

margins - マージン/限界効果 【 評価版 】

マージンというのは、一部の共変量については特定の値に固定し、残りの共変量については平均化したり積分したりした上で、フィットされたモデルに基づき算出される予測値のことを言います。具体的に算出できる統計量には次のようなものがあります。

- estimated marginal means (最小 2 乗平均とも呼ばれる)
- adjusted predictions and adjusted treatment means
- predictive margins (潜在的アウトカム平均とも呼ばれる)
- average and conditional adjusted predictions
- average and conditional marginal effects (限界効果、弾力性)

1. マージンと限界効果

2. 応答変数のマージン

Example 1

Example 2

Example 3

Example 4

Example 5

Example 6

Example 7

Example 8

Example 9

Example 10

Example 11

Example 12

Example 13

Example 14

Example 15

3. 限界効果

3.1 dydx() オプション

3.2 at() オプション

3.3 弾力性

3.4 微分と差分

Example 16

Example 17

Example 18

4. サーベイデータの扱い

Example 19

Example 20

Example 21

5. 標準化マージン

6. asbalanced オプション

6.1 asbalanced によるバランシング

6.2 標準化によるバランシング

6.3 非線形応答のバランシング

6.4 一部の共変量のためのバランシング

6.5 fvset design の利用

6.6 空白セルの扱い

7. ネスト型デザイン

7.1 ネスト型デザインとは

7.2 asbalanced オプションの扱い

7.3 コード付けのスタイル

8. 特記事項

8.1 モデル式に関する要件

8.2 推定可能性

8.3 検定の操作可能性

8.4 estimates use コマンドとの併用

8.5 at() の構文

8.6 例外的な推定コマンド

9. Glossary

補足 1

補足 2

1. マージンと限界効果

`margins` コマンドは `predict` 等と同様の `postestimation` 機能であり、応答変数のマージン、及び応答変数の導関数のマージン — 限界効果 — を算出する機能を提供します。

- マージン (`margins`) – フィットされたモデルの一部、もしくはすべての共変量を固定したときに、そのモデルから導かれる統計量のことを言う。
- 限界効果 (`marginal effects`) – 応答変数の導関数に関するマージン。1 つの共変量の変化に対する応答変数の変化を表し、弾力性 (`elasticity`) と呼ばれることもある。

応答関数を $f(x)$ と書くとき、 $f(\cdot)$ の値を求めるのが前者の機能であり、 $f'(\cdot)$ の値を求めるのが後者の機能と言えます。`margins` の概要については Williams (2012) を参照ください。

2. 応答変数のマージン

応答変数のマージン (`margins of responses`) に対しては `predictive margins`, `adjusted predictions`, `recycled predictions` という表現が使われることがあります。またバランスの取れたデータの場合には `margins of responses` の代りに、周辺平均の推定値 (`estimated marginal means`) とか最小 2 乗平均 (`least-squares means`) という表現も用いられます。

マージンとは一部、もしくはすべての共変量の値を実際とは異なる値に固定したデータセットを対象にする形で、フィットされたモデルに基づき計算される統計量のことを言います。例えば男性と女性を対象に線形回帰モデルをフィットさせたとしたときに、男性についての周辺平均 (`marginal mean`) — すなわち平均値のマージン — というのは、すべての観測データを男性とみなしたときに得られる従属変数の予測される平均値のことを言います。従って実際に男性を表す観測データの他に女性を表す観測データも計算の対象に含まれることになります。女性についての周辺平均も同様に、すべての観測データが女性を表すものとして計算されます。

計算に際してはモデル中に現れる性別に関する項がすべて男性、または女性として扱われます。例えば

```
. regress y age bp i.sex sex#c.age sex#c.bp
```

といったモデルをフィットさせたとして、男性や女性についてのマージンを計算する場合、`margins` は `i.sex` という直接的な効果のみならず、`sex#c.age` や `sex#c.bp` といった間接的な効果についても考慮に入れて計算を行います。

マージンの計算対象となる応答値は `predict` によって算出される任意の統計量 (もしくはそれらを含む任意の数式) であって構いません。

デフォルトの場合、標準誤差はデルタ法によって算出されます。この手法の場合、共変量の評価を行う値は固定値であることが仮定されます。`svy` を用いるか否かによらず、標本がある集団を表す場合に、`vce(unconditional)` オプションを指定すれば、サンプリングに伴う共変量値の変動を考慮した標準誤差が算出されます。これを `predictive margins` と呼ぶ研究者もいます。

margins を理解する上での最良の策は用例を見て行く方法です。以下においては Example データセット margex.dta を用いて数々の用例を紹介します。

```
. use https://www.stata-press.com/data/r19/margex.dta *1
(Artificial data for margins)
```

変数に対してはラベル情報が付加されていないものもありますが、一応それらの一覧を掲載しておく次のようになります。

```
. describe *2
```

. describe				
Contains data from https://www.stata-press.com/data/r19/margex.dta				
Observations:	3,000	Artificial data for margins		
Variables:	11	27 Nov 2024 14:27		
Variable name	Storage type	Display format	Value label	Variable label
y	float	%6.1f		
outcome	byte	%2.0f		Outcome
sex	byte	%6.0f	sexlbl	Sex
group	byte	%2.0f		Group
age	byte	%3.0f		Age
distance	float	%6.2f		
ycn	float	%6.1f		
yc	byte	%6.1f		
treatment	byte	%2.0f		Treatment
agegroup	byte	%8.0g	agelab	Age group
arm	byte	%8.0g		Arm
Sorted by: group				

観測データ件数は 3,000 件であることが示されています。

*1 メニュー操作 : File > Example Datasets > Stata 19 manual datasets と操作、Base Reference Manual [R] の margins の項よりダウンロードする。

*2 メニュー操作 : Data > Describe data > Describe data in memory or in a file

▷ Example 1: 線形回帰

応答変数を y (連続変数)、独立変数を sex , $group$ (共にカテゴリ変数) とする簡単な線形回帰モデルをフィットさせた後、 sex についてのマージンを計算してみます。なお、 $i.$ 演算子については *mwp-028* をご参照ください。

```
. regress y i.sex i.group *3
```

. regress y i.sex i.group						
Source	SS	df	MS	Number of obs = 3,000		
Model	183866.077	3	61288.6923	F(3, 2996) = 152.06		
Residual	1207566.93	2,996	403.059723	Prob > F = 0.0000		
				R-squared = 0.1321		
				Adj R-squared = 0.1313		
Total	1391433.01	2,999	463.965657	Root MSE = 20.076		
y	Coefficient	Std. err.	t	P> t	[95% conf. interval]	
sex						
Female	18.32202	.8930951	20.52	0.000	16.57088	20.07316
group						
2	8.037615	.913769	8.80	0.000	6.245937	9.829293
3	18.63922	1.159503	16.08	0.000	16.36572	20.91272
_cons	53.32146	.9345465	57.06	0.000	51.48904	55.15388



$group$ については値ラベルが設定されていないのでコードの値がそのまま表示されていますが、 sex については値ラベルが表示に用いられている点に注意してください。`nofvlabel` オプションが指定された場合には値ラベルによる表示は抑止されます。

*3 メニュー操作：Statistics ▷ Linear models and related ▷ Linear regression

この状態で sex についてのマージンを計算してみます。

- Statistics ▸ Postestimation ▸ Marginal effects, estimated means, interactions, ...
 ▸ Custom and high-dimensional analyses ▸ Launch と操作
- Main タブ: Factor terms to compute margins for: sex

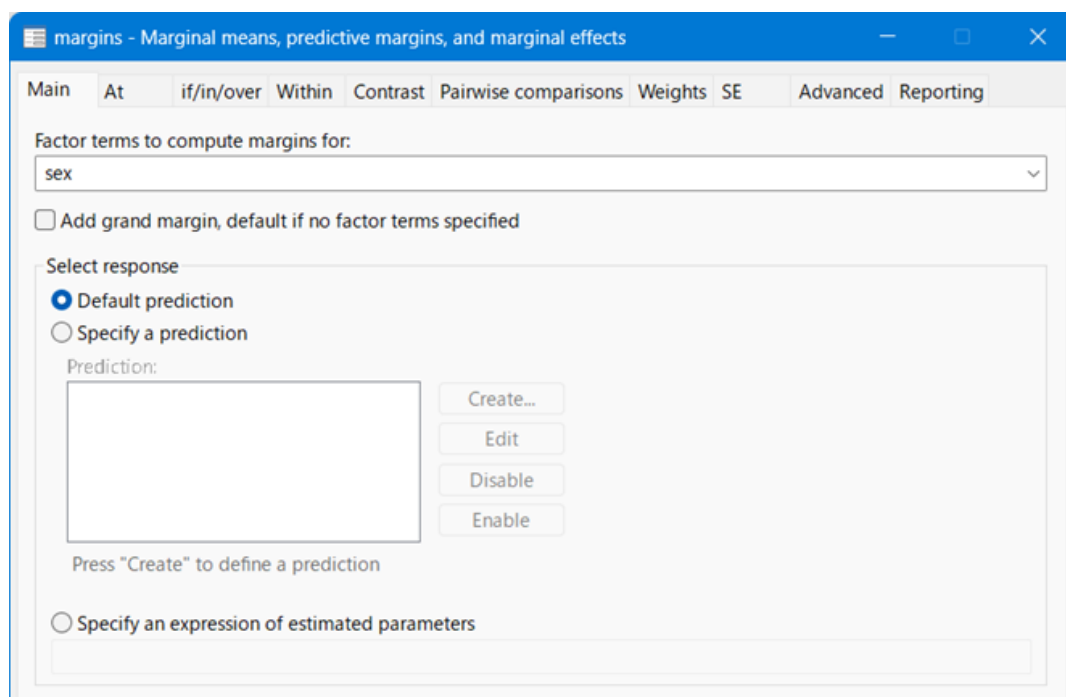


図 1 margins ダイアログ - Main タブ

```
. margins sex

Predictive margins                                Number of obs = 3,000
Model VCE: OLS

Expression: Linear prediction, predict()
```

	Delta-method					
	Margin	std. err.	t	P> t	[95% conf. interval]	
sex						
Male	60.56034	.5781782	104.74	0.000	59.42668	61.69401
Female	78.88236	.5772578	136.65	0.000	77.7505	80.01422

このデータセット中には男性 ($\text{sex} = 0$) のデータが 1,498 件、女性 ($\text{sex} = 1$) のデータが 1,502 件記録されています。マージンとして表示されている 60.56 という値は、3,000 件のデータがすべて男性であるとしたときにモデルで予測される y の値の平均値です。同様に 78.88 という値はすべてが女性であるとしたときにモデルで予測される y の値の平均値です。実際の男性のデータ 1,498 件を平均しても 60.56 という値は出てきませんし、女性のデータ 1,502 件を平均しても 78.88 という値は出てこない点に注意してください。

60.56 とか 78.88 というマージンの値を誘導するにはフィットされたモデル式に立ち返る必要があります。上記 `regress` コマンドの出力から導かれるモデル式は次のように記述できます^{*4}。実際の推定値ではなく `_b[]` という記法によって係数値の参照を行っている点にご注意ください。

$$\hat{y} = _b[1.\text{sex}] \cdot 1.\text{sex} + _b[2.\text{group}] \cdot 2.\text{group} + _b[3.\text{group}] \cdot 3.\text{group} + _b[_\text{cons}]$$

これに基づきすべてが男性 ($\text{sex} = 0$) であるとしたときのモデルによる予測値 y_0 を算出してみます。

```
. generate y0 = _b[1.sex]*0 + _b[2.group]*2.group + _b[3.group]*3.group + _b[_cons] *5
```

同様にすべてが女性 ($\text{sex} = 1$) であるとしたときの予測値 y_1 を算出します。

```
. generate y1 = _b[1.sex]*1 + _b[2.group]*2.group + _b[3.group]*3.group + _b[_cons]
```

これら y_0 と y_1 の平均値を求めてみます。

```
. summarize y0 y1 *6
```

. summarize y0 y1					
Variable	Obs	Mean	Std. dev.	Min	Max
y0	3,000	60.56034	7.126508	53.32146	71.96069
y1	3,000	78.88236	7.12651	71.64348	90.28271

`margins` コマンドによる出力が誘導できていることが確認できました。

◀

▷ Example 2: ロジスティック回帰

評価版では割愛しています。

▷ Example 3: atmeans オプション

評価版では割愛しています。

▷ Example 4: 複数マージン

評価版では割愛しています。

^{*4} `regress` 出力中の `female` というのは値ラベルであり、その実体は `1.sex` です。

^{*5} メニュー操作：Data ▷ Create or change data ▷ Create new variable

^{*6} メニュー操作：Statistics ▷ Summaries, tables and tests ▷ Summary and descriptive statistics ▷ Summary statistics

▷ Example 5: 交互作用項を含むモデル

評価版では割愛しています。

▷ Example 6: 連続変数を含むモデル

評価版では割愛しています。

▷ Example 7: 連続変数のマージン

評価版では割愛しています。

▷ Example 8: 交互作用のマージン

評価版では割愛しています。

▷ Example 9: マージンの分解

評価版では割愛しています。

▷ Example 10: マージンの対比

評価版では割愛しています。

▷ Example 11: predict() オプション

評価版では割愛しています。

▷ Example 12: 数式の指定

評価版では割愛しています。

▷ Example 13: 複数応答に対するマージン

評価版では割愛しています。

▷ Example 14: 複数方程式モデル

評価版では割愛しています。

▷ Example 15: サンプル外予測

評価版では割愛しています。

3. 限界効果

評価版では割愛しています。

4. サーベイデータの扱い

評価版では割愛しています。

5. 標準化マージン

評価版では割愛しています。

6. asbalanced オプション

評価版では割愛しています。

7. ネスト型デザイン

評価版では割愛しています。

8. 特記事項

評価版では割愛しています。

補足 1 – Example 7 に対する補足

評価版では割愛しています。

補足 2 – over() オプションの意味

評価版では割愛しています。

