Today's Agenda

1. When should we use multinomial models?
2. What is the multinomial logit?
    - Review binary logit
    - Intuition
    - Math
3. Estimation, Testing, and Prediction
4. An important assumption: IIA
5. Three alternative multinomial models
    - Nested logit
    - Multinomial probit
    - Conditional logit

Why Use a Multinomial Model?

- Often the type of dependent variable drives the choice of model.

  Continuous DV:

  Binary DV:

  Unordered categories:

  Ordered DV (e.g., Likert scale):

What is the Multinomial Logit?

a generalization of the binary logit for K categories

Imagine we're modeling whether a person likes ice cream

For binary logit, the log odds of Y=1 is a linear function of the x's:

$$\log\left(\frac{P(Y=1 \mid x)}{P(Y=0 \mid x)}\right) = a + bx$$

b measures the change in the log odds of Y=1 associated with a one unit change in x.

exp(b) is an odds ratio. If exp(b)=1.1, then an increase of one unit in x multiplies the odds of Y=1 by 1.1.

What is the Multinomial Logit?

Now imagine we're modeling a person's favorite flavor of ice cream: K=3 and 1=vanilla, 2=chocolate, 3=strawberry

For a multinomial logit, we have K-1 equations. Each equation models the odds of a choice relative to a baseline (in this case strawberry).

$$\log(P(Y=1|x) / P(Y=K|x)) = a_1 + b_1 x$$

...

$$\log(P(Y=K-1|x) / P(Y=K|x)) = a_{K-1} + b_{K-1} x$$

b_1 measures the change in the log odds of Y=1 relative to Y=K associated with a one unit change in x.

$\exp(b_1)$ is an relative risk ratio. If $\exp(b_1)=1.1$, then an increase of one unit in x multiplies the odds of Y=1 relative to Y=K by 1.1.

Predicting Probabilities

It's not hard to solve the previous equations for the probabilities.

$\log\left(\frac{Pr(Y=1|x)}{Pr(Y=0|x)}\right) = a+bx$

For a binary logit:

$P(Y=1|x) = \exp(a+bx) / (1+\exp(a+bx))$

For a multinomial logit:

$P(Y=1|x) = \exp(a_1+b_1x) / (1+\exp(a_1+b_1x) + \ldots + \exp(a_{K-1}+b_{K-1}x)$

…

$P(Y=K-1|x) = \exp(a_{K-1}+b_{K-1}x) / (1+\exp(a_1+b_1x) + \ldots + \exp(a_{K-1}+b_{K-1}x)$

$P(Y=K|x) = 1 – P(Y=1|x) – \ldots – P(Y=K-1|x)$

Example: Ice Cream Flavors

- 200 high school students
- Question: How do the following X's predict favorite flavor (**`fav_flavor`**)?

1 = vanilla, 2 = choc, 3 = straw

  - Sex (**`female`**)
  - Socioeconomic Status (**`ses`**)
  - Writing test score (**`write`**)

```
. tab fav_flavor female ,col chi2 nokey

  Favorite |        female
    Flavor |      male     female |     Total
-----------+----------------------+----------
 Chocolate |        21         24 |        45 
           |     23.08      22.02 |     22.50 
-----------+----------------------+----------
   Vanilla |        47         58 |       105 
           |     51.65      53.21 |     52.50 
-----------+----------------------+----------
Strawberry |        23         27 |        50 
           |     25.27      24.77 |     25.00 
-----------+----------------------+----------
     Total |        91        109 |       200 
           |    100.00     100.00 |    100.00 

          Pearson chi2(2) =   0.0528   Pr = 0.974
```

```
. bysort fav_flavor: summ write

---------------------------------------------------------------
-> fav_flavor = Chocolate

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       write |         45    51.33333    9.397775         31         67

---------------------------------------------------------------
-> fav_flavor = Vanilla

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       write |        105    56.25714    7.943343         33         67

---------------------------------------------------------------
-> fav_flavor = Strawberry

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
       write |         50       46.76    9.318754         31         67
```

```
. mlogit fav_flavor female i.ses write ,base(3) rrr

Iteration 0:   log likelihood = -204.09667
Iteration 1:   log likelihood = -179.76439
Iteration 2:   log likelihood =  -178.8658
Iteration 3:   log likelihood = -178.85898
Iteration 4:   log likelihood = -178.85898

Multinomial logistic regression                   Number of obs   =        200
                                                  LR chi2(8)      =      50.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -178.85898                       Pseudo R2       =     0.1237

------------------------------------------------------------------------------
  fav_flavor |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Chocolate    |
      female |    .590754   .2734778    -1.14   0.256     .2384295    1.463705
             |
         ses |
          2  |   .3961859   .1987013    -1.85   0.065     .1482488    1.058783
          3  |   .7203016   .4774605    -0.49   0.621      .196465    2.640849
             |
       write |   1.067799   .0267719     2.62   0.009     1.016595    1.121581
       _cons |   .0829863   .0983008    -2.10   0.036     .0081418    .8458466
-------------+----------------------------------------------------------------
Vanilla      |
      female |   .5453669   .2304575    -1.43   0.151     .2382285    1.248486
             |
         ses |
          2  |   .6657308   .3238087    -0.84   0.403     .2566125    1.727108
          3  |   2.268685   1.377131     1.35   0.177     .6903676    7.455352
             |
       write |    1.13266   .0270803     5.21   0.000     1.080807    1.186999
       _cons |   .0047821   .0056251    -4.54   0.000     .0004768    .0479589
-------------+----------------------------------------------------------------
Strawberry   |  (base outcome)
------------------------------------------------------------------------------
```

Testing

Numbers identify specific dummy variable

```
. test 2.ses 3.ses

 ( 1)  [Chocolate]2.ses = 0
 ( 2)  [Vanilla]2.ses = 0
 ( 3)  [Strawberry]2o.ses = 0
 ( 4)  [Chocolate]3.ses = 0
 ( 5)  [Vanilla]3.ses = 0
 ( 6)  [Strawberry]3o.ses = 0
       Constraint 3 dropped
       Constraint 6 dropped

           chi2(  4) =   10.55
         Prob > chi2 =    0.0321
```

Brackets identify specific equation

```
. test [Chocolate]female = [Vanilla]female

 ( 1)  [Chocolate]female - [Vanilla]female = 0

           chi2(  1) =    0.04
         Prob > chi2 =    0.8403
```

```
. mlogit fav_flavor female i.ses write ,base(3) rrr

Iteration 0:   log likelihood = -204.09667
Iteration 1:   log likelihood = -179.76439
Iteration 2:   log likelihood =  -178.8658
Iteration 3:   log likelihood = -178.85898
Iteration 4:   log likelihood = -178.85898

Multinomial logistic regression                   Number of obs   =        200
                                                  LR chi2(8)      =      50.48
                                                  Prob > chi2     =     0.0000
Log likelihood = -178.85898                       Pseudo R2       =     0.1237

------------------------------------------------------------------------------
  fav_flavor |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Chocolate    |
      female |    .590754   .2734778    -1.14   0.256     .2384295    1.463705
             |
         ses |
          2  |   .3961859   .1987013    -1.85   0.065     .1482488    1.058783
          3  |   .7203016   .4774605    -0.49   0.621      .196465    2.640849
             |
       write |   1.067799   .0267719     2.62   0.009     1.016595    1.121581
       _cons |   .0829863   .0983008    -2.10   0.036     .0081418    .8458466
-------------+----------------------------------------------------------------
Vanilla      |
      female |   .5453669   .2304575    -1.43   0.151     .2382285    1.248486
             |
         ses |
          2  |   .6657308   .3238087    -0.84   0.403     .2566125    1.727108
          3  |   2.268685   1.377131     1.35   0.177     .6903676    7.455352
             |
       write |    1.13266   .0270803     5.21   0.000     1.080807    1.186999
       _cons |   .0047821   .0056251    -4.54   0.000     .0004768    .0479589
-------------+----------------------------------------------------------------
Strawberry   |  (base outcome)
------------------------------------------------------------------------------
```

Testing

Numbers identify specific dummy variable

```
. test 2.ses 3.ses

 ( 1)  [Chocolate]2.ses = 0
 ( 2)  [Vanilla]2.ses = 0
 ( 3)  [Strawberry]2o.ses = 0
 ( 4)  [Chocolate]3.ses = 0
 ( 5)  [Vanilla]3.ses = 0
 ( 6)  [Strawberry]3o.ses = 0
       Constraint 3 dropped
       Constraint 6 dropped

           chi2(  4) =   10.55
         Prob > chi2 =    0.0321
```

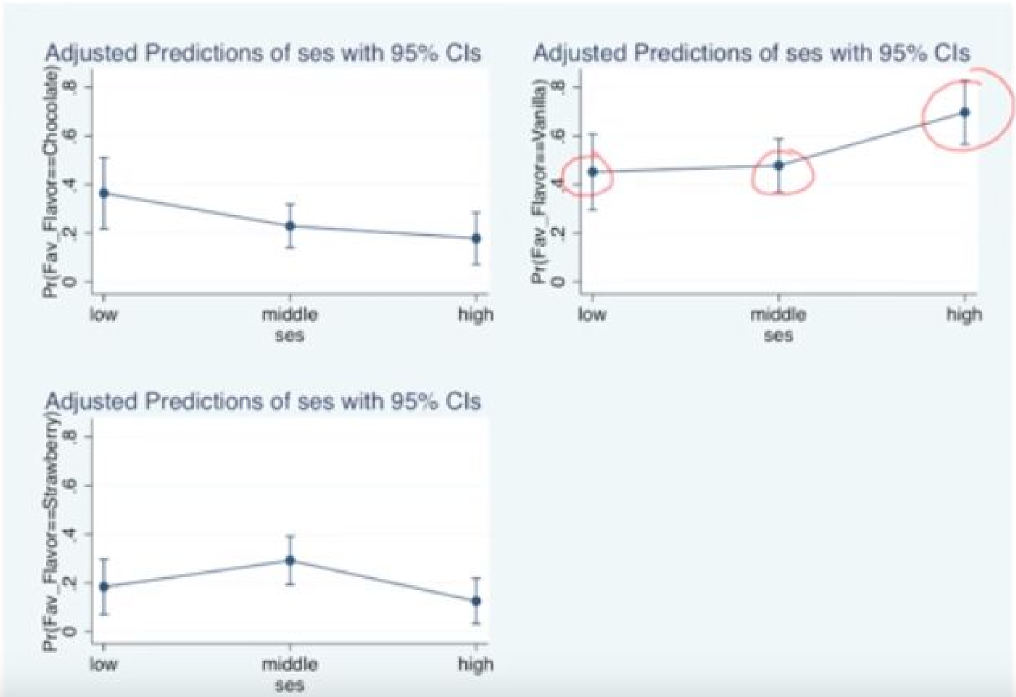
Brackets identify specific equation

```
. test [Chocolate]female = [Vanilla]female

 ( 1)  [Chocolate]female - [Vanilla]female = 0

           chi2(  1) =    0.04
         Prob > chi2 =    0.8403
```

Predicting Probabilities

```
. margins ses ,atmeans predict(outcome(1))

Adjusted predictions                              Number of obs   =        200
Model VCE    : OIM

Expression   : Pr(fav_flavor==Chocolate), predict(outcome(1))
at           : female          =        .545 (mean)
               1.ses           =        .235 (mean)
               2.ses           =        .475 (mean)
               3.ses           =         .29 (mean)
               write           =      52.775 (mean)

------------------------------------------------------------------------------
             |            Delta-method
             |     Margin   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         ses |
          1  |   .3645516   .0747149     4.88   0.000      .218113    .5109902
          2  |   .2296805   .0454905     5.05   0.000     .1405207    .3188403
          3  |   .1784702   .0543075     3.29   0.001     .0720294    .2849109
------------------------------------------------------------------------------
```

Graphing Probabilities



A confession

Data isn't actually about ice cream; It's about choice of high school program

```
. use http://www.ats.ucla.edu/stat/data/hsbdemo, clear
. tab prog

    type of |
    program |      Freq.     Percent        Cum.
------------+-----------------------------------
    general |         45       22.50       22.50
   academic |        105       52.50       75.00
   vocation |         50       25.00      100.00
------------+-----------------------------------
      Total |        200      100.00

. rename prog fav_flavor
. label var fav_flavor "Favorite Flavor"
. label define flavor 1 "Chocolate" 2 "Vanilla" 3 "Strawberry"
. label values fav_flavor flavor
```

For more: `http://www.ats.ucla.edu/stat/stata/dae/mlogit.htm`

Independence of Irrelevant Alternatives (IIA)

- Odds of outcome j versus outcome k do not depend on what other outcomes (l, m, n..) are available.
- Relative changes don't depend on having all of the choices available

Example:

- The relative proportions of students who pick chocolate and strawberry won't change depending on if vanilla is available.
- That is, if an equal number of kids choose chocolate and strawberry when vanilla is available, then an equal number will choose them when vanilla runs out too.

Classic IIA example: Red Bus – Blue Bus

Suppose there are initially only two transportation alternatives:

Option	Probability
Red Bus	1/3
Car	2/3

Odds of Car relative to Red Bus: 2/1

A more realistic pattern would be that the Blue Bus riders would come from the other *bus riders* not the car drivers. But that would change the relative odds of car to red bus, which is not allowed by this model:

Option	Probability
Red Bus	1/6
Blue Bus	1/6
Car	2/3

P(Car)/P(Red Bus) = 4/1

Why does this happen? Because the error terms of the different equations are assumed independent when in real life often not this way. Sometimes they will be highly correlated.

Unfortunately, the multinomial logit doesn't take this into account.

Nested Logit

Decisions are sequential. e.g.,

1. Decide whether to take the car or bus
2. If bus, **then** decide whether to take red or blue

In this example, relative odds of red or blue are invariant to whether car is available. That might be totally reasonable.

Example 2:

1. Decide whether to use birth control
2. Decide between condom or pill

Not clear if weakened IIA assumption is valid here. That is, choice of whether to use contraception might depend on methods available.

Multinomial Probit

Holy grail of choice models as the general form doesn't suffer from IIA at all.

Unfortunately, it's almost never well-identified (i.e., standard errors on coefficients are almost always large even with big samples)

See Keane (1992) "A Note on Identification in the Multinomial Probit Model"

Conditional Logit

- Extension of the multinomial logit
- Allows properties of choices themselves to vary across individuals
- Example:
  - People are choosing bus, car, or bike
  - People vary by age and sex
  - Modes vary by cost

Your turn

alligators.dta contains information about the primary food choice of 219 alligators living in 4 different lakes.

```
  obs:           219                          Alligator data from Agresti (1990)
 vars:             3                          28 Jan 2013 03:33
 size:         2,628
--------------------------------------------------------------------------------
              storage  display     value
variable name   type   format      label      variable label
--------------------------------------------------------------------------------
large           float  %9.0g                  >= 2.3 meters
food            float  %12.0g      foods      Primary Food
lake            float  %9.0g       lakes      Lake
```

Estimate a model that predicts food choice. Be prepared to explain your results.

```
. tab food

Primary Food |      Freq.     Percent        Cum.
-------------+-----------------------------------
        Fish |         94       42.92       42.92
Invertebrate |         61       27.85       70.78
     Reptile |         19        8.68       79.45
        Bird |         13        5.94       85.39
       Other |         32       14.61      100.00
-------------+-----------------------------------
       Total |        219      100.00
```

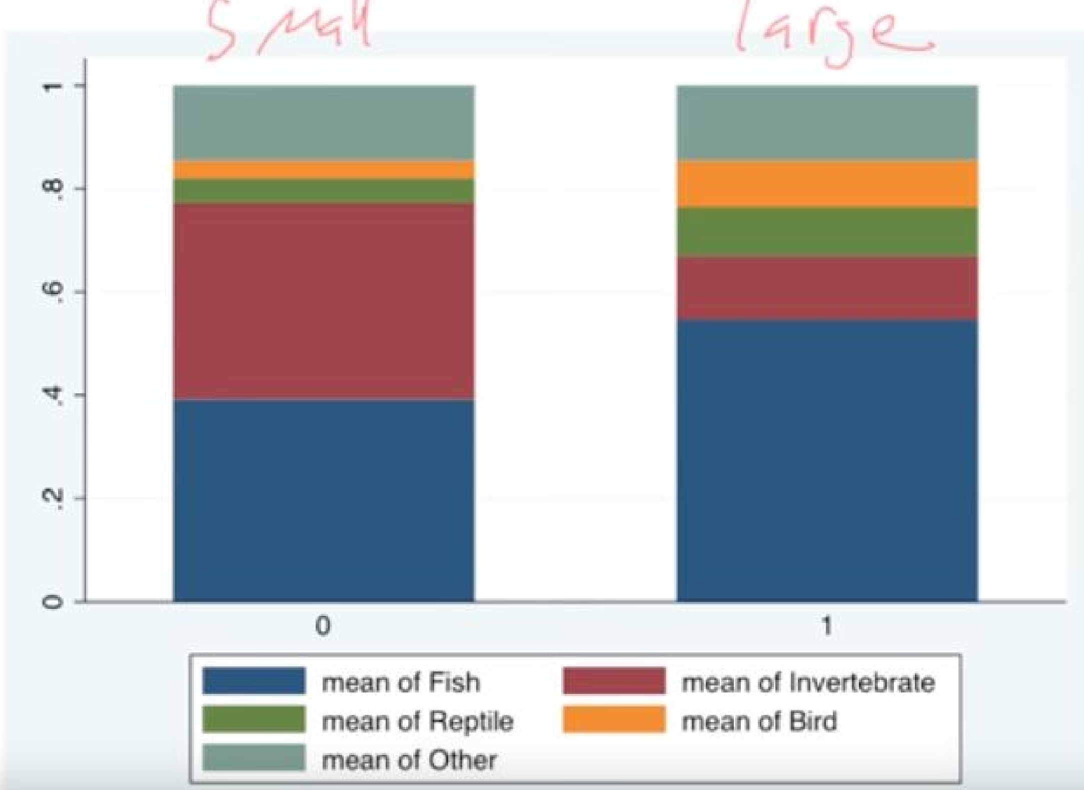
```
. mlogit food large i.lake ,rrr
Multinomial logistic regression                   Number of obs   =        219
                                                  LR chi2(16)     =      64.28
                                                  Prob > chi2     =     0.0000
Log likelihood = -270.04014                       Pseudo R2       =     0.1064

------------------------------------------------------------------------------
        food |        RRR   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
Fish         |  (base outcome)
-------------+----------------------------------------------------------------
Invertebrate |
       large |   .2326536   .0921178    -3.68   0.000     .1070734    .5055196
             |
        lake |
          2  |   13.40433   8.842951     3.93   0.000     3.678751    48.84161
          3  |   16.12456   10.82318     4.14   0.000     4.326548    60.09442
          4  |   5.250685   3.218025     2.71   0.007     1.579552    17.45412
             |
       _cons |   .1739178   .0937736    -3.24   0.001     .0604491    .5003777
-------------+----------------------------------------------------------------
Reptile      |
       large |   1.420861   .8241441     0.61   0.545      .455856     4.42869
             |
        lake |
          2  |   3.373988   2.651998     1.55   0.122     .7229162    15.74704
          3  |    5.43292   4.240103     2.17   0.030     1.176836    25.08133
          4  |   .2885818   .3420941    -1.05   0.294     .0282636    2.946527
             |
       _cons |   .0886536   .0570584    -3.76   0.000     .0251104    .3129961
```

05:52 continued on next slide...


```
Bird         |
       large |    1.87885   1.207123     0.98   0.326       .53335    6.618686
             |
        lake |
          2  |   .2596748   .3021372    -1.16   0.247     .0265484    2.539927
          3  |   1.480899   1.157723     0.50   0.615     .3199494      6.8544
          4  |   .4990158   .3898628    -0.89   0.374     .1079199    2.307422
             |
       _cons |    .131517   .0733936    -3.64   0.000      .044052    .3926432
-------------+----------------------------------------------------------------
Other        |
       large |   .7178101   .3217598    -0.74   0.460     .2981651    1.728074
             |
        lake |
          2  |   .4401925   .3211758    -1.12   0.261     .1053378    1.839506
          3  |    1.99406   1.116021     1.23   0.218        .6658    5.972174
          4  |   .4377111   .2440417    -1.48   0.138     .1467602     1.30547
             |
       _cons |   .4740108    .166843    -2.12   0.034     .2377832    .9449207
------------------------------------------------------------------------------
```

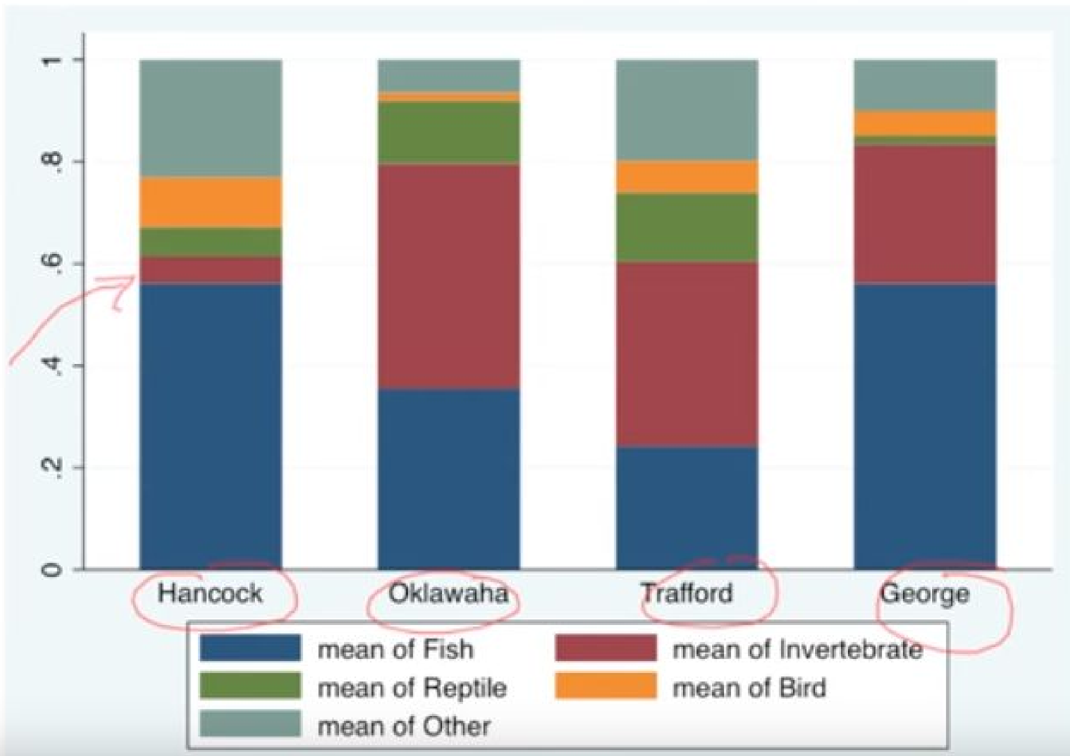
```
tab lake ,gen(lake)

mlogit food large lake2 lake3 lake4 ,rrr

preserve

foreach i of numlist 1/4 {
  egen lake`i'_mean = mean(lake`i')
  replace lake`i' = lake`i'_mean
}

predict Fish Invertebrate Reptile Bird Other

graph bar Fish Invertebrate Reptile Bird Other ,stack
over(large)

restore
```

Predicted probabilities for small and large alligators




```
tab lake ,gen(lake)

mlogit food large lake2 lake3 lake4 ,rrr

preserve

egen large_mean = mean(large)
replace large = large_mean

predict Fish Invertebrate Reptile Bird Other

graph bar Fish Invertebrate Reptile Bird Other ,stack
over(lake)
```

Predicted probabilities by lake



What have we learned?

1. Use multinomial models when you have multinomial dependent variables.
2. A multinomial logit is a straight-forward extension of the binary logit.
   - Binary logit
   - Intuition
   - Math
3. Estimation, Testing, and Prediction
4. An important assumption: IIA
5. Three alternative multinomial models
   - Nested logit
   - Multinomial probit
   - Conditional logit