PLSC 504 – Fall 2024

Regression Models for Binary Responses

September 4, 2024

Binary Outcomes: Quick Review

Latent:

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

Observed:

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

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:

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

CDF:

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

Logistic $\rightarrow$ "Logit"

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

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

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

Normal → "Probit"

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

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

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

Logit and Probit, Explained

Things we talked about at length in PLSC 503 (here and here; code here and here):

- Odds ratios and the random utility model
- Model estimation and interpretation
- Marginal effects, predictions, etc.
- Assessing model fit
- A couple variants (e.g., c-log-log)

Things:

- Models for dealing with "separation"
- Models for *rare events*

Common Focus: Shortage of information on Y

Separation

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

Intuition: House votes on the PPACA (3/21/2010)

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

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

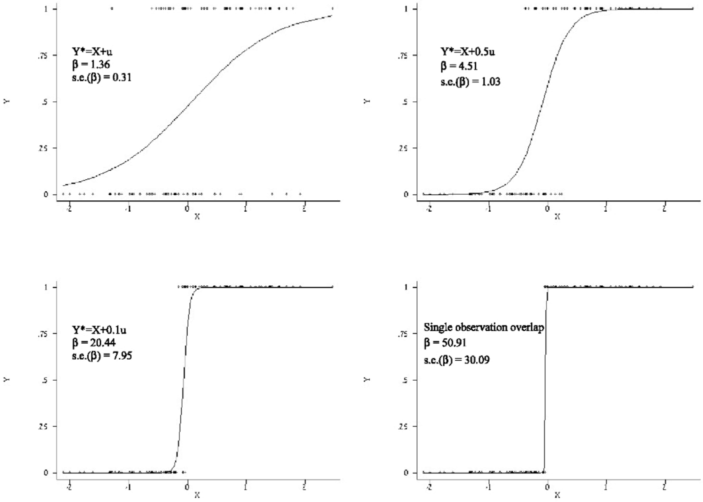
Separation: Effects

"Separation" means that:

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


# Change the maximum # of iterations / convergence tolerance:

> 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

Exact logistic regression (ELR):

- 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 ; `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":

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

- "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

Two directions:

- 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)

Some data, and a silly question:

- 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 $= 15.6\%$

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
```

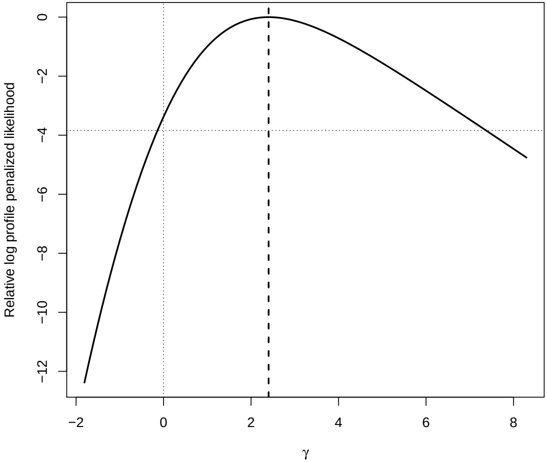
What's Going On?

```
> 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 is an *estimation* problem...
- Separation $\nrightarrow$ dropping covariates!
- Firth's approach $>$ ELR
- Can also be applied to other sparse-data situations:
    - "Fixed effects" logit models (Cook et al. 2020)
    - Multinomial logit (Cook et al. 2018)
    - Survival models (Anderson et al. 2020)

Finally: Read this twitter thread before it's gone.

"Rare" Events

If events ("1s") are rare, we can...

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

Example: Suppose that:

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

then:

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

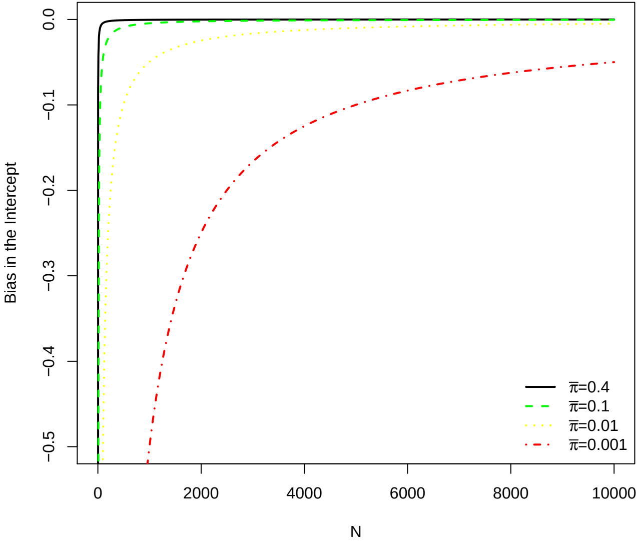
where $\bar{\pi} = \overline{\Pr(Y = 1)}$ is < 0.5.

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...

- τ = 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...

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

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

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

Weighting:

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

Case-Control Data: Prior Correction

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

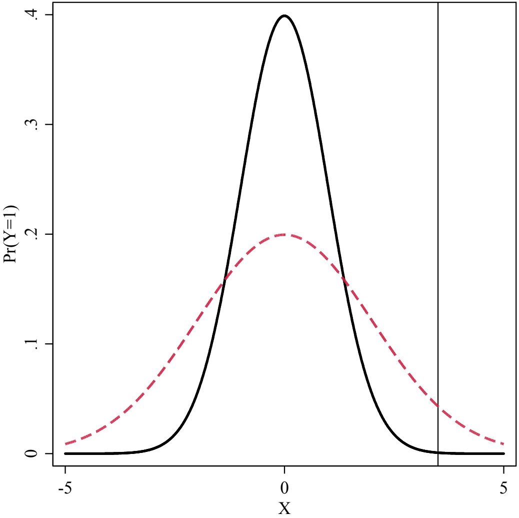
$$\mathrm{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi$$

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

Correction is

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

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



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

Post-Correction Adjustments

Use:

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

where

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

Puhr et al. (2017) note that Firth's method indices bias (toward 0.5) in predicted probabilities, and that the bias is worse when the baseline $\Pr(Y_i = 1)$ is low.

They introduce two modifications to deal with this:

- "Firth's logit with intercept correction" (FLIC)
- "Firth's logit with added covariate" (FLAC)

Through simulations, they show that both remove the bias; they have a slight preference for FLAC, but note that both work well relative to unmodified Firth regression.

An Example

- Washington University's American Panel Study (TAPS)
- $N \approx 1000$ U.S. respondents, 2012-2017
- Outcome: "During the past year, have you ever run out of gas while driving a car or other vehicle?" (`RunOutOfGas`; 0=no, 1=yes)
- Predictors:
  - `Education` – twelve-category ordinal variable with values ranging from 3 to 15;
  - `Income` – a 15-category ordinal variable (each unit roughly corresponds to an increase of $10,000 in annual income);
  - `Age` in years, as of 2016 (divided by 10);
  - `Female` – a binary indicator of sex, naturally-coded;
  - Racial classifications – binary variables for `White`, `Black`, and `Asian` identification;
  - Binary political party variables for `Democrat` and `GOP`; and
  - `Ideology` – a seven-point Likert variable, higher values indicate greater political conservatism

Basic Logit...

```
> table(TAPS$RunOutOfGas)

  0   1
943  28

> prop.table(table(TAPS$RunOutOfGas))

     0      1
0.9712 0.0288

> ROGlogit<-glm(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                      Democrat+GOP+Ideology,data=TAPS,family=binomial)

> summary(ROGlogit)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.661  -0.248  -0.206  -0.170   2.962

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.9347     1.8114   -1.07    0.285
Education    -0.1185     0.1118   -1.06    0.289
Age10        -0.2107     0.1341   -1.57    0.116
Female        0.2911     0.3966    0.73    0.463
White         0.4348     0.7260    0.60    0.549
Black         1.3503     0.7602    1.78    0.076 .
Asian         1.8616     0.8717    2.14    0.033 *
Democrat      0.2743     0.4999    0.55    0.583
GOP          -0.3170     0.5926   -0.53    0.593
Ideology      0.0217     0.1097    0.20    0.843
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 253.77  on 970  degrees of freedom
Residual deviance: 238.13  on 961  degrees of freedom
AIC: 258.1
```

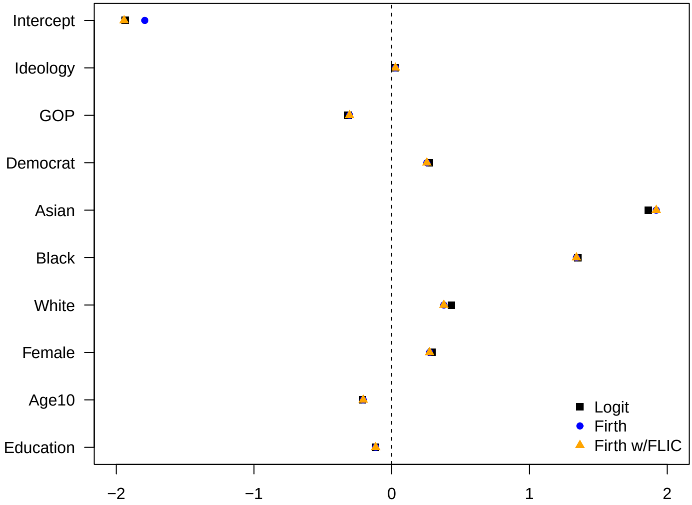
Firth Logit (for comparison)

```
> relogit.firth<-logistf(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                      Democrat+GOP+Ideology,data=TAPS)

> summary(relogit.firth)

logistf(formula = RunOutOfGas ~ Education + Age10 + Female +
    White + Black + Asian + Democrat + GOP + Ideology, data = TAPS)

Model fitted by Penalized ML
Coefficients:
               coef se(coef) lower 0.95 upper 0.95  Chisq      p method
(Intercept) -1.7929    1.657     -5.362     1.6045 1.0457 0.3065      2
Education   -0.1167    0.103     -0.331     0.1009 1.1154 0.2909      2
Age10       -0.2071    0.124     -0.469     0.0498 2.4952 0.1142      2
Female       0.2749    0.367     -0.478     1.0490 0.5124 0.4741      2
White        0.3782    0.646     -1.007     1.7513 0.2769 0.5987      2
Black        1.3409    0.677     -0.182     2.7141 2.9875 0.0839      2
Asian        1.9202    0.766      0.149     3.4429 4.4610 0.0347      2
Democrat     0.2550    0.464     -0.688     1.2418 0.2767 0.5989      2
GOP         -0.3061    0.546     -1.479     0.7889 0.2969 0.5858      2
Ideology     0.0267    0.101     -0.191     0.2333 0.0613 0.8044      2

Method: 1-Wald, 2-Profile penalized log-likelihood, 3-None

Likelihood ratio test=17.5 on 9 df, p=0.0415, n=971
Wald test = 318 on 9 df, p = 0
```

Firth Logit with FLIC

```
> relogit.flic<-logistf(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                           Democrat+GOP+Ideology,data=TAPS,flic=TRUE)

> summary(relogit.flic)

logistf(formula = RunOutOfGas ~ Education + Age10 + Female +
    White + Black + Asian + Democrat + GOP + Ideology, data = TAPS,
    flic = TRUE)

Model fitted by Penalized ML
Coefficients:
               coef se(coef) lower 0.95 upper 0.95  Chisq      p method
(Intercept) -1.9430    1.807     -5.486     1.5995 1.0457 0.3065      1
Education   -0.1167    0.112     -0.331     0.1009 1.1154 0.2909      2
Age10       -0.2071    0.134     -0.469     0.0498 2.4952 0.1142      2
Female       0.2749    0.397     -0.478     1.0490 0.5124 0.4741      2
White        0.3782    0.720     -1.007     1.7513 0.2769 0.5987      2
Black        1.3409    0.756     -0.182     2.7141 2.9875 0.0839      2
Asian        1.9202    0.857      0.149     3.4429 4.4610 0.0347      2
Democrat     0.2550    0.501     -0.688     1.2418 0.2767 0.5989      2
GOP         -0.3061    0.590     -1.479     0.7889 0.2969 0.5858      2
Ideology     0.0267    0.110     -0.191     0.2333 0.0613 0.8044      2

Method: 1-Wald, 2-Profile penalized log-likelihood, 3-None

Likelihood ratio test=17.5 on 9 df, p=0.0415, n=971
Wald test = 299 on 9 df, p = 0
```

Summarizing: $\hat{\beta}$s



Some Final Thoughts

- The key to doing King-Zeng is to be able to conduct C-C sampling *in advance*
- BUT: The R implementation of K&Z (in `Zelig`) is currently a bit buggy (its dependencies are all messed up...)
- In practice: the Firth + FLIC approach is generally superior to King/Zeng (and arguably should *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'll 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)

Things we won't talk about:

- 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)