

GSERM 2022

Regression for Publishing

June 16, 2022

Things that are not counts:

- Ordinal scales/variables
- Grouped Binary Data
 - $\frac{N \text{ of "successes"}}{N \text{ of "trials"}}$
 - Binomial data
 - = counts only if $\Pr(\text{"success"})$ is small

Count properties:

- Discrete / integer-values
- Non-negative
- "Cumulative"

Count Data: Motivation

$$\text{Arrival Rate} = \lambda$$

$$\Pr(\text{Event})_{t,t+h} = \lambda h$$

$$\Pr(\text{No Event})_{t,t+h} = 1 - \lambda h$$

$$\begin{aligned}\Pr(Y_t = y) &= \frac{\exp(-\lambda h) \lambda h^y}{y!} \\ &= \frac{\exp(-\lambda) \lambda^y}{y!}\end{aligned}$$

Poisson Assumptions

- No Simultaneous Events
- Constant Arrival Rate
- Independent Event Arrivals

For M independent Bernoulli trials with (sufficiently small) probability of success π and where $M\pi \equiv \lambda > 0$,

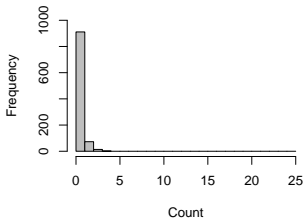
$$\begin{aligned}\Pr(Y_i = y) &= \lim_{M \rightarrow \infty} \left[\binom{M}{y} \left(\frac{\lambda}{M}\right)^y \left(1 - \frac{\lambda}{M}\right)^{M-y} \right] \\ &= \frac{\lambda^y \exp(-\lambda)}{y!}\end{aligned}$$

Poisson: Characteristics

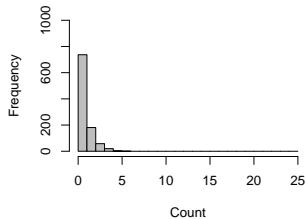
- Discrete
- $E(Y) = \text{Var}(Y) = \lambda$
- Is not preserved under affine transformations...
- For $X \sim \text{Poisson}(\lambda_X)$ and $Y \sim \text{Poisson}(\lambda_Y)$,
 $Z = X + Y \sim \text{Poisson}(\lambda_{X+Y})$ iff X and Y are *independent* but
- ...same is not true for differences.
- $\lambda \rightarrow \infty \iff Y \sim N$

Poissons: Examples

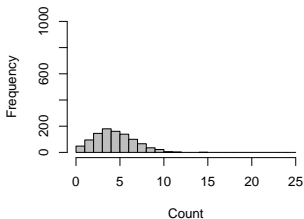
Lambda = 0.5



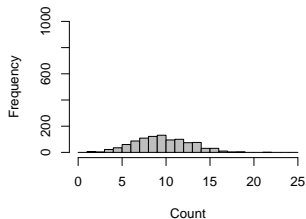
Lambda = 1.0



Lambda = 5



Lambda = 10



Suppose

$$E(Y_i) \equiv \lambda_i = \exp(\mathbf{X}_i\boldsymbol{\beta})$$

then

$$\Pr(Y_i = y | \mathbf{X}_i, \boldsymbol{\beta}) = \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^y}{y!}$$

$$L = \prod_{i=1}^N \frac{\exp[-\exp(\mathbf{X}_i\boldsymbol{\beta})][\exp(\mathbf{X}_i\boldsymbol{\beta})]^{Y_i}}{Y_i!}$$

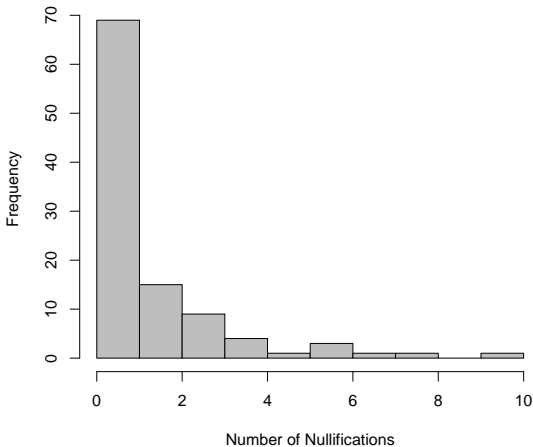
$$\ln L = \sum_{i=1}^N [-\exp(\mathbf{X}_i\boldsymbol{\beta}) + Y_i\mathbf{X}_i\boldsymbol{\beta} - \ln(Y_i!)]$$

Example: Judicial Review

- Y_i = number of Acts of Congress overturned by the Supreme Court in each Congress,
- The *mean tenure* (tenure) of the Supreme Court's justices ($\bar{X} = 10.4, \sigma = 3.4, E(\hat{\beta}) > 0$).
- Whether (1) or not (0) there was *unified government* (unified) ($\bar{X} = 0.83, E(\hat{\beta}) < 0$).

Supreme Court Nullifications, 1789-1996

```
> hist(Nulls$nulls,main="",xlab="Number of Nullifications",  
      col="grey")
```



```
> nulls.poisson<-glm(nulls~tenure+unified,family="poisson",data=Nulls)
> summary(nulls.poisson)
```

Call:

```
glm(formula = nulls ~ tenure + unified, family = "poisson", data = Nulls)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.367	-1.503	-0.623	0.561	4.153

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-0.8776	0.3713	-2.36	0.01809 *
tenure	0.0959	0.0256	3.74	0.00018 ***
unified	0.1435	0.2327	0.62	0.53747

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 251.80 on 103 degrees of freedom
Residual deviance: 237.52 on 101 degrees of freedom
AIC: 385.1

Number of Fisher Scoring iterations: 6

Interpretation: Incidence Rate Ratios

$$\begin{aligned}\frac{\hat{\lambda}|X_D = 1}{\hat{\lambda}|X_D = 0} &= \frac{\exp(\hat{\beta}_0 + \bar{\mathbf{X}}\hat{\beta} + (X_D = 1)\hat{\beta}_{X_D})}{\exp(\hat{\beta}_0 + \bar{\mathbf{X}}\hat{\beta} + (X_D = 0)\hat{\beta}_{X_D})} \\ &= \exp(\hat{\beta}_{X_D})\end{aligned}$$

- Like ORs
- unified: $\text{IRR} = \exp(0.143) = 1.15$

Incidence Rate Ratios, continued

$$\text{IRR}_{X_k, X_k + \delta} = \exp(\delta \hat{\beta}_k)$$

So, a ten-year difference in tenure:

$$\begin{aligned} \text{IRR} &= \exp(10 \times 0.096) \\ &= \exp(0.96) \\ &= 2.61 \end{aligned}$$

Incidence Rate Ratios

```
> library(mfx)
> nulls.poisson.IRR<-poissonirr(nulls~tenure+unified,
                                data=NULLs)
> nulls.poisson.IRR
```

Call:

```
poissonirr(formula = nulls ~ tenure + unified, data = NULLs)
```

Incidence-Rate Ratio:

	IRR	Std. Err.	z	P> z	
tenure	1.1006	0.0282	3.74	0.00018	***
unified	1.1543	0.2686	0.62	0.53747	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Predicted Values (\hat{Y} s)

Mean predicted Y :

$$\begin{aligned} E(Y|\bar{\mathbf{X}}_i) &= \exp[-0.878 + (0.096 \times 10) + (0.143 \times 1)] \\ &= \exp(0.225) \\ &= 1.25 \end{aligned}$$

In-Sample

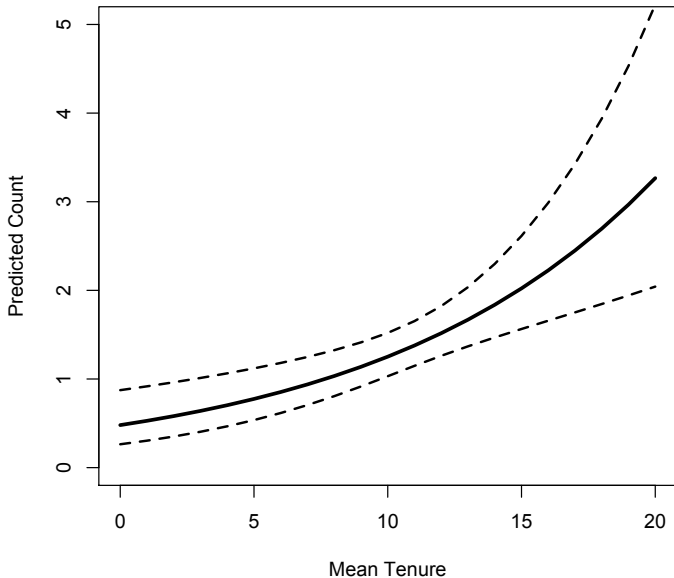
- R : `in $fitted.values`
- Stata : `use predict`

Out-of-Sample: `use predict`

Example: Out-Of-Sample Predicted Values

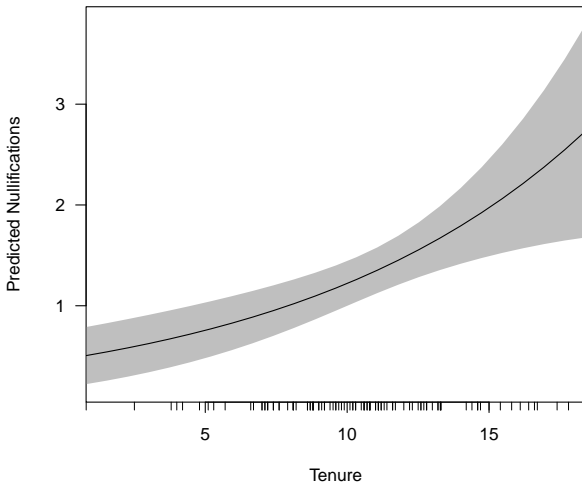
```
> tenure<-seq(0,20,1)
> unified<-1
> simdata<-as.data.frame(cbind(tenure,unified))
> nullhats<-predict(nulls.poisson,newdata=simdata,se.fit=TRUE)
>
> # NOTE: These are XBs, not predicted counts.
> # Transforming:
> nullhats$Yhat<-exp(nullhats$fit)
> nullhats$UB<-exp(nullhats$fit + 1.96*(nullhats$se.fit))
> nullhats$LB<-exp(nullhats$fit - 1.96*(nullhats$se.fit))
> plot(simdata$tenure,nullhats$Yhat,t="l",lwd=3,ylim=c(0,5),ylab=
+       "Predicted Count", xlab="Mean Tenure")
> lines(simdata$tenure,nullhats$UB,lwd=2,lty=2)
> lines(simdata$tenure,nullhats$LB,lwd=2,lty=2)
>
> plot(simdata$tenure,nullhats$Yhat,t="l",lwd=3,ylim=c(0,5),ylab=
+       "Predicted Count", xlab="Mean Tenure")
> lines(simdata$tenure,nullhats$UB,lwd=2,lty=2)
> lines(simdata$tenure,nullhats$LB,lwd=2,lty=2)
```

Plotting Out-Of-Sample Predicted Values



Same Thing, using `-margins-`

```
> cplot(nulls.poisson,"tenure",xlab="Tenure",  
        ylab="Predicted Nullifications")
```



Predicted Probabilities

$$\Pr(\widehat{Y_i = y} | \mathbf{X}_i, \hat{\beta}) = \frac{\exp[-\exp(\mathbf{X}_i \hat{\beta})][\exp(\mathbf{X}_i \hat{\beta})]^y}{y!}$$

$$\begin{aligned} \rightarrow \Pr(\widehat{Y_i = 0} | \bar{\mathbf{X}}_i, \hat{\beta}) &= \frac{[\exp(-1.25)](1.25)^0}{0!} \\ &= \frac{(0.287)(1)}{1} \\ &= 0.287 \end{aligned}$$

$$\begin{aligned} \Pr(\widehat{Y_i = 1} | \bar{\mathbf{X}}_i, \hat{\beta}) &= \frac{[\exp(-1.25)](1.25)^1}{1!} \\ &= \frac{(0.287)(1.25)}{1} \\ &= 0.359 \end{aligned}$$

Predicted Probabilities

$$\begin{aligned}\Pr(\widehat{Y_i = 2} | \bar{\mathbf{X}}_i, \hat{\beta}) &= \frac{[\exp(-1.25)](1.25)^2}{2!} \\ &= \frac{(0.287)(1.563)}{2} \\ &= 0.224\end{aligned}$$

$$\begin{aligned}\Pr(\widehat{Y_i = 3} | \bar{\mathbf{X}}_i, \hat{\beta}) &= \frac{[\exp(-1.25)](1.25)^3}{3!} \\ &= \frac{(0.287)(1.953)}{6} \\ &= 0.093\end{aligned}$$

“Exposure” and “Offsets”

$$E(Y_i | \mathbf{X}_i, M_i) = \lambda_i M_i$$

Same as including $\ln(M_i)$ in \mathbf{X} with $\beta_{\ln M} = 1$.

Example: Data on numbers of interstate disputes by country, 1950-1985...

- $N = 102$, but
- N_{dyads} = number of dyad-years which were aggregated to create each observation, ranging from five to 3249
- disputes = number of (interstate) dispute-years that country experienced during 1950-1985
- allies = number of (dyadic) ally-years each country had during 1950-1985
- $\text{openness} = \frac{1}{36} \left(\frac{\text{Imports}_t + \text{Exports}_t}{\text{GDP}_t} \right)$ across all 36 years in the data.

“Exposure” and “Offsets”: Data

```
# Data are aggregated dyadic data, 1950-1985...
```

```
> IR<-read.csv("Data/offsetIR.csv")
```

```
> summary(IR)
```

ccode	Ndyads	disputes	allies	openness	exposure
Min. : 2	Min. : 5	Min. : 0.00	Min. : 0.0	Min. : 0.032	Min. : 1.61
1st Qu.: 214	1st Qu.: 44	1st Qu.: 0.00	1st Qu.: 0.0	1st Qu.: 0.185	1st Qu.: 3.79
Median : 436	Median : 92	Median : 1.00	Median : 26.0	Median : 0.296	Median : 4.52
Mean : 418	Mean : 179	Mean : 3.55	Mean : 63.9	Mean : 0.392	Mean : 4.42
3rd Qu.: 598	3rd Qu.: 146	3rd Qu.: 4.00	3rd Qu.: 81.0	3rd Qu.: 0.535	3rd Qu.: 4.98
Max. : 900	Max. : 3249	Max. : 52.00	Max. : 1283.0	Max. : 1.659	Max. : 8.09
				NA's : 12	

```
> cor(IR,use="complete.obs")
```

	ccode	Ndyads	disputes	allies	openness	exposure
ccode	1.00000	-0.29623	-0.1399	-0.3983	0.02744	-0.6544
Ndyads	-0.29623	1.00000	0.8626	0.9200	-0.07511	0.6988
disputes	-0.13989	0.86257	1.0000	0.8255	-0.16819	0.6335
allies	-0.39826	0.92004	0.8255	1.0000	-0.12548	0.7003
openness	0.02744	-0.07511	-0.1682	-0.1255	1.00000	-0.1433
exposure	-0.65442	0.69878	0.6335	0.7003	-0.14325	1.0000

Ignoring Exposure

```
> IR.fit1<-glm(disputes~allies+openness,data=IR,family="poisson")
> summary(IR.fit1)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	1.1559498	0.1117581	10.343	< 2e-16 ***
allies	0.0025184	0.0001159	21.734	< 2e-16 ***
openness	-1.1144132	0.2773631	-4.018	5.87e-05 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 732.68 on 101 degrees of freedom
Residual deviance: 392.97 on 99 degrees of freedom
(12 observations deleted due to missingness)
AIC: 588.29

Number of Fisher Scoring iterations: 6

Correcting for Exposure

```
> IR.fit2<-glm(disputes~allies+openness,data=IR,family="poisson",
  offset=log(Ndyads))
> summary(IR.fit2)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-3.2906055	0.1194616	-27.545	< 2e-16 ***
allies	-0.0006058	0.0001333	-4.544	5.52e-06 ***
openness	-1.6040587	0.3167415	-5.064	4.10e-07 ***

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Null deviance: 320.19 on 101 degrees of freedom
Residual deviance: 277.79 on 99 degrees of freedom
(12 observations deleted due to missingness)
AIC: 473.11

Number of Fisher Scoring iterations: 5

Correcting for Exposure (continued)

```
> IR.fit3<-glm(disputes~allies+openness+log(Ndyads),data=IR,  
+             family="poisson")  
> summary(IR.fit3)
```

Call:

```
glm(formula = disputes ~ allies + openness + log(Ndyads), family = "poisson",  
    data = IR)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.838	-1.390	-0.758	0.605	4.731

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.42656676	0.34345252	-7.07	0.00000000000016 ***
allies	-0.00000948	0.00025687	-0.04	0.97
openness	-1.44462460	0.31193821	-4.63	0.0000036368547 ***
log(Ndyads)	0.81097748	0.07095243	11.43	< 0.000000000000002 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 732.68 on 101 degrees of freedom

Residual deviance: 270.59 on 98 degrees of freedom

(12 observations deleted due to missingness)

AIC: 467.9

Number of Fisher Scoring iterations: 5

Test $\beta_{\text{exposure}} = 1.0$

```
> # z-test:
```

```
> 2*pnorm((0.811-1)/.071)
[1] 0.007768438
```

```
> # Wald test:
```

```
> wald.test(b=coef(IR.fit3),Sigma=vcov(IR.fit3),Terms=4,H0=1)
```

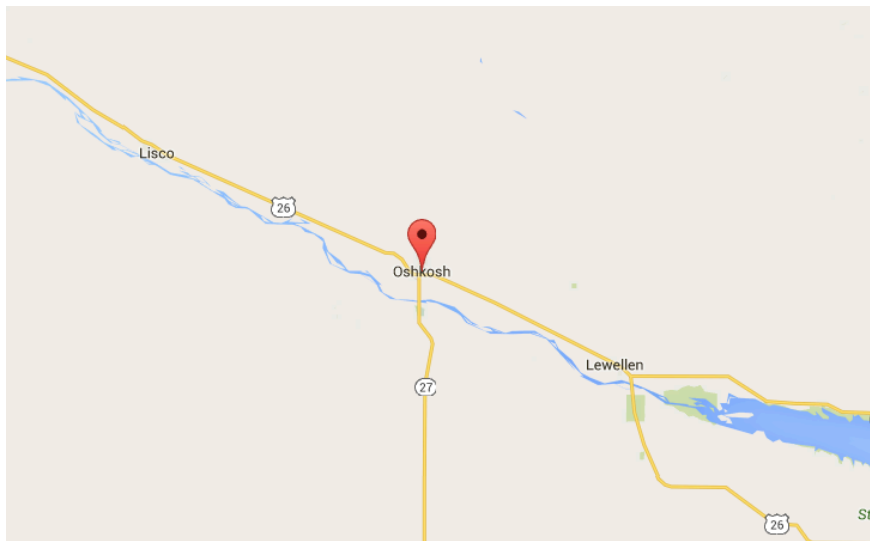
```
Wald test:
```

```
-----
```

```
Chi-squared test:
```

```
X2 = 7.1, df = 1, P(> X2) = 0.0077
```

Contagion, Heterogeneity, and Dispersion





Heterogeneity, Contagion, and Dispersion

Cats (daily values):

$$Y_{cats} = \{0, 1, 1, 0, 2, 0, 1, 0, 3, 1, 2, 1, 0, 2\}$$

$$\bar{Y}_{cats} = 1.0,$$

$$\sigma_{cats} = 0.92.$$

Heterogeneity, Contagion, and Dispersion

$$E(Y_{cats}) = \lambda_{cats}$$

Poisson assumes:

- $Y = 0$ at $t = 0$,
- Exclusive events
- $t_j = t_k \forall j \neq k$
- *Constant, independent* $\Pr(\text{Event})$ over t

Daily values:

$$Y_{antelope} = \{0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 0, 7, 7\}$$

$$\bar{Y}_{antelope} = 1.0,$$

$$\sigma_{antelope} = 6.46.$$

Positive contagion \rightarrow *overdispersion*.

Daily values:

$$Y_{foxes} = \{1, 0, 1, 1, 1, 1, 1, 2, 1, 1, 1, 1, 1, 1\}$$

$$\bar{Y}_{foxes} = 1.0,$$

$$\sigma_{foxes} = 0.15.$$

Negative contagion \rightarrow underdispersion.

Aggregation & Cross-Period Effects

Aggregated two-day measures:

$$\begin{aligned} Y_{cats} &= \{1, 1, 2, 1, 4, 3, 2\} \\ Y_{antelope} &= \{0, 0, 0, 0, 0, 0, 14\} \\ Y_{foxes} &= \{1, 2, 2, 3, 2, 2, 2\} \end{aligned}$$

Poisson requires:

- Correct specification
- Correct distribution for ϵ
- Constant $E(Y|\mathbf{X}, \beta)$

Omitted variables \rightarrow overdispersion:

$$\lambda_i \equiv E(Y_i) = f[\mathbf{X}_i\beta + \textcolor{red}{Z}_i\theta]$$

Examine:

$$\hat{u}_i = \delta \hat{\lambda}_i + \epsilon_i$$

where

$$\hat{u}_i = \frac{(Y_i - \hat{\lambda}_i)^2 - Y_i}{\hat{\lambda}_i \sqrt{2}}$$

- Estimate a Poisson regression of Y_i on \mathbf{X}_i , and generate predicted counts $\hat{\lambda}_i$.
- Calculate \hat{u}_i according to the equation above.
- Estimate δ using OLS, and test $H_0 : \hat{\delta} = 0$.

Overdispersion: Models

$$\begin{aligned} E(Y_i) \equiv \lambda_i &= \exp(\mathbf{X}_i \boldsymbol{\beta} + u_i) \\ &= \exp(\mathbf{X}_i \boldsymbol{\beta}) \exp(u_i) \\ &= \lambda_i \nu_i \end{aligned}$$

$$\nu_i \sim \text{gamma} \left(1, \frac{1}{\alpha} \right)$$

$$\Pr(Y_i = y | \lambda_i, \alpha) = \left(\frac{\Gamma(\alpha^{-1} + Y_i)}{\Gamma(\alpha^{-1})\Gamma(Y_i + 1)} \right) \left(\frac{\alpha^{-1}}{\alpha^{-1} + \lambda_i} \right)^{\alpha^{-1}} \left(\frac{\lambda_i}{\lambda_i + \alpha^{-1}} \right)^{Y_i}$$

where

$$\Gamma(a) = \int_0^{\infty} \exp(-t) t^{a-1} dt$$

Basis:

$$\lambda_i = \exp(\mathbf{X}_i\beta)$$

Model has

$$E(Y) = \lambda$$

$$\text{Var}(Y) = \lambda(1 + \alpha\lambda), \alpha > 0$$

Negative Binomial (log-)Likelihood

$$\ln L_{NB} = \sum_{i=1}^N \left\{ \left(\sum_{j=0}^{Y_i-1} \ln(j + \alpha^{-1}) \right) - \ln Y_i! - \right. \\ \left. (Y_i - \alpha^{-1}) \ln[1 + \alpha \exp(\mathbf{X}_i \boldsymbol{\beta})] + Y_i \ln \alpha + Y_i \mathbf{X}_i \boldsymbol{\beta} \right\}$$

So:

- $\alpha = 0 \iff E(Y) = \text{Var}(Y)$
- LR test for overdispersion:

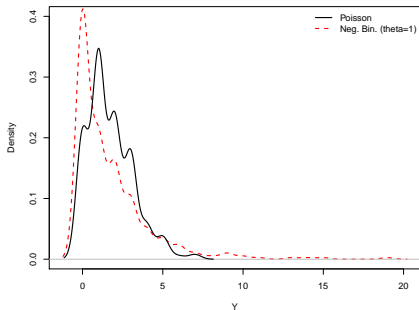
$$-2 \times (\ln \widehat{L_{Poisson}} - \ln \widehat{L_{NB}}) \sim \chi_1^2$$

- $\widehat{E(Y_i)} \equiv \hat{\lambda}_i = \exp(\mathbf{X}_i \hat{\boldsymbol{\beta}})$

What Difference Does It Make?

```
> N<-400
> set.seed(7222009)
> X <- runif(N,min=0,max=1)
> YPois <- rpois(N,exp(0+1*X))          # Poisson
> YNB <- rnbinom(N,size=1,mu=exp(0+1*X)) # NB with theta=1.0
>
> describe(cbind(YPois,YNB))
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
YPois	1	400	1.72	1.41	1	1.56	1.48	0	7	7	0.92	0.84	0.07
YNB	2	400	1.71	2.44	1	1.22	1.48	0	19	19	2.76	11.15	0.12



What Difference Does It Make (cont'd)?

```
> # Regressions:
>
> summary(glm(YPois~X,family="poisson")) # Poisson

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.009637   0.085337  -0.113    0.91
X            1.030573   0.131992   7.808 5.82e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for poisson family taken to be 1)

    Null deviance: 516.06  on 399  degrees of freedom
Residual deviance: 453.55  on 398  degrees of freedom
AIC: 1274.4

> summary(glm.nb(YPois~X)) # NB

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.009629   0.085345  -0.113    0.91
X            1.030557   0.132007   7.807 5.86e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Negative Binomial(7837.699) family taken to be 1)

    Null deviance: 515.96  on 399  degrees of freedom
Residual deviance: 453.46  on 398  degrees of freedom
AIC: 1276.5

            Theta: 7838
        Std. Err.: 135342
Warning while fitting theta: iteration limit reached

2 x log-likelihood: -1270.451
```

What Difference Does It Make (cont'd)?

```
> # More regressions:
>
> summary(glm(YNB~X,family="poisson")) # Poisson

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03170    0.08593  -0.369   0.712
X            1.06109    0.13248   8.009 1.15e-15 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 1118.0  on 399  degrees of freedom
Residual deviance: 1052.1  on 398  degrees of freedom
AIC: 1698.6

> summary(glm.nb(YNB~X)) # NB

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.03525    0.13650  -0.258   0.796
X            1.06773    0.22809   4.681 2.85e-06 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Negative Binomial(0.8499) family taken to be 1)

Null deviance: 436.92  on 399  degrees of freedom
Residual deviance: 414.81  on 398  degrees of freedom
AIC: 1407.4

            Theta: 0.850
        Std. Err.: 0.109

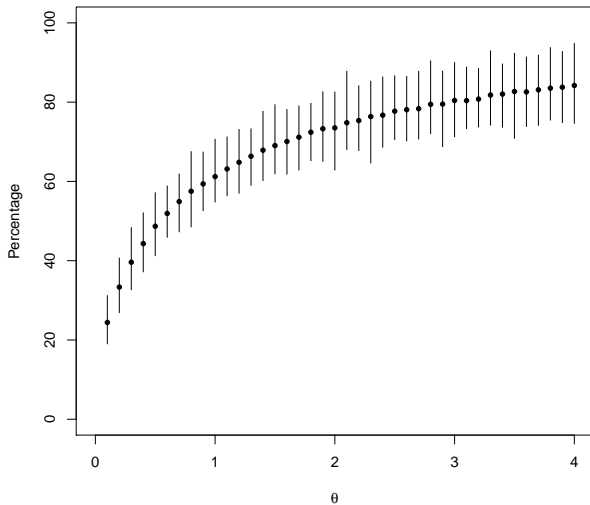
2 x log-likelihood: -1401.354
```

Poisson Regression Underestimates N.B. Variances

```
Sims <- 250 # (250 sims each)
theta <- seq(0.1,4,by=0.1) # values of theta
diffs<-matrix(nrow=Sims,ncol=length(theta))

set.seed(7222009)
for(j in 1:length(theta)) {
  for(i in 1:Sims) {
    X<-runif(N,min=0,max=1)
    Y<-rnbino(N,size=theta[j],mu=exp(0+1*X))
    p<-glm(Y~X,family="poisson")
    nb<-glm.nb(Y~X)
    diffs[i,j]<- ((sqrt(vcov(p))[2,2]) / sqrt(vcov(nb))[2,2])*100
  }
}
```

Percentage of True (Negative Binomial) S.E. From Fitted Poisson Model, by $\theta = \frac{1}{\alpha}$



Negative Binomial In Practice

Model fitting (in R):

- `glm.nb` (in MASS)
- `negbinomial` (in VGAM)
- `negbin` (in aod)
- `glmnb.fit` (in statmod)
- Probably others...

Model interpretation + diagnostics:

- `fitNBP` (in statmod) (dispersion parameter estimation)
- `negbinirr` (in mfx) (IRRs)
- `negbinmfx` (in mfx) (marginal effects)

Underdispersion: COM Poisson

Underdispersion:

- ... is implied by *negative contagion*
- \rightarrow limiting effect on $\max(Y)$.

Conway-Maxwell-Poisson: Allows for either over- or underdispersion:

$$Pr(Y_i = y) = \frac{\lambda^y}{(y!)^\nu} \frac{1}{Z(\lambda, \nu)}$$

where:

$$Z(\lambda, \nu) = \sum_{j=0}^{\infty} \frac{\lambda^j}{(j!)^\nu}.$$

$\nu \in [0, \infty)$, with:

- $\nu = 1 \rightarrow$ Poisson
- $\nu \rightarrow \infty \rightarrow \text{Bernoulli}\left(\frac{\lambda}{1+\lambda}\right)$
- $\nu = 0 \rightarrow$ geometric (with $0 < \lambda < 1$)

And yes, there is an [R package](#)...

Example: SCOTUS Amicus Curiae (1953-85)

- $N = 7157$
- `namici` is the number of amicus curiae briefs filed in each case,
- `term` is the term (i.e., year) of the Court,
- `civlibs` is whether (=1) or not (=0) the case involved a civil rights and liberties issue.

```
> summary(amicus)
```

<code>namici</code>	<code>term</code>	<code>civlibs</code>
Min. : 0.00	Min. :53.0	Min. :0.000
1st Qu.: 0.00	1st Qu.:64.0	1st Qu.:0.000
Median : 0.00	Median :72.0	Median :1.000
Mean : 1.03	Mean :71.1	Mean :0.501
3rd Qu.: 1.00	3rd Qu.:79.0	3rd Qu.:1.000
Max. :53.00	Max. :85.0	Max. :1.000

Amicus Example: Poisson

```
> amici.poisson<-glm(namici~term+civlibs,data=amici,family="poisson")
> summary(amici.poisson)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-4.51196	0.11190	-40.32	<2e-16 ***
term	0.06361	0.00147	43.27	<2e-16 ***
civlibs	-0.29797	0.02350	-12.68	<2e-16 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 22875 on 7156 degrees of freedom
Residual deviance: 20675 on 7154 degrees of freedom
(4 observations deleted due to missingness)
AIC: 26862

Number of Fisher Scoring iterations: 6

Overdispersion Test: "By Hand"

```
> Phats<-fitted.values(amici.poisson)
> Uhats<-((amici$namici-Phats)^2 - amici$namici) / (Phats * sqrt(2))
> summary(lm(Uhats~Phats))
```

```
Call:
lm(formula = Uhats ~ Phats)
```

Residuals:

Min	1Q	Median	3Q	Max
-5.9	-3.0	-2.3	-1.9	1707.0

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.579	0.693	2.28	0.023 *
Phats	1.466	0.591	2.48	0.013 *

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 28.4 on 7155 degrees of freedom
Multiple R-squared: 0.000858, Adjusted R-squared: 0.000718
F-statistic: 6.14 on 1 and 7155 DF, p-value: 0.0132

Negative Binomial Regression

```
> library(MASS)
> amici.NB<-glm.nb(namici~term+civlibs,data=amici)
> summary(amici.NB)

Call:
glm.nb(formula = namici ~ term + civlibs, data = amici, init.theta = 0.256657474,
        link = log)

Coefficients:
              Estimate Std. Error z value      Pr(>|z|)
(Intercept) -4.68314    0.22058  -21.23 < 0.0000000000000002 ***
term          0.06573    0.00304   21.60 < 0.0000000000000002 ***
civlibs      -0.26777    0.05403   -4.96    0.00000072 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for Negative Binomial(0.2567) family taken to be 1)

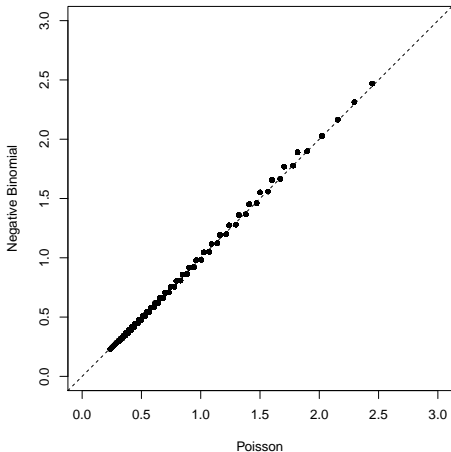
Null deviance: 5442  on 7156  degrees of freedom
Residual deviance: 4968  on 7154  degrees of freedom
AIC: 17378

Number of Fisher Scoring iterations: 1
              Theta:  0.25666
            Std. Err.:  0.00838

> 1 / amici.NB$theta
[1] 3.896
```

Predicted Values: Poisson and NB

```
> plot(amici.poisson$fitted.values,amici.NB$fitted.values,xlab="Poisson",  
      ylab="Negative Binomial",main="Predicted Counts")  
> abline(a=0,b=1,lwd=2)
```



- Models where Over- / Underdispersion = $f(\mathbf{Z}_i\gamma)$
- Models for Censored / Truncated Counts
- “Zero-Inflated” and “Hurdle” Count Models
- Time Series of Counts
- Panel Data Models for Event Counts
- Spatial Count Models
- etc...

Generalized Linear Models (GLMs)

The Exponential Family

$$f(z|\psi) = \Pr(Z = z|\psi)$$

Exponential if:

$$f(z|\psi) = r(z)s(\psi) \exp[q(z)h(\psi)]$$

provided that $r(z) > 0$ and $s(\psi) > 0$.

$$f(z|\psi) = \exp[\underbrace{\ln r(z) + \ln s(\psi)}_{\text{"additive"}} + \underbrace{q(z)h(\psi)}_{\text{"interactive"}}]$$

$$y = q(z)$$

$$\theta = h(\psi)$$

$$f[y|\theta] = \exp[y\theta - b(\theta) + c(y)].$$

- $b(\theta)$ is a “normalizing constant”
- $c(y)$ is a function solely of y
- $y\theta$ is a multiplicative term

A Familiar Family Member: Poisson

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}.$$

$$\begin{aligned} f(y|\lambda) &= \exp \{ \ln [\exp(-\lambda)\lambda^y / y!] \} \\ &= \exp \left[\underbrace{y \ln(\lambda)}_{y\theta} - \underbrace{\lambda}_{b(\theta)} - \underbrace{\ln(y!)}_{c(y)} \right] \end{aligned}$$

$$f(y|\theta, \phi) = \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right]$$

Familiar Family Member II: Normal

$$f(y|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left[\frac{(y - \mu)^2}{2\sigma^2}\right]$$

$$\begin{aligned} f(y|\mu, \sigma^2) &= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}(y^2 - 2y\mu + \mu^2)\right] \\ &= \exp\left[-\frac{1}{2}\ln(2\pi\sigma^2) - \frac{1}{2\sigma^2}y^2 + \frac{1}{2\sigma^2}2y\mu - \frac{1}{2\sigma^2}\mu^2\right] \\ &= \exp\left[\frac{y\mu}{\sigma^2} - \frac{\mu^2}{2\sigma^2} - \frac{y^2}{2\sigma^2} - \frac{1}{2}\ln(2\pi\sigma^2)\right] \\ &= \exp\left\{\frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2}\left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2)\right]\right\} \end{aligned}$$

$$f(y|\mu, \sigma^2) = \exp \left\{ \frac{y\mu - \frac{\mu^2}{2}}{\sigma^2} + \frac{-1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right] \right\}$$

$\theta = \mu$, so:

- $y\theta = y\mu$
- $b(\theta) = \frac{\mu^2}{2}$
- $a(\phi) = \sigma^2$
- $c(y, \phi) = \frac{-1}{2} \left[\frac{y^2}{\sigma^2} + \ln(2\pi\sigma^2) \right]$

Other Family Members

- Binomial (\supset Bernoulli; also Multinomial)
- Exponential
- Gamma
- Logarithmic
- Inverse Gaussian
- Negative Binomial
- others...

$$\begin{aligned}\ln L(\theta, \phi|y) &= \ln f(y|\theta, \phi) \\ &= \ln \left\{ \exp \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \right\} \\ &= \frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi)\end{aligned}$$

$$\begin{aligned}\frac{\partial \ln L(\theta, \phi|y)}{\partial \theta} &\equiv \mathbf{S} = \frac{\partial}{\partial \theta} \left[\frac{y\theta - b(\theta)}{a(\phi)} + c(y, \phi) \right] \\ &= \frac{y - \frac{\partial}{\partial \theta} b(\theta)}{a(\phi)}.\end{aligned}$$

Among family members:

- \mathbf{S} is a sufficient statistic for θ .
- $E(\mathbf{S}) = 0$.
- $\text{Var}(\mathbf{S}) \equiv \mathcal{I}(\theta) = E[(\mathbf{S})^2|\theta]$

$$E(Y) = \frac{\partial}{\partial \theta} b(\theta)$$

and

$$\text{Var}(Y) = a(\phi) \frac{\partial^2}{\partial \theta^2} b(\theta)$$

Example: Poisson Again

$$\begin{aligned} E(Y) &= \frac{\partial}{\partial \theta} \exp(\theta) \\ &= \exp(\theta)|_{\theta=\ln(\lambda)} \\ &= \lambda \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= 1 \times \frac{\partial^2}{\partial \theta^2} \exp(\theta)|_{\theta=\ln(\lambda)} \\ &= \exp[\ln(\lambda)] \\ &= \lambda \end{aligned}$$

Example: Normal Again

$$\begin{aligned} E(Y) &= \frac{\partial}{\partial \theta} \left(\frac{\theta^2}{2} \right) \\ &= \theta|_{\theta=\mu} \\ &= \mu \end{aligned}$$

$$\begin{aligned} \text{Var}(Y) &= \sigma^2 \times \frac{\partial^2}{\partial \theta^2} \left(\frac{\theta^2}{2} \right) \\ &= \sigma^2 \times \frac{\partial}{\partial \theta} \theta \\ &= \sigma^2 \end{aligned}$$

Linear Model(s)

$$Y_i = \mathbf{X}_i\boldsymbol{\beta} + u_i$$

$$E(Y_i) \equiv \boldsymbol{\mu}_i = \mathbf{X}_i\boldsymbol{\beta}$$

The “Generalized” Part

$$g(\mu_i) = \mathbf{X}_i\beta.$$

$$\begin{aligned}\eta_i &= \mathbf{X}_i\beta \\ &= g(\mu_i)\end{aligned}$$

$$\begin{aligned}\mu_i &= g^{-1}(\eta_i) \\ &= g^{-1}(\mathbf{X}_i\beta)\end{aligned}$$

Random component $\sim \text{EF}(\cdot)$ with

$$E(Y_i) = \mu_i.$$

Systematic component:

$$g(\mu_i) = \eta_i$$

“Link function”:

$$g(\mu_i) = \eta_i$$

or

$$g^{-1}(\eta_i) = \mu_i.$$

The Return of The Family

$$\begin{aligned}\theta_i &= g(\mu_i) \\ &= \eta_i \\ &= \mathbf{X}_i\beta\end{aligned}$$

$$g^{-1}(\theta_i) = \eta_i$$

GLM Example: Linear-Normal

$$f(y|\mu, \sigma^2) = \mathcal{N}(\mu, \sigma^2)$$

$$\mu_i = \eta_i$$

$$\begin{array}{rcl} \mu_i \equiv \theta_i & = & \eta_i \\ Y_i & \sim & \mathcal{N}(\mu_i, \sigma^2) \end{array}$$

GLM Example: Binary

$$f(y|\pi) = \pi^y(1 - \pi)^{1-y}$$

$$\theta_i = \ln \left(\frac{\mu_i}{1 - \mu_i} \right)$$

$$\begin{aligned}\mu_i &= g^{-1}(\theta_i) \\ &= \frac{\exp(\eta_i)}{1 + \exp(\eta_i)} \\ Y_i &\sim \text{Bernoulli}(\mu_i)\end{aligned}$$

GLM Example: Counts (Independent Events)

$$f(y|\lambda) = \frac{\exp(-\lambda)\lambda^y}{y!}$$

$$\ln(\lambda_i) = \boldsymbol{\eta}_i$$

$$\boldsymbol{\mu}_i = \boldsymbol{g}^{-1}(\boldsymbol{\theta}_i)$$

$$= \exp(\boldsymbol{\eta}_i)$$

$$Y_i \sim \text{Poisson}(\lambda_i)$$

Common GLM Flavors

Distribution	Range of Y	Link(s) $g(\cdot)$	Inverse Link $g^{-1}(\cdot)$
Normal	$(-\infty, \infty)$	Identity: $\theta = \mu$ (Canonical)	θ
Binomial	$\{0, \dots, n\}$	Logit: $\theta = \ln\left(\frac{\mu}{1-\mu}\right)$ (Canonical) Probit: $\theta = \Phi^{-1}(\mu)$ C-Log-Log: $\theta = \ln[-\ln(1-\mu)]$	$\frac{\exp(\theta)}{1+\exp(\theta)}$ $\Phi(\theta)$ $1 - \exp[-\exp(\theta)]$
Bernoulli	$\{0, 1\}$	(same as Binomial)	(same as Binomial)
Multinomial	$\{0, \dots, J\}$	(same as Binomial)	(same as Binomial)
Poisson	$[0, \infty]$ (integers)	Log: $\theta = \ln(\mu)$ (Canonical)	$\exp(\theta)$
Gamma	$(0, \infty)$	Reciprocal: $\theta = -\frac{1}{\mu}$ (Canonical)	$-\frac{1}{\theta}$

Note: The Bernoulli is a special case of the Binomial with $n = 1$. The multinomial is the J -outcome variant of the Binomial, and is also related to the Poisson (see, e.g., Agresti 2002).

- Pick $f(Y)$
- Pick $g(\cdot)$
- Specify \mathbf{X}
- Estimate

- MLE
- IRLS (\approx MLE):

$$\hat{\boldsymbol{\beta}}^{(t+1)} = [\mathbf{X}'\mathbf{W}^{(t)}\mathbf{X}]^{-1}\mathbf{X}'\mathbf{W}^{(t)}\mathbf{z}^{(t)}$$

with

$$\mathbf{W}_{N \times N}^{(t)} = \text{diag} \left[\frac{\left(\partial \mu_i^{(t)} / \partial \eta_i^{(t)} \right)^2}{\text{Var}(Y_i)} \right]$$

and

$$\mathbf{z}^{(t)} = \boldsymbol{\eta}^{(t)} + (Y - \boldsymbol{\mu}^{(t)}) \left(\frac{\partial \boldsymbol{\eta}^{(t)}}{\partial \boldsymbol{\mu}} \right).$$

At iteration t :

1. Calculate $\mathbf{z}^{(t)}$, $\mathbf{W}^{(t)}$
2. Regress $\mathbf{z}^{(t)}$ on \mathbf{X} , using $\mathbf{W}^{(t)}$ as weights, to obtain $\hat{\boldsymbol{\beta}}^{(t+1)}$
3. Generate $\boldsymbol{\eta}^{(t+1)} = \mathbf{X}\hat{\boldsymbol{\beta}}^{(t+1)}$
4. Generate $\boldsymbol{\mu}^{(t+1)} = g^{-1}(\boldsymbol{\eta}^{(t+1)})$
5. Use $\boldsymbol{\eta}^{(t+1)}$ and $\boldsymbol{\mu}^{(t+1)}$ to calculate $\mathbf{z}^{(t+1)}$ and $\mathbf{W}^{(t+1)}$
6. Repeat until convergence.

“Response” Residuals:

$$\begin{aligned}\hat{u}_i &= Y_i - \hat{\mu}_i \\ &= Y_i - g^{-1}(\mathbf{x}_i \hat{\boldsymbol{\beta}})\end{aligned}$$

“Pearson” Residuals:

$$\hat{p}_i = \frac{\hat{u}_i}{[\text{Var}(\hat{u}_i)]^{1/2}}$$

“Deviance”:

$$\begin{aligned}\hat{d}_i &= -2[\ln L_i(\hat{\theta}) - \ln L_i(\theta_S)] \\ &= 2 \left\{ \left[\frac{Y_i \theta_S - b(\theta_S)}{a(\phi)} + c(Y_i, \phi) \right] - \left[\frac{Y_i \hat{\theta} - b(\hat{\theta})}{a(\phi)} + c(Y_i, \phi) \right] \right\} \\ &= 2 \left[\frac{Y_i(\theta_S - \hat{\theta}) - b(\theta_S) + b(\hat{\theta})}{a(\phi)} \right]\end{aligned}$$

“Deviance” Residuals:

$$\hat{r}_{Di} = \left(\frac{\hat{u}_i}{|\hat{u}_i|} \right) \sqrt{\hat{d}_i^2}$$

Toy Example: Linear-Normal

$$X = \{1, 1, 2, 2, 3, 3, 4, 4, 5, 5\}$$

$$Y = \{0, 2, 1, 3, 2, 4, 3, 5, 4, 6\}$$

$$Y_i = 0 + 1X_i + u_i$$

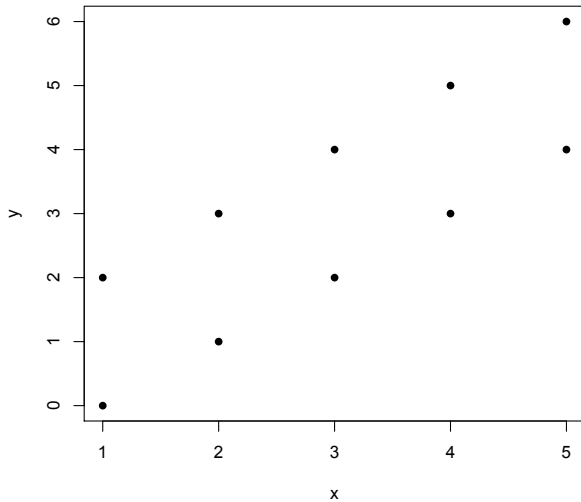
$$\hat{u}_i^2 = 1 \quad \forall i$$

$$\text{"TSS"} \equiv \sum (Y_i - \bar{Y})^2 = 30$$

$$\text{"RSS"} \equiv \sum \hat{u}_i^2 = 10$$

$$\text{"MSS"} / \text{"ESS"} = 20$$

Toy Example: Plot



Toy Example: OLS

```
> linmod<-lm(y~x)
> summary(linmod)
```

```
Call:
lm(formula = y ~ x)
```

Residuals:

Min	1Q	Median	3Q	Max
-1.000e+00	-1.000e+00	1.110e-16	1.000e+00	1.000e+00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.617e-16	8.292e-01	-6.77e-16	1.00000
x	1.000e+00	2.500e-01	4	0.00395 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

Residual standard error: 1.118 on 8 degrees of freedom
Multiple R-squared: 0.6667, Adjusted R-squared: 0.625
F-statistic: 16 on 1 and 8 DF, p-value: 0.00395

Toy Example: Linear-Normal GLM

```
> linalgm<-glm(y~x,family="gaussian")
> summary(linalgm)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-1.000e+00	-1.000e+00	-5.551e-17	1.000e+00	1.000e+00

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	-5.617e-16	8.292e-01	-6.77e-16	1.00000
x	1.000e+00	2.500e-01	4	0.00395 **

Signif. codes: 0 *** 0.001 ** 0.01 * 0.05 . 0.1 1

(Dispersion parameter for gaussian family taken to be 1.25)

Null deviance: 30 on 9 degrees of freedom
Residual deviance: 10 on 8 degrees of freedom
AIC: 34.379

Number of Fisher Scoring iterations: 2

Better GLM Example: Political Knowledge

- 2008 NES political knowledge
- Identify Speaker of the House, VP, British PM, and Chief Justice
- Y_i = number of correct answers (out of four)

$$f(Y_i, p_i) = \binom{4}{Y_i} p_i^{Y_i} (1 - p_i)^{4 - Y_i}$$

$$Y \sim \text{Binomial}(4, p)$$

$$E(Y_i) = \frac{\exp(\mathbf{X}_i \beta)}{1 + \exp(\mathbf{X}_i \beta)}$$

GLM Example Data (2008 NES)

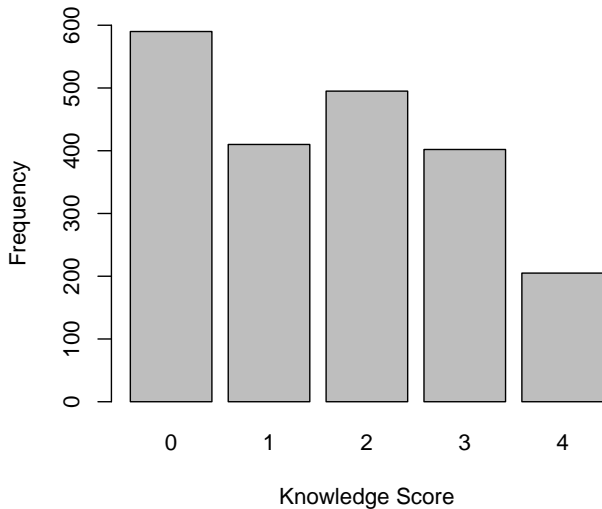
```
> summary(NES08[,4:16])
```

knowledge		sex		race	
Min. :0.00	1. Male respondent selected : 999	1. White	:1442		
1st Qu.:1.00	2. Female respondent selected:1324	2. Black/African-American:	583		
Median :2.00		4. Other race	: 262		
Mean :2.37		5. White and another race:	16		
3rd Qu.:4.00		6. Black and another race:	6		
Max. :4.00		(Other)	: 2		
NA's :221		NA's	: 12		

age		female		white		oftenvote		conservative	
Min. :17	Min. :0.00	Min. :0.0000	Seldom	:621	Min. :1.00				
1st Qu.:33	1st Qu.:0.00	1st Qu.:0.0000	Part of the Time:	287	1st Qu.:3.00				
Median :46	Median :1.00	Median :1.0000	Nearly Always	:612	Median :4.00				
Mean :47	Mean :0.57	Mean :0.6207	Always	:788	Mean :4.14				
3rd Qu.:59	3rd Qu.:1.00	3rd Qu.:1.0000	NA's	: 15	3rd Qu.:5.00				
Max. :90	Max. :1.00	Max. :1.0000			Max. :7.00				
NA's :22					NA's :697				

prayfreq		heterosexual		married		yrsofschool		income	
Never	:235	Min. :0.0000	Min. :0.0000	Min. :0.00	Min. :1.00				
Once/week	:321	1st Qu.:1.0000	1st Qu.:0.0000	1st Qu.:12.00	1st Qu.:5.00				
Few times a week:	416	Median :1.0000	Median :0.0000	Median :13.00	Median :11.00				
Daily	:525	Mean :0.9591	Mean :0.4224	Mean :13.08	Mean :10.52				
Several/Day	:806	3rd Qu.:1.0000	3rd Qu.:1.0000	3rd Qu.:15.00	3rd Qu.:15.00				
NA's	: 20	Max. :1.0000	Max. :1.0000	Max. :17.00	Max. :25.00				
		NA's :49	NA's :15	NA's :11	NA's :151				

Political Knowledge (2008 NES)



GLM Results

```
> nes08.binom<-glm(cbind(knowledge,4-knowledge)~age+female+white+oftenvote+conservative
+prayfreq+heterosexual+married+yrschool+income,data=nes2008,family=binomial)
> summary(nes08.binom)
```

Deviance Residuals:

	Min	1Q	Median	3Q	Max
	-3.59683	-1.01716	0.03124	1.34899	2.85336

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	2.097696	0.248976	8.425	< 2e-16 ***
age	-0.010789	0.001910	-5.650	1.60e-08 ***
female	0.213865	0.059534	3.592	0.000328 ***
white	-0.154109	0.064613	-2.385	0.017073 *
oftenvote	-0.097272	0.027511	-3.536	0.000407 ***
conservative	0.019421	0.019317	1.005	0.314704
prayfreq	0.048818	0.022248	2.194	0.028216 *
heterosexual	0.070894	0.138471	0.512	0.608665
married	-0.166501	0.058363	-2.853	0.004333 **
yrschool	-0.090790	0.013116	-6.922	4.45e-12 ***
income	-0.009015	0.005259	-1.714	0.086492 .

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 3181.4 on 1359 degrees of freedom
Residual deviance: 2952.9 on 1349 degrees of freedom
(963 observations deleted due to missingness)

AIC: 4563.1

Number of Fisher Scoring iterations: 4

GLMs: Other Topics + Extensions

Other Topics:

- Generalizations for Overdispersion (binomial)
- Diagnostics (leverage, etc.)
- Joint Mean-Dispersion Models

Extensions:

- Bias-reduced models (a la Firth 1993)
- “Generalized additive models” (GAMs)
- “Generalized estimating equations” (GEEs)
- “Vector” GLMs (Yee and Wild 1996; Yee and Hastie 2003)

McCullagh, P., and J. A. Nelder. 1989. *Generalized Linear Models*, 2nd Ed. London: Chapman & Hall.

Dobson, Annette J., and and Adrian G. Barnett. 2008. *An Introduction to Generalized Linear Models*, 3rd Ed. London: Chapman & Hall.

Faraway, Julian J. 2006. *Extending the Linear Model with R: Generalized Linear, Mixed Effects, and Nonparametric Regression Models*. London: Chapman & Hall / CRC.

Dunn, Peter K., and Gordon K. Smyth. 2018. *Generalized Linear Models With Examples in R*. New York: Springer.

Hardin, James W., and Joseph W. Hilbe. 2012. *Generalized Linear Models and Extensions*, 3rd Ed. College Station, TX: Stata Press.