PLSC 503 – Spring 2023
Regression Models for Nominal and Ordinal Outcomes

April 10, 2023

Motivation: Discrete *Outcomes*

Outcome variable has $J > 2$ *unordered* categories:

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

Write:

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

Means that:

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

And set:

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

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*

Utility:

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

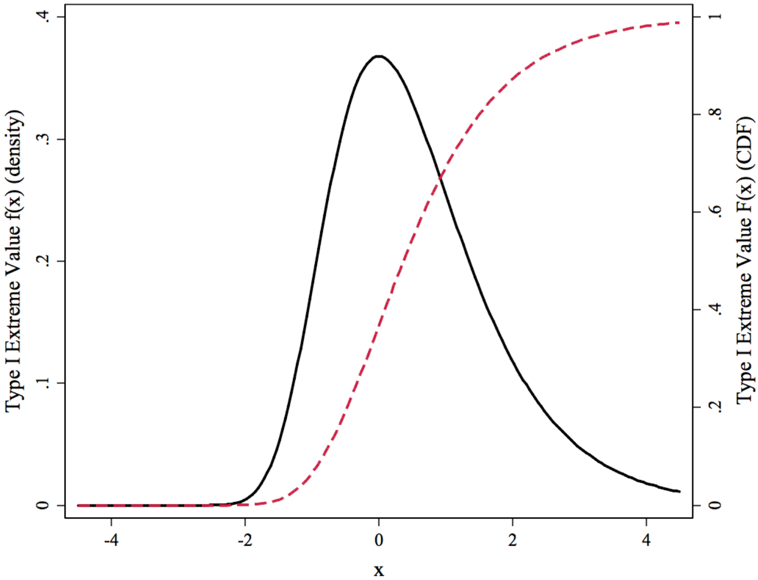
$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}_j$$

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

$\epsilon \sim$???

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

Type I Extreme Value



The probability of choosing choice j is:

$$
\begin{aligned}
\Pr(Y_i = j) &= \Pr(U_j > U_1, U_j > U_2, ...U_j > U_J) \\
&= \int f(\epsilon_j) \left[\int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_1} f(\epsilon_1) d\epsilon_1 \times \int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_2} f(\epsilon_2) d\epsilon_2 \times ... \right] d\epsilon_j \\
&= \int f(\epsilon_j) \times \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_1)] \times \\
&\quad \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \boldsymbol{\beta}_j - \mathbf{X}_i \boldsymbol{\beta}_2)] \times ... d\epsilon_j \\
&= \frac{\exp(\mathbf{X}_i \boldsymbol{\beta}_j)}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \boldsymbol{\beta}_j)}
\end{aligned}
$$

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):

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

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  1.4e-13 ***
(Intercept):2   3.0273     0.1783   16.98  < 2e-16 ***
partyid:1       0.4827     0.0476   10.15  < 2e-16 ***
partyid:2      -0.6805     0.0478  -14.25  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Names of linear predictors: log(mu[,1]/mu[,3]), log(mu[,2]/mu[,3])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  3  of the response
```

MNL: 1992 Election ("Baseline" = Bush)

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

Call:
vglm(formula = presvote ~ partyid, family = multinomial(refLevel = 1),
    data = nes92)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1   4.8425     0.2373   20.41  < 2e-16 ***
(Intercept):2   1.8152     0.2456    7.39  1.4e-13 ***
partyid:1      -1.1632     0.0546  -21.32  < 2e-16 ***
partyid:2      -0.4827     0.0476  -10.15  < 2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  1  of the response
```

MNL: 1992 Election ("Baseline" = Clinton)

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

Call:
vglm(formula = presvote ~ partyid, family = multinomial(refLevel = 2),
    data = nes92)

Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept):1  -4.8425     0.2373   -20.4   <2e-16 ***
(Intercept):2  -3.0273     0.1783   -17.0   <2e-16 ***
partyid:1       1.1632     0.0546    21.3   <2e-16 ***
partyid:2       0.6805     0.0478    14.2   <2e-16 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1
Names of linear predictors: log(mu[,1]/mu[,2]), log(mu[,3]/mu[,2])
Residual deviance: 2167 on 2942 degrees of freedom
Log-likelihood: -1083 on 2942 degrees of freedom
Number of Fisher scoring iterations: 5
Reference group is level  2  of the response
```

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	–

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

Choice-Specific Covariates: Data Structure

```
> nes92CL<-mlogit.data(nes92,shape="wide",choice="PVote",varying=4:6)
> head(nes92CL,6)
~~~~~~~
 first 6 observations out of 4419
~~~~~~~
  caseid presvote partyid PVote     alt  FT chid    idx
1   3001        1       6  TRUE    Bush  85    1 1:Bush
2   3001        1       6 FALSE Clinton  30    1 1:nton
3   3001        1       6 FALSE   Perot   0    1 1:erot
4   3002        1       7  TRUE    Bush 100    2 2:Bush
5   3002        1       7 FALSE Clinton   0    2 2:nton
6   3002        1       7 FALSE   Perot   0    2 2:erot

~~~ indexes ~~~~
  chid     alt
1    1    Bush
2    1 Clinton
3    1   Perot
4    2    Bush
5    2 Clinton
6    2   Perot
indexes:  1, 2
```

Note that:

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

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

- "Fixed effects" (choice-specific intercepts), plus
- Observation-specific $\mathbf{X}$s, plus
- Interactions...

CL in R : Estimation

```
> nes92.clogit<-mlogit(PVote~FT|partyid,data=nes92CL)
> summary(nes92.clogit)

Call:
mlogit(formula = PVote ~ FT | partyid, data = nes92CL, method = "nr",
    print.level = 0)

Frequencies of alternatives:
   Bush Clinton   Perot
  0.339   0.469   0.191

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 0.00293
successive function values within tolerance limits

Coefficients :
                    Estimate Std. Error t-value             Pr(>|t|)
Clinton:(intercept)  2.81272    0.26880   10.46 < 0.0000000000000002 ***
Perot:(intercept)    0.94353    0.28563    3.30              0.00096 ***
FT                   0.06299    0.00322   19.58 < 0.0000000000000002 ***
Clinton:partyid     -0.63187    0.06225  -10.15 < 0.0000000000000002 ***
Perot:partyid       -0.19212    0.05703   -3.37              0.00076 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -736
McFadden R^2:  0.519
Likelihood ratio test : chisq = 1590 (p.value = <0.0000000000000002)
```

Interpretation: Example Data Redux

Back to 1992 again:

- 1992 ANES ($N = 1473$)
- Outcome + predictors:
  - `PresVote` $\in$ {Bush, Clinton, Perot} (factor)
  - `presvote`: 1=Bush, 2=Clinton, 3=Perot (numeric)
  - `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  < 2e-16 ***
(Intercept):2  1.98008    0.52454    3.77  0.00016 ***
partyid:1     -1.13561    0.05486  -20.70  < 2e-16 ***
partyid:2     -0.50132    0.04870  -10.29  < 2e-16 ***
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

Names of linear predictors: log(mu[,2]/mu[,1]), log(mu[,3]/mu[,1])
Residual deviance: 2107 on 2936 degrees of freedom
Log-likelihood: -1054 on 2936 degrees of freedom
Number of Fisher scoring iterations: 5
No Hauck-Donner effect found in any of the estimates
Reference group is level  1  of the response
```

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
```

In-Sample Predicted Outcomes

```
> 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.

Interpretation: Marginal Effects

$$\frac{\partial \mathrm{Pr}(Y_i = j)}{\partial X_k} = \mathrm{Pr}(Y_i = j|\mathbf{X}) \left[\hat{\boldsymbol{\beta}}_{jk} - \sum_{j=1}^{J} \hat{\boldsymbol{\beta}}_{jk} \times \mathrm{Pr}(Y_i = j|\mathbf{X})\right]$$

Depends on:

- $\mathrm{Pr}\widehat{(Y_i =} j)$
- $\hat{\beta}_{jk}$
- $\sum_{j=1}^{J} \hat{\boldsymbol{\beta}}_{jk}$

Available for `-multinom-` (in the `-nnet-` package) via the `-margins-` package...

Marginal Effects: Illustrated

```
> Re-fit the model using -multinom-:
>
> BigNES92$PresVote<-cut(BigNES92$presvote,3,labels=c("Bush","Clinton","Perot")
> BigNES92$White<-ifelse(BigNES92$white=="White",1,0) # numeric
> MNL.alt<-multinom(PresVote~partyid+age+White+female,data=BigNES92,
+                   Hess=TRUE)
# weights:  18 (10 variable)
initial  value 1618.255901
iter  10 value 1077.315546
final  value 1053.650587
converged

> summary(marginal_effects(MNL.alt))
  dydx_partyid       dydx_age           dydx_White         dydx_female
 Min.   :0.0104   Min.   :0.00003   Min.   :-0.1482   Min.   :0.0013
 1st Qu.:0.0578   1st Qu.:0.00032   1st Qu.:-0.0608   1st Qu.:0.0125
 Median :0.1069   Median :0.00093   Median : 0.0190   Median :0.0344
 Mean   :0.1060   Mean   :0.00130   Mean   :-0.0044   Mean   :0.0450
 3rd Qu.:0.1490   3rd Qu.:0.00234   3rd Qu.: 0.0402   3rd Qu.:0.0801
 Max.   :0.2612   Max.   :0.00329   Max.   : 0.1805   Max.   :0.1093
```

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
```

Odds Ratios: Interpretation

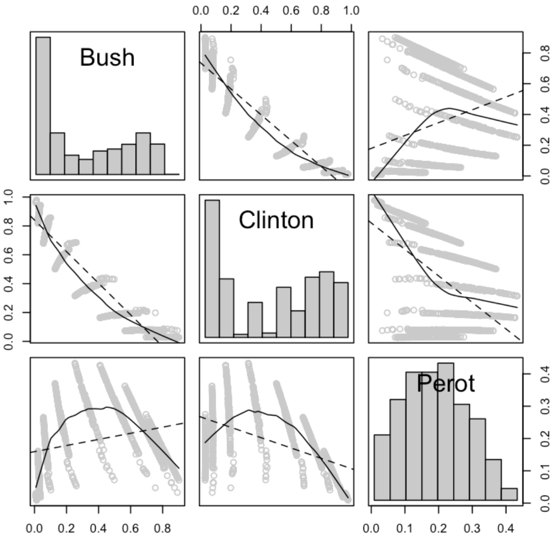
- A one unit increase in `partyid` corresponds to:
  - A decrease in the odds of a Clinton vote, versus a vote for Bush, of $\exp(-1.136) = 0.321$ (or about 68 percent), and
  - A decrease in the odds of a Perot vote, versus a vote for Bush, of $\exp(-0.501) = 0.606$ (or about 40 percent).
  - These 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

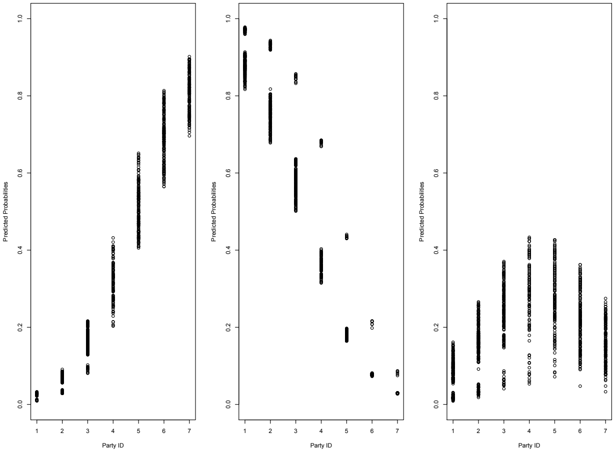
$$
\begin{aligned}
\widehat{\Pr(\texttt{presvote}_i} = \text{Bush}) &= \frac{\exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_{\text{Bush}})}{\sum_{j=1}^{J} \exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_j)} \\
&= \frac{1}{1 + \sum_{j=2}^{J} \exp(\mathbf{X}_i \hat{\boldsymbol{\beta}}_j)}
\end{aligned}
$$

In-Sample Predicted Probabilities

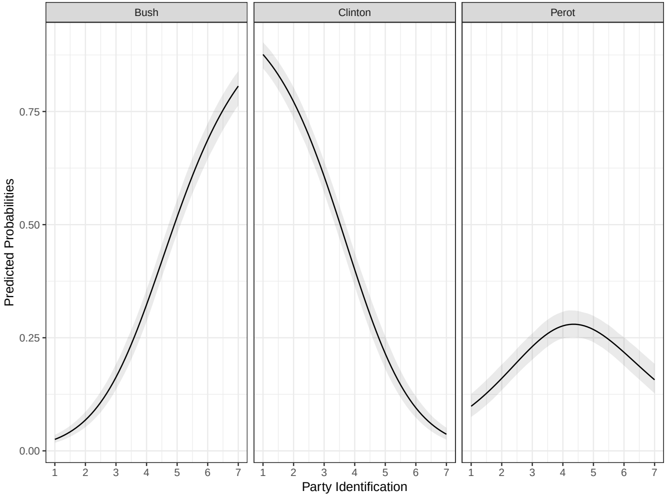
```
> hats<-as.data.frame(fitted.values(NES.MNL))
> names(hats)[3]<-"Perot" # nice names...
> names(hats)[2]<-"Clinton"
> names(hats)[1]<-"Bush"
> attach(hats)

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


Bush
Clinton
Perot
0.0 0.2 0.4 0.6 0.8 1.0
0.0 0.2 0.4 0.6 0.8
0.0 0.1 0.2 0.3 0.4

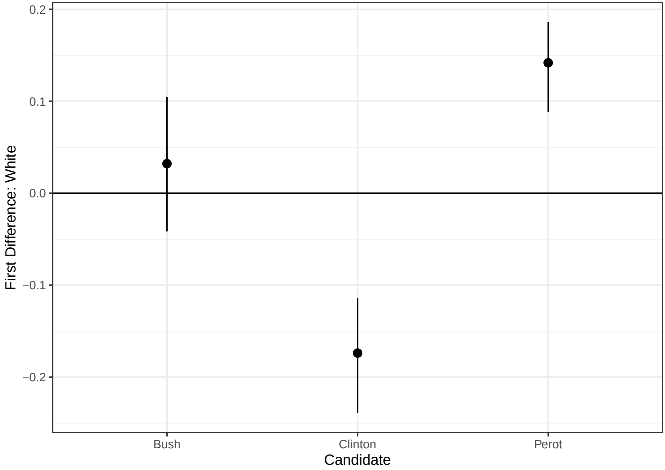
In-Sample $\widehat{\Pr}$s vs. `partyid`



Out-Of-Sample Predictions (using `MNLpred`)



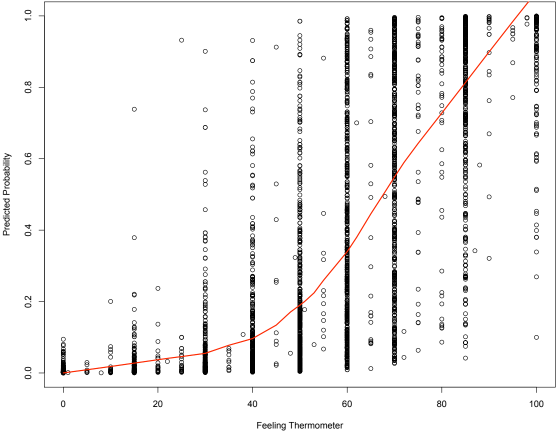
OOS First Differences (using `MNLpred`)



```
> nes92.clogit<-mlogit(PVote~FT|partyid,data=nes92CL)
> summary(nes92.clogit)

Call:
mlogit(formula = PVote ~ FT | partyid, data = nes92CL, method = "nr")

Frequencies of alternatives:choice
   Bush Clinton   Perot
  0.339   0.469   0.191

nr method
6 iterations, 0h:0m:0s
g'(-H)^-1g = 0.00293
successive function values within tolerance limits

Coefficients :
                    Estimate Std. Error z-value Pr(>|z|)
(Intercept):Clinton  2.81272    0.26880   10.46  < 2e-16 ***
(Intercept):Perot    0.94353    0.28563    3.30  0.00096 ***
FT                   0.06299    0.00322   19.58  < 2e-16 ***
partyid:Clinton     -0.63187    0.06225  -10.15  < 2e-16 ***
partyid:Perot       -0.19212    0.05703   -3.37  0.00076 ***
---
Signif. codes:  0 *** 0.001 ** 0.01 * 0.05 . 0.1   1

Log-Likelihood: -736
McFadden R^2:  0.519
Likelihood ratio test : chisq = 1590 (p.value = <2e-16)
```

Predicted Probabilities (In-Sample)

```
> CLhats<-predict(NES.CL,type="expected")
> plot(cldata$FT,CLhats,xlab="Feeling Thermometer",ylab="Predicted Probability")
> lines(lowess(CLhats~cldata$FT),lwd=2,col="red")
```



Other Topics (for PLSC 504)

- "Independence of Irrelevant Alternatives"
- $\rightarrow$ Multinomial Probit
- $\rightarrow$ Heteroscedastic Extreme Value model
- "Mixed" Logit
- Nested Logit

Models for Ordinal Outcomes

Ordinal data are:

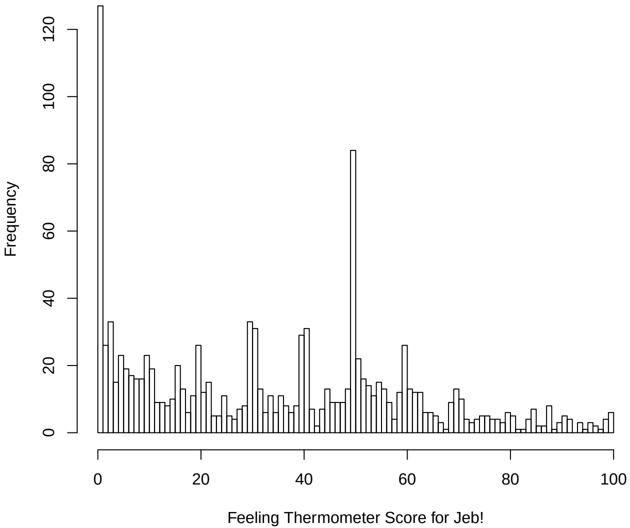
- Discrete: $Y \in \{1, 2, ...\}$
- *Grouped Continuous* Data
- *Assessed Ordered* Data

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

"I'd like to get your feelings toward some of our political leaders and other people who are in the news these days. I'll read the name of a person and I'd like you to rate that person using something we call the feeling thermometer. Ratings between 50 and 100 degrees mean that you feel favorably and warm toward the person; ratings between 0 and 50 degrees mean that you don't feel favorably toward the person and that you don't care too much for that person. You would rate the person at the 50 degree mark if you don't feel particularly warm or cold toward the person."

Thermometer Scores for Jeb! (2016)



A Fake-Data Example

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

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

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

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

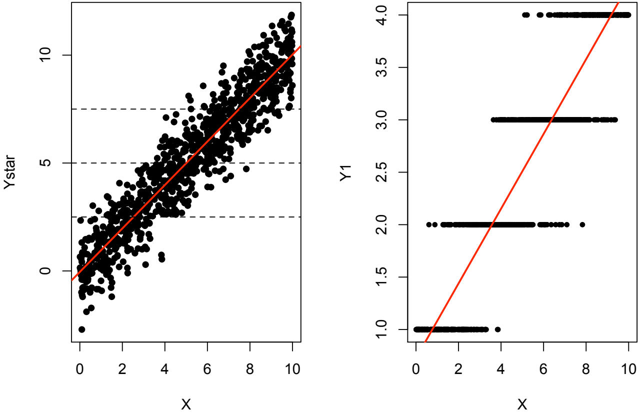
$$\begin{aligned}
Y_{2i} &= 1 \quad \text{if} \quad Y_i^* < 2 \\
&= 2 \quad \text{if} \quad 2 \leq Y_i^* < 8 \\
&= 3 \quad \text{if} \quad 8 \leq Y_i^* < 9 \\
&= 4 \quad \text{if} \quad Y_i^* > 9
\end{aligned}$$

```
> 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
```

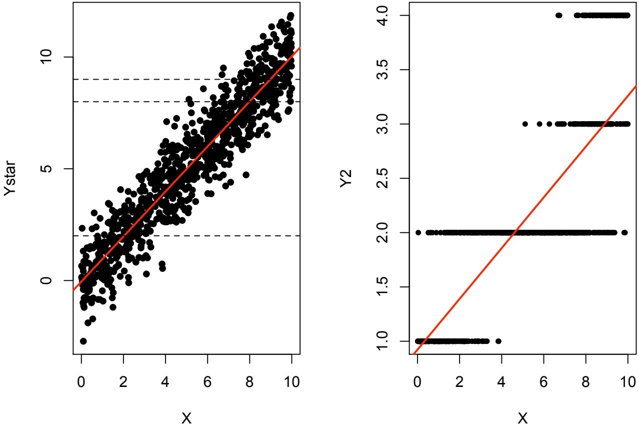
What That Looks Like



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

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

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

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

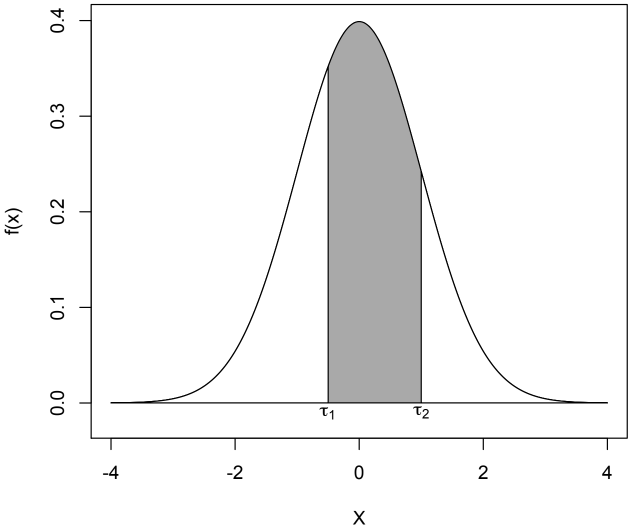
Ordinal Response Models: Probabilities

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

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

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

What That Looks Like



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

Define:

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

Likelihood:

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

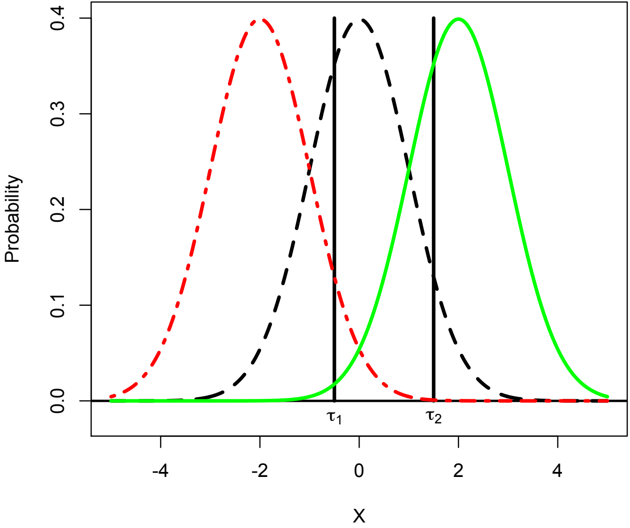
Log-Likelihood, probit:

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

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

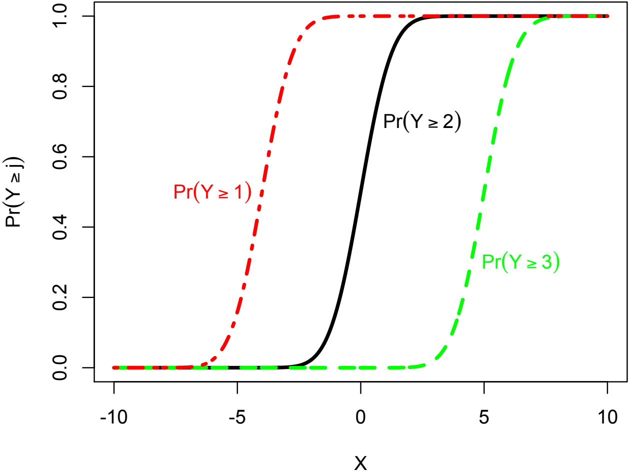


- (Usual) Assumption about $\sigma^2_{Y^*}$
- β_0 vs. the τs...
- Must either omit β_0 or drop one of the $J-1$ τs
- In practice: Stata & R omit β_0

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

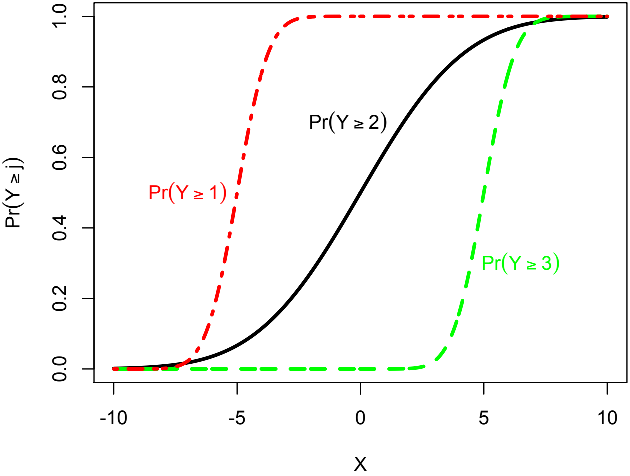
(aka "proportional odds"...)

Parallel Regressions Envisioned



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

Nonparallel Regressions Envisioned



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

1996 Consumer Reports Beer Survey:

```
> 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
```

Ordered Probit

```
> 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
```

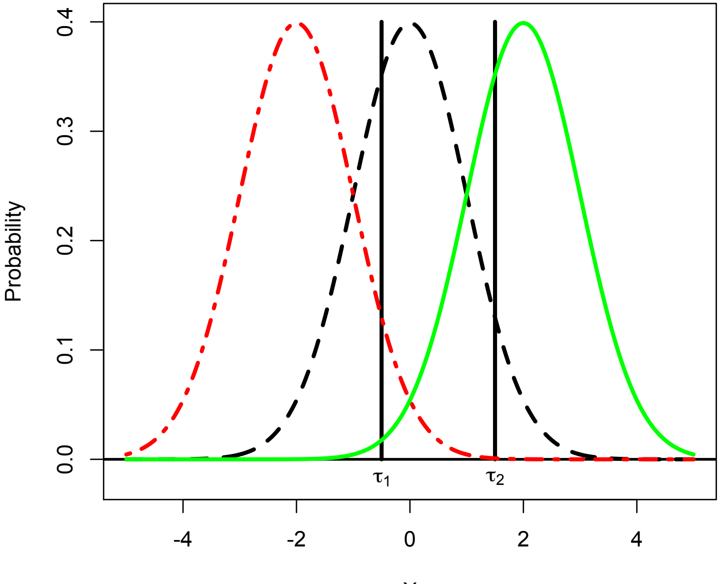
Interpretation: Marginal Effects

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

So:

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

Marginal Effects, Illustrated



For a δ-unit change in X_k:

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

Calculating Odds Ratios

```
> 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.

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

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

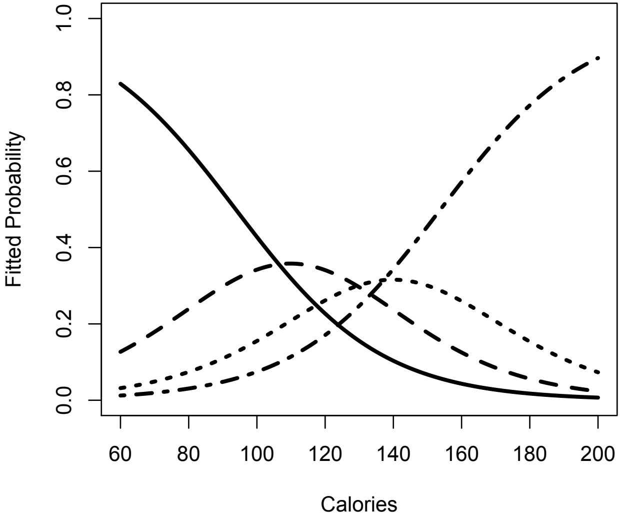
For `craftbeer`=1:

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

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

- Can be category-specific or “cumulative”
- In-sample in `$fitted.values`
- `polr` class supports `predict`, `confint`, etc.

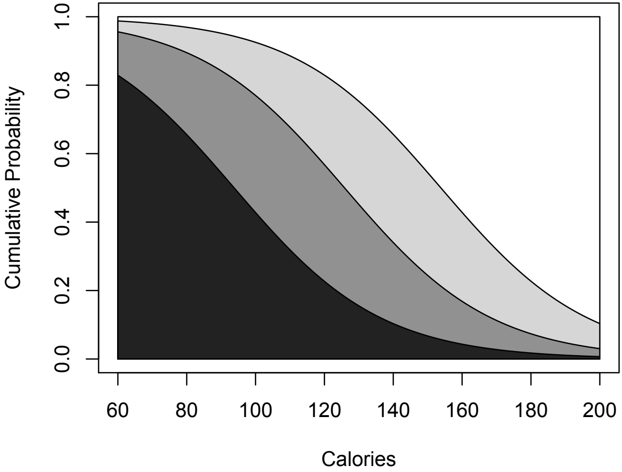
Plot by Outcome



(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 / Extensions (also for PLSC 504...)

- *Generalized* models (relax parallel regressions; Brant (1990))
- *Heteroscedastic* models
- 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.)