

# PLSC 503 – Spring 2024

## Regression Models for Nominal and Ordinal Outcomes

April 8, 2024

## Motivation: Discrete *Outcomes*

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

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

Write:

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

Means that:

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

And set:

$$P_{ij} = \exp(\mathbf{X}_i \beta_j)$$

Rescale:

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

Ensures

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

Constrain  $\beta_1 = \mathbf{0}$ ; then:

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

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

where  $\beta'_j = \beta_j - \beta_1$ .

# Alternative Motivation: Discrete *Choice*

Utility:

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

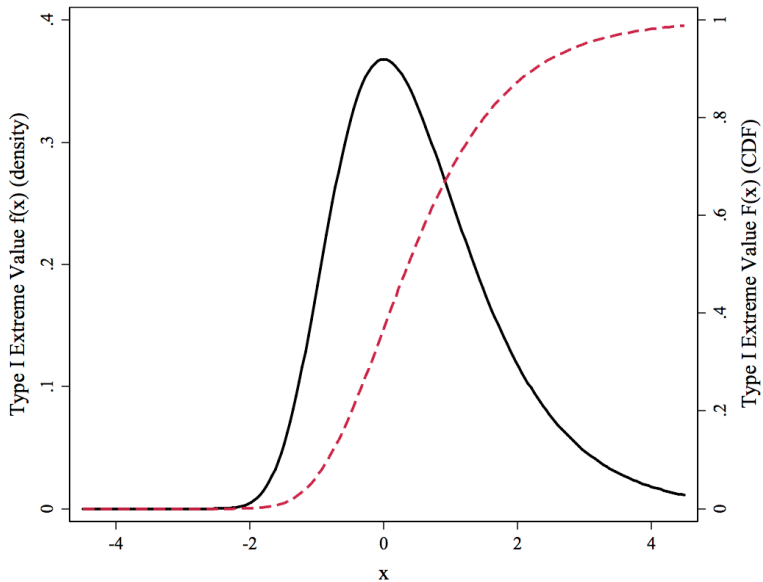
$$\mu_i = \mathbf{X}_i \boldsymbol{\beta}_j$$

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

$\epsilon \sim ???$

- *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:

$$\begin{aligned}
 \Pr(Y_i = j) &= \Pr(U_j > U_1, U_j > U_2, \dots, U_j > U_J) \\
 &= \int f(\epsilon_j) \left[ \int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \beta_j - \mathbf{X}_i \beta_1} f(\epsilon_1) d\epsilon_1 \times \int_{-\infty}^{\epsilon_{ij} + \mathbf{X}_i \beta_j - \mathbf{X}_i \beta_2} f(\epsilon_2) d\epsilon_2 \times \dots \right] d\epsilon_j \\
 &= \int f(\epsilon_j) \times \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \beta_j - \mathbf{X}_i \beta_1)] \times \\
 &\quad \exp[-\exp(\epsilon_{ij} + \mathbf{X}_i \beta_j - \mathbf{X}_i \beta_2)] \times \dots d\epsilon_j \\
 &= \frac{\exp(\mathbf{X}_i \beta_j)}{\sum_{j=1}^J \exp(\mathbf{X}_i \beta_j)}
 \end{aligned}$$



Define:

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

Then:

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

So:

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

and (of course):

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

## Example: The 1992 U.S. Presidential Election



# 1992 American National Election Study

## Data:

- $Y$  (PresVote)  $\in \{\text{Bush}(= 1), \text{Clinton}(= 2), \text{Perot}(= 3)\}$
- $\mathbf{X}$  = political demographic characteristics + “feeling thermometers”

```
> describe(NES92)
```

	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
ID	1	1473	4671.15	1104.02	5113	4681.28	1546.35	3001	6251	3250	-0.11	-1.66	28.77
VotedFor*	2	1473	1.85	0.71	2	1.82	1.48	1	3	2	0.22	-1.03	0.02
PresVote	3	1473	1.85	0.71	2	1.82	1.48	1	3	2	0.22	-1.03	0.02
PartyID	4	1473	3.75	2.11	3	3.69	2.97	1	7	6	0.15	-1.39	0.06
Age	5	1473	45.89	16.67	43	44.85	17.79	18	91	73	0.50	-0.72	0.43
FamIncome	6	1473	15.53	5.76	16	16.10	5.93	1	24	23	-0.78	-0.17	0.15
Female	7	1473	0.51	0.50	1	0.52	0.00	0	1	1	-0.06	-2.00	0.01
White*	8	1473	1.88	0.33	2	1.97	0.00	1	2	1	-2.31	3.36	0.01
FT.Bush	9	1473	51.75	27.26	60	52.99	29.65	0	100	100	-0.30	-0.72	0.71
FT.Clinton	10	1473	55.77	25.08	60	57.35	29.65	0	100	100	-0.45	-0.37	0.65
FT.Perot	11	1473	44.85	26.51	50	44.90	29.65	0	100	100	-0.16	-0.68	0.69

## Model:

$$\text{PresVote}_i = f(\beta_0 + \beta_1 \times \text{PartyID}_i + \beta_2 \times \text{Age}_i + \beta_3 \times \text{White}_i + \beta_4 \times \text{Female}_i)$$

# MNL #1, using vglm (“Baseline” = Perot)

```
> NES92.mlogit<-vglm(PresVote~PartyID+Age+White+Female,multinomial,data=NES92)
> summary(NES92.mlogit)
```

Call:

```
vglm(formula = PresVote ~ PartyID + Age + White + Female, family = multinomial,
      data = NES92)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept):1	-1.98008	0.52454	-3.77	0.00016	***
(Intercept):2	3.82657	0.46402	8.25	< 2e-16	***
PartyID:1	0.50132	0.04870	10.29	< 2e-16	***
PartyID:2	-0.63429	0.04918	-12.90	< 2e-16	***
Age:1	0.01556	0.00504	3.09	0.00203	**
Age:2	0.01296	0.00510	2.54	0.01096	*
WhiteWhite:1	-0.87918	0.43605	-2.02	0.04377	*
WhiteWhite:2	-1.86826	0.38611	-4.84	0.0000013	***
Female:1	0.50928	0.16266	3.13	0.00174	**
Female:2	0.38427	0.16267	2.36	0.01816	*

---

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: 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 3 of the response

## MNL #2, using multinom ("Baseline" = Perot)

```
> NES92$PresVote2<-factor(NES92$PresVote,
+                           levels = c("3", "1", "2"),
+                           labels = c("Perot", "Bush", "Clinton"))

> NES92.mlogit2<-multinom(PresVote2~PartyID+Age+White+Female,data=NES92)
# weights:  18 (10 variable)
initial value 1618.255901
iter  10 value 1080.908630
final value 1053.650588
converged

> summary(NES92.mlogit2)
Call:
multinom(formula = PresVote2 ~ PartyID + Age + White + Female,
  data = NES92)

Coefficients:
      (Intercept) PartyID      Age WhiteWhite Female
Bush           -1.98   0.501 0.0156    -0.879   0.509
Clinton         3.83  -0.634 0.0130    -1.868   0.384

Std. Errors:
      (Intercept) PartyID      Age WhiteWhite Female
Bush           0.525  0.0487 0.00504      0.436   0.163
Clinton        0.464  0.0492 0.00510      0.386   0.163

Residual Deviance: 2107
AIC: 2127
```

## MNL #3, using mlogit

First, we have to “reshape” the data:

```
> head(NES92)
  ID VotedFor PresVote PartyID Age FamIncome Female   White FT.Bush FT.Clinton FT.Perot PresVote2
1 3001     Bush        1      6  31          20      0   White      85         30         0      Bush
2 3002     Bush        1      7  89           9      1   White     100          0         0      Bush
3 3003     Bush        1      7  35          17      1   White      85         30        60      Bush
4 3005 Clinton        2      6  27           3      1 Non-White    40         60        60  Clinton
5 3006 Clinton        2      2  54          15      1   White      30         70        50  Clinton
6 3007 Clinton        2      1  45           2      1 Non-White    15         70        50  Clinton
```

```
> AltNES92<-dfidx(NES92,varying=9:11,shape="wide",choice="VotedFor")
```

```
> head(AltNES92)
```

```
~~~~~
```

```
first 10 observations out of 4419
```

```
~~~~~
```

	ID	VotedFor	PresVote	PartyID	Age	FamIncome	Female	White	PresVote2	FT	idx
1	3001	TRUE	1	6	31	20	0	White	Bush	85	1:Bush
2	3001	FALSE	1	6	31	20	0	White	Bush	30	1:nton
3	3001	FALSE	1	6	31	20	0	White	Bush	0	1:erot
4	3002	TRUE	1	7	89	9	1	White	Bush	100	2:Bush
5	3002	FALSE	1	7	89	9	1	White	Bush	0	2:nton
6	3002	FALSE	1	7	89	9	1	White	Bush	0	2:erot
7	3003	TRUE	1	7	35	17	1	White	Bush	85	3:Bush
8	3003	FALSE	1	7	35	17	1	White	Bush	30	3:nton
9	3003	FALSE	1	7	35	17	1	White	Bush	60	3:erot

# MNL #3, using mlogit (continued)

Now, fit the model:

```
> NES92.mlogit3<-mlogit(VotedFor~0|PartyID+Age+White+Female,data=AltNES92,reflevel="Perot")
> summary(NES92.mlogit3)
```

```
Call:
mlogit(formula = VotedFor ~ 0 | PartyID + Age + White + Female,
      data = AltNES92, reflevel = "Perot", method = "nr")
```

Frequencies of alternatives:choice

	Perot	Bush	Clinton
	0.191	0.339	0.469

```
nr method
5 iterations, 0h:0m:0s
g'(-H)^-1g = 4.94E-08
gradient close to zero
```

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z )
(Intercept):Bush	-1.98008	0.52454	-3.77	0.00016 ***
(Intercept):Clinton	3.82657	0.46403	8.25	2.2e-16 ***
PartyID:Bush	0.50132	0.04870	10.29	< 2e-16 ***
PartyID:Clinton	-0.63429	0.04918	-12.90	< 2e-16 ***
Age:Bush	0.01556	0.00504	3.09	0.00203 **
Age:Clinton	0.01296	0.00510	2.54	0.01096 *
WhiteWhite:Bush	-0.87918	0.43606	-2.02	0.04378 *
WhiteWhite:Clinton	-1.86826	0.38612	-4.84	1.3e-06 ***
Female:Bush	0.50928	0.16266	3.13	0.00174 **
Female:Clinton	0.38427	0.16267	2.36	0.01816 *

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -1050

McFadden R<sup>2</sup>: 0.311

Likelihood ratio test : chisq = 952 (p.value = <2e-16)



# MNL: 1992 Election (“Baseline” = Bush)

```
> Bush.nes92.mlogit<-vglm(PresVote~PartyID+Age+White+Female,  
+                           data=NES92,family=multinomial(refLevel=1))  
> summary(Bush.nes92.mlogit)
```

Call:

```
vglm(formula = PresVote ~ PartyID + Age + White + Female, family = multinomial(refLevel = 1),  
     data = NES92)
```

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

# MNL: 1992 Election (“Baseline” = Clinton)

```
> Clinton.nes92.mlogit<-vglm(PresVote~PartyID+Age+White+Female,  
+                             data=NES92,family=multinomial(refLevel=2))  
> summary(Clinton.nes92.mlogit)
```

Call:

```
vglm(formula = PresVote ~ PartyID + Age + White + Female, family = multinomial(refLevel = 2),  
     data = NES92)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept):1	-5.80665	0.44301	-13.11	< 2e-16 ***
(Intercept):2	-3.82657	0.46402	-8.25	< 2e-16 ***
PartyID:1	1.13561	0.05486	20.70	< 2e-16 ***
PartyID:2	0.63429	0.04918	12.90	< 2e-16 ***
Age:1	0.00260	0.00514	0.51	0.6128
Age:2	-0.01296	0.00510	-2.54	0.0110 *
WhiteWhite:1	0.98908	0.31346	3.16	0.0016 **
WhiteWhite:2	1.86826	0.38611	4.84	0.0000013 ***
Female:1	0.12500	0.16895	0.74	0.4594
Female:2	-0.38427	0.16267	-2.36	0.0182 *

---

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: 2107 on 2936 degrees of freedom

Log-likelihood: -1054 on 2936 degrees of freedom

Number of Fisher scoring iterations: 5

Warning: Hauck-Donner effect detected in the following estimate(s):

'(Intercept):2'

Reference group is level 2 of the response

## PartyID Coefficient Estimates and “Baselines”

Note: PartyID is 1 (strong Democrat)  $\rightarrow$  7 (strong Republican)

		<u>“Baseline” category</u>		
		Clinton	Perot	Bush
Comparison	Clinton	–	-0.63	-1.14
Category	Perot	0.63	–	-0.50
	Bush	1.14	0.50	–

Consider the choice of Bush vs. Perot:

```
> NES92$PickBush<-NA
> NES92$PickBush<-ifelse(NES92$VotedFor=="Bush",1,NES92$PickBush)
> NES92$PickBush<-ifelse(NES92$VotedFor=="Perot",0,NES92$PickBush)
> BushBinary<-glm(PickBush~PartyID+Age+White+Female,data=NES92,family="binomial")
> summary(BushBinary)
```

Call:

```
glm(formula = PickBush ~ PartyID + Age + White + Female, family = "binomial",
    data = NES92)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )	
(Intercept)	-1.9024	0.5372	-3.54	0.00040	***
PartyID	0.5106	0.0505	10.12	< 2e-16	***
Age	0.0143	0.0052	2.75	0.00595	**
WhiteWhite	-0.9817	0.4586	-2.14	0.03230	*
Female	0.5768	0.1683	3.43	0.00061	***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1022.50 on 781 degrees of freedom  
Residual deviance: 880.28 on 777 degrees of freedom  
(691 observations deleted due to missingness)  
AIC: 890.3

Number of Fisher Scoring iterations: 4

# MNL and Binary Logit (continued)

What about Clinton vs. Perot?:

```
> NES92$PickClinton<-NA
> NES92$PickClinton<-ifelse(NES92$VotedFor=="Clinton",1,NES92$PickClinton)
> NES92$PickClinton<-ifelse(NES92$VotedFor=="Perot",0,NES92$PickClinton)
> ClintonBinary<-glm(PickClinton~PartyID+Age+White+Female,data=NES92,family="binomial")
> summary(ClintonBinary)
```

Call:

```
glm(formula = PickClinton ~ PartyID + Age + White + Female, family = "binomial",
    data = NES92)
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z )
(Intercept)	3.92614	0.48490	8.10	5.6e-16 ***
PartyID	-0.68125	0.05301	-12.85	< 2e-16 ***
Age	0.01381	0.00537	2.57	0.0101 *
WhiteWhite	-1.91056	0.39879	-4.79	1.7e-06 ***
Female	0.51690	0.17024	3.04	0.0024 **

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

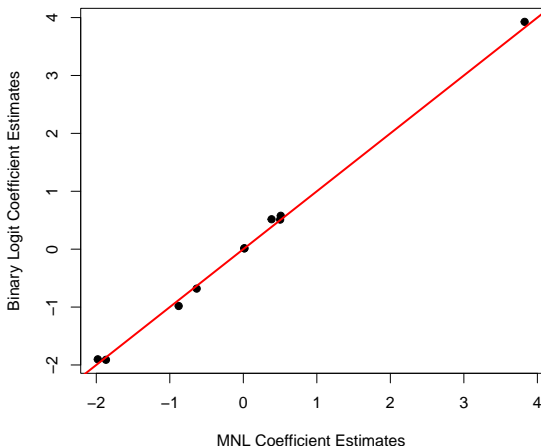
(Dispersion parameter for binomial family taken to be 1)

Null deviance: 1171.48 on 972 degrees of freedom  
Residual deviance: 861.57 on 968 degrees of freedom  
(500 observations deleted due to missingness)  
AIC: 871.6

Number of Fisher Scoring iterations: 5

## MNL and Binary Logit (continued)

Are the  $\hat{\beta}$ s the same? (A: Yes, basically...)



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

# Choice-Specific Covariates: Data Structure

```
> head(AltNES92)
```

	ID	VotedFor	PresVote	PartyID	Age	FamIncome	Female	White	PresVote2	FT	idx\$id1	\$id2
*	<dbl>	<lgl>	<dbl>	<dbl>	<dbl>	<dbl>	<dbl>	<chr>	<fct>	<dbl>	<int>	<fct>
1	3001	TRUE	1	6	31	20	0	White	Bush	85	1	Bush
2	3001	FALSE	1	6	31	20	0	White	Bush	30	1	Clinton
3	3001	FALSE	1	6	31	20	0	White	Bush	0	1	Perot
4	3002	TRUE	1	7	89	9	1	White	Bush	100	2	Bush
5	3002	FALSE	1	7	89	9	1	White	Bush	0	2	Clinton
6	3002	FALSE	1	7	89	9	1	White	Bush	0	2	Perot
7	3003	TRUE	1	7	35	17	1	White	Bush	85	3	Bush
8	3003	FALSE	1	7	35	17	1	White	Bush	30	3	Clinton
9	3003	FALSE	1	7	35	17	1	White	Bush	60	3	Perot
10	3005	FALSE	2	6	27	3	1	Non-White	Clinton	40	4	Bush

# 4,409 more rows  
# Use 'print(n = ...)' to see more rows



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 (Feeling Thermometers only)

```
> NES92.clogit<-mlogit(VotedFor~FT,data=AltNES92,reflevel="Perot")
> summary(NES92.clogit)
```

Call:

```
mlogit(formula = VotedFor ~ FT, data = AltNES92, reflevel = "Perot",
       method = "nr")
```

Frequencies of alternatives:choice

	Perot	Bush	Clinton
	0.191	0.339	0.469

nr method

6 iterations, 0h:0m:0s

g'(-H)<sup>-1</sup>g = 0.00219

successive function values within tolerance limits

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z )
(Intercept):Bush	0.03307	0.10039	0.33	0.74
(Intercept):Clinton	0.45841	0.09253	4.95	0.00000073 ***
FT	0.07512	0.00314	23.89	< 2e-16 ***

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -801

McFadden R<sup>2</sup>: 0.476

Likelihood ratio test : chisq = 1460 (p.value = <2e-16)

# CL in R ("Full" Model)

```
> NES92.clogit2<-mlogit(VotedFor~FT|PartyID+Age+White+Female,data=AltNES92,reflevel="Perot")
> summary(NES92.clogit2)
```

Frequencies of alternatives:choice

Perot	Bush	Clinton
0.191	0.339	0.469

nr method

6 iterations, 0h:0m:0s

g'(-H)^-1g = 3.06E-08

gradient close to zero

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z )
(Intercept):Bush	-0.39416	0.58730	-0.67	0.50214
(Intercept):Clinton	2.69235	0.50403	5.34	9.2e-08 ***
FT	0.06231	0.00325	19.17	< 2e-16 ***
PartyID:Bush	0.22298	0.05842	3.82	0.00014 ***
PartyID:Clinton	-0.40620	0.05798	-7.01	2.4e-12 ***
Age:Bush	0.00868	0.00612	1.42	0.15639
Age:Clinton	0.00839	0.00598	1.40	0.16040
WhiteWhite:Bush	-1.31961	0.47725	-2.77	0.00569 **
WhiteWhite:Clinton	-1.57156	0.41404	-3.80	0.00015 ***
Female:Bush	0.39271	0.19995	1.96	0.04953 *
Female:Clinton	0.28585	0.19474	1.47	0.14213

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

Log-Likelihood: -723

McFadden R^2: 0.527

Likelihood ratio test : chisq = 1610 (p.value = <2e-16)

# Interpretation: Baseline MNL Results

```
> NES.MNL<-vglm(PresVote~PartyID+Age+White+Female,data=NES92,  
+ multinomial(refLevel=1)) # Bush is comparison category  
> summaryvglm(NES.MNL)
```

Call:

```
vglm(formula = PresVote ~ PartyID + Age + White + Female, family = multinomial(refLevel = 1),  
data = NES92)
```

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 In  $LR$  statistic  $Q$  tests:

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

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

# Test H: No Effect of Age

Is the effect of Age across the three candidates equal to zero?

```
> 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

Are the estimated coefficients for Clinton (vs. Bush) jointly equal to zero?

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

Wald test:

-----

Chi-squared test:

$X^2 = 444.6$ ,  $df = 5$ ,  $P(> X^2) = 0.0$

## Interpretation: Marginal Effects

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

Depends on:

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

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



# Marginal Effects: Illustrated

```
> MNL.alt<-multinom(PresVote2~PartyID+Age+White+Female,data=NES92,Hess=TRUE)
# weights:  18 (10 variable)
initial  value 1618.255901
iter    10 value 1080.908630
final    value 1053.650588
converged
```

```
> summary(marginal_effects(MNL.alt))
```

dydx_PartyID	dydx_Age	dydx_Female	dydx_WhiteWhite
Min.      :-0.0908	Min.      :-0.00362	Min.      :-0.1158	Min.      :0.0416
1st Qu.: -0.0439	1st Qu.: -0.00282	1st Qu.: -0.0892	1st Qu.: 0.0943
Median   : 0.0185	Median   :-0.00215	Median   :-0.0674	Median   :0.1352
Mean     : 0.0083	Mean     :-0.00207	Mean     :-0.0648	Mean     :0.1435
3rd Qu.: 0.0618	3rd Qu.: -0.00138	3rd Qu.: -0.0420	3rd Qu.: 0.1848
Max.     : 0.1070	Max.     :-0.00011	Max.     :-0.0034	Max.     :0.2926

# Odds ( “Relative Risk” ) Ratios

MNL has:

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

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

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

One-Unit Change in  $X_k$ :

$$RRR_{jk} = \exp(\beta_{jk})$$

$\delta$ -Unit Change in  $X_k$ :

$$RRR_{jk} = \exp(\beta_{jk} \times \delta)$$

# Odds (“Relative Risk”) Ratios

```
> mnl.or <- function(model) {  
  coeffs <- c(t(coef(NES.MNL)))  
  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	139.540	332.504	792.309
(Intercept):2	2.591	7.243	20.250
PartyID:1	0.288	0.321	0.358
PartyID:2	0.551	0.606	0.666
Age:1	0.987	0.997	1.008
Age:2	0.975	0.985	0.994
WhiteWhite:1	0.201	0.372	0.688
WhiteWhite:2	1.025	2.409	5.662
Female:1	0.634	0.882	1.229
Female:2	0.437	0.601	0.827

# Odds Ratios: Interpretation

## Odds ratio interpretations:

- 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

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

# In-Sample Predicted Outcomes

Generate predicted vote choices:

```
> NES92$Predictions<-" "  
> NES92$Predictions<-ifelse(fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,2]  
+ & fitted.values(NES.MNL)[,1]>fitted.values(NES.MNL)[,3],  
+ paste("Bush"),NES92$Predictions) # Bush  
> NES92$Predictions<-ifelse(fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,1]  
+ & fitted.values(NES.MNL)[,2]>fitted.values(NES.MNL)[,3],  
+ paste("Clinton"),NES92$Predictions) # Clinton  
> NES92$Predictions<-ifelse(fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,1]  
+ & fitted.values(NES.MNL)[,3]>fitted.values(NES.MNL)[,2],  
+ paste("Perot"),NES92$Predictions) # Perot  
  
> # "Confusion Table":  
>  
> table(NES92$VotedFor,NES92$Predictions)
```

	Bush	Clinton	Perot
Bush	415	77	8
Clinton	56	619	16
Perot	135	133	14

Model fit:

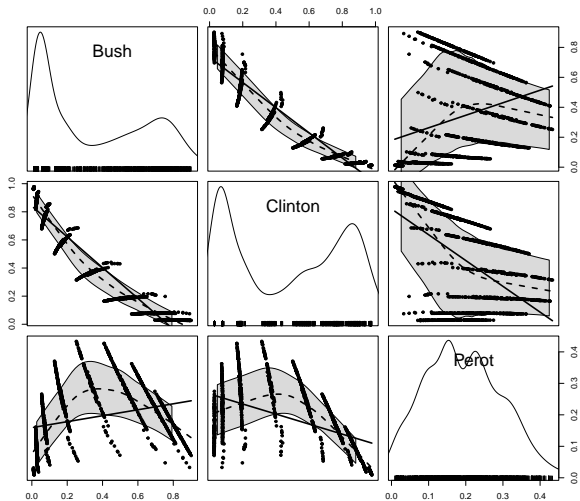
- “Null” Model:  $\left(\frac{691}{1473}\right) = 46.9\%$  correct.
- Estimated model:  $\frac{(415+619+14)}{1473} = \frac{1048}{1473} = 71.2\%$  correct.
- $PRE = \frac{1048-691}{1473-691} = \frac{357}{782} = 45.7\%$ .
- Correct predictions: 90% Clinton, 83% Bush, 5% Perot.

# In-Sample Predicted Probabilities

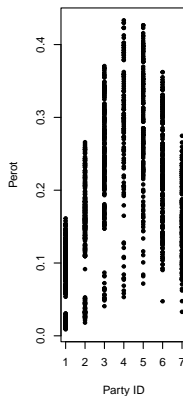
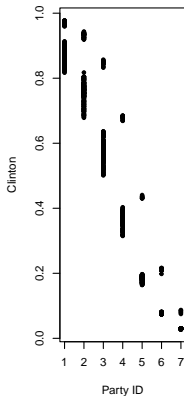
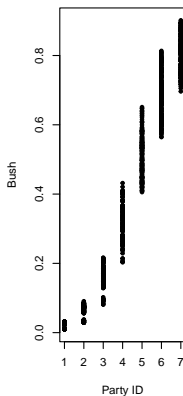
```
> hats<-as.data.frame(fitted.values(NES.MNL))
> names(hats)[3]<-"Perot" # nice names...
> names(hats)[2]<-"Clinton"
> names(hats)[1]<-"Bush"
> attach(hats)

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

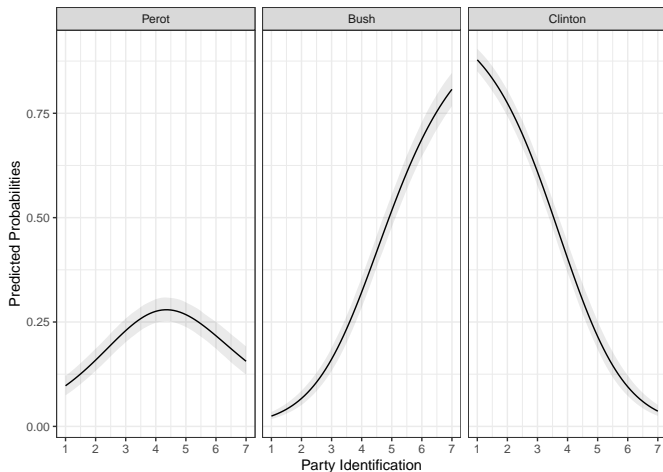




# In-Sample $\hat{P}$ rs vs. partyid

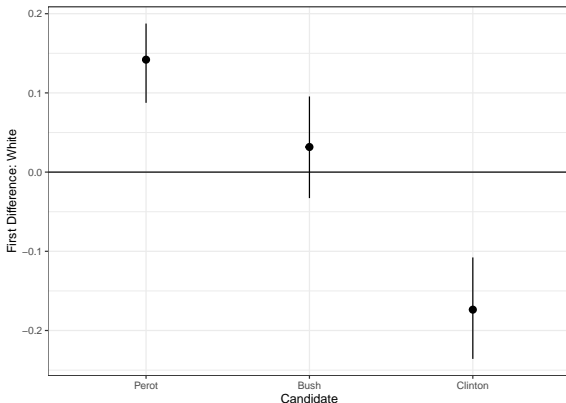


# Out-Of-Sample Predictions (using MNLpred)



# OOS First Differences (using MNLpred)

First differences in probabilities associated with White:



# Conditional Logit: Example

```
> summary(NES92.clogit2)
```

Call:

```
mlogit(formula = VotedFor ~ FT | PartyID + Age + White + Female,  
data = AltNES92, reflevel = "Perot", method = "nr")
```

Frequencies of alternatives:choice

```
Perot    Bush Clinton  
0.191    0.339    0.469
```

nr method

6 iterations, 0h:0m:0s

$g'(-H)^{-1}g = 3.06E-08$

gradient close to zero

Coefficients :

	Estimate	Std. Error	z-value	Pr(> z )
(Intercept):Bush	-0.39416	0.58730	-0.67	0.50214
(Intercept):Clinton	2.69235	0.50403	5.34	9.2e-08 ***
FT	0.06231	0.00325	19.17	< 2e-16 ***
PartyID:Bush	0.22298	0.05842	3.82	0.00014 ***
PartyID:Clinton	-0.40620	0.05798	-7.01	2.4e-12 ***
Age:Bush	0.00868	0.00612	1.42	0.15639
Age:Clinton	0.00839	0.00598	1.40	0.16040
WhiteWhite:Bush	-1.31961	0.47725	-2.77	0.00569 **
WhiteWhite:Clinton	-1.57156	0.41404	-3.80	0.00015 ***
Female:Bush	0.39271	0.19995	1.96	0.04953 *
Female:Clinton	0.28585	0.19474	1.47	0.14213

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

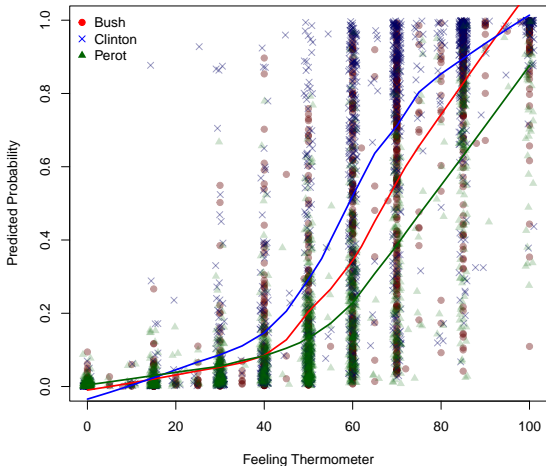
Log-Likelihood: -723

McFadden R<sup>2</sup>: 0.527

Likelihood ratio test : chisq = 1610 (p.value = <2e-16)

# Conditional Logit: In-Sample Predicted Probabilities

```
> CLhats<-predict(NES92.clogit2,AltNES92)
```



- “Independence of Irrelevant Alternatives”
- → Multinomial Probit
- → Heteroscedastic Extreme Value model
- “Mixed” Logit
- Nested Logit