MLE: Categorical and Limited Dependent Variables

Unit 2: Ordered, Multinomial, Count, and Limited Dependent Variables

2a. Multinomial Logit

PS2730-2021

Weeks 5-6

Professor Steven Finkel

Modeling Nominal Variables

- Nominal outcomes have multiple categories that cannot be ranked
  - Vote for Democratic, Republican, Independent candidate
  - Career choice post-PhD: academia, government, private sector, etc.
  - Regime preference: democracy, autocracy, theocracy, military dictatorship
- Here you can't treat categories as higher/lower and hence separated by thresholds which need to be crossed to get to the "next" category. Need to model the separate probabilities of obtaining outcomes of each of the unrelated categories
- There is no conceivable "continuity" for the variable; OLS will be absolutely inappropriate (unlike the ordered case where it can approximate the "true" effects)
- So we estimate instead with models derived for nominal outcomes, most basic of which is the Multinomial Logit Model (MNL)
- These models *may* also be appropriate for ordinal outcomes where the proportional odds/parallel regression assumption is violated

Multinomial Logit

- Assume we have 3 outcomes – e.g., well-known 1992 election in US with incumbent George H.W. Bush (Republican) versus Bill Clinton (Democratic) versus Ross Perot (Independent). Primary IVs: age, male, partisanship, trust in government
- Given three outcomes, we'll need to estimate the effect of the IVs on voting for Bush over Clinton; the effect of the IVs on Clinton over Perot; and the effect of the IVs on Bush over Perot.
- Since we have unordered categories, no reason to expect the effects to be the same, so we need potentially 3 sets of estimates
- Possible procedure:

  estimate 3 separate binary logits

$$\ln\left[\frac{\Pr(B \mid x)}{\Pr(C \mid x)}\right] = XB_{B|C}$$

$$\ln\left[\frac{\Pr(C \mid x)}{\Pr(P \mid x)}\right] = XB_{C|P}$$

$$\ln\left[\frac{\Pr(B \mid x)}{\Pr(P \mid x)}\right] = XB_{B|P}$$

But: three equations are redundant

Since ln(a/b)=ln(a)-ln(b), then it must be the case that the sum of the first two equal the third

$$\ln\left[\frac{\Pr(B\mid x)}{\Pr(C\mid x)}\right]=\ln\Pr(B\mid x)-\ln\Pr(C\mid x)$$

$$\ln\left[\frac{\Pr(C\mid x)}{\Pr(P\mid x)}\right]=\ln\Pr(C\mid x)-\ln\Pr(P\mid x)$$

$$\ln\left[\frac{\Pr(B\mid x)}{\Pr(C\mid x)}\right]+\ln\left[\frac{\Pr(C\mid x)}{\Pr(P\mid x)}\right]=\ln\Pr(B\mid x)-\ln\Pr(P\mid x)$$

$$\ln\left[\frac{\Pr(B\mid x)}{\Pr(C\mid x)}\right]+\ln\left[\frac{\Pr(C\mid x)}{\Pr(P\mid x)}\right]=\ln\left[\frac{\Pr(B\mid x)}{\Pr(P\mid x)}\right]$$

- This means that intercept and slope you get from first two binary logit estimates can be manipulated to get the third, and that any two sets of coefficients you get can be manipulated to the get the last
- This is not fatal but problematic, because when running binary logits you are basing the coefficients on different numbers of cases (i.e., Bush and Clinton voters in equation 1, Clinton and Perot voters in equation 2, etc.). Therefore the equalities that are implicit in the model will not necessarily hold.
- More problematic, the resultant probability calculations in separate logits will not necessarily add to 1. Will get relative Ps, Bush vs. Clinton, Clinton vs. Perot, but estimating the final Bush over Perot will not **necessarily** give you what 1 minus the two others should give
- What we need is a model that incorporates these constraints while giving the same extensive information that series of binary logits would provide. This leads to the MNL model.

- So:
  - We want the probability for outcome *m* to be a non-linear function of the XB, where we have a different β for each set of outcomes (i.e., the β for predicting the probability of Bush based on age will be different than the β for predicting the probability of Clinton or Perot based on age)
  - We need the probabilities to be non-negative, so we exponentiate the β to get P(y=m|x)= exp^{XB_m}
  - We need the probabilities to sum to 1, so we divide the P for each outcome by the sum of all the outcome probabilities

$$P(y=m\,|\,x)=\frac{\exp^{XB_m}}{\sum_{1}^{J}\exp^{XB_j}}$$

- Problem: The model as written is not identified. Suppose you took the equation for the probability and multiplied it by some constant value $\frac{exp^{X\tau}}{exp^{X\tau}}$. Then it can be shown that the probability can be re-written as $P(y = m|x) = \frac{\exp(X(\beta_m+\tau])}{\sum_{j=1}^{J}\exp(X((\beta_m+\tau])}$
- So there is a different value of the regression coefficient linking X to P(y=m) for any nonzero τ you might pick. Hence *underidentified* model without further constraint.
- Solution: **Pick one category and assume that βs for that category are 0.** This means that one category will be the baseline category against which everything else is estimated.
- This choice is purely **arbitrary** - the estimated Ps for each category will be the same regardless of which category we pick as the baseline
- Let's pick Bush as the baseline, so that β_B, for example, all equal 0 (note: we can't add τ to β_B since that would violate the assumption)

- So the Ps equal

$$P(y=C\,|\,x)=\frac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+\exp^{XB_B}}=\frac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+1}$$

$$P(y=P\,|\,x)=\frac{\exp^{XB_P}}{\exp^{XB_C}+\exp^{XB_P}+\exp^{XB_B}}=\frac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+1}$$

$$P(y=B\,|\,x)=\frac{\exp^{XB_B}}{\exp^{XB_C}+\exp^{XB_P}+\exp^{XB_B}}=\frac{1}{\exp^{XB_C}+\exp^{XB_P}+1}$$

or generically:

$$P(y=m|x,\text{not baseline category } b)=\frac{exp^{XB_m}}{\sum exp^{XB_{j\neq b}}+1}$$

$$P(y=m|x,\text{baseline category } b)=\frac{1}{\sum exp^{XB_{j\neq b}}+1}$$

MNL as an Odds Model

- Can transform the probability model of MNL into familiar odds and log-odds framework as well

$$Odds\text{ Clinton/Bush} = \frac{P(y=C|x)}{P(y=B|x)} = \frac{\dfrac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+1}}{\dfrac{1}{\exp^{XB_C}+\exp^{XB_P}+1}} = \exp^{XB_C}$$

$$Odds\text{ Perot/Bush} = \frac{P(y=P|x)}{P(y=B|x)} = \frac{\dfrac{\exp^{XB_P}}{\exp^{XB_C}+\exp^{XB_P}+1}}{\dfrac{1}{\exp^{XB_C}+\exp^{XB_P}+1}} = \exp^{XB_P}$$

$$Odds\text{ Clinton/Perot} = \frac{P(y=C|x)}{P(y=P|x)} = \frac{\dfrac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+1}}{\dfrac{\exp^{XB_P}}{\exp^{XB_C}+\exp^{XB_P}+1}} = \frac{\exp^{XB_C}}{\exp^{XB_P}}$$

$$LnOdds \text{ Clinton/Bush} = \ln\left(\frac{P(y=C|x)}{P(y=B|x)}\right) = XB_C$$

$$LnOdds \text{ Perot/Bush} = \ln\left(\frac{P(y=P|x)}{P(y=B|x)}\right) = XB_P$$

$$LnOdds \text{ Clinton/Perot} = \ln\left(\frac{P(y=C|x)}{P(y=P|x)}\right) = X(B_C - B_P)$$

- So the MNL model is **linear in the log-odds** of outcome *m* versus outcome *n*; when outcome *n* equals the baseline category, this reduces to the "logit" of XB from the outcome's β.

- When outcome n is not equal to the baseline category, the logit is X time the difference in the two β.
- Generically: $$LnOdds \text{ m|n} = X(B_m - B_n)$$

- Where one category *n* is the baseline category (B_b=0), in which case

$$LnOdds \text{ m|n}_{n=b} = XB_m$$

- Side note #1: This is also the **Generalized Linear Model** form of the MNL model: it is linear in the log-odds of category m over category n

$$LnOdds\ \mathrm{m|n} = \eta = X(B_m - B_n)$$

with one category *n* set as baseline with B=0 for identification purposes

- Side note #2: see on slide 9 how the odds for any outcome relative to any other have absolutely nothing to do with a fourth or fifth alternative that would come into play? It would all cancel out in the denominator. So inherent in MNL is the idea that the relative odds of alternative A versus alternative B has nothing to do with whether alternatives C, D, or E are also available. This is the Independence of Irrelevant Alternatives (IIA) assumption of MNL which we'll discuss further. This assumption might not hold and need to estimate an alternative model!

Maximum Likelihood Estimation of the MNL Model

- Steps:

- Assume a probability distribution for Y – e.g., categorical (multinomial) in this case
- Express the joint probability of the data (i.e., all of the Y) using the assumed probability distribution
- Calculate the joint probability of the data given the parameters–the "likelihood function" (taking the log of the likelihood to simplify)
- Maximize this function with respect to the unknown parameters (e.g., the Bs in the multinomial logit function)

- Define a set of dummy variables d_{ij} where

$$d_{ij} = \begin{cases} 1 \text{ if } y_{ij} = j \\ 0 \text{ otherwise} \end{cases}$$

- In other words, d is 1 for the observed outcome category *j* for each unit
- Then the probability of observing a given outcome for each individual is

$$P_i = P_{i1}^{d_{i1}} P_{i2}^{d_{i2}} P_{i2}^{d_{i2}} \ldots P_{iJ}^{d_{iJ}} = \prod P_{ij}^{d_{ij}}$$

- For each case that falls in category 1, use P(Y=1) as its probability

  For each case that falls in category 2, take P(Y=2) as its probability

  For each case that falls in category 3, take P(Y=3) as its probability etc.

- For entire sample:

$$LnL = \sum_{i=1}^{N} \sum_{j=1}^{J} d_{ij} \ln P_{ij} \quad \text{with}$$

$$P(y = m|x, \text{not baseline category } b) = \frac{exp^{XB_m}}{\sum exp^{XB_{j\neq b}} + 1}$$

$$P(y = m|x, \text{baseline category } b) = \frac{1}{\sum exp^{XB_{j\neq b}} + 1}$$

So:

$$LnL = \sum_{i=1}^{N} \sum_{j=1}^{J\neq b} d_{ij\neq b} \ln \frac{\exp^{XB_j}}{\sum \exp^{XB_{j\neq b}} + 1} + \sum_{i=1}^{N} d_{ij=b} \ln \frac{1}{\sum \exp^{XB_{j\neq b}} + 1}$$

and then find Bs that maximize this function

```
. mlogit presvote age male, b(1)

Iteration 0:   log likelihood = -1706.6244
Iteration 1:   log likelihood = -1683.8657
Iteration 2:   log likelihood = -1683.5033
Iteration 3:   log likelihood = -1683.5031
Iteration 4:   log likelihood = -1683.5031

Multinomial logistic regression                 Number of obs     =      1,658
                                                LR chi2(4)        =      46.24
                                                Prob > chi2       =     0.0000
Log likelihood = -1683.5031                     Pseudo R2         =     0.0135

------------------------------------------------------------------------------
    presvote |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
bush         |  (base outcome)
-------------+----------------------------------------------------------------
clinton      |
         age |  -.0003201   .0031936    -0.10   0.920    -.0065795    .0059392
        male |  -.2562606   .1111937    -2.30   0.021    -.4741963   -.0383249
       _cons |   .4710838   .1734719     2.72   0.007     .1310851    .8110826
-------------+----------------------------------------------------------------
perot        |
         age |   -.018297   .0044395    -4.12   0.000    -.0269983   -.0095957
        male |   .3797324   .1451582     2.62   0.009     .0952275    .6642372
       _cons |  -.0060474   .2274405    -0.03   0.979    -.4518226    .4397277
------------------------------------------------------------------------------
```

Bush (category 1 in the actual DV) is set as the baseline category. So all coefficients are interpreted as the effects of the variable on outcome Clinton or outcome Perot, relative to Bush

- As age increases by 1 unit, the log-odds of voting for Clinton over Bush decreases by .00032 (not statistically significant)
- As age increases by 1 unit, the log-odds of voting for Perot over Bush decreases by .010 (statistically significant)
- Men have logits for Clinton over Bush that are .25 smaller than women; .38 larger for Perot over Bush, both difference statistically significant
- The intercepts represent the base log-odds of Clinton over Bush, Perot over Bush (i.e. when all independent variables are 0)

presvote	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
bush						
age	.0003201	.0031936	0.10	0.920	-.0059392	.0065795
male	.2562606	.1111937	2.30	0.021	.0383249	.4741963
_cons	-.4710838	.1734719	-2.72	0.007	-.8110826	-.1310851
clinton	(base outcome)					
perot						
age	-.0179769	.0042516	-4.23	0.000	-.0263099	-.0096439
male	.635993	.1382718	4.60	0.000	.3649852	.9070008
_cons	-.4771313	.2153059	-2.22	0.027	-.8991231	-.0551394

- With Clinton as baseline, you can see that the coefficients for Bush are the **negative** of the previous slides, and the coefficients for Perot are: (Perot versus Bush from previous slide *minus* Clinton versus Bush from previous slide)
- This corresponds to the equality we derived on slides 3-4!
- "listcoef" gives you all contrasts and their associated significance so you don't need to run all the different models with different baseline categories!

Variable: age (sd=17.189)

		b	z	P>|z|	e^b	e^bStdX
bush	vs clinton	0.0003	0.100	0.920	1.000	1.006
bush	vs perot	0.0183	4.121	0.000	1.018	1.370
clinton	vs bush	-0.0003	-0.100	0.920	1.000	0.995
clinton	vs perot	0.0180	4.228	0.000	1.018	1.362
perot	vs bush	-0.0183	-4.121	0.000	0.982	0.730
perot	vs clinton	-0.0180	-4.228	0.000	0.982	0.734

Substantive interpretation: older voters significantly more likely to vote for both Clinton and Bush than Perot, but no difference in age for logit Clinton v. Bush

Two Statistical Tests for MNL

1. Is the effect of a given independent variable 0 across all categories? If so, can ignore it altogether

- Test with Likelihood Ratio (LR) or Wald test with reduced model without X (or where all B_j are constrained to be 0) and a full model with X (or with unconstrained estimation of the B_j).
- Difference in 2*LL follows chi-square distribution with J-1 df

```
. mlogtest, lr

LR tests for independent variables (N=1658)

  Ho: All coefficients associated with given variable(s) are 0

                 |      chi2     df    P>chi2
-----------------+---------------------------
             age |    21.240      2     0.000
            male |    22.058      2     0.000
```

- We reject H_0 for both variables in this case

2. Are categories *m* and *n* statistically indistinguishable? That is, the effects of *all* independent variables may be statistically insignificant in the contrast of *m* and *n*. This would mean that we can obtain more efficient estimates with fewer outcome categories on the dependent variable (though it may not make substantive sense to do this). Df here is the number of IVs!

- H_0: $\beta_{1,m|n} = \beta_{2,m|n} \ldots = \beta_{k,m|n} = 0$

```
. mlogtest, combine

Wald tests for combining alternatives (N=1658)

  Ho: All coefficients except intercepts associated with a given pair
      of alternatives are 0 (i.e., alternatives can be combined)
```

	chi2	df	P>chi2
bush & clinton	**5.318**	**2**	**0.070**
bush & perot	**25.093**	**2**	**0.000**
clinton & perot	**41.010**	**2**	**0.000**

So we can statistically combine Bush and Clinton into one category, but we obviously wouldn't do this!

Interpretation of Effects

- Multiple ways of making sense of the effects in MNL
- Interpretations of the logits
  - Following derivation on slide 10, can say that a unit change in X changes the *logit* of observing category m versus baseline by β_m
  - A unit change in X changes the *logit* of observing category m versus category n by $(\beta_m - \beta_n)$
- Interesting, and forms the basis of the individual significance test. But, as with binary logit, nobody understands these numbers!
- Much easier to digest: effects on odds, and effects on probabilities for given changes in X

- Odds interpretation, following derivation on slide 9:
  - Every unit change in X leads to a constant e^{β_m} factor change in the odds of observing category *m* versus baseline category *b* (regardless of the value of all/any other variables)
  - Every unit change in X leads to a constant $e^{\beta_m} / e^{\beta_n}$ factor change in the odds of observing category *m* versus category *n*
  - Can express these changes in terms of unit change in X or *standard deviation* change in X via "listcoef"

```
. listcoef

mlogit (N=1658): Factor change in the odds of presvote

Variable: age (sd=17.189)
```

		b	z	P>|z|	e^b	e^bStdX
bush	vs clinton	0.0003	0.100	0.920	1.000	1.006
bush	vs perot	0.0183	4.121	0.000	1.018	1.370
clinton	vs bush	-0.0003	-0.100	0.920	1.000	0.995
clinton	vs perot	0.0180	4.228	0.000	1.018	1.362
perot	vs bush	-0.0183	-4.121	0.000	0.982	0.730
perot	vs clinton	-0.0180	-4.228	0.000	0.982	0.734

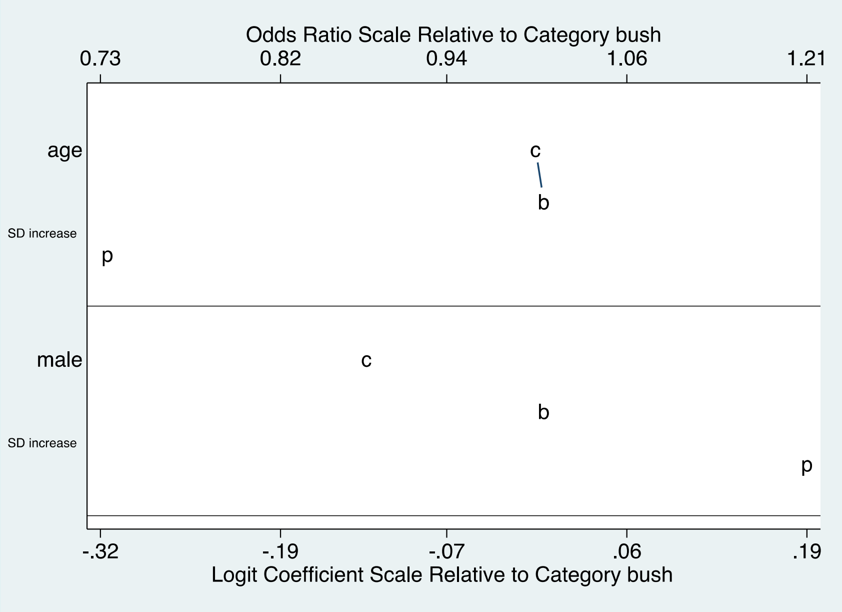
```
Variable: male (sd=0.499)
```

		b	z	P>|z|	e^b	e^bStdX
bush	vs clinton	0.2563	2.305	0.021	1.292	1.136
bush	vs perot	-0.3797	-2.616	0.009	0.684	0.827
clinton	vs bush	-0.2563	-2.305	0.021	0.774	0.880
clinton	vs perot	-0.6360	-4.600	0.000	0.529	0.728
perot	vs bush	0.3797	2.616	0.009	1.462	1.209
perot	vs clinton	0.6360	4.600	0.000	1.889	1.374

Increasing by one unit on age changes the odds of, e.g., Bush v. Perot by 1.02 (or 2%); one s.d. change on age changes the odds by 1.37 (37%)

Men's odds of voting, e.g., for Bush over Clinton are 1.292 greater than women's (29.2%)

- Long and Freese SPOST has flexible routine for plotting these odds ratios to give better visual sense of the effects



- Can see that a standard deviation increase in age leads to a .73 decrease in the odds of voting Perot relative to Bush (which is a 1/.73 or 1.37 increase in odds of Bush versus Perot – see previous slide's result)
- Can see that there is *no* significant different in odds for C/B based on s.d. change in age – this is the connected line in the top part of the graph
- Can see difference in the factor change in the odds for male – higher for Perot, lower for Clinton, relative to Bush (but really should use 0-1, not SD change)

MNL Interpretation: Marginal Effects on P(Y=m)

- Effects on P(Y=m) also common in MNL
- Marginal change

$$\frac{\partial \Pr(y=m \mid x)}{\partial x_k} = \Pr(y=m \mid x)\left[\beta_{km} - \sum_{j=1}^{J} \beta_{kj} \Pr(y=j \mid x)\right]$$

- This means that marginal change depends not only on the value of x and the β for that category, but also on the values of all other variables *and* the coefficients for the other categories (this last part is different from binary logit)
- So the marginal effect for x on category *m* does not need to even have the same sign as the regression coefficient for x on category *m*!
- It also means that the marginal effect can change signs as x changes in magnitude, which is somewhat counterintuitive as well
- Consequently, marginal change not used that frequently in MNL

- Discrete change:

$$\frac{\Delta \Pr(y=m \mid x)}{\Delta x_k (x_k^{start} \rightarrow x_k^{end})} = \Pr(y=m \mid x, x_k = x_k^{end}) - \Pr(y=m \mid x, x_k = x_k^{start})$$

where all other variables are held at *x* --either at their observed values or at the mean – and we vary x_k by a given amount, either a unit, a standard unit, or min/max or any other quantity we want

- The probability for the outcome *m*, given X, follows the derivation above:

$$P(y=m|x, \text{not baseline category } b) = \frac{exp^{XB_m}}{\sum exp^{XB_{j\neq b}} + 1}$$

$$P(y=m|x, \text{baseline category } b) = \frac{1}{\sum exp^{XB_{j\neq b}} + 1}$$

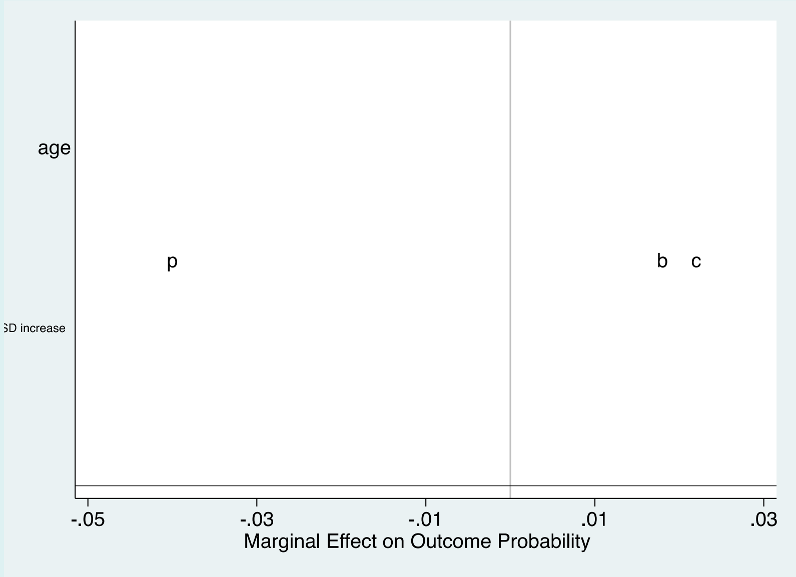
```
. mchange

mlogit: Changes in Pr(y) | Number of obs = 1658

Expression: Pr(presvote), predict(outcome())
```

	bush	clinton	perot
age			
+1	0.001	0.001	-0.003
p-value	0.083	0.042	0.000
+SD	0.018	0.022	-0.041
p-value	0.119	0.066	0.000
Marginal	0.001	0.001	-0.003
p-value	0.081	0.041	0.000
male			
+1	0.009	-0.097	0.087
p-value	0.696	0.000	0.000
+SD	0.007	-0.048	0.041
p-value	0.549	0.000	0.000
Marginal	0.018	-0.095	0.077
p-value	0.426	0.000	0.000

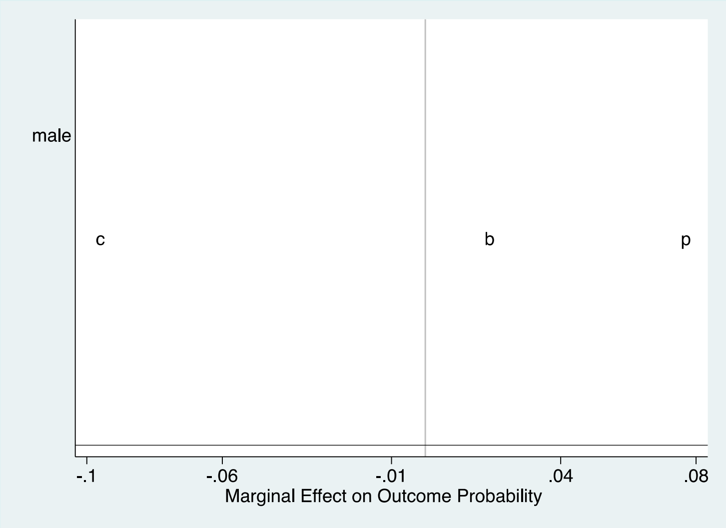
Average predictions

	bush	clinton	perot
Pr(y|base)	0.340	0.478	0.182

- SPOST "mchange" – observed values for other IVs as default
- Changing a standard unit on age increased the probability of voting for Bush by .018, Clinton by .022, and decreases the probability of voting for Perot by .041
- Men have a .01 greater probability of voting for Bush than women; a .088 greater probability of voting for Perot than women; and women have a .098 greater probability of voting for Clinton



SPOST "mchangeplot" – can see how discrete change in X (here a standard deviation change in age) changes the probabilities of observing each outcome



"mchangeplot" after "mchange, amount(bin)" for dichotomous variable male – can see how changing from one category to the other on X changes the probabilities of observing each outcome

Independence of Irrelevant Alternatives

- MNL very straightforward and logical extension of binary logit. However, the whole framework rests on one extremely important but perhaps restrictive assumption, the "independence of irrelevant alternatives", or IIA.
- This assumption states that the odds of observing outcome *m* versus *n* depends solely on outcomes *m* and *n*, regardless of whether outcomes *r s* and *t* may or may not be present
- Can see this with our derivation for odds

$$\textit{Odds}\ \text{Clinton/Perot} = \frac{P(y=C\,|\,x)}{P(y=P\,|\,x)} = \frac{\dfrac{\exp^{XB_C}}{\exp^{XB_C}+\exp^{XB_P}+1}}{\dfrac{\exp^{XB_P}}{\exp^{XB_C}+\exp^{XB_P}+1}} = \frac{\exp^{XB_C}}{\exp^{XB_P}}$$

$$\textit{Odds}\ \text{m|n} = \frac{\exp^{XB_m}}{\exp^{XB_n}}$$

- This means the odds for outcomes *m* versus *n* are only related to the coefficients for *m* and *n* categories (where *n* might be the baseline category in which case the denominator would be 1); the coefficients and the odds are not affected by whether another choice is also available
- Classic illustration: the "red bus, blue bus" problem. Suppose individuals have 3 transportation choices (Red Bus, Blue Bus, Car)
- Assume that they treat the two buses as equivalent and are indifferent between travelling by bus or car
- This implies a .50 probability of choosing Car, and a .50 probability of choosing a Bus
- So: if choice set is Car versus Red Bus, odds of Car is 1:1
- Adding the Blue Bus to the set, though, odds of Car to Red Bus will be 2:1 (.50 probability versus .25 probability for Red Bus – since there is also a .25 probability of choosing the Blue Bus).
- This violates IIA!!! The odds were assumed to still be 1:1 when adding Blue Bus to the mix. But Blue and Red buses are "close substitutes"

- In political science, IIA is often problematic. IIA assumes, e.g., that third parties should not affect the relative probabilities (odds) of choosing between mainstream parties. Only way that happens is if they takes votes in equal proportions from mainstream. But third parties are often "close(r) substitutes" of some parties compared to others and thus violate IIA.
- We can test for IIA (though Long and Freese are dubious about the validity of the tests). Intuition: Run Full model, then leave out one alternative and refit a Restricted model. Obtain β_F and β_R.
- If IIA holds, then the estimates for the effects of the remaining alternatives in the Full model and in the Restricted model should be the same, given sampling error

- Hausman test: If IIA holds, both estimates should be the same, but restricted model will be inefficient since it does not use all the available data

$$H_{IIA} = (\hat{\beta}_R - \hat{\beta}_F)\left[Var\hat{\beta}_R - Var\hat{\beta}_F\right]^{-1}(\hat{\beta}_R - \hat{\beta}_F)$$

- Need to restrict the coefficients tested in the full model to those corresponding only to the coefficients in the restricted model so that the two sets of coefficients have same dimension
- Alternative test: Small and Hsiao (1985), also implemented in SPOST
- Long and Freese express doubts about empirical properties of both tests (especially since they can sometimes implausibly be negative!) and recommend that theory be the guide for accepting or not accepting IIA.
- If not accepted, alternative models include nested logit and multinomial probit with correlated errors