

選択モデル

はじめに

- 選択モデル(Choice Models)は、選択の結果を利用するデータのモデルです。個人や企業などの意思決定者によって、ある選択肢のセットから1つの物や手段が選択されます。
- 単一の選択をする離散選択モデルと、選択肢をランク付けする順序選択モデルの2種類があります。
- この資料では離散選択モデルを例に、選択データの操作と要約、モデルのフィット、およびそれらのモデルの結果の解釈に関するコマンドについて解説します。
- ✧ 本文中のコマンドをコピーし、Stata のコマンドウィンドウに貼り付けて実行できます。全ての操作のコマンドは、do ファイル `choice_models.do` にまとめられています。

係数の解釈

- 交通機関の選択を例にデモンストレーションします。2つの都市間の交通手段（飛行機、電車、バス、自動車）について、210人に調査を行いました。
- 各交通手段にかかる時間、各個人の所得、一緒に移動する人数に関する情報があります。

データセットをインポートします。交通機関での移動時間を表す変数 `traveltime` と待ち時間を表す変数 `termtime` から、合計移動時間を表す時間変数 `time` を作成します。

```
use https://www.stata-press.com/data/r16/travel, clear

generate time = traveltime + termtime
```

`cmset` コマンドを使用して、選択モデルのデータであることを宣言します。変数 `id` は各個人を識別し、変数 `mode` は交通手段の選択肢であることを指定します。

```
cmset id mode
```

```
caseid variable: id
alternatives variable: mode
```

次に、被説明変数を `choice` として、選択された交通手段を示す条件付きロジスティック回

帰モデルをフィットします。共変量のうち変数 **time** は唯一、4つの交通手段の選択肢間で変化します。**income** と **partysize** は、選択肢ごとの値ではなく、個人（またはケース）ごとに1つの値しかいないため、**casevars()** オプションにリストします。

```
cmclogit choice time, casevars(income partysize)
```

```
Iteration 0: log likelihood = -249.36629
Iteration 1: log likelihood = -236.01608
Iteration 2: log likelihood = -235.65162
Iteration 3: log likelihood = -235.65065
Iteration 4: log likelihood = -235.65065
```

```
Conditional logit choice model          Number of obs    =      840
Case ID variable: id                    Number of cases  =      210

Alternatives variable: mode              Alts per case:  min =      4
                                           avg =      4.0
                                           max =      4

Log likelihood = -235.65065              Wald chi2(7)     =      71.14
                                           Prob > chi2      =      0.0000
```

choice	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]	
mode						
time	-.0041641	.0007588	-5.49	0.000	-.0056512	-.002677
air	(base alternative)					
train						
income	-.0613414	.0122637	-5.00	0.000	-.0853778	-.0373051
partysize	.4123606	.2406358	1.71	0.087	-.0592769	.883998
_cons	3.39349	.6579166	5.16	0.000	2.103997	4.682982
bus						
income	-.0363345	.0134318	-2.71	0.007	-.0626605	-.0100086
partysize	-.1370778	.3437092	-0.40	0.690	-.8107354	.5365798
_cons	2.919314	.7658496	3.81	0.000	1.418276	4.420351
car						
income	-.0096347	.0111377	-0.87	0.387	-.0314641	.0121947
partysize	.7350802	.2184636	3.36	0.001	.3068993	1.163261
_cons	.7471042	.6732971	1.11	0.267	-.5725338	2.066742

結果から、次のことが分かります。

- **time** の係数は負であるため、移動時間が長くなると、その交通手段を選択する確率が低くなります。

- 今回は飛行機を選択肢のベースとして比較しています。「電車の場合の所得の係数が負である」ことは、所得が増加するにつれて、飛行機よりも電車を選択する可能性が低くなることを示しています。
- 自動車の場合の移動人数の係数は正です。一緒に移動する人数が多くなると、飛行機よりも自動車を選択する可能性が高くなることを示しています。

マージンからの推論

上記の結果から、次のことに興味を持ったとします。

1. 飛行機を選択する人の割合は全体の何パーセントか？
2. 所得の影響はどのようなものか？所得が\$10,000 増加すると、自動車や電車を選択する確率はどのように変化するか？
3. 空港の安全対策が強化され、待ち時間が 60 分増加した場合、各交通手段の選択の確率にどのように影響するか？

`margins` コマンドを使用して、これらについて考察していきます。

予想される選択確率 — 飛行機を選択する人の割合は全体の何パーセントか？

フィット後に `margins` コマンドを実行すると、各交通手段を選択する平均予測確率が得られます。

```
margins
```

```
Predictive margins                                Number of obs    =          840
Model VCE      : OIM

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

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
<code>_outcome</code>					
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

- このモデルと 2 つの都市間の移動者のランダムなサンプルに基づき、28%の人が飛行機で移動すると予想しています。
- また、30%が電車で、14%がバスで、28%が自動車で移動すると予想しています。

連続共変量の効果 — 所得の影響はどのようなものか？

はじめに、自動車を選択する確率に対する所得の影響を評価します。`margins` の `at()` オプションを使用して、\$30,000 から \$70,000 まで \$10,000 刻みの所得レベルでの自動車移動の予測確率を推定します。

```
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

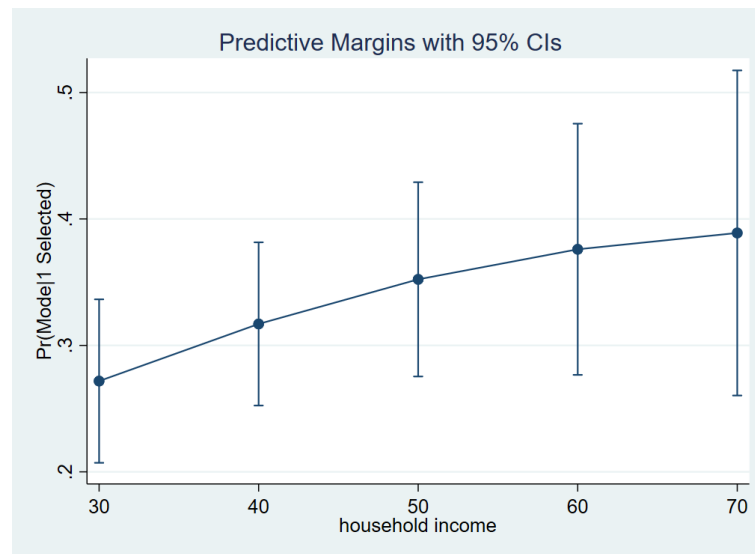
1._at         : income           =       30
2._at         : income           =       40
3._at         : income           =       50
4._at         : income           =       60
5._at         : income           =       70
```

	Delta-method				
	Margin	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

これらの確率をプロットして、所得の影響を視覚化します。

```
marginsplot
```

Variables that uniquely identify margins: income



- 結果から、自動車を選択する予想確率は、所得が増えるにつれて高くなります。

これらの違いが統計的に有意であるかを調べます。所得が\$10,000 増えるごとに、予想される確率の違いをテストします。`margins` の `contrast()` オプションを使用して逆隣接 (`ar`) コントラストを求めます。また、`nowald` オプションと `effects` オプションを使用して、テスト結果の出力を簡略化します。

```
margins, at(income=(30(10)70)) outcome(car) contrast(atcontrast(ar)
nowald effects)
```

```

Contrasts of 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

```

	Delta-method		z	P> z	[95% Conf. Interval]	
	Contrast	Std. Err.				
_at						
(2 vs 1)	.0451903	.016664	2.71	0.007	.0125296	.0778511
(3 vs 2)	.0352574	.017903	1.97	0.049	.0001681	.0703466
(4 vs 3)	.0237702	.0190387	1.25	0.212	-.013545	.0610854
(5 vs 4)	.0129204	.0200549	0.64	0.519	-.0263866	.0522273

- この表の一行目から、所得が\$30,000 から\$40,000 が増えると、自動車での移動を選択する確率が 0.045 増加することがわかります。
- 所得が\$40,000 から\$50,000 が増えると、自動車での移動の予想確率が 0.035 増加します。これらの効果は両方とも 5%の有意水準で有意です。
- 所得が\$50,000 から\$60,000 に、および\$60,000 から\$70,000 に増加しても、自動車での移動の確率に大きな違いはありません。

次に、電車を選択する確率に対する所得の影響を評価します。`margins` コマンドのオプションで `outcome(car)` の代わりに `outcome(train)` を使用します。

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

```

Predictive margins                                Number of obs   =       840
Model VCE    : OIM

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

1._at       : income           =       30
2._at       : income           =       40
3._at       : income           =       50
4._at       : income           =       60
5._at       : income           =       70

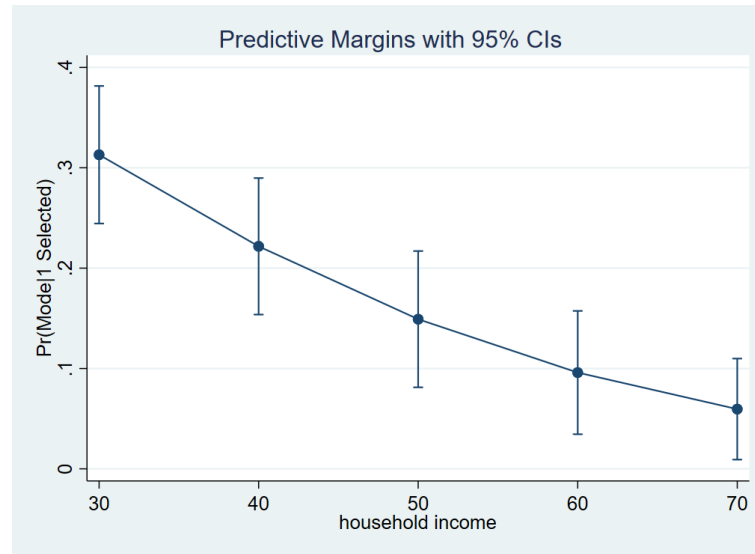
```

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
_at						
1	.3129898	.034968	8.95	0.000	.2444538	.3815257
2	.2217291	.0346969	6.39	0.000	.1537244	.2897338
3	.1491281	.0346442	4.30	0.000	.0812268	.2170294
4	.0959391	.0313489	3.06	0.002	.0344965	.1573818
5	.0595511	.0256786	2.32	0.020	.009222	.1098802

これらの確率をプロットして、所得の影響を視覚化します。

```
marginsplot
```

Variables that uniquely identify margins: income



- 結果から、所得レベルが上がると、電車を選択する確率は低くなります。

逆隣接コントラストを使用して、確率の違いを再度テストします。

```
margins, at(income=(30(10)70)) outcome(train) contrast(atcontrast(ar)
nowald effects)
```



```

Contrasts of predictive margins          Number of obs   =       840
Model VCE      : OIM

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

1._at        : income      =       30
2._at        : income      =       40
3._at        : income      =       50
4._at        : income      =       60
5._at        : income      =       70

```

	Delta-method		z	P> z	[95% Conf. Interval]	
	Contrast	Std. Err.				
_at						
(2 vs 1)	-.0912606	.0174009	-5.24	0.000	-.1253659	-.0571554
(3 vs 2)	-.072601	.0104846	-6.92	0.000	-.0931505	-.0520516
(4 vs 3)	-.053189	.0066782	-7.96	0.000	-.0662779	-.0401
(5 vs 4)	-.036388	.0064958	-5.60	0.000	-.0491195	-.0236566

- 結果から、所得が\$10,000 増えるごとに、電車を選択する予想確率が大幅に低下することがわかりました。

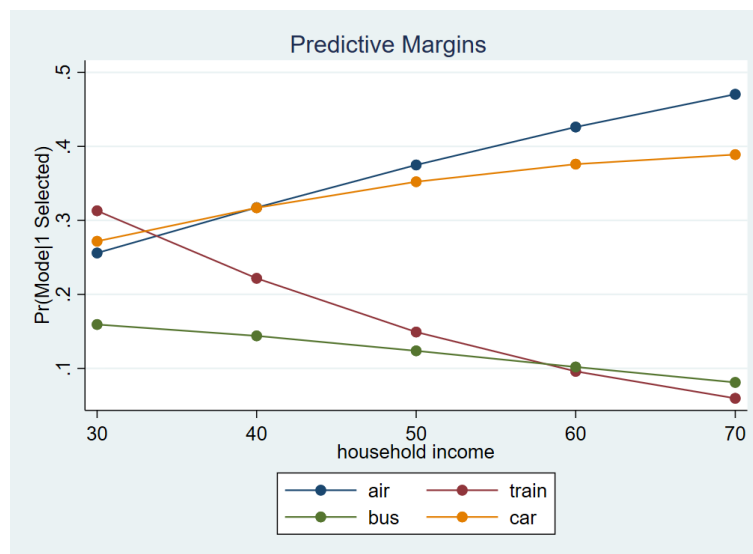
対象を自動車や電車に限定せず、各交通手段の選択確率が所得によってどのように変化するかを調べる際は、次のコマンドを実行します。ここでは出力結果は省略します。

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

結果をグラフ化すると次のようになります。ここでは **noci** オプションを使用して信頼区間の表示を省略します。

```
marginsplot, noci
```

Variables that uniquely identify margins: income _outcome



- このグラフにより、各所得レベルでの交通手段の選択確率を視覚的に比較できます。所得が増えると、交通手段としてバスや電車を選択する確率が低下し、飛行機や自動車を選択する確率が高くなります。

交通手段の選択予想確率の違いを正式にテストします。ここでは、\$30,000の所得レベルでは、4つの交通手段すべてが等しく選択される可能性があるかを調べます。`margins` コマンドで `contrast(outcomejoint)` オプションを使用します。

```
margins, at(income=30) contrast(outcomejoint)
```

```
Contrasts of predictive margins          Number of obs   =       840
Model VCE      : OIM

Expression   : Pr(mode|1 selected), predict()
at           : income           =       30
```

	df	chi2	P>chi2
_outcome	3	12.80	0.0051

- 結果から、これらの予想確率の少なくとも1つは、他の確率とは大幅に異なることがわかります。よって、\$30,000の所得レベルでは、4つの交通手段すべてが等しく選択されることはありません。

さらに具体的に、所得が\$50,000 の場合、電車とバスの交通手段を選択する予想確率に違いがあるかを調べます。`margins` コマンドの `outcome()` オプションでこれら 2 つの交通手段を指定し、`contrast(outcomecontrast(r))` オプションで 2 つの確率の差を推定します。

```
margins, at(income=50) outcome(bus train) contrast(outcomecontrast(r)
nowald effects)
```

```
Contrasts of predictive margins          Number of obs   =          840
Model VCE      : OIM

Expression    : Pr(mode|1 selected), predict()
at            : income          =          50
```

	Delta-method				
	Contrast	Std. Err.	z	P> z	[95% Conf. Interval]
_outcome (bus vs train)	-.0254125	.0504996	-0.50	0.615	-.1243899 .0735648

- 結果から、この所得レベルではバスと電車の交通手段を選択する確率に有意差は見られません。

カテゴリ共変量の影響

- これまでは連続変数 `income` を使用しましたが、代わりにカテゴリ変数の効果を評価する場合の方法を紹介します。

はじめに `xtile` コマンドを使用して、`income` の四分位数を表す新規変数 `income_cat` を作成し、適当なラベルをつけます。

```
xtile income_cat = income, nquantiles(4)

label define quartiles 1 "Quartile 1" 2 "Quartile 2" 3 "Quartile 3" 4
"Quartile 4"

label values income_cat quartiles
```

`income_cat` に因子変数を示す `i.` 演算子をつけてモデルをフィットします。

```
cmlogit choice time, casevars(i.income_cat partysize)
```

(結果省略)

次のコマンドを実行して、所得の四分位数ごとに交通手段として電車を選択する予想確率を推定します。

```
margins income_cat, outcome(train)
```

```
Predictive margins                                Number of obs    =          840
Model VCE      : OIM

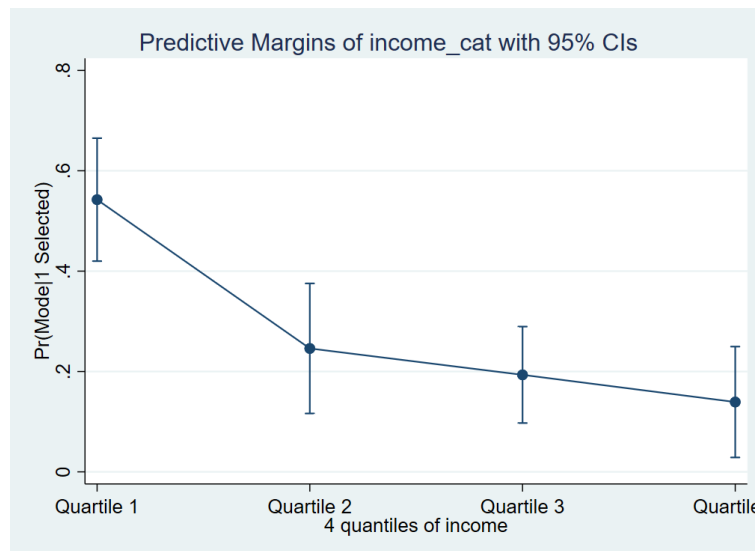
Expression    : Pr(mode|1 selected), predict()
Outcome       : train
```

	Delta-method		z	P> z	[95% Conf. Interval]	
	Margin	Std. Err.				
<code>income_cat</code>						
Quartile 1	.5424554	.0624929	8.68	0.000	.4199715	.6649392
Quartile 2	.2459475	.0661098	3.72	0.000	.1163746	.3755203
Quartile 3	.1933788	.0490343	3.94	0.000	.0972733	.2894843
Quartile 4	.1391895	.0563599	2.47	0.014	.0287262	.2496529

結果をグラフ化します。

```
marginsplot
```

Variables that uniquely identify margins: income_cat



ar. 演算子を使用して、隣接する所得の四分位数間で予想される確率の違いをテストします。

```
margins ar.income_cat, outcome(train) contrast(nowald effects)
```

```
Contrasts of predictive margins          Number of obs    =          840
Model VCE      : OIM

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

	Delta-method				
	Contrast	Std. Err.	z	P> z	[95% Conf. Interval]
income_cat (Quartile 2 vs Quartile 1) (Quartile 3 vs Quartile 2) (Quartile 4 vs Quartile 3)	-.2965079	.0905087	-3.28	0.001	-.4739018 -.1191141
	-.0525687	.082642	-0.64	0.525	-.214544 .1094067
	-.0541893	.0744351	-0.73	0.467	-.2000794 .0917009

- 交通手段として電車を選択する予想確率は、第1所得四分位から第2所得四分位に移動するとき大幅に変化することがわかります。

選択肢固有の共変量の影響 — 空港での待ち時間が 60 分増加した場合の影響は？

- これまで、交通手段の選択における所得の影響を検討してきましたが、選択肢固有の変数である移動時間についても検討することができます。
- 空港での待ち時間が全てのフライトで 60 分増加すると、交通手段の選択はどのように変化するかを調べます。

次のコマンドを実行します。margins の at() オプションを使用します。

```
margins, at(time=generate(time+60)) alternative(air)
```

```
Predictive margins                                Number of obs   =       840
Model VCE      : OIM

Expression    : Pr(mode|1 selected), predict()
Alternative   : air
at            : time                = time+60
```

	Delta-method				
	Margin	Std. Err.	z	P> z	[95% Conf. Interval]
_outcome					
air	.238248	.0261139	9.12	0.000	.1870657 .2894303
train	.3138431	.0293829	10.68	0.000	.2562537 .3714325
bus	.1513295	.0243434	6.22	0.000	.1036173 .1990417
car	.2965794	.0289688	10.24	0.000	.2398016 .3533572

- 元の飛行機の選択予想確率は 28% でした。今回は 24% で、移動時間が長くなると飛行機を選択する確率が低くなっています。

次のコマンドを実行すると、元の移動時間と、飛行機での移動時間が 60 分増加した場合の両方を一度に推定できます。

```
margins, at(time=generate(time)) at(time=generate(time+60))
alternative(air)
```

```

Predictive margins                                Number of obs   =       840
Model VCE    : OIM

Expression   : Pr(mode|1 selected), predict()
Alternative  : air

1._at       : time                               = time
2._at       : time                               = time+60

```

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
<u>_outcome#_at</u>						
air#1	.2761905	.0277787	9.94	0.000	.2217453	.3306357
air#2	.238248	.0261139	9.12	0.000	.1870657	.2894303
train#1	.3	.0284715	10.54	0.000	.2441968	.3558032
train#2	.3138431	.0293829	10.68	0.000	.2562537	.3714325
bus#1	.1428571	.0232375	6.15	0.000	.0973125	.1884018
bus#2	.1513295	.0243434	6.22	0.000	.1036173	.1990417
car#1	.2809524	.0280657	10.01	0.000	.2259446	.3359602
car#2	.2965794	.0289688	10.24	0.000	.2398016	.3533572

- #1 がついているものが元の移動時間、#2 がついているものが飛行機での移動時間が 60 分増加した場合です。

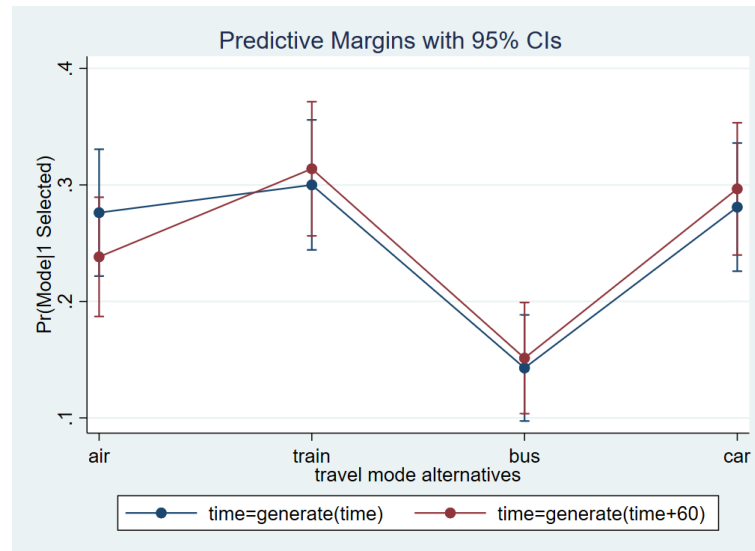
結果をグラフ化します。X軸に各選択肢を配置する `xdimension()` オプションを使用します。

```

marginsplot, xdimension(_outcome)

```

Variables that uniquely identify margins: _atopt _outcome
 Multiple at() options specified:
 _atoption=1: time=generate(time)
 _atoption=2: time=generate(time+60)



- 飛行機を選択する確率が減少したのに対し、他の交通手段の選択確率はやや増加しています。

これらの差が統計的に有意であるかを調べます。`margins` の `contrast(atcontrast(r))` オプションを使用して、元の移動時間との比較を行います。

```
margins, at(time=generate(time)) at(time=generate(time+60))
alternative(air) contrast(atcontrast(r) nowald effects)
```


Contrasts of predictive margins Number of obs = 840
 Model VCE : OIM

Expression : Pr(mode|1 selected), predict()
 Alternative : air

1._at : time = time

2._at : time = time+60

	Delta-method		z	P> z	[95% Conf. Interval]	
	Contrast	Std. Err.				
._at@_outcome						
(2 vs 1) air	-.0379425	.0056615	-6.70	0.000	-.0490387	-.0268462
(2 vs 1)						
train	.0138431	.0027216	5.09	0.000	.0085089	.0191773
(2 vs 1) bus	.0084724	.001936	4.38	0.000	.004678	.0122668
(2 vs 1) car	.015627	.0026002	6.01	0.000	.0105308	.0207232

- 結果から、これらの差は全て統計的に有意です。
- 飛行機を選択予想確率は3.8パーセントポイント減少しています。
- さらに発展させて、「飛行機の移動時間が60分増加すると同時に、電車の移動時間が60分減少した場合、各交通手段の選択の確率にどのように影響するか」を調べます。

変数 **time** を複製して新規変数 **newtime** を作成します。飛行機(**mode==1**)の移動時間を60分増加、電車(**mode==2**)の移動時間を60分減少させます。

```
generate newtime = time

replace newtime = time+60 if mode==1

replace newtime = time-60 if mode==2
```

newtime を使用して **margins** を実行します。飛行機と電車の移動時間の変化が同時に発生することを指定するために、**alternative(simultaneous)** オプションを使用します。

```
margins, at(time=generate(time)) at(time=generate(newtime))
alternative(simultaneous)
```

Predictive margins Number of obs = 840
 Model VCE : OIM

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

1._at : time = time

2._at : time = newtime

	Delta-method				[95% Conf. Interval]	
	Margin	Std. Err.	z	P> z		
_outcome#_at						
air#1	.2761905	.0277787	9.94	0.000	.2217453	.3306357
air#2	.224539	.0256826	8.74	0.000	.1742021	.274876
train#1	.3	.0284715	10.54	0.000	.2441968	.3558032
train#2	.3578287	.0328487	10.89	0.000	.2934464	.4222111
bus#1	.1428571	.0232375	6.15	0.000	.0973125	.1884018
bus#2	.1392549	.0228972	6.08	0.000	.0943773	.1841325
car#1	.2809524	.0280657	10.01	0.000	.2259446	.3359602
car#2	.2783773	.0281348	9.89	0.000	.2232342	.3335205

結果をグラフ化します。

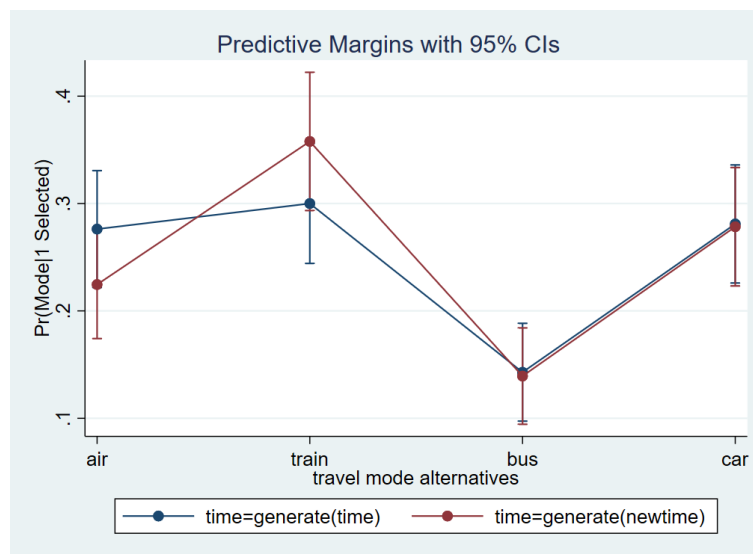
```
marginsplot, xdimension(_outcome)
```

Variables that uniquely identify margins: _atopt _outcome

Multiple at() options specified:

_atoption=1: time=generate(time)

_atoption=2: time=generate(newtime)



- 結果から、飛行機の選択予想確率は減少し、電車の選択予想確率は増加しました。
- 他の交通手段の選択予想確率にはあまり変化がありません。

これらの差が統計的に有意であるかを調べます。

```
margins, at(time=generate(time)) at(time=generate(newtime))
alternative(simultaneous) contrast(atcontrast(r) nowald effects)
```

```
Contrasts of predictive margins          Number of obs   =          840
Model VCE      : OIM
```

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

```
1._at       : time           = time
```

```
2._at       : time           = newtime
```

	Delta-method					[95% Conf. Interval]
	Contrast	Std. Err.	z	P> z		
_at@_outcome						
(2 vs 1) air	-.0516514	.0080654	-6.40	0.000	-.0674594	-.0358435
(2 vs 1) train	.0578287	.0109404	5.29	0.000	.0363859	.0792716
(2 vs 1) bus	-.0036023	.0022959	-1.57	0.117	-.0081021	.0008976
(2 vs 1) car	-.002575	.0036587	-0.70	0.482	-.009746	.0045959

- 結果から、飛行機および電車の差は統計的に有意です。
- 飛行機の選択予想確率は元の移動時間より 5.2 パーセントポイント減少しています。
- 電車の選択予想確率は元の移動時間より 5.7 パーセントポイント増加しています。