GSERM 2017
Regression III
Practical GLMs

June 23, 2017 (morning session)

Models for Binary Responses

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

$$\begin{aligned} Y_i &= 0 \text{ if } Y_i^* < 0 \\ Y_i &= 1 \text{ if } Y_i^* \geq 0 \end{aligned}$$

So:

$$\begin{aligned} \Pr(Y_i = 1) &= \Pr(Y_i^* \geq 0) \\ &= \Pr(\mathbf{X}_i \boldsymbol{\beta} + u_i \geq 0) \\ &= \Pr(u_i \geq -\mathbf{X}_i \beta) \\ &= \Pr(u_i \leq \mathbf{X}_i \beta) \\ &= \int_{-\infty}^{\mathbf{X}_i \boldsymbol{\beta}} f(u) du \end{aligned}$$

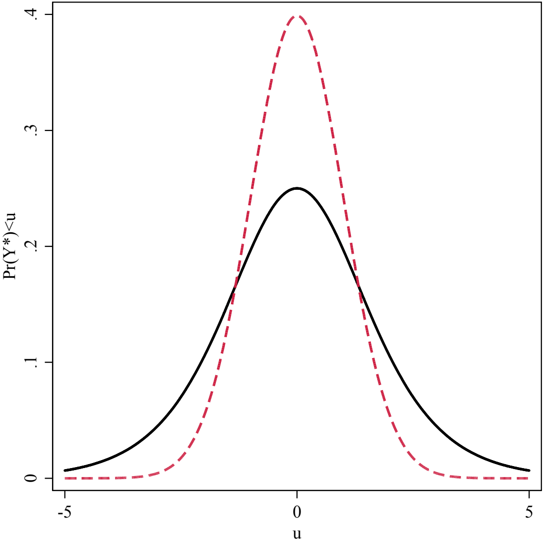
"Standard logistic" PDF:

$$\mathsf{Pr}(u) \equiv \lambda(u) = \frac{\exp(u)}{[1+\exp(u)]^2}$$

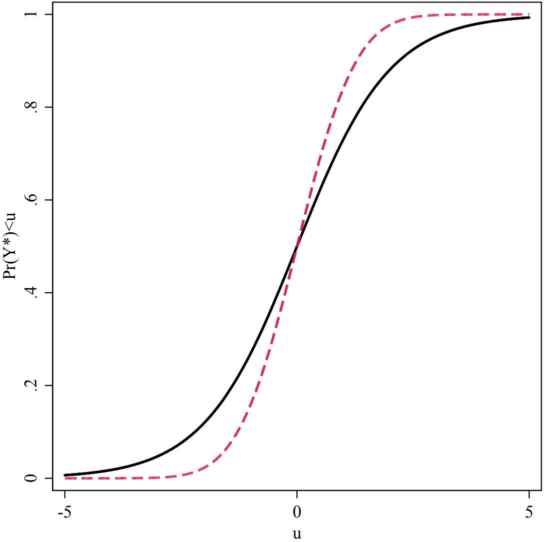
CDF:

$$\begin{aligned}\Lambda(u) &= \int \lambda(u)du \\ &= \frac{\exp(u)}{1+\exp(u)} \\ &= \frac{1}{1+\exp(-u)}\end{aligned}$$

Standard Normal and Logistic PDFs



Standard Normal and Logistic CDFs



- $\lambda(u) = 1 - \lambda(-u)$
- $\Lambda(u) = 1 - \Lambda(-u)$
- $\mathrm{Var}(u) = \frac{\pi^2}{3} \approx 3.29$

Logistic → "Logit"

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Pr(Y_i^* > 0) \\
&= \Pr(u_i \leq \mathbf{X}_i\boldsymbol{\beta}) \\
&= \Lambda(\mathbf{X}_i\boldsymbol{\beta}) \\
&= \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})} \\
\text{(equivalently)} &= \frac{1}{1 + \exp(-\mathbf{X}_i\boldsymbol{\beta})}
\end{aligned}
$$

Likelihoods

$$L_i = \left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)^{Y_i}\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]^{1-Y_i}$$

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

$$\begin{aligned}\ln L \quad = \quad & \sum_{i=1}^{N} Y_i \ln\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right) + \\ & (1-Y_i)\ln\left[1-\left(\frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1+\exp(\mathbf{X}_i\boldsymbol{\beta})}\right)\right]\end{aligned}$$

Be Normal?

$$\mathsf{Pr}(u) \equiv \phi(u) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right)$$

$$\Phi(u) = \int_{-\infty}^{u} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{u^2}{2}\right) du$$

Normal → "Probit"

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Phi(\mathbf{X}_i\boldsymbol{\beta}) \\
&= \int_{-\infty}^{\mathbf{X}_i\boldsymbol{\beta}} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{(\mathbf{X}_i\boldsymbol{\beta})^2}{2}\right) d\mathbf{X}_i\boldsymbol{\beta}
\end{aligned}
$$

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

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

Digression I: Logit as an Odds Model

$$\text{Odds}(Z) \equiv \Omega(Z) = \frac{\Pr(Z)}{1 - \Pr(Z)}.$$

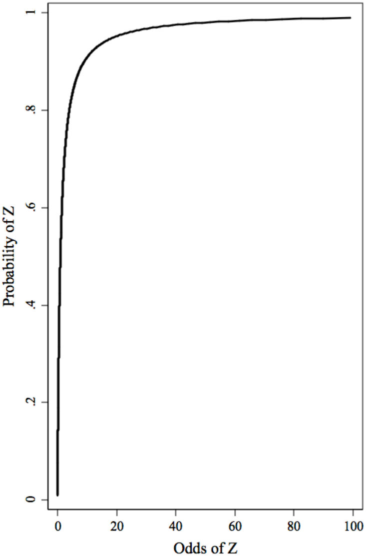
$$\ln[\Omega(Z)] = \ln\left[\frac{\Pr(Z)}{1 - \Pr(Z)}\right]$$

$$\ln[\Omega(Z_i)] = \mathbf{X}_i\boldsymbol{\beta}$$

$$\begin{aligned}\Omega(Z_i) &= \frac{\Pr(Z)}{1 - \Pr(Z)} \\ &= \exp(\mathbf{X}_i\boldsymbol{\beta})\end{aligned}$$

$$\Pr(Z_i) = \frac{\exp(\mathbf{X}_i\boldsymbol{\beta})}{1 + \exp(\mathbf{X}_i\boldsymbol{\beta})}$$

Visualizing Log-Odds



Example: House Voting on NAFTA

- `vote` – Whether (=1) or not (=0) the House member in question voted in favor of NAFTA.
- `democrat` – Whether the House member in question is a Democrat (=1) or a Republican (=0).
- `pcthispc` – The percentage of the House member's district who are of Latino/hispanic origin.
- `cope93` – The 1993 AFL-CIO (COPE) voting score of the member in question; this variable ranges from 0 to 100, with higher scores indicating more pro-labor positions.
- `DemXCOPE` – The multiplicative interaction of `democrat` and `cope93`.

$$\Pr(\texttt{vote}_i = 1) = f[\beta_0 + \beta_1(\texttt{democrat}_i) + \beta_2(\texttt{pcthispc}_i) + \beta_3(\texttt{cope93}_i) + \beta_4(\texttt{democrat}_i \times \texttt{cope93}_i) + u_i]$$

```
> summary(nafta)
      vote            democrat          pcthispc         cope93           DemXCOPE
 Min.   :0.0000   Min.   :0.0000   Min.   : 0.0   Min.   :  0.00   Min.   :  0.00
 1st Qu.:0.0000   1st Qu.:0.0000   1st Qu.: 1.0   1st Qu.: 17.00   1st Qu.:  0.00
 Median :1.0000   Median :1.0000   Median : 3.0   Median : 81.00   Median : 75.00
 Mean   :0.5392   Mean   :0.5853   Mean   : 8.8   Mean   : 60.18   Mean   : 51.65
 3rd Qu.:1.0000   3rd Qu.:1.0000   3rd Qu.:10.0   3rd Qu.:100.00   3rd Qu.:100.00
 Max.   :1.0000   Max.   :1.0000   Max.   :83.0   Max.   :100.00   Max.   :100.00
```

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

or

$$\Pr(Y_i = 1) = \Phi(\mathbf{X}_i\boldsymbol{\beta})$$

Probit Estimates

```
> NAFTA.GLM.probit<-glm(vote~democrat+pcthispc+cope93+DemXCOPE,
  family=binomial(link="probit"))
> summary(NAFTA.GLM.probit)

Call:
glm(formula = vote ~ democrat + pcthispc + cope93 + DemXCOPE,
    family = binomial(link = "probit"))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.07761    0.15339    7.03  2.1e-12 ***
democrat     3.03359    0.73884    4.11  4.0e-05 ***
pcthispc     0.01279    0.00467    2.74   0.0062 **
cope93      -0.02201    0.00440   -5.00  5.8e-07 ***
DemXCOPE    -0.02888    0.00903   -3.20   0.0014 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 598.99  on 433  degrees of freedom
Residual deviance: 441.06  on 429  degrees of freedom
AIC: 451.1
```

Logit Estimates

```
> NAFTA.GLM.logit<-glm(vote~democrat+pcthispc+cope93+DemXCOPE,family=binomial)
> summary(NAFTA.GLM.logit)

Call:
glm(formula = vote ~ democrat + pcthispc + cope93 + DemXCOPE,
    family = binomial)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.79164    0.27544    6.50  7.8e-11 ***
democrat     6.86556    1.54729    4.44  9.1e-06 ***
pcthispc     0.02091    0.00794    2.63  0.00846 **
cope93      -0.03650    0.00760   -4.80  1.6e-06 ***
DemXCOPE    -0.06705    0.01820   -3.68  0.00023 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 598.99  on 433  degrees of freedom
Residual deviance: 436.83  on 429  degrees of freedom
  (1 observation deleted due to missingness)
AIC: 446.8
```

Log-Likelihoods, "Deviance," etc.

- Reports "deviances":
  - "Residual" deviance $= 2(\ln L_S - \ln L_M)$
  - "Null" deviance $= 2(\ln L_S - \ln L_N)$
  - stored in `object$deviance` and `object$null.deviance`
- So:

$$
\begin{aligned}
LR_{\boldsymbol{\beta}=\mathbf{0}} &= 2(\ln L_M - \ln L_N) \\
&= \text{"Null" deviance} - \text{"Residual" deviance}
\end{aligned}
$$

```
> NAFTA.GLM.logit$null.deviance - NAFTA.GLM.logit$deviance
[1] 162.1577
```

Interpretation: "Signs-n-Significance"

For both logit and probit:

- $\hat{\beta}_k > 0 \leftrightarrow \frac{\partial \Pr(Y=1)}{\partial X_k} > 0$
- $\hat{\beta}_k < 0 \leftrightarrow \frac{\partial \Pr(Y=1)}{\partial X_k} < 0$
- $\frac{\hat{\beta}_k}{\hat{\sigma}_k} \sim N(0,1)$

Interactions:

$$\hat{\beta}_{\texttt{cope93}|\texttt{democrat=1}} \equiv \hat{\phi}_{\texttt{cope93}} = \hat{\beta}_3 + \hat{\beta}_4$$

$$\text{s.e.}(\hat{\beta}_{\texttt{cope93}|\texttt{democrat=1}}) = \sqrt{\text{Var}(\hat{\beta}_3) + (\texttt{democrat})^2 \text{Var}(\hat{\beta}_4) + 2\,(\texttt{democrat})\,\text{Cov}(\hat{\beta}_3, \hat{\beta}_4)}$$

$\hat{\phi}_{\texttt{cope93}}$ point estimate:

```
> NAFTA.GLM.logit$coeff[4]+ NAFTA.GLM.logit$coeff[5]

    cope93
-0.1035551
```

z-score ("by hand"):

```
> (NAFTA.GLM.logit $coeff[4]+ NAFTA.GLM.logit $coeff[5]) / (sqrt(vcov(NAFTA.GLM.logit)[4,4] +
  (1)^2*vcov(NAFTA.GLM.logit)[5,5] + 2*1*vcov(NAFTA.GLM.logit)[4,5]))

   cope93
-6.245699
```

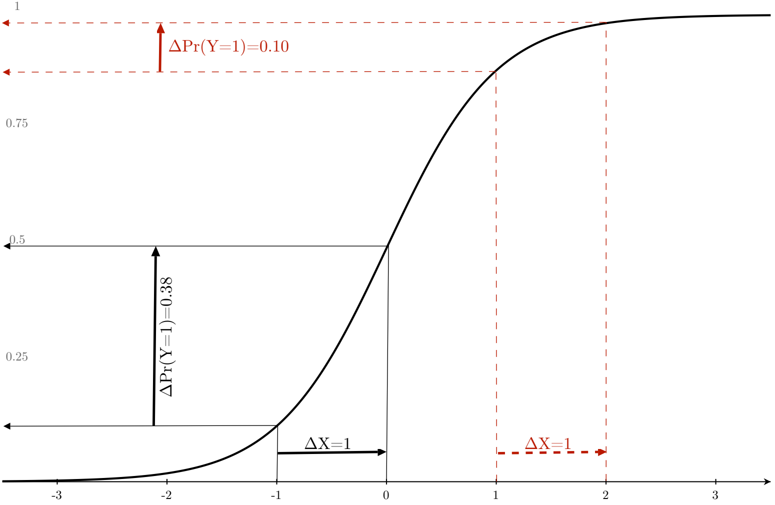
(Or use car...)

```
> library(car)
> linear.hypothesis(NAFTA.GLM.logit,"cope93+DemXCOPE=0")
Linear hypothesis test

Hypothesis:
cope93 + DemXCOPE = 0

Model 1: vote ~ democrat + pcthispc + cope93 + DemXCOPE
Model 2: restricted model

  Res.Df Df  Chisq Pr(>Chisq)
1    429
2    430 -1 39.009  4.219e-10 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
```

Predicted Probabilities

$$
\begin{aligned}
\widehat{\Pr(Y_i = 1)} &= F(\mathbf{X}_i\hat{\boldsymbol{\beta}}) \\
&= \frac{\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})}{1 + \exp(\mathbf{X}_i\hat{\boldsymbol{\beta}})} \text{ for logit,} \\
&= \Phi(\mathbf{X}_i\hat{\boldsymbol{\beta}}) \text{ for probit.}
\end{aligned}
$$

Predicted Probabilities Illustrated



Predicted Probabilities: Standard Errors

$$\begin{aligned} \text{Var}[\widehat{\Pr(Y_i = 1)})] &= \left[\frac{\partial F(\mathbf{X}_i \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}}\right]' \hat{\mathbf{V}} \left[\frac{\partial F(\mathbf{X}_i \hat{\boldsymbol{\beta}})}{\partial \hat{\boldsymbol{\beta}}}\right] \\ &= [f(\mathbf{X}_i \hat{\boldsymbol{\beta}})]^2 \mathbf{X}_i' \hat{\mathbf{V}} \mathbf{X}_i \end{aligned}$$

So,

$$\text{s.e.}[\widehat{\Pr(Y_i = 1)})] = \sqrt{[f(\mathbf{X}_i \hat{\boldsymbol{\beta}})]^2 \mathbf{X}_i' \hat{\mathbf{V}} \mathbf{X}_i}$$

$$\hat{\Delta}\Pr(Y=1)_{\mathbf{X}_A \rightarrow \mathbf{X}_B} = \frac{\exp(\mathbf{X}_B\hat{\boldsymbol{\beta}})}{1+\exp(\mathbf{X}_B\hat{\boldsymbol{\beta}})} - \frac{\exp(\mathbf{X}_A\hat{\boldsymbol{\beta}})}{1+\exp(\mathbf{X}_A\hat{\boldsymbol{\beta}})}$$

or

$$= \Phi(\mathbf{X}_B\hat{\boldsymbol{\beta}}) - \Phi(\mathbf{X}_A\hat{\boldsymbol{\beta}})$$

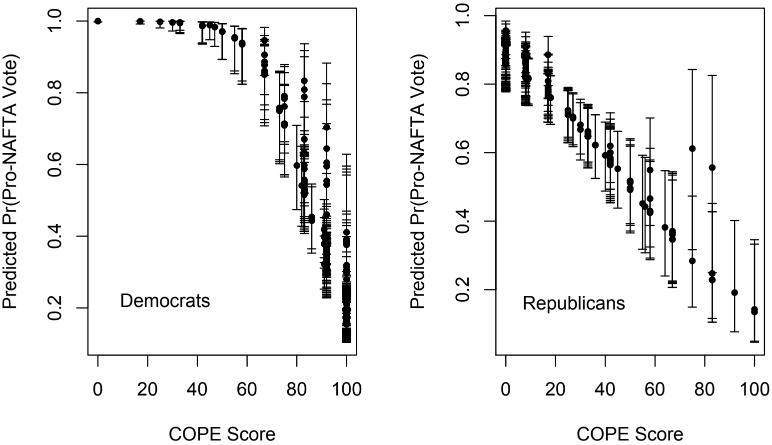
Standard errors obtainable via delta method, bootstrap, etc...

In-Sample Predictions

```
> preds<-NAFTA.GLM.logit$fitted.values

> hats<-predict(NAFTA.GLM.logit,se.fit=TRUE)
> hats
$fit
         1           2           3           4 ...
 9.01267619  7.25223902  6.11013844  5.57444635 ...
 ...
 $se.fit
       1         2         3         4 ...
1.5331506 1.2531475 1.1106989 0.9894208 ...


> XBUB<-hats$fit + (1.96*hats$se.fit)
> XBLB<-hats$fit - (1.96*hats$se.fit)
> plotdata<-cbind(as.data.frame(hats),XBUB,XBLB)
> plotdata<-data.frame(lapply(plotdata,binomial(link="logit")$linkinv))
```

```
...
> par(mfrow=c(1,2))
> library(plotrix)
> plotCI(cope93[democrat==1],plotdata$fit[democrat==1],
  ui=plotdata$XBUB[democrat==1],li=plotdata$XBLB[democrat==1],pch=20,
  xlab="COPE Score",ylab="Predicted Pr(Pro-NAFTA Vote)")
> text(locator(1),label="Democrats")
> plotCI(cope93[democrat==0],plotdata$fit[democrat==0],
  ui=plotdata$XBUB[democrat==0],li=plotdata$XBLB[democrat==0],pch=20,
  xlab="COPE Score",ylab="Predicted Pr(Pro-NAFTA Vote)")
> text(locator(1),label="Republicans")
```

In-Sample Predictions



Out-of-Sample Predictions

"Fake" data:

```
> sim.data<-data.frame(pcthispc=mean(nafta$pcthispc),democrat=rep(0:1,101),
  cope93=seq(from=0,to=100,length.out=101))
> sim.data$DemXCOPE<-sim.data$democrat*sim.data$cope93
```

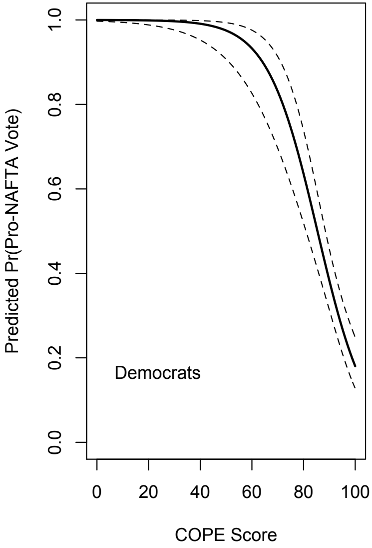
Generate predictions:

```
> OutHats<-predict(NAFTA.GLM.logit,se.fit=TRUE,newdata=sim.data)
> OutHatsUB<-OutHats$fit+(1.96*OutHats$se.fit)
> OutHatsLB<-OutHats$fit-(1.96*OutHats$se.fit)
> OutHats<-cbind(as.data.frame(OutHats),OutHatsUB,OutHatsLB)
> OutHats<-data.frame(lapply(OutHats,binomial(link="logit")$linkinv))
```

Plotting...

```
> par(mfrow=c(1,2))
> both<-cbind(sim.data,OutHats)
> both<-both[order(both$cope93,both$democrat),]

> plot(both$cope93[democrat==1],both$fit[democrat==1],t="l",lwd=2,ylim=c(0,1),
  xlab="COPE Score",ylab="Predicted Pr(Pro-NAFTA Vote)")
> lines(both$cope93[democrat==1],both$OutHatsUB[democrat==1],lty=2)
> lines(both$cope93[democrat==1],both$OutHatsLB[democrat==1],lty=2)
> text(locator(1),label="Democrats")

> plot(both$cope93[democrat==0],both$fit[democrat==0],t="l",lwd=2,ylim=c(0,1),
  xlab="COPE Score",ylab="Predicted Pr(Pro-NAFTA Vote)")
> lines(both$cope93[democrat==0],both$OutHatsUB[democrat==0],lty=2)
> lines(both$cope93[democrat==0],both$OutHatsLB[democrat==0],lty=2)
> text(locator(1),label="Republicans")
```

Out-of-Sample Predictions



Odds Ratios

$$\ln \Omega(\mathbf{X}) = \ln \left[\frac{\frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})}}{1 - \frac{\exp(\mathbf{X}\boldsymbol{\beta})}{1+\exp(\mathbf{X}\boldsymbol{\beta})}} \right] = \mathbf{X}\boldsymbol{\beta}$$

$$\frac{\partial \ln \Omega}{\partial \mathbf{X}} = \boldsymbol{\beta}$$

Odds Ratios

Means:

$$\frac{\Omega(X_k+1)}{\Omega(X_k)} = \exp(\hat{\beta}_k)$$

More generally,

$$\frac{\Omega(X_k+\delta)}{\Omega(X_k)} = \exp(\hat{\beta}_k\delta)$$

$$\text{Percentage Change} = 100[\exp(\hat{\beta}_k\delta) - 1]$$

Odds Ratios Implemented

```
> lreg.or <- function(model)
+           {
+           coeffs <- coef(summary(NAFTA.GLM.logit))
+           lci <- exp(coeffs[ ,1] - 1.96 * coeffs[ ,2])
+           or <- exp(coeffs[ ,1])
+           uci <- exp(coeffs[ ,1] + 1.96 * coeffs[ ,2])
+           lreg.or <- cbind(lci, or, uci)
+           lreg.or
+           }

> lreg.or(NAFTA.GLM.fit)
                lci       or       uci
(Intercept)  3.4966   5.9993 1.029e+01
democrat    46.1944 958.6783 1.990e+04
pcthispc     1.0054   1.0211 1.037e+00
cope93       0.9499   0.9642 9.786e-01
DemXCOPE     0.9024   0.9351 9.691e-01
```

- **Proportional reduction in error (PRE)**
- Pseudo-R^2,
- ROC curves.

PRE

$$\text{PRE} = \frac{N_{MC} - N_{NC}}{N - N_{NC}}$$

- N_{NC} = number correct under the "null model,"
- N_{MC} = number correct under the estimated model,
- N = total number of observations.

PRE: Example

```
> table(NAFTA.GLM.logit$fitted.values>0.5,nafta$vote==1)

       FALSE TRUE
  FALSE   148   49
  TRUE     52  185
```

$$\begin{aligned}\text{PRE} &= \frac{N_{MC} - N_{NC}}{N - N_{NC}} \\ &= \frac{(148 + 185) - 234}{434 - 234} \\ &= \frac{99}{200} \\ &= \mathbf{0.495}\end{aligned}$$

Chi-Square Test (Prediction)

```
> chisq.test(NAFTA.GLM.logit$fitted.values>0.5,nafta$vote==1)

Pearson's Chi-squared test with Yates' continuity correction

data:  NAFTA.GLM.logit$fitted.values > 0.5 and nafta$vote == 1
X-squared = 120.3453, df = 1, p-value < 2.2e-16
```

Models for Event Counts

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

Count Properties

- Discrete / integer-values
- Non-negative
- “Cumulative”

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

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

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

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

Poisson Assumptions

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

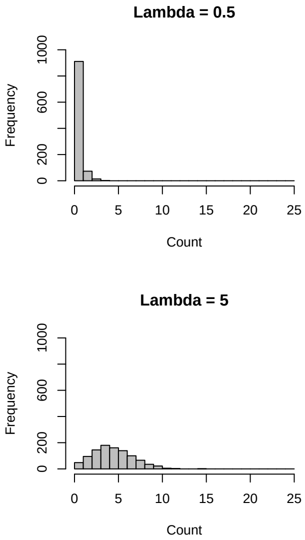
For M independent Bernoulli trials with (sufficiently small) probability of success π and where $M\pi \equiv \lambda > 0$,

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

Poisson: Characteristics

- Discrete
- $\mathrm{E}(Y) = \mathrm{Var}(Y) = \lambda$
- 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.
- $\lambda \to \infty \iff Y \sim N$

Poissons: Examples



Suppose

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

then

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

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

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

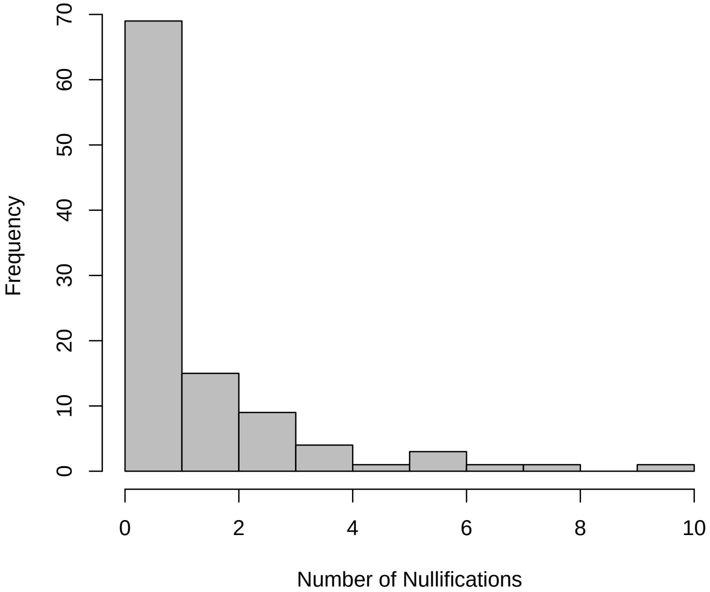
Example: Judicial Review

- 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 ($\bar{X} = 10.4, \sigma = 3.4, \mathrm{E}(\hat{\beta}) > 0$).

- Whether (1) or not (0) there was *unified government* (`unified`) ($\bar{X} = 0.83, \mathrm{E}(\hat{\beta}) < 0$).

Supreme Court Nullifications, 1789-1996



```
> 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

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

- Like ORs
- `unified`: 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

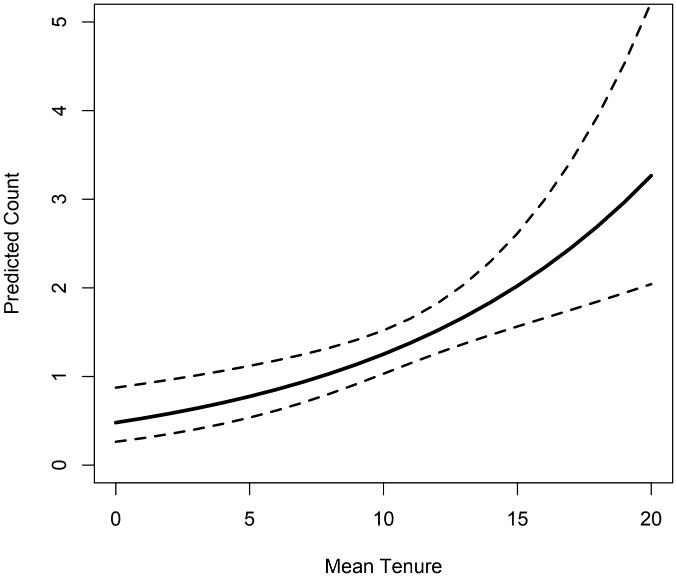
- R : in `$fitted.values`
- Stata : use `predict`

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



Predicted Probabilities

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

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

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

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

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

“Exposure” and “Offsets”

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

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
- `Ndyads` = 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
- `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...

> 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)

Coefficients:
               Estimate  Std. Error z value              Pr(>|z|)
(Intercept) -2.42656676  0.34345252   -7.07      0.0000000000016 ***
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 < 0.0000000000000002 ***
---
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 $\beta_{\texttt{exposure}} = 1.0$

```
> # 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
```

GLM Interpretation, Generally

- Nonlinearity means $\frac{\partial E(Y)}{\partial X} \neq c$.
- $\rightarrow$ obtaining marginal effects requires "holding all else constant"
- Model fit is usually best thought of in predictive terms
- Deviance residuals can be used for diagnostics