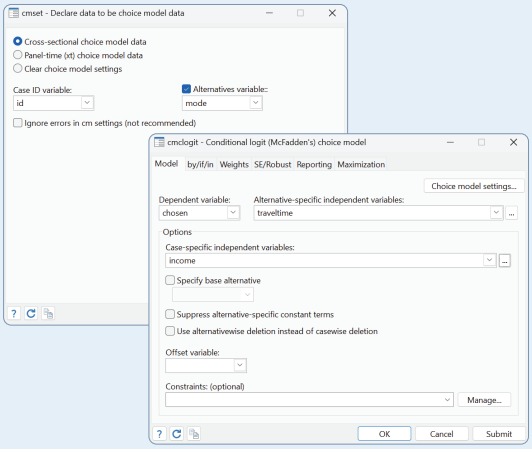
Choice models

Stata's choice-modeling suite makes it easy to explore discrete choice data, fit choice models, and interpret the results. Get answers to real research questions.

- Summarize choice data
- Model discrete choices
  - Conditional logit
  - Mixed logit
  - Multinomial probit
  - Rank-ordered logit
  - Rank-ordered probit
  - Panel-data mixed logit
- Truly interpret the results
  - Expected probabilities
    - For any alternative
    - For any subpopulation
    - At specific covariate levels
  - Differences in probabilities (effects)
    - As a covariate changes for an alternative
      Increased airfare decreases probability of flying
    - As a covariate changes for another alternative
      Increased airfare increases probability of car travel
  - Marginal effects
  - Tests and confidence intervals for everything



Prepare your data

Declare variables that identify individuals and alternatives

```
. cmset id mode
```

Summarize data

Tabulate chosen alternatives

```
. cmtab, choice(chosen)
```

Summarize variables (**traveltime** and **cost**) across chosen alternatives

```
. cmsummarize traveltime cost, choice(chosen)
```

Tabulate choice sets

```
. cmchoiceset
```

Fit a discrete choice model

Conditional logit (McFadden's choice) model; **traveltime** varies across alternatives; **income** is constant within **id**

```
. cmclogit chosen traveltime, casevars(income)
```

Multinomial probit

```
. cmmprobit chosen traveltime, casevars(income)
```

Mixed logit with random coefficients for **cost**

```
. cmmixlogit chosen traveltime,
              random(cost) casevars(income)
```

Fit a model for a rank-ordered outcome

Rank-ordered probit

```
. cmroprobit rank traveltime, casevars(income)
```

Rank-ordered logit

```
. cmrologit rank traveltime cost
```

Fit a model to panel data

Mixed logit model

```
. cmset id time mode
. cmxtmixlogit chosen traveltime,
              random(cost) casevars(income)
```

Viewer - view cm1.smcl

view cm1.smcl

Dialog | Also see | Jump to

```
. cmxtmixlogit chosen traveltime, random(cost) casevars(income)
```

chosen	Coefficient	Std. err.	z	P>|z|	[95% conf.	interval]
mode						
traveltime	-.837606	.0437603	-19.14	0.000	-.9233746	-.7518374
cost	-1.560057	.2667461	-5.85	0.000	-2.082869	-1.037244
/Normal						
sd(cost)	2.015974	.2594489			1.566529	2.594369
Car	(base alternative)					
Public						
income	-.3681157	.034001	-10.83	0.000	-.4347564	-.301475
_cons	-.0095711	.2526377	-0.04	0.970	-.5047318	.4855896
Bicycle						
income	-.5083127	.0457894	-11.10	0.000	-.5980583	-.4185671
_cons	-.3506506	.3112727	-1.13	0.260	-.9607339	.2594326
Walk						
income	-.8844826	.0681116	-12.99	0.000	-1.017979	-.7509864
_cons	.792664	.3787151	2.09	0.036	.050396	1.534932

CAP NUM INS

After fitting a choice model with any **cm** command, you can easily answer interesting research questions.

What proportion of individuals do we expect will select air travel? Train travel? Bus travel? Car travel?

Viewer - view cm2.smcl

view cm2.smcl

Dialog | Also see | Jump to

```
. margins

Predictive margins                                      Number of obs = 840
Model VCE: OIM

Expression: Pr(mode|1 selected), predict()
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
_outcome						
Air	.2761905	.0275268	10.03	0.000	.2222389	.330142
Train	.3	.0284836	10.53	0.000	.2441731	.3558269
Bus	.1428571	.0234186	6.10	0.000	.0969576	.1887567
Car	.2809524	.028043	10.02	0.000	.2259891	.3359156

CAP NUM INS

We expect 28% to select air, 30% to select train, 14% to select bus, and 28% to select car.

What proportion of individuals with income levels ranging from $30,000 to $70,000 per year are expected to select car travel?

Viewer - view cm3.smcl

view cm3.smcl

Dialog | Also see | Jump to

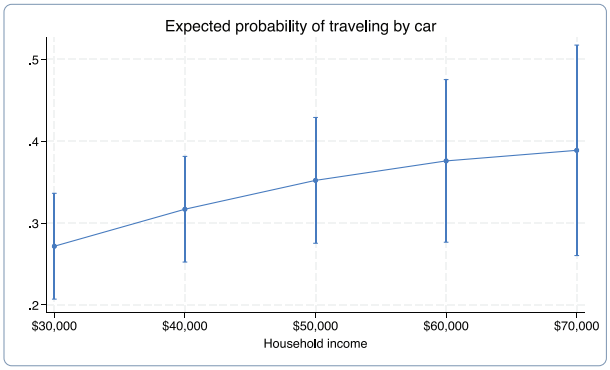
```
. margins, at(income=(30(10)70)) outcome(Car)

Predictive margins                                      Number of obs = 840
Model VCE: OIM

Expression: Pr(mode|1 selected), predict()
Outcome:    Car

1._at: income = 30

2._at: income = 40

3._at: income = 50

4._at: income = 60

5._at: income = 70
```

	Margin	Delta-method std. err.	z	P>|z|	[95% conf.	interval]
_at						
1	.2717914	.0329811	8.24	0.000	.2071497	.3364331
2	.3169817	.0329227	9.63	0.000	.2524544	.3815091
3	.3522391	.0391994	8.99	0.000	.2754097	.4290684
4	.3760093	.050679	7.42	0.000	.2766802	.4753383
5	.3889296	.0655865	5.93	0.000	.2603825	.5174768

CAP NUM INS

Easily visualize the result:

```
. marginsplot
```

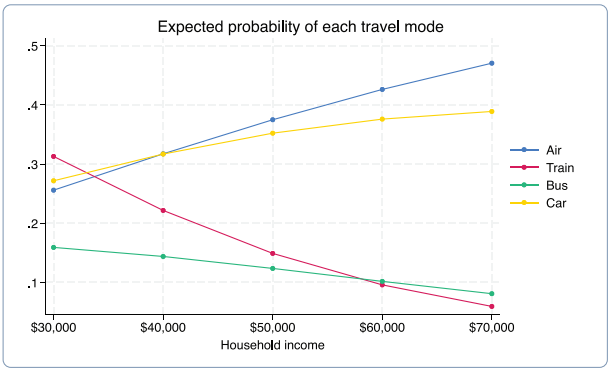


As income levels increase, what happens to the expected proportions of each travel method? Type

```
. margins, at(income=(30(10)70))
```

(output omitted)

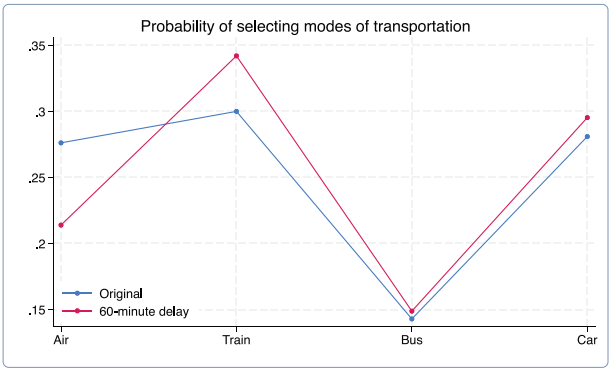
```
. marginsplot
```



What if wait times at airports increase by an hour? How do we expect this to affect the probability of selecting air travel? How does it affect the probability of selecting car travel? Train travel? Bus travel?

```
. margins, alternative(Air)
  at(traveltime=generate(traveltime))
  at(traveltime=generate(traveltime+60))
```

```
. marginsplot
```



What would we expect if air travel time increases by an hour while car travel time decreases by 30 minutes?

What would we expect if the price of train travel increases by 20%?

What would we expect if ...?

You can now answer questions like these and many others.