

makespline - スプラインの生成 【 評価版 】

makespline コマンドは既存の変数からスプライン基底関数を生成します。

1. 機能概要

Example 1

Example 2

Example 3

Example 4

補足 1

補足 2

補足 3

1. 機能概要

makespline コマンドは既存の変数からスプライン関数を構成する新変数を生成します。スプライン関数としては次の 3 種類に対応しています。

- B スプライン (B-spline)
- 区分的多項式スプライン (piecewise polynomial spline)
- 制限付き 3 次スプライン (restricted cubic spline)

これらのスプラインは元々のデータを平滑な曲線によって近似しようとするものです。



データを平滑化したスプライン曲線のプロットが目的なのであれば `twoway mspline` コマンドの使用を推奨します。補足 1 を参照ください。makespline コマンドによって生成された基底関数を利用してスプライン曲線をプロットする方法は後述しますが、少々テクニックを要します。

コマンド構文上は

```
. makespline basis varlist ...
```

という形式を取るわけですが、このうち *basis* は基底関数 (basis functions) の種別を表すもので、次の 4 種類が指定できます。

<i>basis</i>	基底関数
<code>bspline</code>	B スプライン
<code>piecewise</code>	区分的多項式スプライン
<code>rcs</code>	制限付き 3 次スプライン
<code>linear</code>	線形スプライン

評価版では割愛しています。

▷ Example 1: B スプライン

評価版では割愛しています。

▷ Example 2: 区分的多項式スプライン

評価版では割愛しています。

▷ Example 3: セミパラメトリックな推定

評価版では割愛しています。

▷ Example 4: スプラインの描画

ここでは線形スプラインを用いたときの推定結果をどのようにしてグラフ化するかを紹介します。まず使用する Example データセットを `auto.dta` に戻します。

```
. sysuse auto.dta, clear
(1978 automobile data)
```

今、価格 `price` と燃費 `mpg` の間にどのような関係があるかについて分析を行うものとします。その際、ノット数 3 の線形スプラインを用いて基底関数の生成を行います。この場合、`mpg` の 25, 50, 75 パーセンタイルの値がノットの位置となります。

```
. summarize mpg, detail
```

Mileage (mpg)					
Percentiles		Smallest			
1%	12	12			
5%	14	12			
10%	14	14	Obs		74
25%	18	14	Sum of wgt.		74
50%	20		Mean		21.2973
			Std. dev.		5.785503
		Largest			
75%	25	34			
90%	29	35	Variance		33.47205
95%	34	35	Skewness		.9487176
99%	41	41	Kurtosis		3.975005

`summarize` の出力からすると、区分的多項式スプラインの境界が `mpg = 18, 20, 25` の位置に置かれることがわかります。

それでは `makespline linear` コマンドによって基底関数の生成を行います。その際、基底関数用の `stub` としては `mpg` を指定することにします。

```
. makespline linear mpg, knots(3) basis(mpg)
```

この操作により次の3つの基底項が生成されます。なお、makespline linear の場合、rescaling は行われな
い点に注意してください。

```
. describe mpg_*
```

Variable name	Storage type	Display format	Value label	Variable label
mpg_1_1	double	%10.0g		Piecewise polynomial basis term 1 for mpg
mpg_1_2	double	%10.0g		Piecewise polynomial basis term 2 for mpg
mpg_1_3	double	%10.0g		Piecewise polynomial basis term 3 for mpg

この状態で regress による推定を行います。共変量としては mpg と mpg_* の他に国産車か外国車の別を示す
0/1 変数 i.foreign を指定することになります。

```
. regress price mpg mpg_* i.foreign
```

Source	SS	df	MS	Number of obs	=	74
Model	316201619	5	63240323.8	F(5, 68)	=	13.49
Residual	318863777	68	4689173.19	Prob > F	=	0.0000
Total	635065396	73	8699525.97	R-squared	=	0.4979
				Adj R-squared	=	0.4610
				Root MSE	=	2165.4

price	Coefficient	Std. err.	t	P> t	[95% conf. interval]
mpg	-1330.299	213.0425	-6.24	0.000	-1755.419 -905.1798
mpg_1_1	1698.953	622.5153	2.73	0.008	456.7432 2941.163
mpg_1_2	-622.9298	651.5686	-0.96	0.342	-1923.115 677.2551
mpg_1_3	139.4188	277.0762	0.50	0.616	-413.4783 692.3158
foreign					
Foreign	1676.381	609.4723	2.75	0.008	460.1983 2892.564
_cons	28796.47	3449.408	8.35	0.000	21913.28 35679.65

フィットされた回帰直線は次のように表現できます。

```
generate double xb = _b[_cons] + _b[1.foreign]*foreign + ///
                    mpg*_b[mpg] + (mpg>18)*(mpg-18)*_b[mpg_1_1] + ///
                    (mpg>20)*(mpg-20)*_b[mpg_1_2] + ///
                    (mpg>25)*(mpg-25)*_b[mpg_1_3]
```

mpg の効果はノットの位置で次のように変化します。なお、`_b[]` は推定された係数値を意味する点に注意してください。

- $mpg \leq 18$ の場合、その効果は `_b[mpg]` のみ。
- $18 < mpg \leq 20$ の場合、その効果は `(_b[mpg] + _b[mpg_1_1])`。
- $20 < mpg \leq 25$ の場合、その効果は `(_b[mpg] + _b[mpg_1_1] + _b[mpg_1_2])`。
- $25 < mpg$ の場合、その効果は `(_b[mpg] + _b[mpg_1_1] + _b[mpg_1_2] + _b[mpg_1_3])`。

国産車が外国車かによって分けたときの予測値は次のようなコマンドによって表現できます。

```
. generate xb_domestic = _b[_cons] + mpg*_b[mpg]
> + (mpg>18)*(mpg-18)*_b[mpg_1_1]
> + (mpg>20)*(mpg-20)*_b[mpg_1_2]
> + (mpg>25)*(mpg-25)*_b[mpg_1_3]

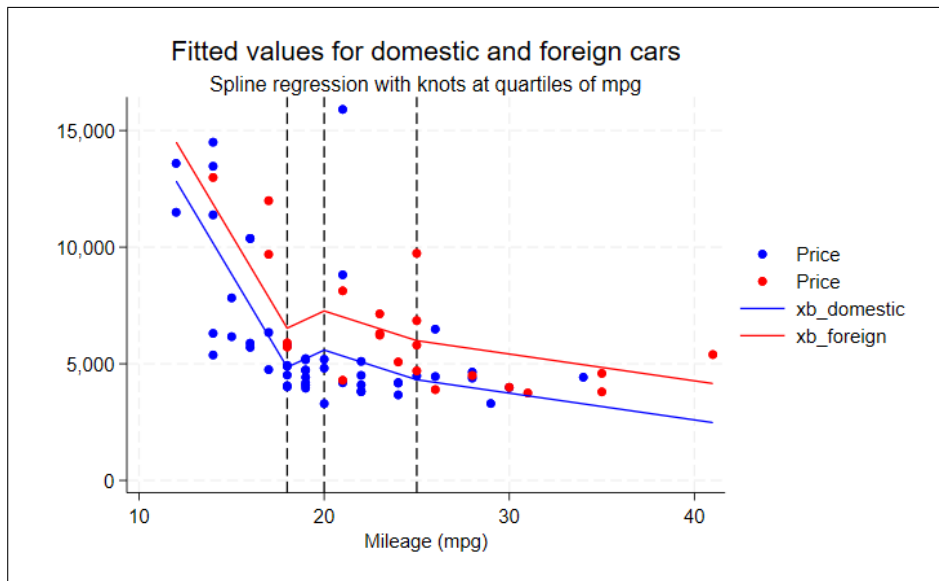
. generate xb_foreign = _b[_cons] + _b[1.foreign] + mpg*_b[mpg]
> + (mpg>18)*(mpg-18)*_b[mpg_1_1]
> + (mpg>20)*(mpg-20)*_b[mpg_1_2]
> + (mpg>25)*(mpg-25)*_b[mpg_1_3]
```

これらを使ってグラフを作成すると次のようになります。mpg の効果が区分ごとにどう変化するかが示されています。なお、このグラフは

- 実際に観測された点の分布を示す scatter plot と
- モデルによって予測される点を直線で結んだ line plot とを

重畳させる形で作成したものです。ダイアログ操作については補足 3 を参照ください。

```
. twoway (scatter price mpg if foreign==0, mcolor(blue))
> (scatter price mpg if foreign==1, mcolor(red))
> (line xb_domestic mpg, sort lcolor(blue))
> (line xb_foreign mpg, sort lcolor(red)),
> xline(18 20 25)
> title("Fitted values for domestic and foreign cars")
> subtitle("Spline regression with knots at quartiles of mpg")
```



この用例では線形スプラインを用いたときのグラフ作成方法が示されているわけですが、3次多項式スプライン使用時の例については補足2を参照ください。 <

補足 1 – twoway mspline コマンド

評価版では割愛しています。

補足 2 – predict を用いた描画

評価版では割愛しています。

補足 3 – ダイアログ操作

評価版では割愛しています。

