GSERM 2020
Regression for Publishing

June 18, 2020 (first session)

Ordinal Data

Ordinal data are:

- Discrete: $Y \in \{1, 2, ...\}$
- *Grouped Continuous* Data
- *Assessed Ordered* Data

In general:

- Some things can be ordered, but shouldn't be
- Some things are ordered in some circumstances but not others
- Orderings can differ across applications

A Fake-Data Example

$$Y_i^* = 0 + 1.0X_i + u_i,$$

$$X_i \sim U[0, 10]$$

$$u_i \sim N(0, 1)$$

$$\begin{aligned} Y_{1i} &= 1 \text{ if } Y_i^* < 2.5 \\ &= 2 \text{ if } 2.5 \leq Y_i^* < 5 \\ &= 3 \text{ if } 5 \leq Y_i^* < 7.5 \\ &= 4 \text{ if } Y_i^* > 7.5 \end{aligned}$$

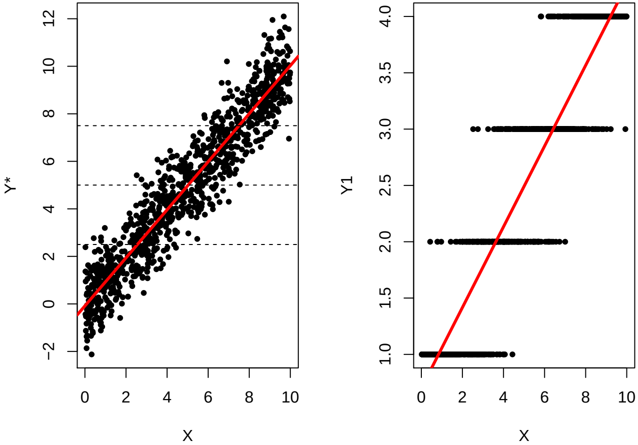
$$\begin{aligned} Y_{2i} &= 1 \text{ if } Y_i^* < 2 \\ &= 2 \text{ if } 2 \leq Y_i^* < 8 \\ &= 3 \text{ if } 8 \leq Y_i^* < 9 \\ &= 4 \text{ if } Y_i^* > 9 \end{aligned}$$

```
> summary(lm(Ystar~X))

Call:
lm(formula = Ystar ~ X)

Residuals:
   Min     1Q Median     3Q    Max
-3.006 -0.654 -0.049  0.643  3.298

Coefficients:
            Estimate Std. Error t value              Pr(>|t|)
(Intercept)  -0.0830     0.0609   -1.36                  0.17
X             1.0110     0.0106   95.48 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.988 on 998 degrees of freedom
Multiple R-squared:  0.901,Adjusted R-squared:  0.901
F-statistic: 9.12e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

Also A Pretty Good Regression

```
> summary(lm(Y1~X))

Call:
lm(formula = Y1 ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-1.2889 -0.2439  0.0158  0.2592  1.3968

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  0.69979    0.02639    26.5 <0.0000000000000002 ***
X            0.35825    0.00459    78.0 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.428 on 998 degrees of freedom
Multiple R-squared:  0.859,Adjusted R-squared:  0.859
F-statistic: 6.09e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

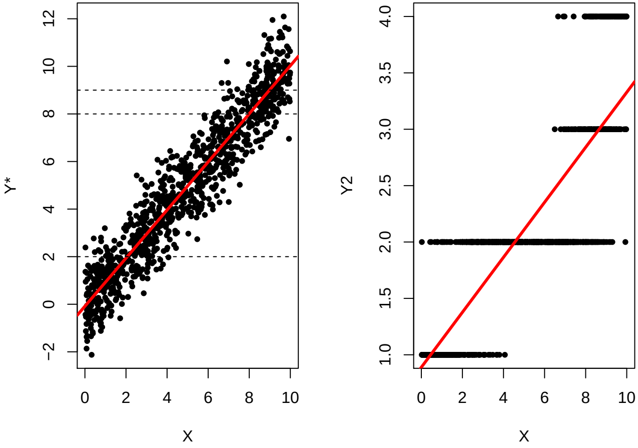
What That Looks Like



A Not-So-Good Regression

```
> summary(lm(Y2~X))

Call:
lm(formula = Y2 ~ X)

Residuals:
    Min      1Q  Median      3Q     Max
-1.3115 -0.3205 -0.0405  0.2914  1.4876

Coefficients:
            Estimate Std. Error t value             Pr(>|t|)
(Intercept)  0.88919    0.03069    29.0 <0.0000000000000002 ***
X            0.24383    0.00534    45.7 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Residual standard error: 0.498 on 998 degrees of freedom
Multiple R-squared:  0.676,Adjusted R-squared:  0.676
F-statistic: 2.09e+03 on 1 and 998 DF,  p-value: <0.0000000000000002
```

What That Looks Like



Models for Ordinal Responses

$$Y_i^* = \mu + u_i$$

$$Y_i = j \text{ if } \tau_{j-1} \leq Y_i^* < \tau_j,\ j \in \{1, ...J\}$$

$$\begin{aligned} Y_i &= 1 \text{ if } -\infty \leq Y_i^* < \tau_1 \\ &= 2 \text{ if } \tau_1 \leq Y_i^* < \tau_2 \\ &= 3 \text{ if } \tau_2 \leq Y_i^* < \tau_3 \\ &= 4 \text{ if } \tau_3 \leq Y_i^* < \infty \end{aligned}$$

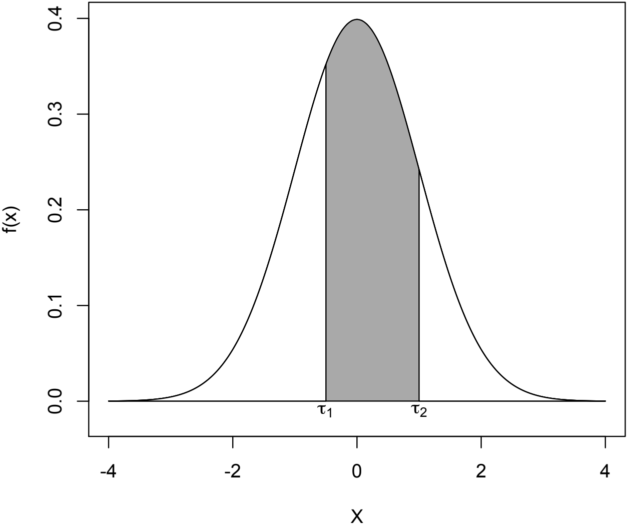
Ordinal Response Models: Probabilities

$$\begin{aligned}
\Pr(Y_i = j) &= \Pr(\tau_{j-1} \leq Y^* < \tau_j) \\
&= \Pr(\tau_{j-1} \leq \mu_i + u_i < \tau_j) \qquad (1)
\end{aligned}$$

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

$$\begin{aligned}
\Pr(Y_i = j | \mathbf{X}, \boldsymbol{\beta}) &= \Pr(\tau_{j-1} \leq Y_i^* < \tau_j | \mathbf{X}) \\
&= \Pr(\tau_{j-1} \leq \mathbf{X}_i \boldsymbol{\beta} + u_i < \tau_j) \\
&= \Pr(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta} \leq u_i < \tau_j - \mathbf{X}_i \boldsymbol{\beta}) \\
&= \int_{-\infty}^{\tau_j - \mathbf{X}_i \boldsymbol{\beta}} f(u_i) du - \int_{-\infty}^{\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta}} f(u_i) du \\
&= F(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})
\end{aligned}$$

What That Looks Like



Probabilities (here, probit)

$$
\begin{aligned}
\Pr(Y_i = 1) &= \Phi(\tau_1 - \mathbf{X}_i\beta) - 0 \\
\Pr(Y_i = 2) &= \Phi(\tau_2 - \mathbf{X}_i\beta) - \Phi(\tau_1 - \mathbf{X}_i\beta) \\
\Pr(Y_i = 3) &= \Phi(\tau_3 - \mathbf{X}_i\beta) - \Phi(\tau_2 - \mathbf{X}_i\beta) \\
\Pr(Y_i = 4) &= 1 - \Phi(\tau_3 - \mathbf{X}_i\beta)
\end{aligned}
$$

Define:

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

Likelihood:

$$L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \prod_{i=1}^{N} \prod_{j=1}^{J} [F(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - F(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]^{\delta_{ij}}$$

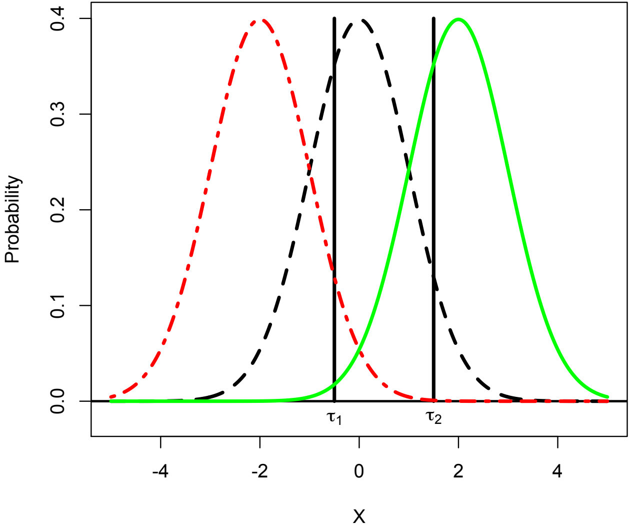
Log-Likelihood, probit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Phi(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Phi(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

Log-Likelihood, logit:

$$\ln L(Y|\mathbf{X}, \boldsymbol{\beta}, \tau) = \sum_{i=1}^{N} \sum_{j=1}^{J} \delta_{ij} \ln[\Lambda(\tau_j - \mathbf{X}_i \boldsymbol{\beta}) - \Lambda(\tau_{j-1} - \mathbf{X}_i \boldsymbol{\beta})]$$

The Intuition

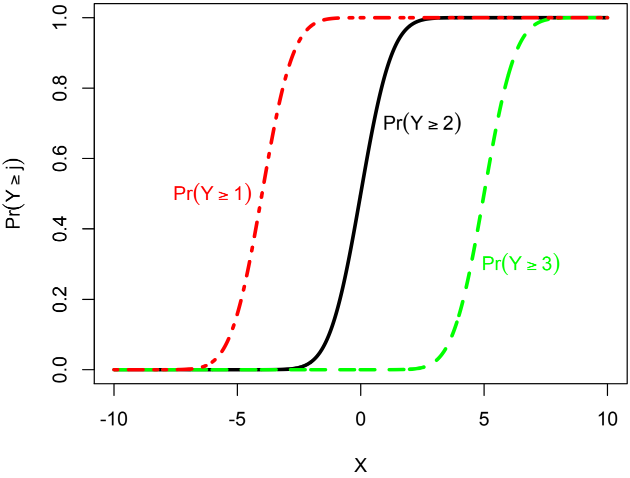


- (Usual) Assumption about $\sigma^2_{Y^*}$
- β_0 vs. the τs...
- Must either omit β_0 or drop one of the $J-1$ τs
- In practice: Stata & R omit β_0

$$\frac{\partial \mathrm{Pr}(Y_i \geq j)}{\partial X} = \frac{\partial \mathrm{Pr}(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

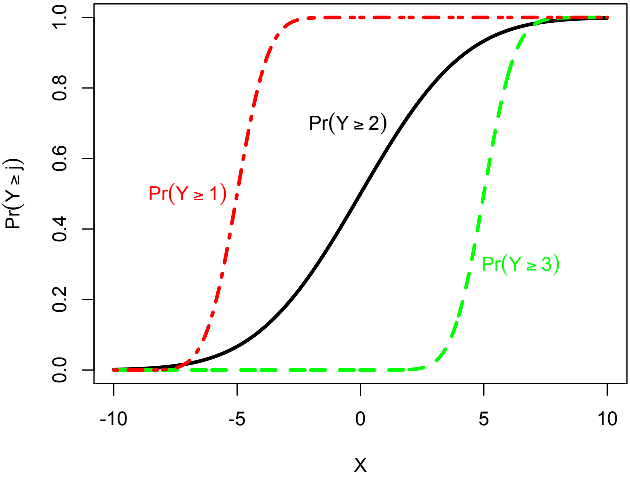
(aka "proportional odds"...)

Parallel Regressions Envisioned



$$\frac{\partial \text{Pr}(Y_i \geq j)}{\partial X} \neq \frac{\partial \text{Pr}(Y_i \geq j')}{\partial X} \; \forall \; j \neq j'$$

Nonparallel Regressions Envisioned



- R :
  - `polr` (in `MASS`)
  - `ologit`/`oprobit` (in `Zelig`; calls `polr`)
  - `vglm` (in `VGAM`)

- Stata : `ologit`, `oprobit`

Best Example Ever

```
> summary(beer)

    name              contqual        quality          price          calories    
 Length:69          Min.   :24.00   Min.   :1.000   Min.   :2.360   Min.   : 58.0  
 Class :character   1st Qu.:49.00   1st Qu.:2.000   1st Qu.:3.900   1st Qu.:142.0  
 Mode  :character   Median :70.00   Median :3.000   Median :4.790   Median :148.0  
                    Mean   :64.78   Mean   :2.536   Mean   :4.963   Mean   :142.3  
                    3rd Qu.:80.00   3rd Qu.:4.000   3rd Qu.:6.240   3rd Qu.:160.0  
                    Max.   :98.00   Max.   :4.000   Max.   :7.800   Max.   :201.0  

    alcohol        craftbeer           bitter          malty                        class   
 Min.   :0.500   Min.   :0.0000   Min.   : 8.00   Min.   : 5.00   Craft Lager         :13  
 1st Qu.:4.400   1st Qu.:0.0000   1st Qu.:21.00   1st Qu.:12.00   Craft Ale           :17  
 Median :4.900   Median :0.0000   Median :31.00   Median :23.00   Imported Lager      :10  
 Mean   :4.471   Mean   :0.4348   Mean   :35.44   Mean   :33.13   Regular or Ice Beer:16  
 3rd Qu.:5.100   3rd Qu.:1.0000   3rd Qu.:52.50   3rd Qu.:50.50   Light Beer          : 6  
 Max.   :6.000   Max.   :1.0000   Max.   :80.50   Max.   :86.00   Nonalcoholic        : 7  
```

```
> library(MASS)
> beer.logit<-polr(as.factor(quality)~price+calories+craftbeer+bitter
  +malty,data=beer)
> summary(beer.logit)

Call:
polr(formula = as.factor(quality) ~ price + calories + craftbeer +
    bitter + malty)

Coefficients:
           Value Std. Error t value
price     -0.451      0.293    -1.5
calories   0.047      0.012     3.8
craftbeer -1.705      0.942    -1.8
bitter    -0.030      0.042    -0.7
malty      0.051      0.025     2.1

Intercepts:
    Value  Std. Error t value
1|2  2.771  1.674      1.655
2|3  4.270  1.725      2.475
3|4  5.578  1.760      3.170
```

```
> beer.probit<-polr(as.factor(quality)~price+calories+craftbeer+bitter+malty,
+  data=beer,method="probit")
> summary(beer.probit)

Call:
polr(formula = as.factor(quality) ~ price + calories + craftbeer +
    bitter + malty, method = "probit")

Coefficients:
             Value Std. Error t value
price     -0.27914   0.172012 -1.6228
calories   0.02800   0.007184  3.8979
craftbeer -0.98427   0.559020 -1.7607
bitter    -0.01737   0.024719 -0.7025
malty      0.02855   0.014321  1.9937

Intercepts:
    Value  Std. Error t value
1|2  1.647  1.018      1.619
2|3  2.508  1.034      2.426
3|4  3.290  1.049      3.136
```

Confidence Intervals Around $\hat{\beta}$

```
> # Profile-likelihood-based CIs:
>
> CIs.logit <- confint(beer.logit)

> # Compare to normal CIs:
>
> CIs.alt <- cbind(beer.logit$coefficients-1.96*sqrt(diag(vcov(beer.logit)))[1:5],
+                  beer.logit$coefficients+1.96*sqrt(diag(vcov(beer.logit)))[1:5])

> CIs.logit
             2.5 % 97.5 %
price     -1.04459 0.1131
calories   0.02407 0.0724
craftbeer -3.58749 0.1367
bitter    -0.11511 0.0528
malty      0.00295 0.1005

> CIs.alt
              [,1]   [,2]
price     -1.02561 0.1246
calories   0.02282 0.0707
craftbeer -3.55086 0.1415
bitter    -0.11285 0.0535
malty      0.00231 0.0987
```

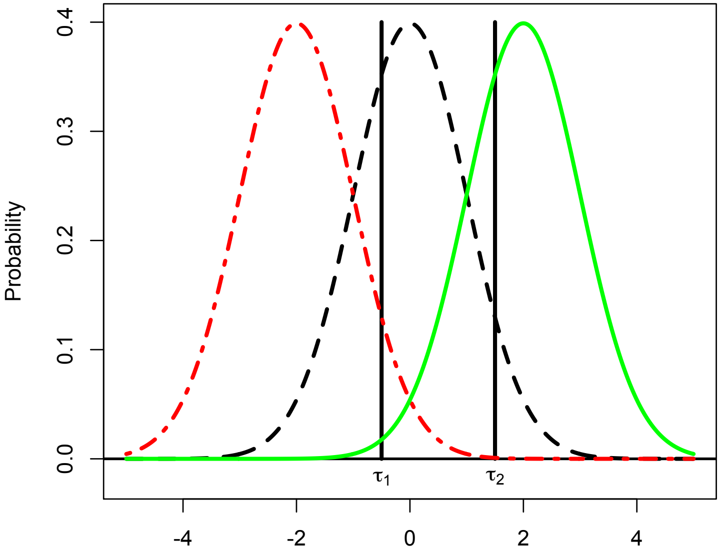
Interpretation: Marginal Effects

$$\begin{aligned}\frac{\partial \Pr(Y=j)}{\partial X_k} &= \frac{\partial F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} - \frac{\partial F(\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta})}{\partial X_k} \\ &= \hat{\beta}_k[f(\hat{\tau}_{j-1} - \bar{\mathbf{X}}\hat{\beta}) - f(\hat{\tau}_j - \bar{\mathbf{X}}\hat{\beta})]\end{aligned}$$

So:

- sign$\left(\frac{\partial \Pr(Y=1)}{\partial X_k}\right)$ = -sign$(\hat{\beta}_k)$
- sign$\left(\frac{\partial \Pr(Y=J)}{\partial X_k}\right)$ = sign$(\hat{\beta}_k)$
- $\frac{\partial \Pr(Y=\ell)}{\partial X_k}$, $\ell \in \{2, 3, ...J-1\}$ are non-monotonic

Marginal Effects, Illustrated



For a δ-unit change in X_k:

$$\begin{aligned}
\mathrm{OR}_{X_k} &= \frac{\frac{\Pr(Y>j|\mathbf{X},X_k+\delta)}{\Pr(Y\leq j|\mathbf{X},X_k+\delta)}}{\frac{\Pr(Y>j|\mathbf{X},X_k)}{\Pr(Y\leq j|\mathbf{X},X_k)}} \\
&= \exp(\delta\hat{\beta}_k)
\end{aligned}$$

```
> olreg.or <- function(model)
+  {
+  coeffs <- coef(summary(model))
+  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
+  }

> olreg.or(beer.logit)

             lci       or      uci
price     0.3586   0.6373    1.133
calories  1.0231   1.0479    1.073
craftbeer 0.0287   0.1818    1.152
bitter    0.8933   0.9707    1.055
malty     1.0023   1.0518    1.104
1|2       0.6003  15.9748  425.133
2|3       2.4319  71.4963 2101.961
3|4       8.4053 264.4357 8319.319
```

Odds Ratios: Explication

- `craftbeer`:
    - $\exp(-1.705) = 0.18$
    - "The odds of being rated "Good" or better (versus "Fair") are more than 80 percent lower for a craft beer than for a regular beer."
    - "The odds of being rated "Very Good" or better (versus "Fair" or "Good") are more than 80 percent lower for a craft beer than for a regular beer."
- `calories`:
    - $\exp(0.047) = 1.05$
    - "A one-calorie increase raises the odds of being in a higher set of categories (versus all lower ones) by about five percent."
    - etc.

Predicted Probabilities: Basics

$$\Pr(\widehat{Y_i = j}|\mathbf{X}) = F(\hat{\tau}_j - \bar{\mathbf{X}}_i\hat{\beta}) - F(\hat{\tau}_{j-1} - \bar{\mathbf{X}}_i\hat{\beta})$$

Means:

- `price` $= 4.96$, `calories` $= 142$, `craftbeer` $= 0$, `bitter` $= 35.4$, `malty` $= 33.1$.
- Yields:

$$\begin{aligned}
\sum_{k=1}^{K} \bar{\mathbf{X}}_k \hat{\beta}_k &= -0.45 \times 4.96 + 0.047 \times 142 - 1.70 \times 0 - \\
&\quad 0.03 \times 35.4 + 0.05 \times 33.1 \\
&= -2.23 + 6.67 - 0 - 1.06 + 1.66 \\
&= \mathbf{5.04}.
\end{aligned}$$

Predicted Probabilities: "By Hand"

$$\begin{aligned}
\Pr(Y=1) &= \Lambda(2.77-5.04)-0 \\
&= \frac{\exp(-2.27)}{1+\exp(-2.27)} \\
&= \mathbf{0.09}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=2) &= \Lambda(4.27-5.04)-\Lambda(2.77-5.04) \\
&= \Lambda(-0.77)-\Lambda(-2.27) \\
&= 0.32-0.09 \\
&= \mathbf{0.23}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=3) &= \Lambda(5.58-5.04)-\Lambda(4.27-5.04) \\
&= \Lambda(0.54)-\Lambda(-0.77) \\
&= 0.63-0.32 \\
&= \mathbf{0.31}.
\end{aligned}$$

$$\begin{aligned}
\Pr(Y=4) &= 1-\Lambda(5.58-5.04) \\
&= 1-\Lambda(0.54) \\
&= 1-0.63 \\
&= \mathbf{0.37}.
\end{aligned}$$

Changes in Predicted Probabilities

For `craftbeer`=1:

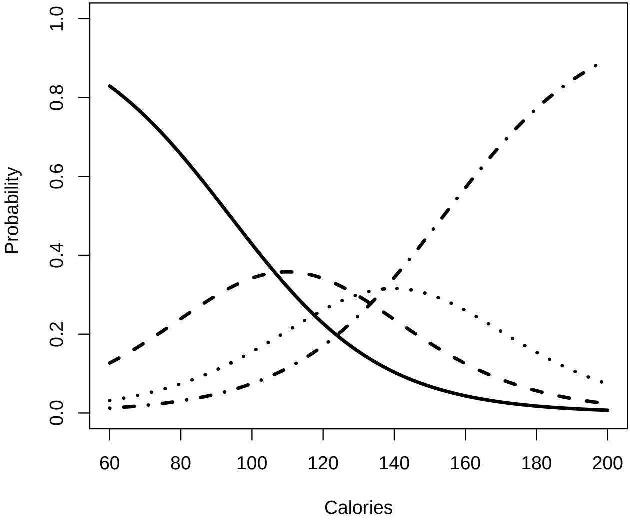
- $\Pr(Y = 1) = \Lambda(2.77 - 3.34) - 0 = \mathbf{0.36}$.
- $\Pr(Y = 2) = \Lambda(4.27 - 3.34) - \Lambda(2.77 - 3.34) = 0.72 - 0.36 = \mathbf{0.36}$.
- $\Pr(Y = 3) = \Lambda(5.58 - 3.34) - \Lambda(4.27 - 3.34) = 0.90 - 0.72 = \mathbf{0.18}$.
- $\Pr(Y = 4) = 1 - 0.90 = \mathbf{0.10}$.

Outcome	Change in Probability
ΔPr(Fair)	0.27
ΔPr(Good)	0.13
ΔPr(Very Good)	-0.13
ΔPr(Excellent)	-0.27

Predicted Probability Plots

- Can be category-specific or "cumulative"
- `polr`:
    - In-sample in `$fitted.values`
    - `polr` class supports `predict`, `confint`, etc.
- `ologit` / `oprobit`: using `predict`

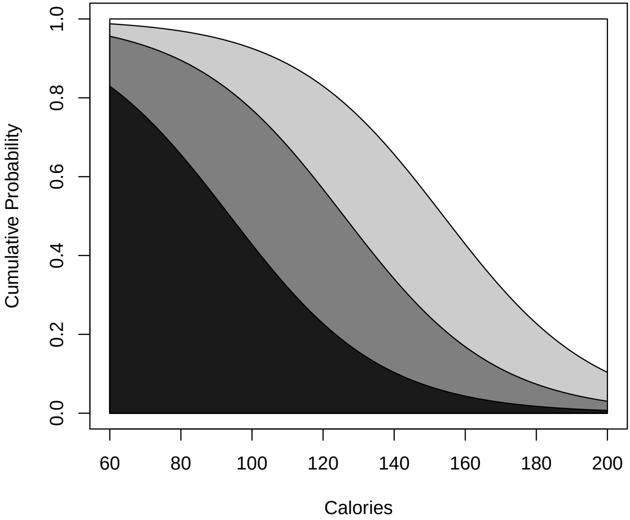
Plot by Outcome



(How'd He Do That?)

```
> calories<-seq(60,200,1)
> price<-mean(beer$price)
> craftbeer<-median(beer$craftbeer)
> bitter<-mean(beer$bitter)
> malty<-mean(beer$malty)
> beersim<-cbind(calories,price,craftbeer,bitter,malty)
> beer.hat<-predict(beer.logit,beersim,type='probs')
> plot(c(60,200), c(0,1), type='n', xlab="Calories", ylab='Fitted
  Probability')
> lines(60:200, beer.hat[1:141, 1], lty=1, lwd=3)
> lines(60:200, beer.hat[1:141, 2], lty=2, lwd=3)
> lines(60:200, beer.hat[1:141, 3], lty=3, lwd=3)
> lines(60:200, beer.hat[1:141, 4], lty=4, lwd=3)
```

Cumulative Predicted Probabilities



(code...)

```
> xaxis<-c(60,60:200,200)
> yaxis1<-c(0,beer.hat[,1],0)
> yaxis2<-c(0,beer.hat[,2]+beer.hat[,1],0)
> yaxis3<-c(0,beer.hat[,3]+beer.hat[,2]+beer.hat[,1],0)
> yaxis4<-c(0,beer.hat[,4]+beer.hat[,3]+beer.hat[,2]+beer.hat[,1],0)
>
> plot(c(60,200), c(0,1), type='n', xlab="Calories", ylab="Cumulative
  Probability")
> polygon(xaxis,yaxis4,col="white")
> polygon(xaxis,yaxis3,col="grey80")
> polygon(xaxis,yaxis2,col="grey50")
> polygon(xaxis,yaxis1,col="grey10")
```

Variants

- Generalized (relaxes parallel regressions; Brant (1990))
- Heteroscedastic
- Varying τs (Maddala, Terza, Sanders)
- Models for "balanced" scales (Jones & Sobel)
- Compound Ordered Hierarchical Probit ("`chopit`") (Wand & King)
- "Zero-Inflated" Ordered Models (Hill, Bagozzi, Moore & Mukherjee)
- Latent class/mixture models (Winkelmann, etc.)

Nominal / Unordered Outcomes

$$Y_i = j,\ j \in \{1, 2, ...J\}$$

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

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

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

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

Define:

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

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):

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

A (Descriptive) Example: 1992 Election

- 1992 National Election Study
- $Y \in \{\text{Bush} = 1, \text{Clinton} = 2, \text{Perot} = 3\}$
- $N = 1473$.
- X = Party ID:
  $\{$"Strong Democrats" $= 1 \rightarrow$ "Strong Republicans" $= 7\}$

MNL: 1992 Election ("Baseline" = Perot)

```
> nes92.mlogit<-vglm(presvote~partyid, multinomial, nes92)
> summary(nes92.mlogit)

Call:
vglm(formula = presvote ~ partyid, family = multinomial, data = nes92)

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)    
(Intercept):1  -1.8152     0.2456   -7.39     0.00000000000014 ***
(Intercept):2   3.0273     0.1783   16.98 < 0.0000000000000002 ***
partyid:1       0.4827     0.0476   10.15 < 0.0000000000000002 ***
partyid:2      -0.6805     0.0478  -14.25 < 0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

MNL: 1992 Election ("Baseline" = Bush)

```
> Bush.nes92.mlogit<-vglm(formula = presvote~partyid,
          family=multinomial(refLevel=1),data=nes92)
> summary(Bush.nes92.mlogit)

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)
(Intercept):1   4.8425     0.2373   20.41 < 0.0000000000000002 ***
(Intercept):2   1.8152     0.2456    7.39     0.00000000000014 ***
partyid:1      -1.1632     0.0546  -21.32 < 0.0000000000000002 ***
partyid:2      -0.4827     0.0476  -10.15 < 0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

MNL: 1992 Election ("Baseline" = Clinton)

```
> Clinton.nes92.mlogit<-vglm(formula=presvote~partyid,
                   family=multinomial(refLevel=2),data=nes92)
> summary(Clinton.nes92.mlogit)

Coefficients:
              Estimate Std. Error z value            Pr(>|z|)
(Intercept):1  -4.8425     0.2373   -20.4 <0.0000000000000002 ***
(Intercept):2  -3.0273     0.1783   -17.0 <0.0000000000000002 ***
partyid:1       1.1632     0.0546    21.3 <0.0000000000000002 ***
partyid:2       0.6805     0.0478    14.2 <0.0000000000000002 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of iterations: 5
```

Coefficient Estimates and "Baselines"

		"Baseline" category		
		Clinton	Perot	Bush
Comparison	Clinton	–	-0.68	-1.16
Category	Perot	0.68	–	-0.48
	Bush	1.16	0.48	–

Interpretation: Example Data Redux

- 1992 ANES ($N = 1473$)
- Variables:
  - `presvote`: 1=Bush, 2=Clinton, 3=Perot
  - `partyid`: (seven-point scale, 7=GOP)
  - `age` (in years)
  - `white` (naturally coded)
  - `female` (ditto)

Baseline MNL Results: 1992 Election

```
> NES.MNL<-vglm(presvote~partyid+age+white+female,data=BigNES92,
+            multinomial(refLevel=1))
> summaryvglm(NES.MNL)

Call:
vglm(formula = presvote ~ partyid + age + white + female, family = multinomial(refLevel = 1),
    data = BigNES92)

Coefficients:
              Estimate Std. Error z value              Pr(>|z|)
(Intercept):1  5.80665    0.44301   13.11 < 0.0000000000000002 ***
(Intercept):2  1.98008    0.52454    3.77              0.00016 ***
partyid:1     -1.13561    0.05486  -20.70 < 0.0000000000000002 ***
partyid:2     -0.50132    0.04870  -10.29 < 0.0000000000000002 ***
age:1         -0.00260    0.00514   -0.51              0.61276
age:2         -0.01556    0.00504   -3.09              0.00203 **
whiteWhite:1  -0.98908    0.31346   -3.16              0.00160 **
whiteWhite:2   0.87918    0.43605    2.02              0.04377 *
female:1      -0.12500    0.16895   -0.74              0.45936
female:2      -0.50928    0.16266   -3.13              0.00174 **
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Number of linear predictors:  2
Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Dispersion Parameter for multinomial family:   1
Residual deviance: 2107 on 2936 degrees of freedom
Log-likelihood: -1054 on 2936 degrees of freedom
Number of iterations: 5
```

Global $\ln LR$ statistic Q tests:

$$\hat{\boldsymbol{\beta}} = \mathbf{0} \, \forall \, j, k$$

$$Q \sim \chi^2_{(J-1)(k-1)}$$

Test H: No Effect of `age`

```
> library(aod)
> wald.test(b=c(t(coef(NES.MNL))),Sigma=vcov(NES.MNL),Terms=c(5,6))

Wald test:
----------

Chi-squared test:
X2 = 11.0, df = 2, P(> X2) = 0.0042
```

Test H: No Difference – Clinton vs. Bush

```
> wald.test(b=c(t(coef(NES.MNL))),Sigma=vcov(NES.MNL),Terms=c(1,3,5,7,9))

Wald test:
----------

Chi-squared test:
X2 = 444.6, df = 5, P(> X2) = 0.0
```

```
> PickBush<-ifelse(fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,2]
        & fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,3], 1,0)
> PickWJC<-ifelse(fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,1]
        & fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,3], 2, 0)
> PickHRP<-ifelse(fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,1]
        & fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,2], 3, 0)

> OutHat<-PickBush+PickWJC+PickHRP
> table(BigNES92$presvote,OutHat)

   OutHat
      1   2   3
  1 415  77   8
  2  56 619  16
  3 135 133  14
```

- "Null" Model: $\left(\frac{691}{1473}\right) = 46.9\%$ correct.
- Estimated model: $\frac{(415+619+14)}{1473} = \frac{1048}{1473} = 71.2\%$ correct.
- PRE $= \frac{1048-691}{1473-691} = \frac{357}{782} = 45.7\%$.
- Correct predictions: 90% Clinton, 83% Bush, 5% Perot.

Odds ("Relative Risk") Ratios

$$\ln\left[\frac{\Pr(Y_i = j|\mathbf{X})}{\Pr(Y_i = j'|\mathbf{X})}\right] = \mathbf{X}(\hat{\boldsymbol{\beta}}_j - \hat{\boldsymbol{\beta}}_{j'})$$

Setting $\hat{\boldsymbol{\beta}}_{j'} = \mathbf{0}$:

$$\ln\left[\frac{\Pr(Y_i = j|\mathbf{X})}{\Pr(Y_i = j'|\mathbf{X})}\right] = \mathbf{X}\hat{\boldsymbol{\beta}}_j$$

One-Unit Change in X_k:

$$RRR_{jk} = \exp(\boldsymbol{\beta}_{jk})$$

δ-Unit Change in X_k:

$$RRR_{jk} = \exp(\boldsymbol{\beta}_{jk} \times \delta)$$

Odds ("Relative Risk") Ratios

```
> mnl.or <- function(model) {
    coeffs <- c(t(coef(model)))
    lci <- exp(coeffs - 1.96 * diag(vcov(NES.MNL))^0.5)
    or <- exp(coeffs)
    uci <- exp(coeffs + 1.96* diag(vcov(NES.MNL))^0.5)
    lreg.or <- cbind(lci, or, uci)
    lreg.or
  }

> mnl.or(NES.MNL)
                     lci       or      uci
(Intercept):1 139.5398 332.5036 792.3088
(Intercept):2   2.5909   7.2433  20.2504
partyid:1       0.2885   0.3212   0.3577
partyid:2       0.5506   0.6057   0.6664
age:1           0.9874   0.9974   1.0075
age:2           0.9749   0.9846   0.9943
whiteWhite:1    0.2012   0.3719   0.6875
whiteWhite:2    1.0248   2.4089   5.6623
female:1        0.6337   0.8825   1.2289
female:2        0.4369   0.6009   0.8266
```

- A one unit increase in `partyid` corresponds to:
  - A decrease in the odds of a Clinton vote, versus a vote for Bush, of $\exp(-1.136) = 0.321$ (or about 68 percent), and
  - A decrease in the odds of a Perot vote, versus a vote for Bush, of $\exp(-0.501) = 0.606$ (or about 40 percent).
  - These are *large* decreases in the odds – not surprisingly, more Republican voters are *much* more likely to vote for Bush than for Perot or Clinton.
- Similarly, `female` voters are:
  - No more or less likely to vote for Clinton vs. Bush (OR=0.88), but
  - Roughly 40 percent less likely to have voted for Perot (OR=0.60).

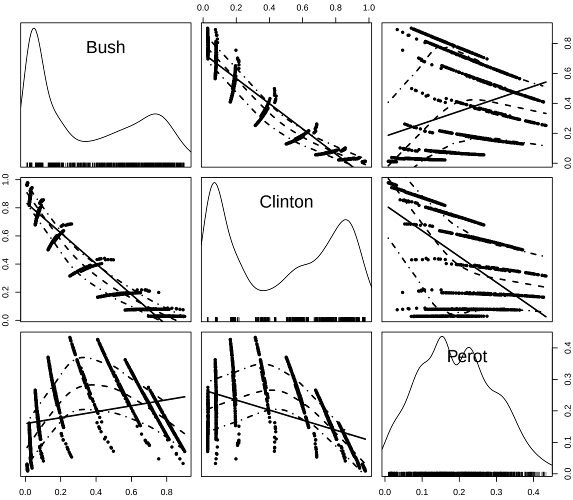
Predicted Probabilities

$$\begin{aligned}\Pr(\widehat{\texttt{presvote}_i} = \text{Bush}) &= \frac{\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}}_{\text{Bush}})}{\sum_{j=1}^{J}\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}}_j)} \\ &= \frac{1}{1+\sum_{j=2}^{J}\exp(\mathbf{X}_i\hat{\boldsymbol{\beta}}_j)}\end{aligned}$$

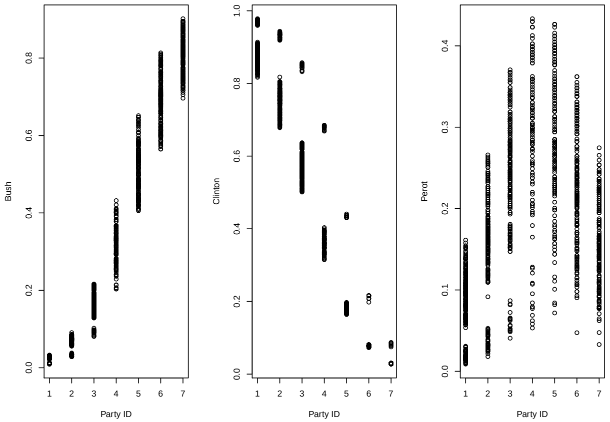
In-Sample Predicted Probabilities

```
> hats<-as.data.frame(fitted.values(NES.MNL))
> names(hats)[3]<-"Perot" # nice names...
> names(hats)[2]<-"Clinton"
> names(hats)[1]<-"Bush"
> attach(hats)

> library(car)
> scatterplot.matrix(~Bush+Clinton+Perot,
  diagonal="histogram",col=c("black","grey"))
```

In-Sample $\widehat{\Pr}$s



In-Sample $\widehat{\text{Pr}}$s vs. `partyid`



- Conditional logit (MNL with choice-specific covariates)
- "Independence of Irrelevant Alternatives"
- $\rightarrow$ Multinomial Probit
- $\rightarrow$ Heteroscedastic Extreme Value model
- "Mixed" Logit
- Nested Logit

Stata things...

MNL in Stata

```
. mlogit presvote partyid, baseoutcome(3)

Multinomial logistic regression                   Number of obs   =       1473
                                                  LR chi2(2)      =     891.93
                                                  Prob > chi2     =     0.0000
Log likelihood = -1083.4749                       Pseudo R2       =     0.2916
------------------------------------------------------------------------------
    presvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
1            |
     partyid |   .4826514   .0475563    10.15   0.000     .3894427      .57586
       _cons |  -1.815236    .245572    -7.39   0.000    -2.296548   -1.333923
-------------+----------------------------------------------------------------
2            |
     partyid |  -.6805163     .04777   -14.25   0.000    -.7741438   -.5868889
       _cons |   3.027259   .1782539    16.98   0.000     2.677888     3.37663
------------------------------------------------------------------------------
(presvote==3 is the base outcome)
```

CL vs. Binary Logit

```
. clogit vote FT, group(caseid)
Conditional (fixed-effects) logistic regression   Number of obs   =       4419
                                                  LR chi2(1)      =    1600.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -817.96837                       Pseudo R2       =     0.4945
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          FT |   .0766544   .0031035    24.70   0.000     .0705716    .0827372
------------------------------------------------------------------------------
```

```
. logit vote FT
Logistic regression                               Number of obs   =       4419
                                                  LR chi2(1)      =    1777.33
                                                  Prob > chi2     =     0.0000
Log likelihood = -1924.0926                       Pseudo R2       =     0.3159
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          FT |   .0744263   .0024393    30.51   0.000     .0696453    .0792073
       _cons |   -5.00402   .1566234   -31.95   0.000    -5.310997   -4.697044
------------------------------------------------------------------------------
```

```
. mlogit presvote partyid if candid==1, baseoutcome(1)

Multinomial logistic regression                   Number of obs   =       1473
                                                  LR chi2(2)      =     891.93
                                                  Prob > chi2     =     0.0000
Log likelihood = -1083.4749                       Pseudo R2       =     0.2916
------------------------------------------------------------------------------
    presvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
2            |
     partyid | -1.163168   .0545618   -21.32   0.000    -1.270107   -1.056229
       _cons |  4.842495   .2373171    20.41   0.000     4.377362    5.307628
-------------+----------------------------------------------------------------
3            |
     partyid | -.4826514   .0475563   -10.15   0.000      -.57586   -.3894427
       _cons |  1.815236    .245572     7.39   0.000     1.333923    2.296548
------------------------------------------------------------------------------
(presvote==1 is the base outcome)
```

Example Redux: CL

```
. clogit vote FT, group(caseid)

Conditional (fixed-effects) logistic regression   Number of obs   =       4419
                                                  LR chi2(1)      =    1600.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -817.96837                       Pseudo R2       =     0.4945
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          FT |   .0766544   .0031035    24.70   0.000     .0705716    .0827372
------------------------------------------------------------------------------
```

Example Redux: CL

```
. gen clintondummy=(candid==2)
. gen perotdummy=(candid==3)
. gen PIDxClinton = partyid * clintondummy
. gen PIDxPerot = partyid * perotdummy

. clogit vote clintondummy perotdummy PIDxClinton PIDxPerot, group(caseid)

Conditional (fixed-effects) logistic regression   Number of obs   =       4419
                                                  LR chi2(4)      =    1069.56
                                                  Prob > chi2     =     0.0000
Log likelihood = -1083.4749                       Pseudo R2       =     0.3305
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
clintondummy |   4.842495   .2373171    20.41   0.000     4.377362    5.307628
  perotdummy |   1.815236    .245572     7.39   0.000     1.333923    2.296548
 PIDxClinton |  -1.163168   .0545618   -21.32   0.000    -1.270107   -1.056229
   PIDxPerot |  -.4826514   .0475563   -10.15   0.000      -.57586   -.3894427
------------------------------------------------------------------------------
```

```
. clogit vote FT clintondummy perotdummy PIDxClinton PIDxPerot, group(caseid)

Conditional (fixed-effects) logistic regression   Number of obs   =       4419
                                                  LR chi2(5)      =    1764.49
                                                  Prob > chi2     =     0.0000
Log likelihood =  -736.0092                       Pseudo R2       =     0.5452
------------------------------------------------------------------------------
        vote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          FT |   .0629875   .0032175    19.58   0.000     .0566813    .0692937
clintondummy |   2.812737   .2687999    10.46   0.000     2.285899    3.339576
  perotdummy |   .9435437   .2856252     3.30   0.001     .3837286    1.503359
 PIDxClinton |  -.6318723    .062255   -10.15   0.000    -.7538899   -.5098548
   PIDxPerot |  -.1921175    .057032    -3.37   0.001    -.3038981   -.0803369
------------------------------------------------------------------------------
```

MNL (Stata version)

```
. mlogit presvote partyid age white female, baseoutcome(1)
Multinomial logistic regression                   Number of obs   =       1473
                                                  LR chi2(8)      =     951.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -1053.6506                       Pseudo R2       =     0.3111
------------------------------------------------------------------------------
    presvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
2            |
     partyid |  -1.135615   .0548618   -20.70   0.000    -1.243142   -1.028088
         age |  -.0026013   .0051396    -0.51   0.613    -.0126746     .007472
       white |    -.98908   .3134669    -3.16   0.002    -1.603464   -.3746961
      female |   -.125005   .1689499    -0.74   0.459    -.4561406    .2061307
       _cons |   5.806651   .4430144    13.11   0.000     4.938358    6.674943
-------------+----------------------------------------------------------------
3            |
     partyid |  -.5013218   .0486977   -10.29   0.000    -.5967675   -.4058761
         age |   -.015565   .0050436    -3.09   0.002    -.0254503   -.0056796
       white |   .8791807   .4360556     2.02   0.044     .0245275    1.733834
      female |   -.509278   .1626614    -3.13   0.002    -.8280884   -.1904676
       _cons |   1.980081   .5245439     3.77   0.000     .9519936    3.008168
------------------------------------------------------------------------------
```

“Relative Risk Ratios” (Stata Remix)

```
. mlogit, rrr

Multinomial logistic regression                   Number of obs   =       1473
                                                  LR chi2(8)      =     951.58
                                                  Prob > chi2     =     0.0000
Log likelihood = -1053.6506                       Pseudo R2       =     0.3111

------------------------------------------------------------------------------
    presvote |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
2            |
     partyid |   .3212245    .017623   -20.70   0.000     .2884764    .3576903
         age |   .9974021   .0051262    -0.51   0.613     .9874054      1.0075
       white |   .3719187   .1165842    -3.16   0.002     .2011984    .6874982
      female |   .8824925    .149097    -0.74   0.459     .6337247    1.228914
-------------+----------------------------------------------------------------
3            |
     partyid |   .6057295   .0294976   -10.29   0.000     .5505886    .6663927
         age |   .9845555   .0049657    -3.09   0.002     .9748708    .9943365
       white |   2.408925   1.050425     2.02   0.044     1.024831    5.662322
      female |   .6009293    .097748    -3.13   0.002     .4368836    .8265726
------------------------------------------------------------------------------
```

Conditional Logit: Odds Ratios (Stata)

```
. clogit, or

Conditional (fixed-effects) logistic regression   Number of obs   =       4419
                                                  LR chi2(5)      =    1764.49
                                                  Prob > chi2     =     0.0000
Log likelihood =  -736.0092                       Pseudo R2       =     0.5452
------------------------------------------------------------------------------
        vote | Odds Ratio   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          FT |   1.065014   .0034267    19.58   0.000     1.058318    1.071751
clintondummy |   16.65545   4.476983    10.46   0.000     9.834527    28.20715
  perotdummy |   2.569069   .7337909     3.30   0.001     1.467747    4.496767
 PIDxClinton |   .5315956   .0330945   -10.15   0.000     .4705327    .6005828
   PIDxPerot |   .8252099   .0470633    -3.37   0.001     .7379361    .9228054
------------------------------------------------------------------------------
```