PLSC 503 – Spring 2021
Binary Response Models, II

March 31, 2021

Running Example: House Vote on NAFTA (1993)

Response / Outcome

- `vote` – Whether (=1) or not (=0) the House member in question voted in favor of NAFTA.

Predictors

- `pcthispc` – The percentage of the House member's district who are of Latino/hispanic origin.
- `democrat` – Whether the House member in question is a Democrat (=1) or a Republican (=0).
- `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`.

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

```
> 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
```

Basic Model(s)

Logit:

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

or probit:

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

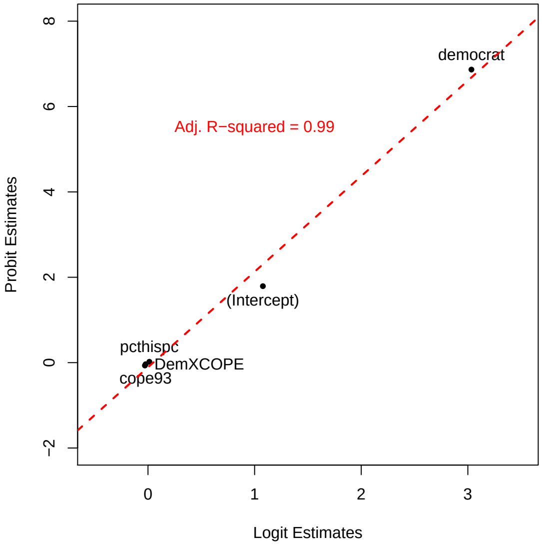
Probit Estimates

```
> 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
```

$\hat{\beta}_{\text{probit}}$ vs. $\hat{\beta}_{\text{logit}}$



Log-Likelihoods, "Deviance," etc.

- `R` / `lm` 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
```

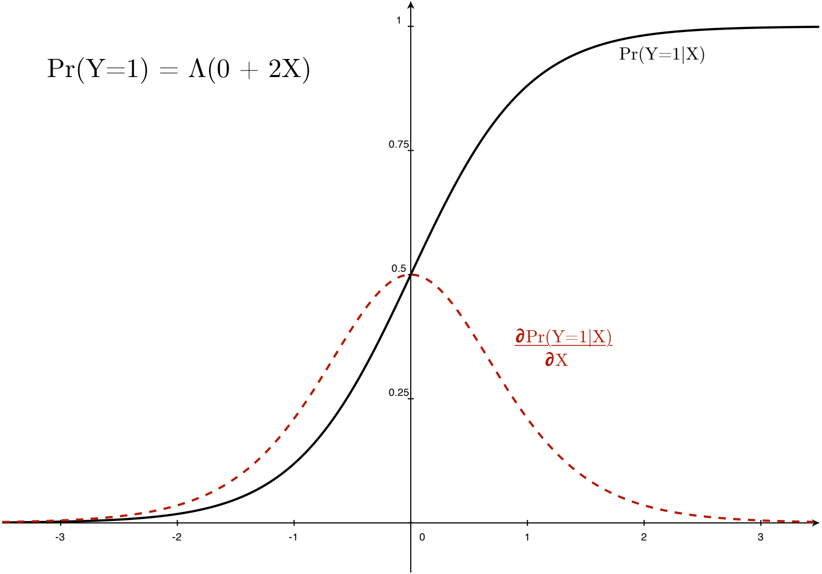
(Or use car...)

```
> 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
```

Marginal Effects

$$\begin{aligned}
\frac{\partial \Pr(\hat{Y}_i = 1)}{\partial X_k} &= \frac{\partial F(\mathbf{X}_i \hat{\boldsymbol{\beta}})}{\partial X_k} \\
&= f(\mathbf{X}_i \hat{\boldsymbol{\beta}}) \hat{\beta}_k \\
&= \Lambda(\mathbf{X}_i \hat{\boldsymbol{\beta}})[1 - \Lambda(\mathbf{X}_i \hat{\boldsymbol{\beta}})] \hat{\beta}_k \quad \text{(logit) or} \\
&= \phi(\mathbf{X}_i \hat{\boldsymbol{\beta}}) \hat{\beta}_k \quad \text{(probit)}
\end{aligned}$$

Marginal Effects Illustrated



Odds Ratios

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

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

Means:

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

More generally,

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

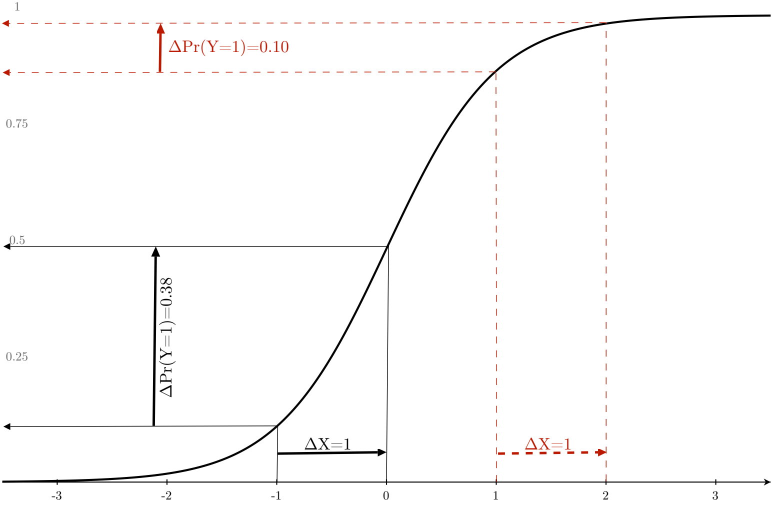
$$\text{Percentage Change} = 100[\exp(\hat{\beta}_k \delta) - 1]$$

```
> 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
```

Predicted Probabilities

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

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,

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

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

or

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

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

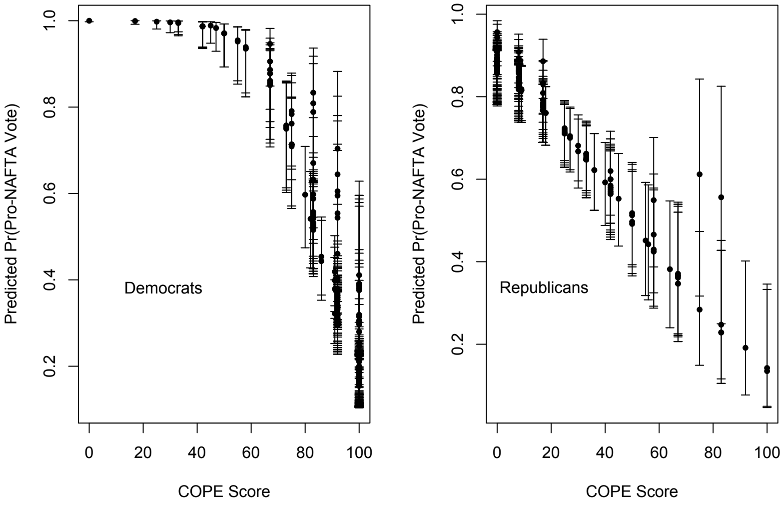
In-Sample Predictions

```
> 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))
```

Plotting

```
...

> 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



"Fake" data:

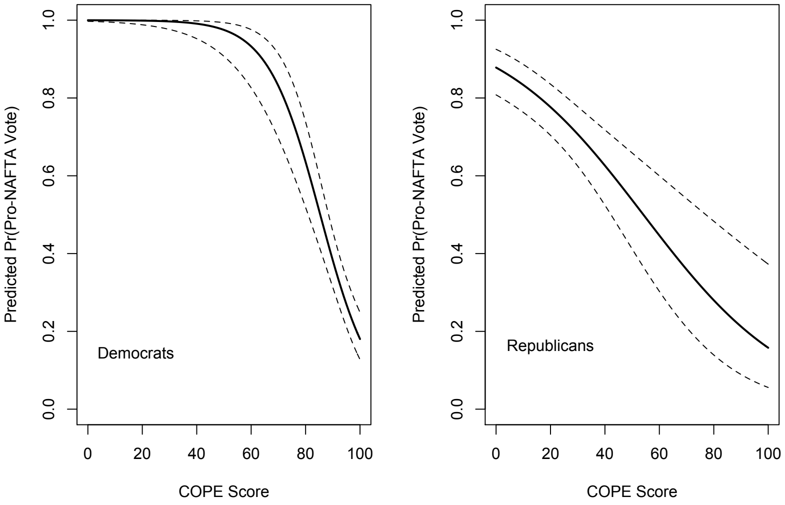
```
> sim.data<-data.frame(pcthispc=mean(nafta$pcthispc),democrat=rep(0:1,101),
  cope93=seq(from=0,to=100,length.out=101))
> sim.data$DemXCOPE<-sim.data$democrat*sim.data$cope93
```

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))
```

```
> 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



Goodness-of-Fit

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

Model Fit: 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:

```
> 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
```

- *Sensitivity*
  - $\Pr(\widehat{Y = 1)}|Y = 1$
  - "true positives"

- *Specificity*
  - $\Pr(\widehat{Y = 0)}|Y = 0$
  - "true negatives"

- $1-$*Specificity* $=$ "false positives"
- $1-$*Sensitivity* $=$ "false negatives"

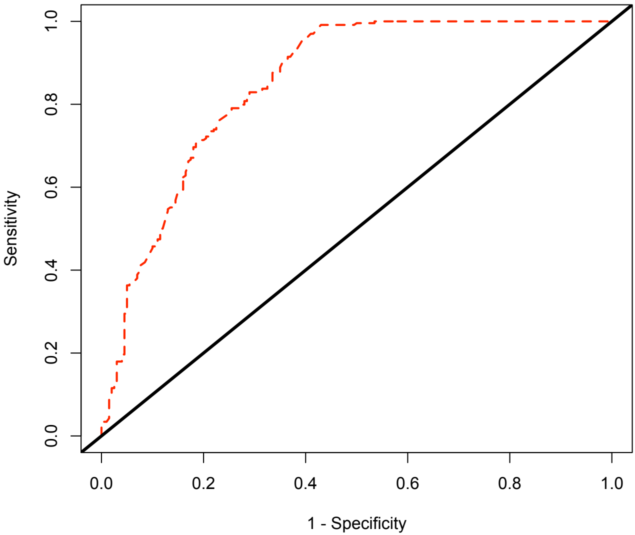
“Receiver Operating Characteristic” (ROC) Curves

- Plot: true positive rate vs. false positive rate (i.e., specificity vs. 1 - sensitivity)
- “aROC”: Area under the curve
- $\rightarrow$ assessment of model fit

ROC Curves Implemented

```
> library(ROCR)

> NAFTA.GLM.logithats<-predict(NAFTA.GLM.logit,
+   type="response")

> preds<-prediction(NAFTA.GLM.logithats,NAFTA$vote)

> plot(performance(preds,"tpr","fpr"),lwd=2,lty=2,
+   col="red",xlab="1 - Specificity",ylab="Sensitivity")

> abline(a=0,b=1,lwd=3)
```

ROC Curve: Example



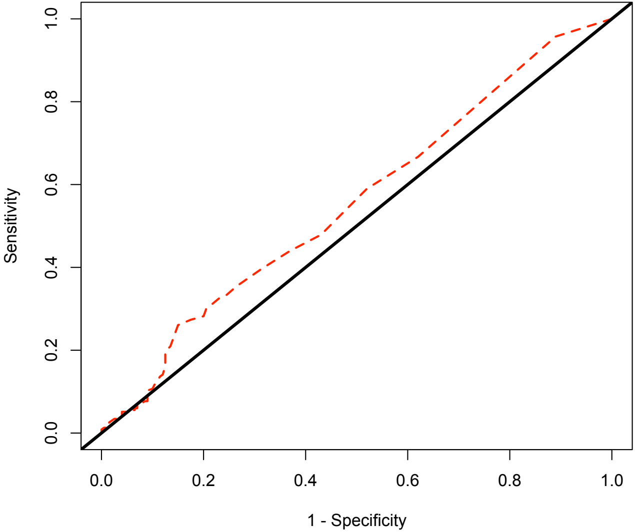
Interpreting ROC Curves

- Area under ROC $=$ 0.90-1.00 $\rightarrow$ Excellent (A)
- Area under ROC $=$ 0.80-0.90 $\rightarrow$ Good (B)
- Area under ROC $=$ 0.70-0.80 $\rightarrow$ Fair (C)
- Area under ROC $=$ 0.60-0.70 $\rightarrow$ Poor (D)
- Area under ROC $=$ 0.50-0.60 $\rightarrow$ Total Failure (F)

ROC Curve: A Poorly-Fitting Model

```
> NAFTA.bad<-glm(vote~pcthispc,family=binomial(link="logit"))
> NAFTA.bad.hats<-predict(NAFTA.bad,type="response")
> bad.preds<-prediction(NAFTA.bad.hats,nafta$vote)

> plot(performance(bad.preds,"tpr","fpr"),lwd=2,lty=2,
+    col="red",xlab="1 - Specificity",ylab="Sensitivity")
> abline(a=0,b=1,lwd=3)
```

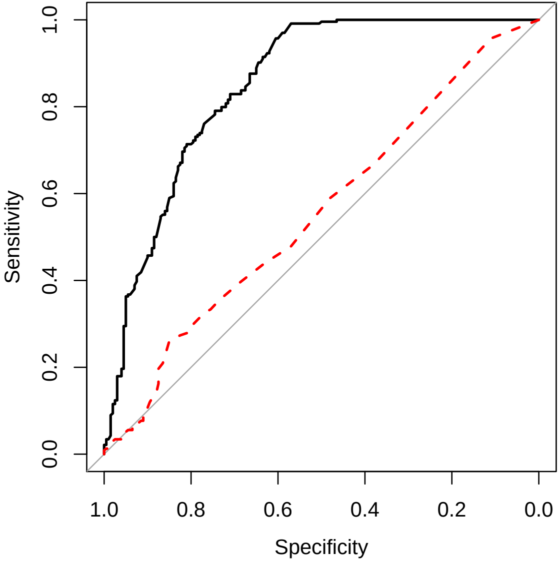
Bad ROC!



Comparing ROCs

```
> install.packages("pROC")
> library(pROC)

> GoodROC<-roc(nafta$vote,NAFTA.GLM.logithats,ci=TRUE)
> GoodAUC<-auc(GoodROC)
> BadROC<-roc(nafta$vote,NAFTA.bad.hats)
> BadAUC<-auc(BadROC)
> GoodAUC
Area under the curve: 0.85
> BadAUC
Area under the curve: 0.556
```

Combined Plot



Extensions: Two Topics, One Theme

- Models for dealing with "separation"
- Models for *rare events*
- Common Focus: Shortage of information on Y

"Separation" = "perfect prediction" = "monotone likelihood"

Intuition:

```
     Dems
Yeas   0   1
   0 178  34
   1   0 219
```

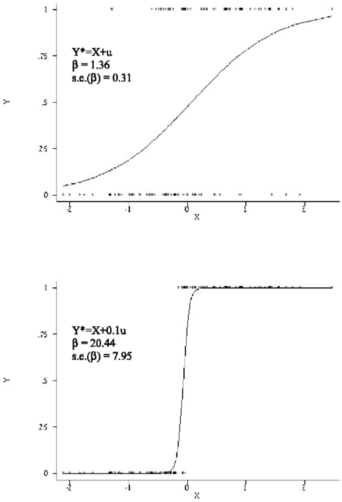
$\Pr(Y = 1 | X = 0) = ?$

Separation: Effects

- $\hat{\beta}_X = \pm\infty$
- $\widehat{\text{s.e.}}_\beta = \infty$
- $\left.\frac{\partial^2 \ln L}{\partial X^2}\right|_{\hat{\beta}} = 0$ (monotone likelihood)

Separation Illustrated

Figure 1: Actual and Predicted Values, Simulated Logistic Regressions



Separation: What Happens

```
> set.seed(7222009)
> Z<-rnorm(500)
> W<-rnorm(500)
> Y<-rbinom(500,size=1,prob=plogis((0.2+0.5*W-0.5*Z)))
> X<-rbinom(500,1,(pnorm(Z)))
> X<-ifelse(Y==0,0,X)  # Induce separation of Y on X


> summary(glm(Y~W+Z+X,family="binomial"))

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   -0.638      0.133   -4.81  1.5e-06 ***
W              0.653      0.140    4.67  3.0e-06 ***
Z             -1.134      0.146   -7.76  8.3e-15 ***
X             20.915    861.458    0.02     0.98
---
Number of Fisher Scoring iterations: 18


> summary(glm(Y~W+Z+X,family="binomial",maxit=100,epsilon=1e-16))

Coefficients:
               Estimate  Std. Error z value Pr(>|z|)
(Intercept)      -0.638       0.133   -4.81  1.5e-06 ***
W                 0.653       0.140    4.67  3.0e-06 ***
Z                -1.134       0.146   -7.76  8.3e-15 ***
X                34.915 5978532.779    0.00        1
---
Number of Fisher Scoring iterations: 32

Warning message:
glm.fit: fitted probabilities numerically 0 or 1 occurred
```

One Solution: Exact Logistic Regression

- Cox (1970, Ch. 4); Hirji et al. (1987 *JASA*); Mehta & Patel (1995 *Stat. Med.*); Forster et al. (2003 *Stat. & Comp.*); Zamar and Graham (2007 *J. Stat. Soft.*).

- Conditions on permutations of covariate patterns

- $\longrightarrow$ Always has finite solutions for $\hat{\beta}$

- Implementation:
    - `elrm` in R (package deprecated); `exlogistic` in Stata
    - Fitted via MCMC; see Forster et al. for details
    - In practice, there are often computational issues...

Firth's (1993) Correction

Firth proposed:

$$L(\boldsymbol{\beta}|Y)^* = L(\boldsymbol{\beta}|Y)\,|\mathbf{I}(\boldsymbol{\beta})|^{\frac{1}{2}}$$

$$\ln L(\boldsymbol{\beta}|Y)^* = \ln L(\boldsymbol{\beta}|Y) + 0.5 \ln |\mathbf{I}(\boldsymbol{\beta})|$$

"Penalized likelihood":

- Is consistent
- Eliminates small-sample bias
- Exist given separation
- To Bayesians, it's "Jeffreys' prior":

$$P(\theta) = \sqrt{\det\left[I(\theta)\right]}$$

Potential Drawbacks

- "Profile" (= "concentrated") likelihood
- $\hat{\beta}$ can be asymmetrical...
- $\rightarrow$ can affect "normal" inference...
- Plotting the profile likelihood and calculating alternative C.I.s is recommended

Software

- R
  - `elrm` (exact logistic regression via MCMC)
  - `brlr` ("bias-reduced logistic regression")
  - `logistf` ("Firth's logistic regression")
- Stata
  - `exlogistic` (exact logistic regression)
  - `firthlogit` (Firth corrected logit)

Example: Pets as Family

- CBS/NYT Poll, April 1997
- Standard political/demographics, plus
- "Do you consider your pet to be a member of your family, or not?"
- Yes $= 84.4\%$, No $= 15.6\%$

Pets as Family: Data

```
> summary(Pets)

   petfamily         female              married              partyid              education
 Min.   :0.000   Min.   :0.000   Married      :442   Democrat   :225   < HS          : 71
 1st Qu.:1.000   1st Qu.:0.000   Widowed      : 46   Independent:214   HS diploma   :244
 Median :1.000   Median :1.000   Divorced/Sep:118   GOP        :229   Some college:184
 Mean   :0.844   Mean   :0.556   NBM          :118   NA's       : 58   College Grad:131
 3rd Qu.:1.000   3rd Qu.:1.000   NA's         :  2                     Post-Grad   : 96
 Max.   :1.000   Max.   :1.000
```

Pets as Family: Basic Model

```
> Pets.1<-glm(petfamily~female+as.factor(married)+as.factor(partyid)
+             +as.factor(education),data=Pets,family=binomial)
> summary(Pets.1)

Coefficients:
                                 Estimate Std. Error z value Pr(>|z|)
(Intercept)                        2.0133     0.5388    3.74  0.00019 ***
femaleMale                        -0.6959     0.2142   -3.25  0.00116 **
as.factor(married)Married         -0.0657     0.2911   -0.23  0.82147
as.factor(married)NBM              0.4599     0.3957    1.16  0.24504
as.factor(married)Widowed         -0.1568     0.4921   -0.32  0.75007
as.factor(partyid)Democrat        -0.1241     0.4286   -0.29  0.77213
as.factor(partyid)GOP             -0.0350     0.4321   -0.08  0.93537
as.factor(partyid)Independent     -0.1521     0.4299   -0.35  0.72338
as.factor(education)College Grad   0.2511     0.4121    0.61  0.54228
as.factor(education)HS diploma     0.0595     0.3685    0.16  0.87182
as.factor(education)Post-Grad      0.1946     0.4331    0.45  0.65321
as.factor(education)Some college   0.0587     0.3867    0.15  0.87928
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 627.14  on 723  degrees of freedom
Residual deviance: 612.76  on 712  degrees of freedom
AIC: 636.8

Number of Fisher Scoring iterations: 4
```

Pets as Family: More Complicated Model

```
> Pets.2<-glm(petfamily~female+as.factor(married)*female+as.factor(partyid)+
+               as.factor(education),data=Pets,family=binomial)

> summary(Pets.2)

Coefficients:
                                      Estimate Std. Error z value Pr(>|z|)
(Intercept)                             2.2971     0.6166    3.73   0.0002 ***
femaleMale                             -1.1833     0.5305   -2.23   0.0257 *
as.factor(married)Married              -0.3218     0.4470   -0.72   0.4716
as.factor(married)NBM                   0.1854     0.6140    0.30   0.7628
as.factor(married)Widowed              -0.7415     0.5780   -1.28   0.1995
as.factor(partyid)Democrat             -0.1575     0.4297   -0.37   0.7140
as.factor(partyid)GOP                  -0.0445     0.4334   -0.10   0.9182
as.factor(partyid)Independent          -0.1757     0.4312   -0.41   0.6837
as.factor(education)College Grad        0.2332     0.4137    0.56   0.5730
as.factor(education)HS diploma          0.0558     0.3703    0.15   0.8801
as.factor(education)Post-Grad           0.2171     0.4342    0.50   0.6171
as.factor(education)Some college        0.0358     0.3890    0.09   0.9266
femaleMale:as.factor(married)Married    0.4853     0.5908    0.82   0.4114
femaleMale:as.factor(married)NBM        0.5260     0.8051    0.65   0.5136
femaleMale:as.factor(married)Widowed   15.2516   549.3719    0.03   0.9779
---

    Null deviance: 627.14  on 723  degrees of freedom
Residual deviance: 607.42  on 709  degrees of freedom
AIC: 637.4

Number of Fisher Scoring iterations: 14
```

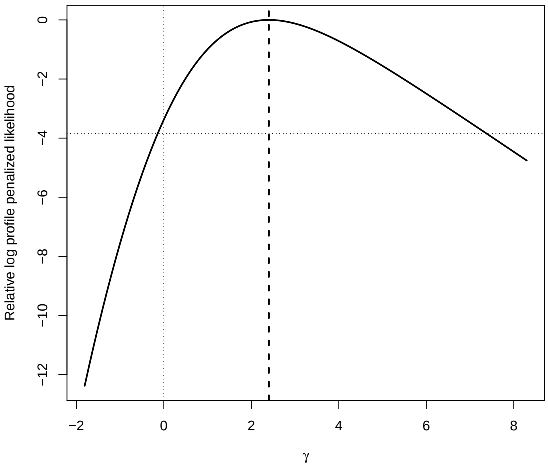
What's Going On?

```
> xtabs(~petfamily+as.factor(married)+female)
, , female = 0

         as.factor(married)
petfamily Married Widowed Divorced/Sep NBM
        0      47       0           11   8
        1     168       7           33  47

, , female = 1

         as.factor(married)
petfamily Married Widowed Divorced/Sep NBM
        0      28       7            7   5
        1     199      32           67  58
```

Pets as Family: Firth Model

```
> Pets.Firth<-logistf(petfamily~female+
+                       as.factor(married)*female+as.factor(partyid)+
+                       as.factor(education),data=Pets)

> Pets.Firth

logistf(formula = petfamily ~ female + as.factor(married) * female +
    as.factor(partyid) + as.factor(education), data = Pets)
Model fitted by Penalized ML
Confidence intervals and p-values by Profile Likelihood

                                          coef se(coef) lower 0.95 upper 0.95    Chisq         p
(Intercept)                            2.15893    0.597      1.054      3.404 16.17636 0.0000577
femaleMale                            -1.13866    0.517     -2.187     -0.145  5.04186 0.0247420
as.factor(married)Married             -0.27387    0.433     -1.192      0.531  0.41518 0.5193531
as.factor(married)NBM                  0.15888    0.588     -0.991      1.367  0.07322 0.7867048
as.factor(married)Widowed             -0.72627    0.561     -1.839      0.384  1.67233 0.1959467
as.factor(partyid)Democrat            -0.11818    0.418     -0.992      0.661  0.08159 0.7751592
as.factor(partyid)GOP                 -0.00776    0.422     -0.888      0.780  0.00034 0.9852893
as.factor(partyid)Independent         -0.13643    0.419     -1.013      0.646  0.10813 0.7422784
as.factor(education)College Grad       0.23904    0.405     -0.574      1.024  0.34480 0.5570689
as.factor(education)HS diploma         0.07531    0.362     -0.667      0.763  0.04289 0.8359331
as.factor(education)Post-Grad          0.21837    0.425     -0.627      1.050  0.26307 0.6080189
as.factor(education)Some college       0.05240    0.380     -0.721      0.781  0.01888 0.8906980
femaleMale:as.factor(married)Married   0.45582    0.577     -0.661      1.613  0.63550 0.4253467
femaleMale:as.factor(married)NBM       0.52329    0.779     -1.023      2.050  0.45133 0.5017022
femaleMale:as.factor(married)Widowed   2.40167    1.684     -0.139      7.374  3.37453 0.0662116

Likelihood ratio test=17.3 on 14 df, p=0.242, n=724
```

Profile Likelihood Plot



Note: Plot shows estimated profile likelihood for different values of the parameter estimate for the interaction term `femaleMale:as.factor(married)Widowed`. Horizontal dotted line is the likelihood associated with $P \leq 0.05$. Vertical dashed line is $\hat{\gamma}$; vertical dotted line indicates $\hat{\gamma} = 0$.

Wrap-Up

- Separation $\nrightarrow$ dropping covariates!
- Firth's approach $>$ ELR
- Can also be applied to other sparse-data situations (e.g., Cox's (1972) proportional hazards model)...

"Rare" Events

- Collect lots of "0s" for a few "1s"
- Classification bias...

Suppose

$$\Pr(Y_i) = \Lambda(0 + 1X_i)$$

Then

$$E(\hat{\beta}_0 - \beta_0) \approx \frac{\bar{\pi} - 0.5}{N\bar{\pi}(1 - \bar{\pi})}$$

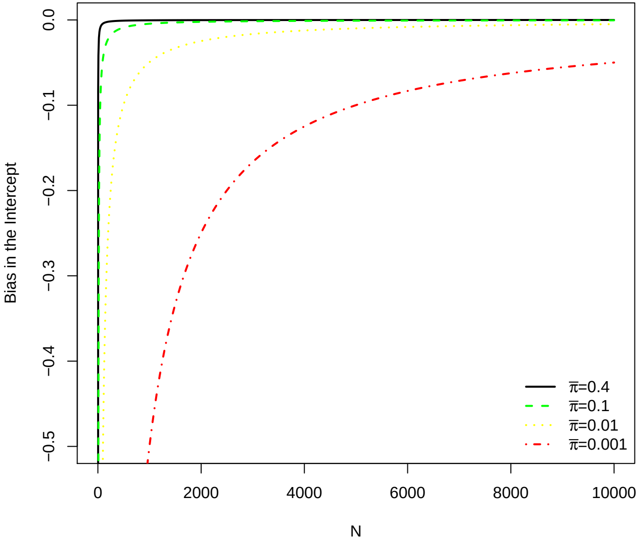
where $\bar{\pi} = \overline{\Pr(Y = 1)}$ is < 0.5.

Bias is:

- always negative,
- worse as $\bar{\pi} \to 0$ (for fixed N),
- disappearing as $N \to \infty$.

Implication: *Logit/probit "work best" around* $\bar{\pi} = 0.5$.

Rare Event Bias, Illustrated



- Calculate $\tau = \frac{N_1 s}{N}$
- Collect data on all "1s"
- Sample from the "0s"
- Estimate a logit*
- *Correct* the estimates ex post...

Sampling...

- $\tau =$ fraction of "1s" in the population
- $\bar{Y} =$ fraction of '1s" in the sample
- K&Z suggest $\bar{Y} \in [0.2, 0.5]$

Weighting...

$$w_1 = \frac{\tau}{\bar{Y}} \text{ (weights for "1s")}$$

$$w_0 = \frac{1-\tau}{1-\bar{Y}} \text{ (weights for "0s")}$$

$$\ln L(\beta|Y) = \sum_{i=1}^{N} w_1 Y_i \ln \Lambda(\mathbf{X}_i \boldsymbol{\beta}) + w_0(1 - Y_i) \ln[1 - \Lambda(\mathbf{X}_i \boldsymbol{\beta})]$$

- Good under (possible) misspecification, but
- Not as efficient as “prior correction,” and
- Gets s.e.s wrong...

Case-Control Data: Prior Correction

$$\hat{\beta}_{0\mathsf{pc}} = \hat{\beta}_0 - \ln\left[\left(\frac{1-\tau}{\tau}\right)\left(\frac{\bar{Y}}{1-\bar{Y}}\right)\right]$$

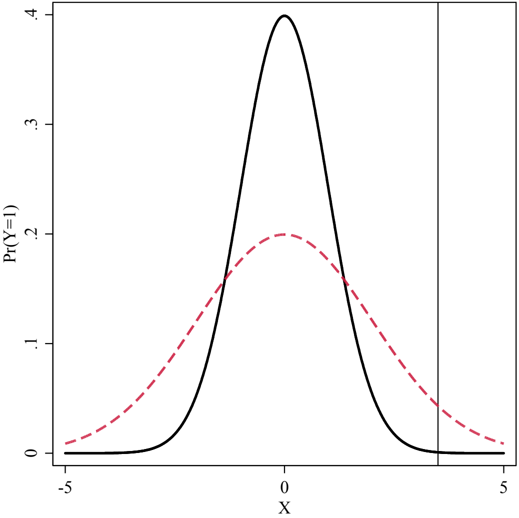
$$\mathsf{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi$$

where $\xi = f[w_i, \hat{\pi}_i, \mathbf{X}]$.

Correction is

$$\tilde{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}} - \mathsf{bias}(\hat{\boldsymbol{\beta}})$$

- Bias correction introduces additional variability...
- Ignoring it yields underpredictions (again).



Pr(Y=1)
0
.1
.2
.3
.4
-5
0
5
X

Use:

$$\Pr(Y_i = 1) \approx \tilde{\pi}_i + C_i$$

where

$$C_i = (0.5 - \tilde{\pi}_i)\tilde{\pi}_i(1 - \tilde{\pi}_i)\mathbf{X}_i\mathbf{V}(\tilde{\beta})\mathbf{X}_i'$$

An Example

- Oneal and Russett 1997; also Beck/Katz/Tucker (1998) etc.
- International disputes
- Politically-relevant dyad-years, 1950-1985
- NT=20448, 405 dyad-years of disputes.

```
> baselogit<-glm(dispute~democracy+growth+allies+contig+capratio+trade,
+                data=RE,family=binomial)
> summary(baselogit)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.32668    0.11451 -37.785  < 2e-16 ***
dembkt       -0.40120    0.10063  -3.987 6.70e-05 ***
grobkt       -3.42753    1.25181  -2.738  0.00618 **
allies       -0.47969    0.11275  -4.255 2.09e-05 ***
contig        1.35358    0.12091  11.195  < 2e-16 ***
capbkt       -0.19620    0.05011  -3.916 9.01e-05 ***
trade       -21.07611   11.30396  -1.864  0.06225 .
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

    Null deviance: 3978.5  on 20447  degrees of freedom
Residual deviance: 3693.8  on 20441  degrees of freedom
AIC: 3707.8

Number of Fisher Scoring iterations: 9
```

```
> set.seed(7222009)
> REones<-RE[dispute==1,]
> REzeros<-RE[dispute==0,]
> RSzeros<-REzeros[sample(1:nrow(REzeros),1000,replace=FALSE),]
> REsample<-data.frame(rbind(REones,RSzeros))
> table(REsample$dispute)

   0    1
1000  405

> sample.logit<-glm(dispute~democracy+growth+allies+contig+capratio+trade,
+                   data=REsample,family=binomial)
> summary(sample.logit)

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.38613    0.12864 -10.776  < 2e-16 ***
democracy    -0.48919    0.11994  -4.078 4.53e-05 ***
growth       -2.18686    1.58474  -1.380 0.167601
allies       -0.33980    0.14240  -2.386 0.017021 *
contig        1.22052    0.14648   8.333  < 2e-16 ***
capratio     -0.18556    0.05149  -3.604 0.000314 ***
trade       -14.63815   11.01629  -1.329 0.183923
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1687.6  on 1404  degrees of freedom
Residual deviance: 1496.2  on 1398  degrees of freedom
AIC: 1510.2
```

Rare Events Logit, Prior Correction

```
> relogit.pc<-zelig(dispute~democracy+growth+allies+contig+capratio+trade,
+                   data=REsample,model="relogit",tau=405/20448,case.control=c("prior"))

> summary(relogit.pc)
Model:

Call:
z5$zelig(formula = dispute ~ democracy + growth + allies + contig +
    capratio + trade, tau = 405/20448, case.control = c("prior"),
    data = REsample)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.4227  -0.1854  -0.1345   2.4056   3.7820

Coefficients:
             Estimate Std. Error z value Pr(>|z|)
(Intercept)  -4.38653    0.12864 -34.100  < 2e-16
democracy    -0.48918    0.11994  -4.078 4.53e-05
growth       -2.13931    1.58474  -1.350 0.177034
allies       -0.33824    0.14240  -2.375 0.017535
contig        1.21645    0.14648   8.305  < 2e-16
capratio     -0.18509    0.05149  -3.595 0.000325
trade       -14.63975   11.01629  -1.329 0.183875

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 1687.6  on 1404  degrees of freedom
Residual deviance: 1496.2  on 1398  degrees of freedom
AIC: 1510.2
```

Rare Events Logit, Weighting Correction

```
> relogit.wc<-zelig(dispute~democracy+growth+allies+contig+capratio+trade,
+                   data=REsample,model="relogit",tau=405/20448,case.control=c("weighting"))

> summary(relogit.wc)
Model:

Call:
relogit(formula = cbind(dispute, 1 - dispute) ~ democracy + growth +
    allies + contig + capratio + trade, data = as.data.frame(.),
    tau = 0.019806338028169, bias.correct = TRUE, case.control = "weighting")

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-0.5285  -0.2185  -0.1578   0.6278   0.9919

Coefficients:
             Estimate Std. Error (robust) z value Pr(>|z|)
(Intercept)  -4.34259             0.13124 -33.089  < 2e-16 ***
democracy    -0.45186             0.11965  -3.776 0.000159 ***
growth       -2.85339             1.67500  -1.704 0.088473 .
allies       -0.41101             0.15008  -2.739 0.006169 **
contig        1.23671             0.15810   7.822 5.18e-15 ***
capratio     -0.18146             0.06188  -2.932 0.003364 **
trade       -12.44992            13.23500  -0.941 0.346868
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 273.37  on 1404  degrees of freedom
Residual deviance: 254.80  on 1398  degrees of freedom
AIC: 53.703
```

A Warning...

From the R documentation:

> *Differences with Stata Version*
>
> *"The Stata version of ReLogit and the R implementation differ slightly in their coefficient estimates due to differences in the matrix inversion routines implemented in R and Stata. Zelig uses orthogonal-triangular decomposition (through lm.influence()) to compute the bias term, which is more numerically stable than standard matrix calculations."*

Some Final Thoughts

- `Zelig` also implements functions for interpreting rare-events logistic regression (marginal effects, etc.)
- Key: be able to conduct C-C sampling *in advance*
- BUT: `Zelig` is currently removed from CRAN (its dependencies are all messed up...)
- In practice: Firth's approach is generally superior to King/Zeng (and should arguably *always* be used for binary-response regressions, especially with small-to-medium Ns)
- Also: Remember that as your N gets big, the problem goes away; Paul Allison has a (old, but useful) blog post on that topic.

Other Binary-Response Extensions

Things we might talk about later:

- Binary responses in panel / longitudinal data
- Multilevel / hierarchical models for binary responses
- Models with (binary) sample selection
- Measurement models for binary outcomes (e.g., item response models)
- Semi- and non-parametric models (see, e.g., Horowitz and Savin 2001)
- "Heteroscedastic" models (where $\sigma_i^2 \neq \sigma^2 \, \forall \, i$) (see, e.g., Alvarez and Brehm 1995, 1997; Tutz 2018)
- "Bivariate" probit models, where

$$\{Y_{1i}, Y_{2i}\} \sim BVN(0, 0, 1, 1, \rho)$$

(e.g., Zorn 2002)