# PLSC 504 – Fall 2023

# Regression Models for Nominal and Binary Responses

August 30, 2023

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 → "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”

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

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

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

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 (c-log-log, scobit)

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
```

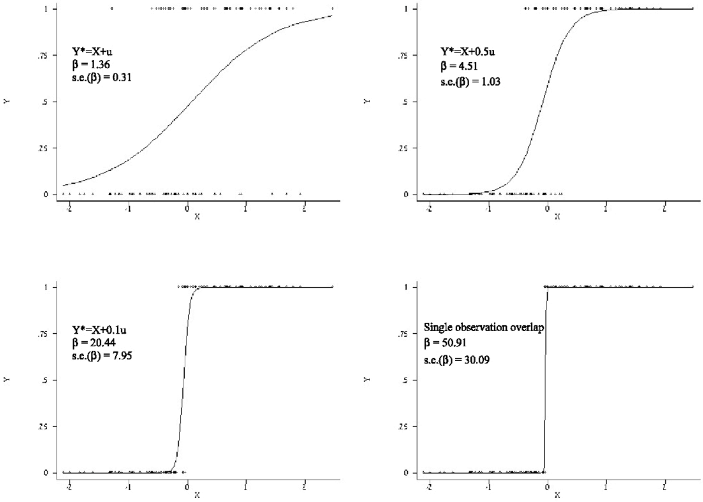
$\Pr(Y = 1 | X = 0) = ?$

"Separation" 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
```

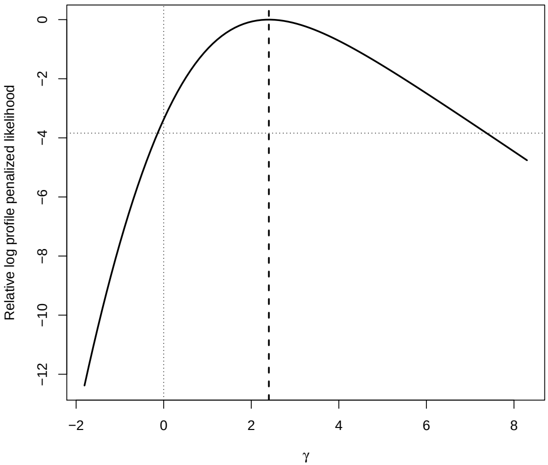
# What's Going On?

```
> 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.

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:

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

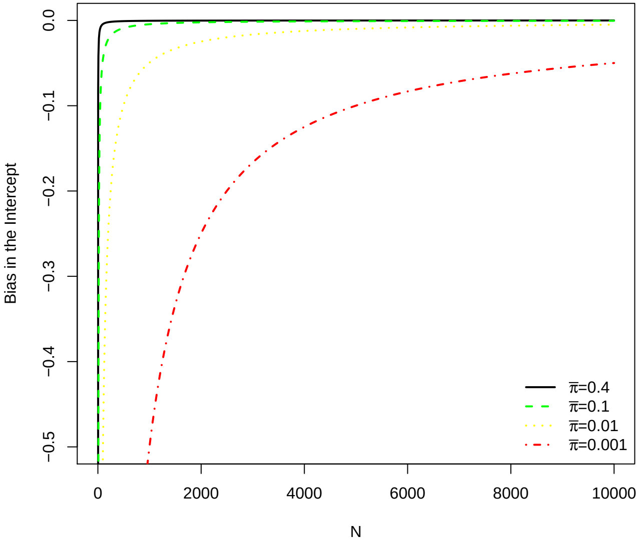
where $\bar{\pi} = \overline{\Pr(Y = 1)}$ is $< 0.5$.

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...

- $\tau =$ 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...

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

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

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

Weighting:

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

# Case-Control Data: Prior Correction

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

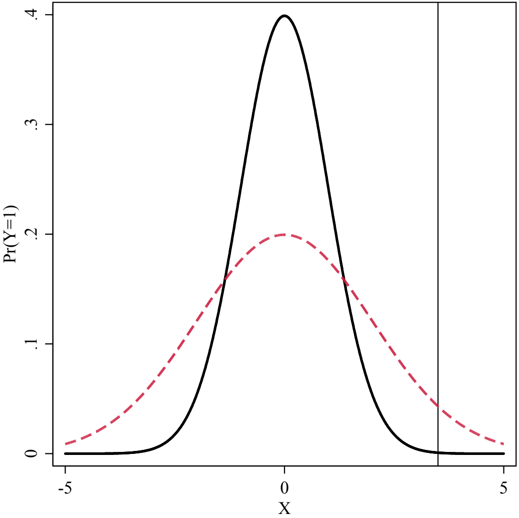
$$\text{bias}(\hat{\boldsymbol{\beta}}) = (\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\xi$$

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

Correction is

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

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


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

Use:

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

where

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

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."*

# 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
```

# Faking It: Case-Control Sampling

```
> set.seed(7222009)
> ROGones<-TAPS[TAPS$RunOutOfGas==1,]
> ROGzeros<-TAPS[TAPS$RunOutOfGas==0,]
> ROGSzeros<-ROGzeros[sample(1:nrow(ROGzeros),100,replace=FALSE),]
> ROGsample<-data.frame(rbind(ROGones,ROGSzeros))
> table(ROGsample$RunOutOfGas)

  0   1
100  28
>
> sample.logit<-glm(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                         Democrat+GOP+Ideology,data=ROGsample,family=binomial)
> summary(sample.logit)


Deviance Residuals:
   Min      1Q  Median      3Q     Max
-1.260  -0.714  -0.577  -0.414   2.140

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.1876     1.9782    0.60     0.55
Education    -0.1185     0.1379   -0.86     0.39
Age10        -0.1569     0.1475   -1.06     0.29
Female        0.1869     0.4710    0.40     0.69
White        -0.1219     0.7916   -0.15     0.88
Black         0.6012     0.8597    0.70     0.48
Asian         1.1924     1.0475    1.14     0.25
Democrat      0.0282     0.5879    0.05     0.96
GOP          -0.4268     0.6566   -0.65     0.52
Ideology     -0.0711     0.1247   -0.57     0.57

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 134.48  on 127  degrees of freedom
Residual deviance: 124.81  on 118  degrees of freedom
AIC: 144.8

Number of Fisher Scoring iterations: 4
```

# Rare Events Logit, Prior Correction

```
> relogit.pc<-zelig(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                    Democrat+GOP+Ideology,data=ROGsample,model="relogit",
+                    tau=28/971,case.control=c("prior"))

> summary(relogit.pc)
Model:

Call:
z5$zelig(formula = RunOutOfGas ~ Education + Age10 + Female +
    White + Black + Asian + Democrat + GOP + Ideology, tau = 28/971,
    case.control = c("prior"), data = ROGsample)

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.433  -0.242  -0.194  -0.139   2.978

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  -1.2215     1.9782   -0.62     0.54
Education    -0.1097     0.1379   -0.80     0.43
Age10        -0.1502     0.1475   -1.02     0.31
Female        0.1926     0.4710    0.41     0.68
White        -0.1416     0.7916   -0.18     0.86
Black         0.5006     0.8597    0.58     0.56
Asian         0.9438     1.0475    0.90     0.37
Democrat      0.0300     0.5879    0.05     0.96
GOP          -0.4105     0.6566   -0.63     0.53
Ideology     -0.0668     0.1247   -0.54     0.59

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 134.48  on 127  degrees of freedom
Residual deviance: 124.81  on 118  degrees of freedom
AIC: 144.8
```

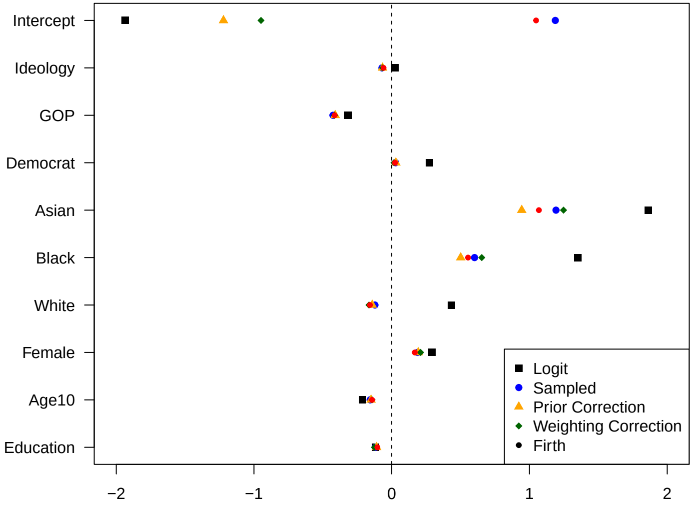
# Rare Events Logit, Weighting Correction

```
> relogit.wc<-zelig(RunOutOfGas~Education+Age10+Female+White+Black+Asian+
+                   Democrat+GOP+Ideology,data=ROGsample,model="relogit",
+                   tau=28/971,case.control=c("weighting"))

> summary(relogit.wc)
Model:

Call:
relogit(formula = cbind(RunOutOfGas, 1 - RunOutOfGas) ~ Education +
    Age10 + Female + White + Black + Asian + Democrat + GOP +
    Ideology, data = as.data.frame(.), tau = 0.0288362512873326,
    bias.correct = TRUE, case.control = "weighting")

Deviance Residuals:
   Min      1Q  Median      3Q     Max
-0.584  -0.290  -0.229  -0.163   1.066

Coefficients:
            Estimate Std. Error (robust) z value Pr(>|z|)
(Intercept)  -0.9491              2.4105   -0.39     0.69
Education    -0.1254              0.1289   -0.97     0.33
Age10        -0.1431              0.1634   -0.88     0.38
Female        0.2091              0.5419    0.39     0.70
White        -0.1650              1.2079   -0.14     0.89
Black         0.6535              1.1905    0.55     0.58
Asian         1.2471              1.5888    0.78     0.43
Democrat      0.0164              0.6920    0.02     0.98
GOP          -0.4153              0.7143   -0.58     0.56
Ideology     -0.0654              0.1323   -0.49     0.62

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 33.452  on 127  degrees of freedom
Residual deviance: 31.650  on 118  degrees of freedom
AIC: 25.77
```

# Firth Logit (for comparison)

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

> summary(relogit.firth)
logistf(formula = RunOutOfGas ~ Education + Age10 + Female +
    White + Black + Asian + Democrat + GOP + Ideology, data = ROGsample)

Model fitted by Penalized ML
Coefficients:
               coef se(coef) lower 0.95 upper 0.95   Chisq     p method
(Intercept)  1.0480    1.930     -2.578      4.731 0.32489 0.569      2
Education   -0.1040    0.134     -0.365      0.151 0.64036 0.424      2
Age10       -0.1428    0.143     -0.423      0.131 1.04461 0.307      2
Female       0.1674    0.456     -0.709      1.057 0.14081 0.707      2
White       -0.1580    0.776     -1.559      1.314 0.04879 0.825      2
Black        0.5540    0.848     -0.999      2.122 0.51072 0.475      2
Asian        1.0684    1.035     -0.820      2.955 1.29570 0.255      2
Democrat     0.0238    0.571     -1.082      1.131 0.00183 0.966      2
GOP         -0.4148    0.628     -1.681      0.792 0.45248 0.501      2
Ideology    -0.0593    0.121     -0.300      0.167 0.25857 0.611      2

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

Likelihood ratio test=9.6 on 9 df, p=0.384, n=128
Wald test = 8.15 on 9 df, p = 0.519
```

# Summarizing: $\hat{\beta}$s



# 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: The R implementation of `Zelig` is currently a bit buggy (its dependencies are all messed up...)
- In practice: Firth's approach is generally superior to King/Zeng (and arguably should *always* be used for binary-response regressions, especially with small-to-medium $N$s)
- Also: Remember that as your $N$ gets big, the problem goes away; Paul Allision 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)

# Nominal Outcomes

# Motivation: Discrete *Outcomes*

$$\Pr(Y_i = j) = P_{ij}$$

$$\sum_{j=1}^{J} P_{ij} = 1$$

$$P_{ij} = \exp(\mathbf{X}_i \boldsymbol{\beta}_j)$$

Rescale:

$$\Pr(Y_i = j) \equiv P_{ij} = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)}$$

Ensures

- $\Pr(Y_i = j) \in (0, 1)$
- $\sum_{j=1}^{J} \Pr(Y_i = j) = 1.0$

Constrain $\boldsymbol{\beta}_1 = \mathbf{0}$; then:

$$\Pr(Y_i = 1) = \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}$$

$$\Pr(Y_i = j) = \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}'_j)}$$

where $\boldsymbol{\beta}'_j = \boldsymbol{\beta}_j - \boldsymbol{\beta}_1$.

# Alternative Motivation: Discrete *Choice*

$$U_{ij} = \mu_i + \epsilon_{ij}$$

$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}_j$$

$$\begin{aligned}
\Pr(Y_i = j) &= \Pr(U_{ij} > U_{i\ell} \, \forall \, \ell \neq j \in J) \\
&= \Pr(\mu_i + \epsilon_{ij} > \mu_i + \epsilon_{i\ell} \, \forall \, \ell \neq j \in J) \\
&= \Pr(\mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} > \mathbf{X}_i \boldsymbol{\beta}_\ell + \epsilon_{i\ell} \, \forall \, \ell \neq j \in J) \\
&= \Pr(\epsilon_{ij} - \epsilon_{i\ell} > \mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j \, \forall \, \ell \neq j \in J)
\end{aligned}$$

$\epsilon \sim$ ???

- *Type I Extreme Value*
- Density: $f(\epsilon) = \exp[-\epsilon - \exp(-\epsilon)]$
- CDF: $\int f(\epsilon) \equiv F(\epsilon) = \exp[-\exp(-\epsilon)]$
- $\rightarrow$ Multinomial Logit

# Estimation

Define:

$$\begin{aligned} \delta_{ij} &= 1 \text{ if } Y_i = j, \\ &= 0 \text{ otherwise.} \end{aligned}$$

Then:

$$\begin{aligned} L_i &= \prod_{j=1}^{J} [\Pr(Y_i = j)]^{\delta_{ij}} \\ &= \prod_{j=1}^{J} \left[ \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]^{\delta_{ij}} \end{aligned}$$

So:
$$L = \prod_{i=1}^{N} \prod_{j=1}^{J} \left[ \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]^{\delta_{ij}}$$

and (of course):

$$\ln L = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln \left[ \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)} \right]$$

*It is exactly the same as the multinomial logit model. Period.*

CL with choice-varying predictors $\mathbf{Z}_{ij}\gamma$ is:

$$\Pr(Y_{ij} = j) = \frac{\exp(\mathbf{Z}_{ij}\gamma)}{\sum_{j=1}^{J} \exp(\mathbf{Z}_{ij}\gamma)}$$

Combinations: $\mathbf{X}_i\boldsymbol{\beta}$ and $\mathbf{Z}_{ij}\gamma$:

- "Fixed effects" for each possible outcome / choice
- Observation-specific $\mathbf{X}$s
- Interactions...

# MNL and CL: Practical Things

The PLSC 503 slides and code include some additional detail, plus a running example (the three-candidate 1992 U.S. presidential election), with discussions of:

- Model estimation (including choosing the baseline/reference outcome),
- Model interpretation and discussion (odds ratios, predicted probabilities, etc.),
- Model fit, and
- Diagnostics.

I've included most of the code for those examples in today's code as well.

# Independence of Irrelevant Alternatives ("IIA")

**"An individual's choice does not depend on the availability or characteristics of unavailable alternatives."**

# IIA, Statistically

$$
\begin{aligned}
\frac{\Pr(Y_i = k)}{\Pr(Y_i = \ell)} &= \frac{\frac{\exp(\mathbf{X}_i\beta_k)}{\sum_{j=1}^{J}\exp(\mathbf{X}_i\beta_j)}}{\frac{\exp(\mathbf{X}_i\beta_\ell)}{\sum_{j=1}^{J}\exp(\mathbf{X}_i\beta_j)}} \\
&= \frac{\exp(\mathbf{X}_i\beta_k)}{\exp(\mathbf{X}_i\beta_\ell)} \\
&= \exp[\mathbf{X}_i(\beta_k - \beta_\ell)]
\end{aligned}
$$

Alternatively:

$$
\frac{\Pr(Y_i = k|S_J)}{\Pr(Y_i = \ell|S_J)} = \frac{\Pr(Y_i = k|S_M)}{\Pr(Y_i = \ell|S_M)} \; \forall \; k, \ell, J, M
$$

# IIA, Intuitively

- Initially: $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.5$, $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$.

- Enter the Blue Bus...
  - Intuitively: $\Pr(\text{Car}) = 0.5$, $\Pr(\text{Red Bus}) = 0.25$, $\Pr(\text{Blue Bus}) = 0.25$
  - IIA requires that $\frac{\Pr(\text{Car})}{\Pr(\text{Red Bus})} = 1$.
  - So, that could be $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = \Pr(\text{Blue Bus}) = 0.33$, or
  - $\Pr(\text{Car}) = \Pr(\text{Red Bus}) = 0.4$ and $\Pr(\text{Blue Bus}) = 0.2$...

Random utility model:

$$\begin{aligned} U_{ij} &= \mu_{ij} + \epsilon_{ij} \\ &= \mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} \end{aligned}$$

... means that:

$$\begin{aligned} \Pr(Y_i = j) &= \Pr(U_{ij} > U_{i\ell}) \forall \ell \neq j \in J \\ &= \Pr(\mathbf{X}_i \boldsymbol{\beta}_j + \epsilon_{ij} > \mathbf{X}_i \boldsymbol{\beta}_\ell + \epsilon_{i\ell}) \forall \ell \neq j \in J \\ &= \Pr(\epsilon_{ij} - \epsilon_{i\ell} > \mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j) \forall \ell \neq j \in J \end{aligned}$$

# IIA Tests: Hausman/McFadden and Small/Hsiao

$$HM = (\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u)'[\hat{\mathbf{V}}_r - \hat{\mathbf{V}}_u]^{-1}(\hat{\boldsymbol{\beta}}_r - \hat{\boldsymbol{\beta}}_u)$$

$$\widehat{HM} \sim \chi^2_{(J-2)k}$$

$$SH = -2\left[L_r(\hat{\boldsymbol{\beta}}_u^{AB}) - L_r(\hat{\boldsymbol{\beta}}_r^{B})\right]$$

$$\widehat{SH} \sim \chi^2_{k_r}$$

## IIA Freedom: Multinomial Probit

$\epsilon_{ij} \sim MVN(0, \Sigma)$, where:

$$\underset{J \times J}{\boldsymbol{\Sigma}} = \begin{bmatrix} \sigma_1^2 & \dots & \sigma_{1J} \\ \vdots & \ddots & \vdots \\ \sigma_{J1} & \dots & \sigma_J^2 \end{bmatrix}$$

Define $\eta_{ij\ell} = \epsilon_{ij} - \epsilon_{i\ell}$. Then:

$$\begin{aligned} \Pr(Y_i = j) &= \Pr(\eta_{ij\ell} > \mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j) \, \forall \, \ell \neq j \in J \\ &= \int_{-\infty}^{\mathbf{X}_i \boldsymbol{\beta}_1 - \mathbf{X}_i \boldsymbol{\beta}_j} \dots \int_{-\infty}^{\mathbf{X}_i \boldsymbol{\beta}_\ell - \mathbf{X}_i \boldsymbol{\beta}_j} \phi_J(\eta_{ij1}, \eta_{ij2}, \dots \eta_{ij\ell}) d\eta_{ij1}, \eta_{ij2}, \dots \eta_{ij\ell} \end{aligned}$$

# MNP: Issues and Estimation

- Identification: (Potentially) Fragile
- Estimation:
  - Always hard
  - Via "GHK" algorithm, or
  - Gaussian quadrature, or
  - Simulation (MCMC) (preferred)
- Software:
  - `mlogit` with `probit = TRUE` (Geweke-Hajivassiliou-Keane algorithm)
  - `MNP` package (Bayesian/MCMC)
  - `endogMNP` package (Bayesian with endogenous switching)
  - Others?

# IIA Freedom: HEV

$$
\begin{aligned}
f(\epsilon_{ij}) &= \lambda(\epsilon_{ij}) \\
&= \frac{1}{\theta_j} \exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right) \exp\left[-\exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right)\right]
\end{aligned}
$$

$$
\begin{aligned}
F(\epsilon_{ij}) &= \Lambda(\epsilon_{ij}) \\
&= \int_{-\infty}^{z} f(\epsilon_{ij}) d\,\epsilon_{ij} \\
&= \exp\left[-\exp\left(-\frac{\epsilon_{ij}}{\theta_j}\right)\right]
\end{aligned}
$$

Means:

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda \left( \frac{\mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_\ell + \epsilon_{ij}}{\theta_\ell} \right) \frac{1}{\theta_j} \lambda \left( \frac{\epsilon_{ij}}{\theta_j} \right) d\, \epsilon_{ij}$$

With $w = \frac{\epsilon_{ij}}{\theta_j}$:

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda \left( \frac{\mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_\ell + \theta_j w}{\theta_\ell} \right) \lambda(w) d\, w$$

MNL $\subset$ HEV: When $\theta_j = 1 \;\; \forall \;\; j \rightarrow$

$$\Pr(Y_i = j) = \int_{-\infty}^{\infty} \prod_{\ell \neq j} \Lambda(\mathbf{X}_i \beta_j - \mathbf{X}_i \beta_\ell + \epsilon_{ij}) \lambda(\epsilon_{ij}) d\, \epsilon_{ij}$$

# IIA Freedom: "Mixed Logit"

$$U_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta} + \epsilon_{ij},$$

$$\epsilon_{ij} = \eta_i + \xi_{ij}$$

$$\Pr(Y_i = j|\eta) \equiv \Pr(Y_{ij} = 1|\eta) = \frac{\exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}{\sum_{j=1}^{J} \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}$$

# What to do with the $\eta$s?

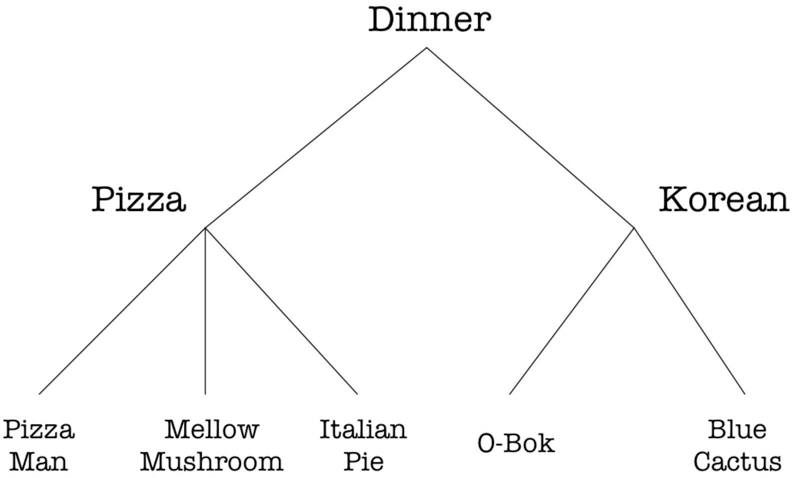
Assume:

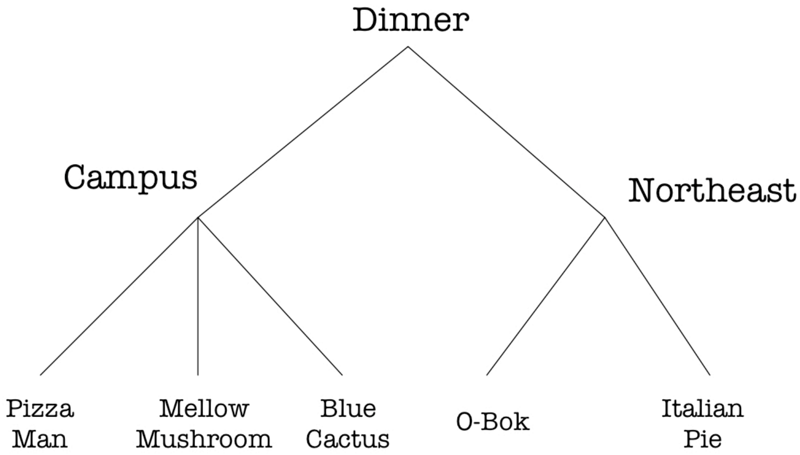
$$\eta_i \sim g(\mathbf{0}, \mathbf{\Omega})$$

Yields:

$$\Pr(Y_i = j) = \int \left[ \frac{\exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)}{\sum_{j=1}^{J} \exp(\mathbf{X}_{ij}\boldsymbol{\beta} + \eta_i)} \right] g(\eta|\mathbf{\Omega})\, d\eta$$

- "Nested" choices
- A priori information about "subsets"
- IIA holds *within* (but not *across*) subsets...


Dinner
Pizza
Korean
Pizza Man
Mellow Mushroom
Italian Pie
O-Bok
Blue Cactus


Dinner
Campus
Northeast
Pizza Man
Mellow Mushroom
Blue Cactus
O-Bok
Italian Pie

# Example: 2002 Swedish Election ($N = 6610$)

```
> summary(Sweden)

         partychoice          female           union          leftright    
 Conservatives   :1469   Min.   :0.0000   Min.   :1.000   Min.   :1.000  
 Liberals        :1212   1st Qu.:0.0000   1st Qu.:1.000   1st Qu.:2.000  
 Social Democrats:2975   Median :0.0000   Median :3.000   Median :3.000  
 Left Party      : 954   Mean   :0.4882   Mean   :2.709   Mean   :2.868  
                         3rd Qu.:1.0000   3rd Qu.:4.000   3rd Qu.:4.000  
                         Max.   :1.0000   Max.   :4.000   Max.   :5.000  
      age       
 Min.   :17.00  
 1st Qu.:29.00  
 Median :42.00  
 Mean   :42.93  
 3rd Qu.:55.00  
 Max.   :90.00  
```

# Swedish Election: MNL

```
> library(mlogit)
> Sweden.Long<-mlogit.data(Sweden,choice="partychoice",shape="wide")
> Sweden.MNL<-mlogit(partychoice~1|female+union+leftright+age,data=Sweden.Long)
> summary(Sweden.MNL)

Frequencies of alternatives:
   Conservatives       Left Party         Liberals Social Democrats
         0.22224          0.14433          0.18336          0.45008

Coefficients :
                               Estimate Std. Error  t-value  Pr(>|t|)
altLeft Party                13.3907039  0.3788540  35.3453 < 2.2e-16 ***
altLiberals                   4.4121638  0.2928137  15.0682 < 2.2e-16 ***
altSocial Democrats          11.3821332  0.3289066  34.6060 < 2.2e-16 ***
altLeft Party:female          0.7211951  0.1218437   5.9190 3.239e-09 ***
altLiberals:female            0.5585172  0.0848597   6.5817 4.652e-11 ***
altSocial Democrats:female    0.3881456  0.0945266   4.1062 4.022e-05 ***
altLeft Party:union          -0.4334637  0.0513499  -8.4414 < 2.2e-16 ***
altLiberals:union            -0.0563136  0.0388720  -1.4487 0.1474228
altSocial Democrats:union    -0.4145682  0.0408153 -10.1572 < 2.2e-16 ***
altLeft Party:leftright      -4.0917135  0.0930610 -43.9681 < 2.2e-16 ***
altLiberals:leftright        -1.1274488  0.0593125 -19.0086 < 2.2e-16 ***
altSocial Democrats:leftright -2.7555009  0.0719411 -38.3022 < 2.2e-16 ***
altLeft Party:age            -0.0277444  0.0038808  -7.1491 8.737e-13 ***
altLiberals:age              -0.0064185  0.0025768  -2.4909 0.0127410 *
altSocial Democrats:age      -0.0105052  0.0029196  -3.5982 0.0003204 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -5627.5
McFadden R^2:  0.33693
Likelihood ratio test : chisq = 5719 (p.value=< 2.22e-16)
```

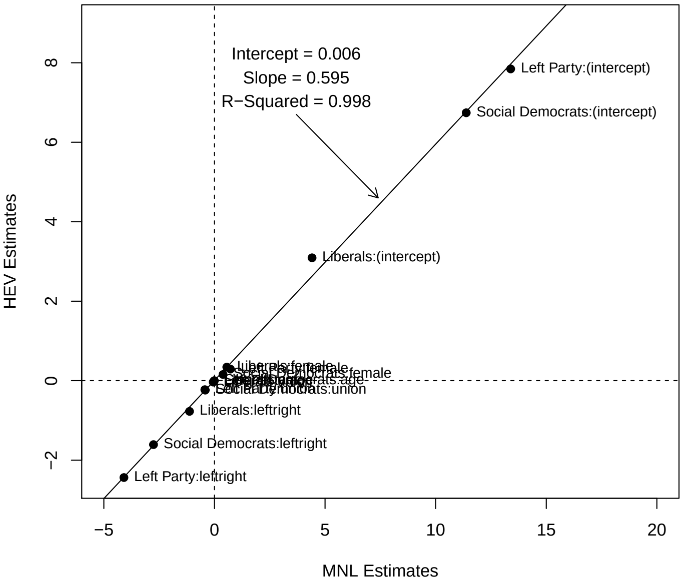
# Hausman-McFadden IIA Test

```
> # Restricted model (omitting Social Democrats)
> Sweden.MNL.Restr<-mlogit(partychoice~1|female+union+leftright+age,
+ Sweden.Long,alt.subset=c("Conservatives","Liberals","Left Party"))
>
> hmftest(Sweden.MNL,Sweden.MNL.Restr)

Hausman-McFadden test

data:  Sweden.Long
chisq = 19.1137, df = 10, p-value = 0.03884
alternative hypothesis: IIA is rejected
```

# Swedish Election: HEV

```
> Sweden.Het<-mlogit(partychoice~1|female+union+leftright+
+                       age,data=Sweden.Long,heterosc=TRUE)
> summary(Sweden.Het)

Coefficients :
                             Estimate Std. Error z-value Pr(>|z|)
Left Party:(intercept)        7.84569    0.42849   18.31  < 2e-16 ***
Liberals:(intercept)          3.09199    0.30607   10.10  < 2e-16 ***
Social Democrats:(intercept) 6.74242    0.32038   21.04  < 2e-16 ***
Left Party:female             0.29096    0.08057    3.61   0.0003 ***
Liberals:female               0.34113    0.06510    5.24  1.6e-07 ***
Social Democrats:female       0.15572    0.05718    2.72   0.0065 **
Left Party:union             -0.22645    0.03704   -6.11  9.7e-10 ***
Liberals:union               -0.03498    0.02685   -1.30   0.1926
Social Democrats:union       -0.23786    0.03319   -7.17  7.8e-13 ***
Left Party:leftright         -2.43814    0.17450  -13.97  < 2e-16 ***
Liberals:leftright           -0.77255    0.04629  -16.69  < 2e-16 ***
Social Democrats:leftright   -1.60927    0.09462  -17.01  < 2e-16 ***
Left Party:age               -0.01612    0.00338   -4.77  1.9e-06 ***
Liberals:age                 -0.00200    0.00176   -1.14   0.2543
Social Democrats:age         -0.00267    0.00175   -1.53   0.1258
sp.Left Party                 0.90017    0.14304    6.29  3.1e-10 ***
sp.Liberals                   0.59981    0.09925    6.04  1.5e-09 ***
sp.Social Democrats           0.69163    0.10197    6.78  1.2e-11 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -5840
McFadden R^2:  0.312
Likelihood ratio test : chisq = 5300 (p.value = <2e-16)
```

# $\hat{\beta}$s: MNL vs. HEV



# HEV > MNL?

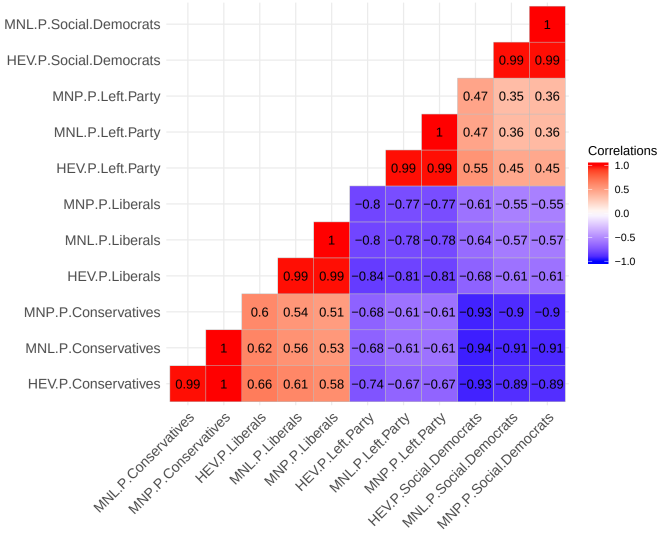
Tests:

```
> MNL.HEV.Wald <- waldtest(Sweden.Het, heterosc = FALSE) # Wald test
> MNL.HEV.Wald

Wald test

data:  homoscedasticity
chisq = 20, df = 3, p-value = 0.0004

> MNL.HEV.LR <- lrtest(Sweden.Het)         # LR test
> MNL.HEV.LR
Likelihood ratio test

Model 1: partychoice ~ 1 | female + union + leftright + age
Model 2: partychoice ~ 1 | female + union + leftright + age
  #Df LogLik Df Chisq Pr(>Chisq)
1  18  -5836
2  15  -5627 -3   416     <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

> MNL.HEV.Score <- scoretest(Sweden.MNL, heterosc = TRUE)   # score test
> MNL.HEV.Score

score test

data:  heterosc = TRUE
chisq = 20, df = 3, p-value = 0.00002
alternative hypothesis: heteroscedastic model
```

# Swedish Election: MNP

```
> library(MNP)
> Sweden.MNP<-mnp(partychoice~female+union+leftright+age, data=Sweden)
> summary(Sweden.MNP)

Coefficients:
                                 mean  std.dev.      2.5%     97.5%
(Intercept):Liberals         3.964677  0.879442  0.983572     4.669
(Intercept):Social Democrats 7.993453  1.495732  3.986961     9.812
(Intercept):Left Party      10.342468  2.082971  4.845935    12.714
female:Liberals              0.293136  0.046373  0.204654     0.382
female:Social Democrats      0.290311  0.079166  0.124746     0.447
female:Left Party            0.613163  0.163673  0.289974     0.944
union:Liberals              -0.083366  0.036782 -0.140052     0.024
union:Social Democrats      -0.275696  0.059260 -0.369943    -0.145
union:Left Party            -0.346922  0.087131 -0.489992    -0.148
leftright:Liberals          -0.913247  0.168331 -1.045781    -0.350
leftright:Social Democrats  -1.920076  0.362403 -2.371245    -0.977
leftright:Left Party        -3.409277  0.750701 -4.308455    -1.576
age:Liberals                -0.003350  0.001490 -0.006264 -0.000409
age:Social Democrats        -0.007171  0.002630 -0.012327    -0.002
age:Left Party              -0.025595  0.007323 -0.039641    -0.011

Covariances:
                                   mean std.dev.   2.5%  97.5%
Liberals:Liberals                1.0000   0.0000 1.0000  1.000
Liberals:Social Democrats        1.4083   0.3925 0.2116  1.830
Liberals:Left Party              2.4450   1.0779 0.6731  3.988
Social Democrats:Social Democrats 2.6696   0.9215 0.5630  3.898
Social Democrats:Left Party      4.4852   2.1846 0.3521  7.524
Left Party:Left Party            9.4811   5.0787 1.1682 17.095

Base category: Conservatives
Number of alternatives: 4
Number of observations: 6610
Number of estimated parameters: 20
Number of stored MCMC draws: 5000
```

# How I Stopped Worrying and Learned To Love MNL...



# Software

| Model | Stata | SAS | R |
|---|---|---|---|
| Multinomial Logit | mlogit | proc catmod | vglm, mlogit, multinom* |
| Conditional Logit | clogit | proc mdc | clogit, mlogit |
| Multinomial Probit | mprobit / asmprobit | proc mdc | mnp*, mlogit |
| Heteroscedastic Extreme Value | No(?) | proc mdc | mlogit |
| Mixed Logit | mixlogit | proc mdc | mlogit |
| Nested Logit | nlogit | proc mdc | mlogit |

* See also bayesm.

# Things To Read


- Bhat, Chandra R. 1995. "A Heteroscedastic Extreme Value Model of Intercity Travel Mode Choice." *Transportation Research Part B: Methodological* 29(6):471-83.
- Colonescu, Constantin. 2016. *Principles of Econometrics with R*. Chapter 16: "Qualitative and LDV Models." Available here.
- Hensher, David A., and William H. Greene. 2002. "Specification and Estimation of the Nested Logit Model: Alternative Normalisations." *Transportation Research Part B* 36:1-17.
- Imai, Kosuke, and D. A. van Dyk. 2005. "A Bayesian Analysis of the Multinomial Probit Model Using Marginal Data Augmentation." *Journal of Econometrics* 124:311-334.
- McFadden, Daniel. 1974. "The Measurement of Urban Travel Demand." *Journal of Public Economics* 3:303-28.
- Sarrias, Mauricio, and Ricardo Daziano. 2017. "Multinomial Logit Models with Continuous and Discrete Individual Heterogeneity in R: The Gmnl Package." *Journal of Statistical Software, Articles* 79(2):1-46.
- Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* New York: Cambridge University Press.