\beta},\mathbf{X}]}{\text{exposure}}\right) &= \mathbf{X}\boldsymbol{\beta} \\
\log(\mathbb{E}[Y|\boldsymbol{\beta},\mathbf{X}]) - \log(\text{exposure}) &= \mathbf{X}\boldsymbol{\beta} \\
\log(\mathbb{E}[Y|\boldsymbol{\beta},\mathbf{X}]) &= \mathbf{X}\boldsymbol{\beta} + \log(\text{exposure}) \\
\mathbb{E}[Y|\boldsymbol{\beta},\mathbf{X}] &= \exp\left[\mathbf{X}\boldsymbol{\beta} + \log(\text{exposure})\right]
\end{aligned}$$

  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:

$$\log\left(\frac{\texttt{ca}}{\texttt{cells}}\right) = \mathbf{X}\boldsymbol{\beta} \qquad\Longrightarrow\qquad \log(\texttt{ca}) = \log(\texttt{cells}) + \mathbf{X}\boldsymbol{\beta}.$$

- 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 $\text{Poisson}(\lambda)$ with probability $1 - \pi$.

Zero-Inflated Poisson Model

- Let $Y_1, \ldots, Y_N$ 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

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

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 $\text{logit}(\pi_i) = \mathbf{u}_i^{\mathsf{T}}\boldsymbol{\gamma}$, where the response variable states zero or nonzero status and $\boldsymbol{\gamma}$ is a regression coefficient vector for covariates $\mathbf{u}_i^{\mathsf{T}}$.

- The second part is a poisson regression, specified by $\log(\lambda_i) = \mathbf{x}_i^{\mathsf{T}}\boldsymbol{\beta}$, 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 $\mathbf{x}_i^{\mathsf{T}}$.

- 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

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

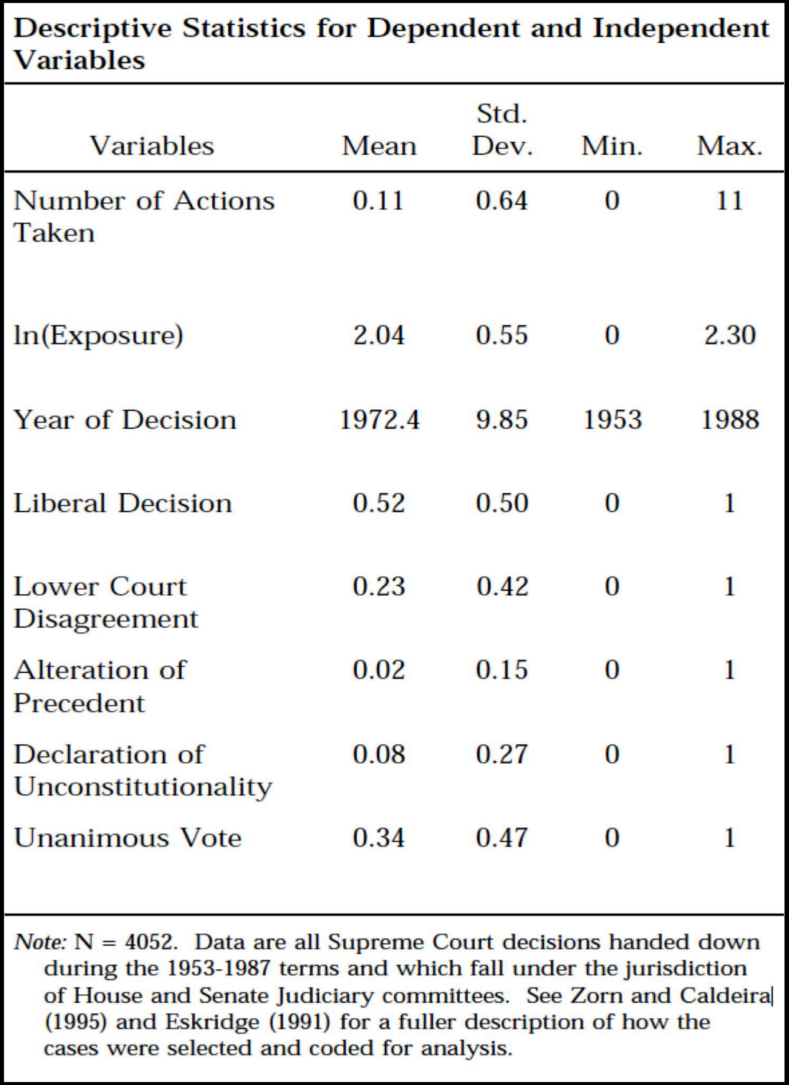
- The hurdle model has the advantage of handling both zero-inflated and zero-deflated count data.

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.

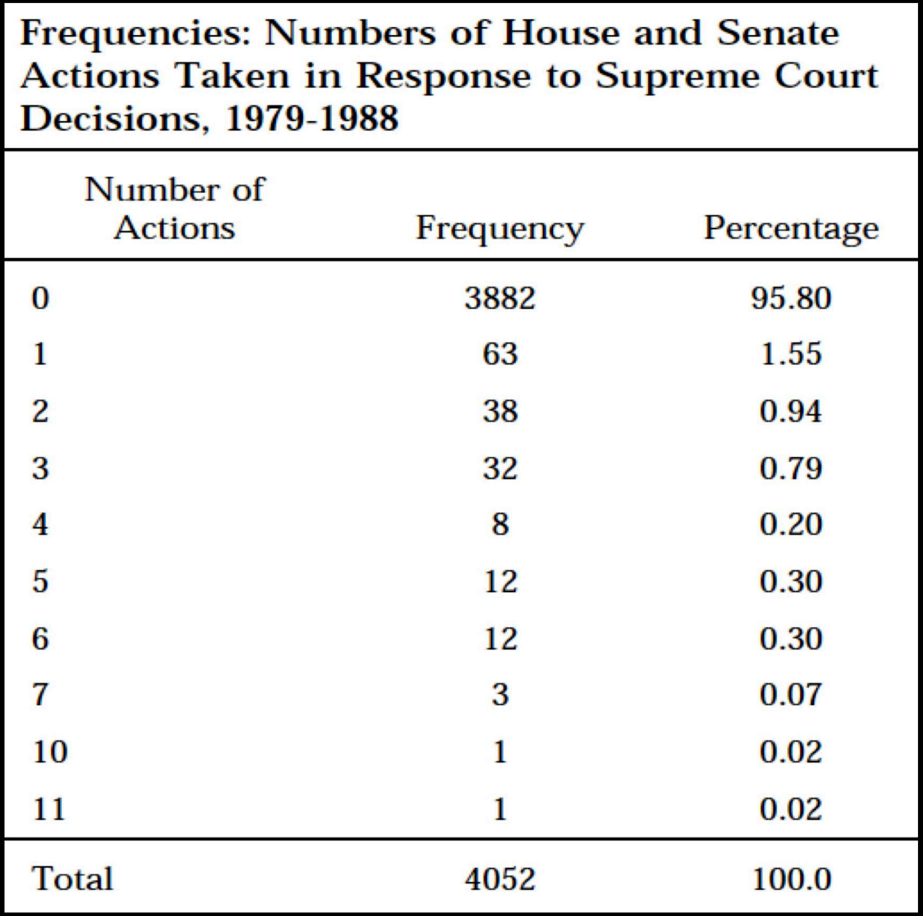
Congress and the Supreme Court

- 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.

- he data contain significantly more zeros than would be predicted by a Poisson with a mean of 0.11.

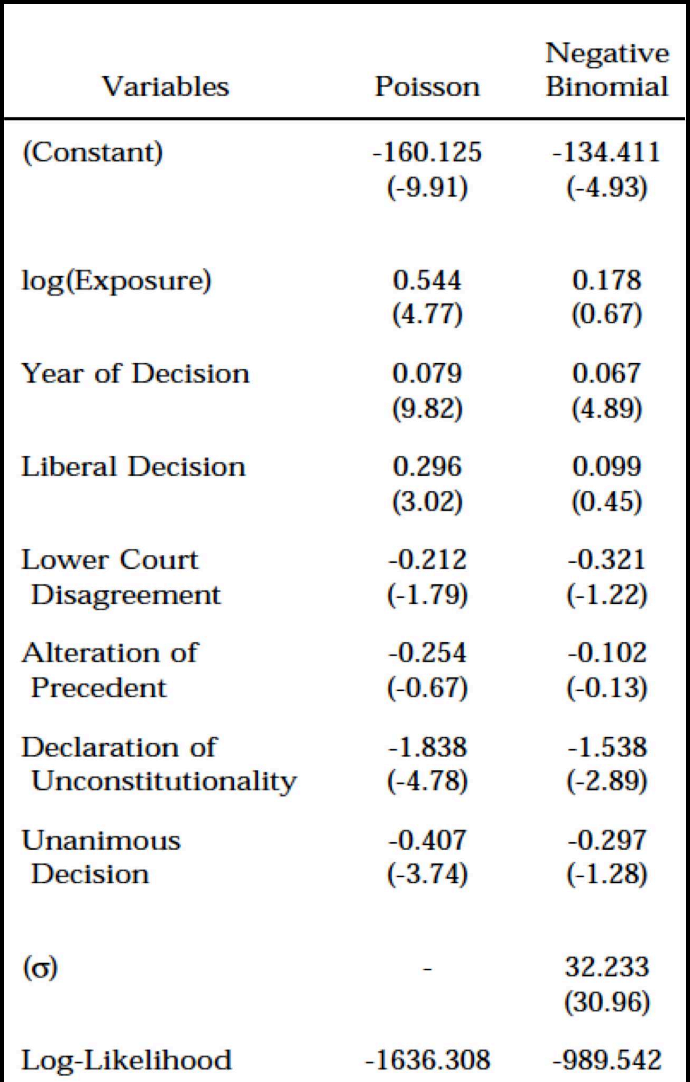
- In nearly 96 percent of all cases analyzed here no Congressional response occurred during the 1979-1988 period.



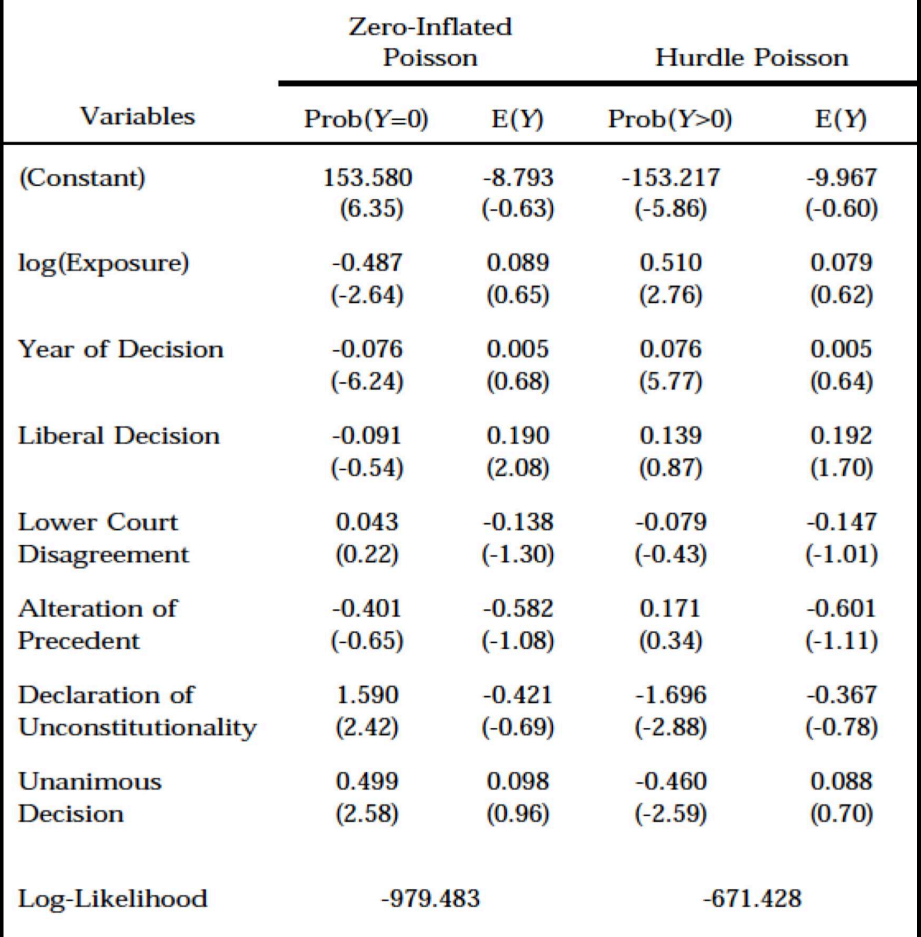
Frequencies: Numbers of House and Senate Actions Taken in Response to Supreme Court Decisions, 1979-1988

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)

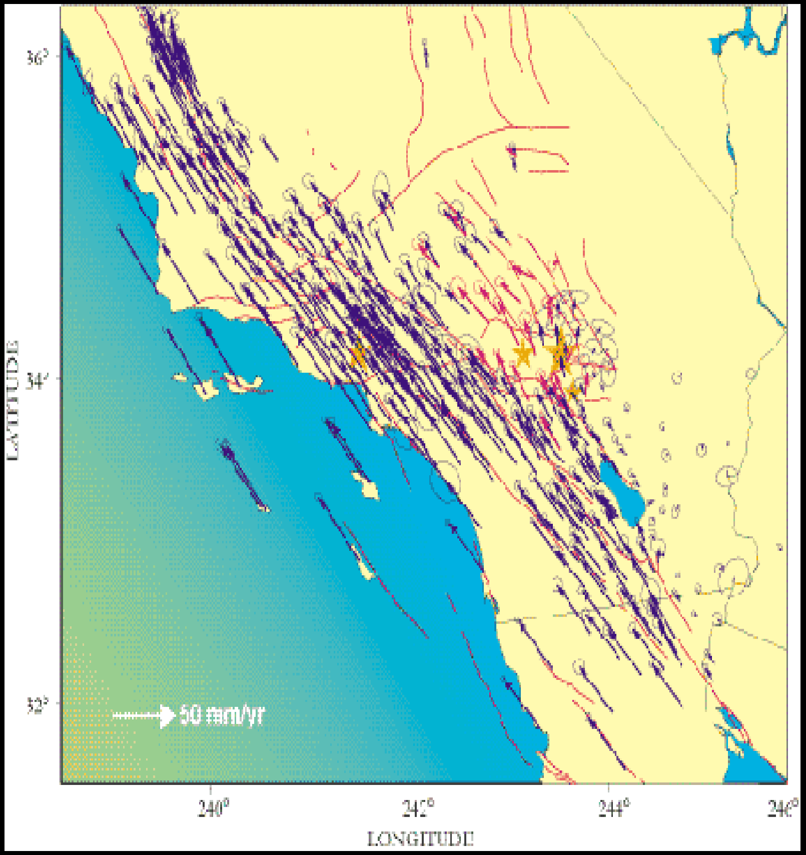


Variables	Poisson	Negative Binomial
(Constant)	-160.125 (-9.91)	-134.411 (-4.93)
log(Exposure)	0.544 (4.77)	0.178 (0.67)
Year of Decision	0.079 (9.82)	0.067 (4.89)
Liberal Decision	0.296 (3.02)	0.099 (0.45)
Lower Court Disagreement	-0.212 (-1.79)	-0.321 (-1.22)
Alteration of Precedent	-0.254 (-0.67)	-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



	Zero-Inflated Poisson		Hurdle Poisson	
Variables	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?

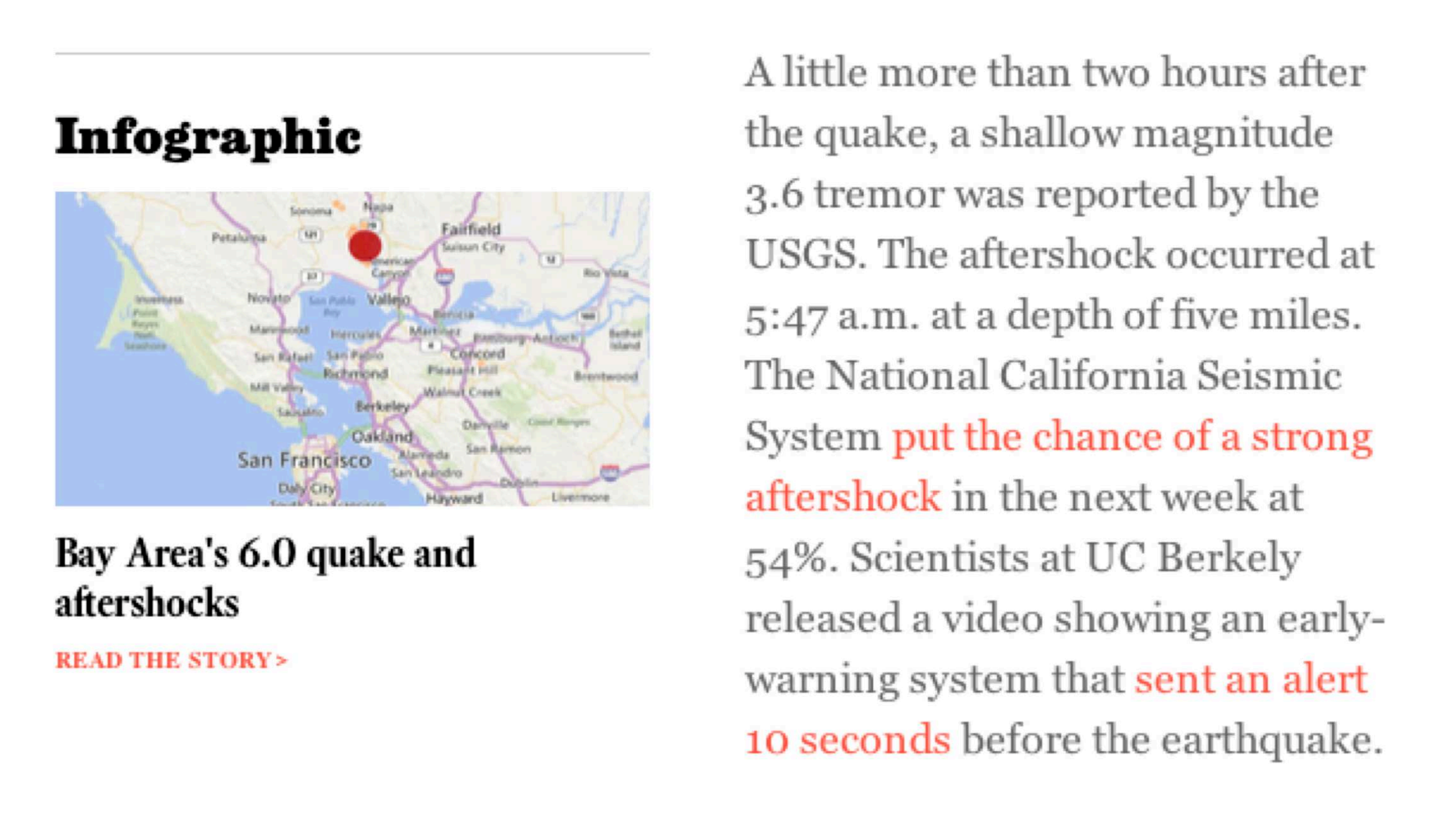
- Some geopolitical events are very hard to predict, but their after-effects may be much more reliably anticipated.

- Examples: terrorist attacks, unannounced nuclear tests, civil wars, coups.

- Bayesian learning may (over time) increase our knowledge.

- The need for real-time analysis parallels necesary 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.

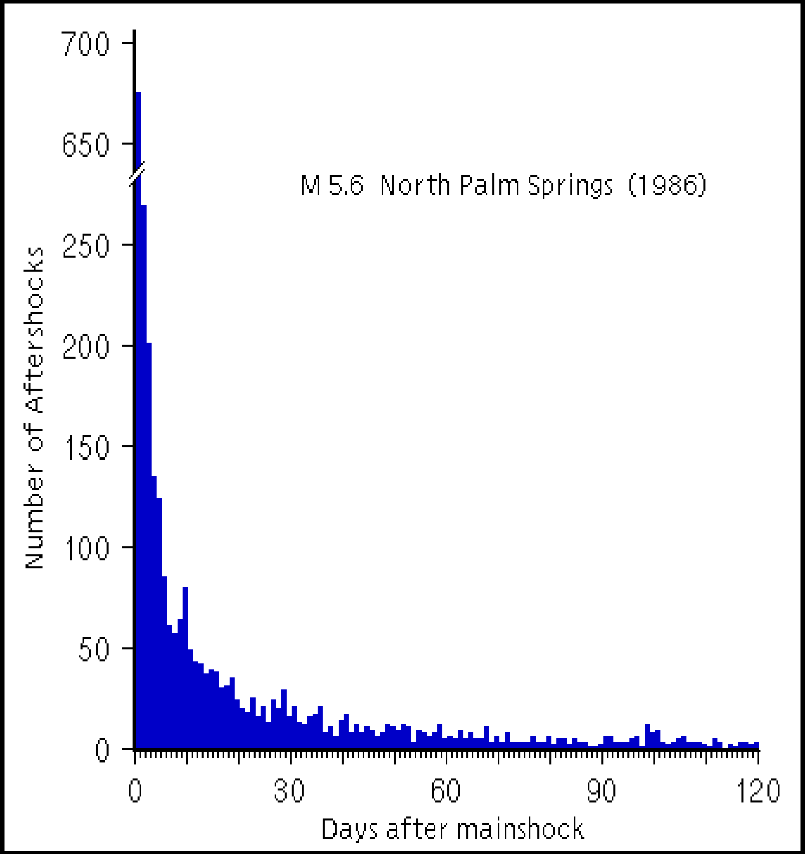
- So the probability of n aftershocks at time period t is: $\boxed{P(n|t) = \frac{N(t)^n e^{-N(t)}}{n!}.}$

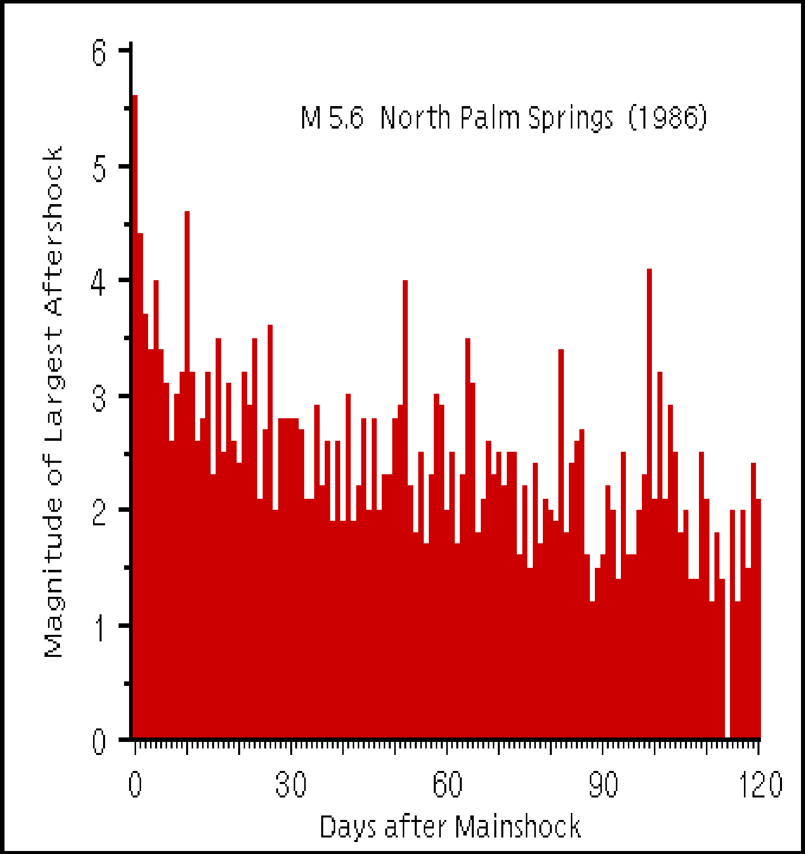
- Use the Gutenberg-Richter relation (an empirical law), aftershock version:

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

  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:

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

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

- This produces posteriors:

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

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

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.
- Reasenberg and Jones (1989) assume $\mathbf{\Sigma}^{-1}$ 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	**Time After Mainshock, Coalinga**								
$t_2 - t_1$	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	**Time After Mainshock, Whittier-Narrows**								
$t_2 - t_1$	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 Whitter-Narrows, if the main shock happened within the last 15 minutes then the probablity of a serious aftershoock in the next 24 hours is 0.393.