第三回　重回帰・ロジスティック・Cox ハザードでのスプライン曲線の作図

mkspline で変数を変換してスプライン回帰をおこなう際，変換した変数以外にも独立変数をモデルに投入したいわゆる重回帰をおこなうと，第二回までにお伝えした内容ではスプライン曲線の作図が容易ではありません．本書では，xblc というユーザ作成の ado ファイルを用いて，この重回帰のスプライン曲線を作図する方法をお伝えします．

xblc のインストール

Stata を立ち上げ，コマンドウィンドウから以下を実行してください．

```
search xblc
```

search コマンドでいくつかの主要な Stata 関連サイトを検索し，xblc コマンドを見つけます．成功すると，次のような画面が表示されます．

```
---------------------------------------------------------------------------
search for xblc                                          (manual:  [R] search)
---------------------------------------------------------------------------

Search of official help files, FAQs, Examples, SJs, and STBs

SJ-11-3 st0215_1  .  Tab. and plot results after flex. modeling of quant. cov.
        (help xblc if installed)  . . . . . . . . . N. Orsini and S. Greenland
        Q3/11   SJ 11(3):478
        bug fix and added graph options for xblc

SJ-11-1 st0215   Tabulate and plot results after flex. modeling of quant. cov.
        (help xblc if installed)  . . . . . . . . . N. Orsini and S. Greenland
        Q1/11   SJ 11(1):1--29
        provides a postestimation command that facilitates the
        presentation of the association between a quantitative
        covariate and the response variable
```

st0215_1 をクリックし，さらに (click here to install) をクリックするとダウンロードおよびインストールを開始します．処理は 10 秒以下で完了することと思います．次のような表示が出れば成功です．

```
installation complete.
---------------------------------------------------------------------------

(click here to return to the previous screen)
```

正常にインストールできた方は，次のコマンドで xblc のバージョンを確認します．

```
. which xblc
c:\ado\plus\x\xblc.ado
*! N.Orsini v.1.0.1 12jul11
*! N.Orsini v.1.0.0 19apr10
```

例題データ

例として用いるデータは第一回，第二回と同じものです．

```
webuse mksp1, clear
```

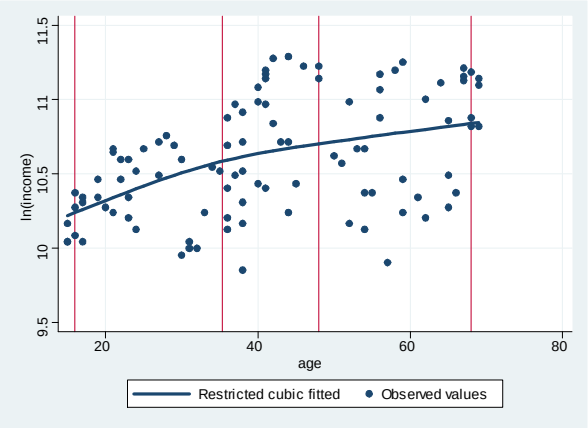
- 変数：
  - id: 調査への参加者の ID
  - age: 参加者の年齢
  - educ: 教育期間（年）
  - income: 所得（米ドル/年）
  - lninc: income の対数

lninc を従属変数，age を独立変数とする制約 3 次スプライン回帰は，以下で実施できます．

```
mkspline agesp=age, cubic nknot(4)
regress lninc agesp*
```

スプライン曲線の作図は，以下のようなコマンドを do ファイルに記述して実行します．

```
predict lninchat, xb
sort age
twoway line lninchat age, clcolor(navy) clwidth(thick) ///
|| scatter lninc age, mcolor(navy) ///
ytitle("ln(income)") ///
legend(label(1 "Restricted cubic fitted") ///
label(2 "Observed values")) ///
xline(16 35.35 48 68) xlabel(, grid)
```



LightStone® 株式会社ライトストーン

予測値の信頼区間の計算

`xblc` では，信頼区間を次のように計算します．まず，従属変数を Y，`mkspline` で展開した後の独立変数を X_1，X_2，X_3 と，`mkspline` で展開しなかった独立変数を Z_4，Z_5 として，重回帰のモデル式を次のように表すことにします．

$$Y = \alpha + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 Z_4 + \beta_5 Z_5 + u \tag{1}$$

`mkspline` で展開しなかった独立変数 Z_4，Z_5 をその平均値 $\bar{Z}_4$，$\bar{Z}_5$ で差し引き，いわゆるセンタリング（中心化）という変数変換を行って，Z_{4c}，Z_{5c} を便宜的に求めます．

$$\begin{aligned} Z_{4c} &= Z_4 - \bar{Z}_4 \\ Z_{5c} &= Z_5 - \bar{Z}_5 \end{aligned} \tag{2}$$

Z_4，Z_5 の代わりに Z_{4c}，Z_{5c} を用いると，モデル式は次のようになります．

$$\begin{aligned} Y &= \alpha^* + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 Z_{4c} + \beta_5 Z_{5c} + u \\ \alpha^* &= \alpha - \beta_4 \bar{Z}_4 - \beta_5 \bar{Z}_5 \end{aligned} \tag{3}$$

この係数 α^*，β_1，β_2，β_3，β_4，β_5 について，重回帰による推定結果を $\hat{\alpha}^*$，$\hat{\beta}_1$，$\hat{\beta}_2$，$\hat{\beta}_3$，$\hat{\beta}_4$，$\hat{\beta}_5$ とします．

$$Y = \hat{\alpha}^* + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2 + \hat{\beta}_3 X_3 + \hat{\beta}_4 Z_{4c} + \hat{\beta}_5 Z_{5c} + u \tag{4}$$

それぞれの独立変数が，特定の値 x_1，x_2，x_3，z_4，z_5 を取ったとき，予測値 $\hat{y}$ は次で書けます．

$$\hat{y} = \hat{\alpha}^* + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 z_{4c} + \hat{\beta}_5 z_{5c} \tag{5}$$

誤差分布に正規分布を仮定し，$\hat{\alpha}^*$，$\hat{\beta}_1$，$\hat{\beta}_2$，$\hat{\beta}_3$，$\hat{\beta}_4$，$\hat{\beta}_5$ が独立の分布に従うとすると，$\hat{y}$ の分散 $V(\hat{y})$ について，次のようにできます．

$$\begin{aligned} V(\hat{y}) &= V\left(\hat{\alpha}^* + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3 + \hat{\beta}_4 z_4 + \hat{\beta}_5 z_5\right) \\ &= V(\hat{\alpha}^*) + V\left(\hat{\beta}_1 x_1\right) + V\left(\hat{\beta}_2 x_2\right) + V\left(\hat{\beta}_3 x_3\right) + V\left(\hat{\beta}_4 z_4\right) + V\left(\hat{\beta}_5 z_5\right) \\ &= V(\hat{\alpha}^*) + x_1 V\left(\hat{\beta}_1\right) + x_2 V\left(\hat{\beta}_2\right) + x_3 V\left(\hat{\beta}_3\right) + z_4 V\left(\hat{\beta}_4\right) + z_5 V\left(\hat{\beta}_5\right) \end{aligned}$$

とくに z_{4c}，z_{5c} がゼロのとき，式 (2) より Z_4，Z_5 それぞれが平均値をとるときの $V(\hat{y})$ を求めることになり，次のように簡単になります．

$$V(\hat{y}) = V(\hat{\alpha}) + x_1 V\left(\hat{\beta}_1\right) + x_2 V\left(\hat{\beta}_2\right) + x_3 V\left(\hat{\beta}_3\right) \tag{6}$$

LightStone® 株式会社ライトストーン

$\hat{y}$ の 95% 信頼区間は次のようになります．

$$\hat{y} \pm z_{0.025}\sqrt{V(\hat{y})} = \hat{y} \pm z_{0.025}\sqrt{V(\hat{\alpha}) + x_1 V\left(\hat{\beta}_1\right) + x_2 V\left(\hat{\beta}_2\right) + x_3 V\left(\hat{\beta}_3\right)}$$

X_1，X_2，X_3 について，mkspline で展開する前の変数を X とすると，ある値 $X = c$ に対応する X_1，X_2，X_3 それぞれの値 c_1，c_2，c_3 を x_1，x_2，x_3 に選べば，スプライン曲線の横軸 X について，$X = c$ での信頼区間が上式から求まります．上式の計算について，Stata では lincom を用いて簡単に結果を求めることができます．

重回帰分析でのスプライン曲線の作図

まず，重回帰分析を実施します．新たに educ を独立変数に加え，センタリングを行ったうえで回帰をおこないます[1]．

```
sum educ
generate educ_c = educ - r(mean)
regress lninc agesp* educ_c
```

次に，xblc を利用してスプライン曲線を作図したいのですが，xblc ではスプライン曲線の横軸にする age の各観測値をコマンド内で指定する必要があります．このため，levelsof という Stata に標準搭載のコマンドを使用して各観測点の指定を手短かにおこないます．

```
. levelsof age
15 16 17 19 20 21 22 23 24 25 27 28 29 30 31 32 33 34 35 36 37 38
>  40 41 42 43 44 45 46 48 50 51 52 53 54 55 56 57 58 59 61 62 64
>  65 66 67 68 69
```

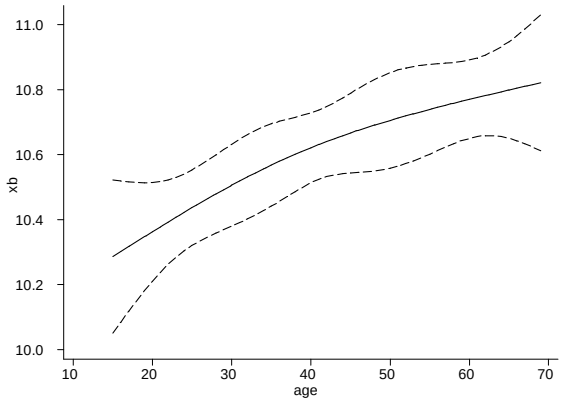
表示された各観測値は，r(levels) というマクロ変数に格納されています．この状態で xblc を実行します．

```
xblc agesp*, covname(age) at(`r(levels)') line
```

上記コマンドについて詳述します，コマンド名 xblc に続き，regress で使用した変数 agesp*を入力します．カンマ（,）を入力し，covname() オプションを入力して，カッコ内に本来の観測値である変数 age を入力します．at() オプションを入力し，age の各観測値を入力するのですが，r(levels) に格納されているのでシングルクォーテーションでくくって at() のカッコ内に記入します．ここで用いるシングルクォーテーションは Stata の特別なルールに従った記号を使用します．シングルクォーテーションの開始は **Shift** + **@**キーで入力できる記号を使用します．シングルクォーテーションの終了は **Shift** + **7** キーで入力できる記号を使用します．line オプションを入力するとスプライン曲線が作図されます．

[1]センタリングをするのは，=0 で平均値を示すようにするためです．スプライン曲線で 2 軸に表さない独立変数（または共変量）は，連続変数の場合その値がゼロのとき，またカテゴリ変数の場合その値が最小値のときのグラフがそれぞれ描かれます．すなわち，連続変数については，センタリング後の変数を用いることで，センタリング前の変数の平均値でのグラフを描くことができます．同様に，カテゴリ変数については、推定コマンド（logit など）で ib#. を用いて基準（ベース）を指定することで，その値でのグラフを描くことができます．

LightStone® 株式会社ライトストーン



levelsof コマンドで指定する変数が小数を含み，且つ double 型（Stata の倍精度型）の場合，その後の xblc で値が存在しないという旨のエラーが起きてしまいます．おそらく xblc コマンドが float 型（Stata の単精度型）にしか対応していないことが原因だと考えられます．この場合，double 型で保存されている値を別の変数に退避させ，元の変数のほうは float 型に型変換をおこなってから levelsof および xblc を実行すると成功します．

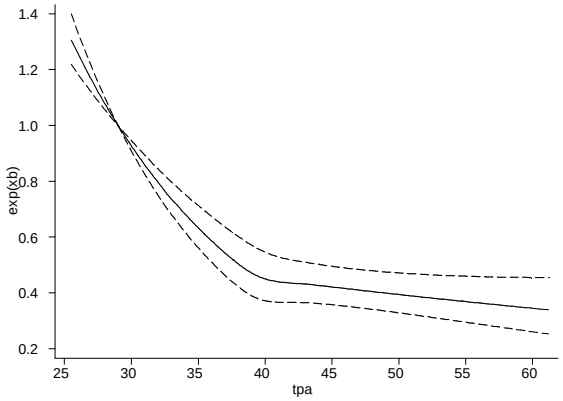
```
generate age_double = age
recast float age, force
levelsof age
xblc agesp*, covname(age) at(`r(levels)') line
```


LightStone® 株式会社ライトストーン

ロジスティック回帰でのスプライン曲線の作図

ロジスティック回帰ではオッズ比のスプライン曲線を作図することになると思います．オッズ比のプロット時は xblc コマンドで eform オプションを入力します．オッズ比は基準のオッズが必要になりますので，reference() オプションで基準となる観測点を指定します．基準となる観測点はデータセット内に存在している必要があります．横軸の値をカッコ内に入力して指定します[2]．

```
use http://nicolaorsini.altervista.org/data/pa_luts, clear
quietly summarize age
generate agec = age - r(mean)
mkspline tpas = tpa, knots(37.2 39.6 42.3 45.6) cubic
logit ipss2 tpas1 tpas2 tpas3 agec
levelsof tpa
xblc tpas1-tpas3, covname(tpa) at(`r(levels)') reference(29) eform line
```



Cox 比例ハザードモデルでのスプライン曲線の作図

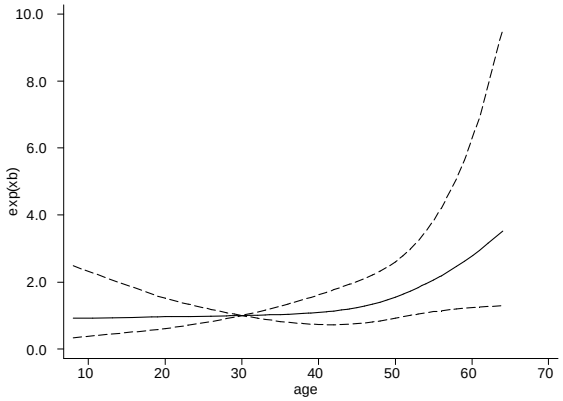
Cox 比例ハザードモデルでは，前述のロジスティック回帰と同様の方法で作図できます．通常は，ハザード比のスプライン曲線を作図することになると思います．ハザード比のプロット時は xblc コマンドで eform オプションを入力します．ハザード比は基準のハザードが必要になりますので，reference() オプションで基準となる観測点を指定します．基準となる観測点はデータセット内に存在している必要があります．横軸の値をカッコ内に入力して指定します[3]．

```
webuse stan3, clear
stset
```

[2]以下の例は xblc のヘルプにある Example のコマンドを少し修正したものです
[3]以下の例は stcox のヘルプにある Example のコマンドを少し修正したものです

LightStone® 株式会社ライトストーン

```
mkspline ages = age, nknot(3) cubic
quietly sum year
generate year_c = year - r(mean)
stcox ages* year_c
levelsof age
xblc ages*, covname(age) at(`r(levels)') reference(30) eform line
```



参考文献

[1] 『*A procedure to tabulate and plot results afterflexible modeling of a quantitative covariate*』, N. Orsini and S. Greenland, The Stata Journal volume 11 number 1 pp.1-29, Stata Press, 2011 (http://www.stata-journal.com/sjpdf.html?articlenum=st0215)

株式会社　ライトストーン

2017 年 2 月

（改訂）2017 年 10 月

LightStone® 株式会社ライトストーン

研究者向けの統計解析ソフトウェア **Stata® 評価版のお申込み**

機能制限なし！30日間無料で使える評価版でまずはお試しください！

https://www.lightstone.co.jp/stata/evaluate.html

※学生の方への評価版の提供はございません。大学の講義で使用するなど単位取得に関わる形でのご利用となる場合は学生版のご購入をご検討ください。

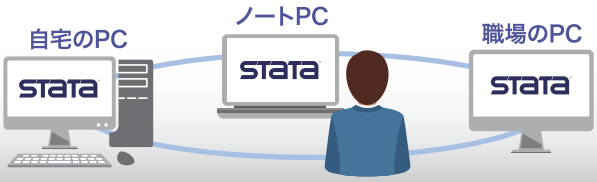
ライセンスはサブスクリプションがおすすめ！

サブスクリプションライセンスを選ぶメリット

- 常にStataの最新バージョンが利用できます
- 技術サポートの対象となります
- 毎年の経費としてStataを導入できます
- 利用者の増減に柔軟に対応可能
- 初期導入費を抑えたい場合にもおすすめ

シングルライセンス（サブスクリプション）

職場のPC・ノートPC・自宅のPC等、個人所有のPCにインストール可能です。
同時に使用できるのは1台のみとなります。



マルチユーザライセンス（サブスクリプション）　2ユーザ以上

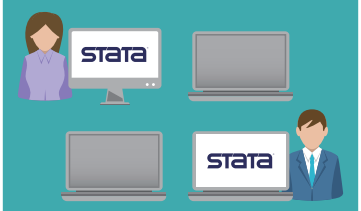
ボリュームライセンス

シングルライセンスを複数人でまとめてご購入いただく際のボリュームディスカウント商品です。

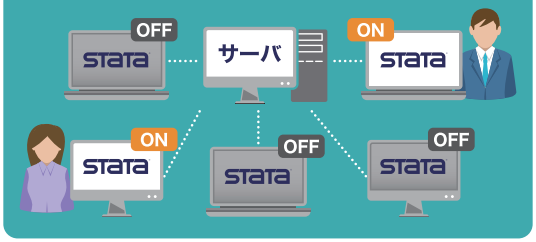
同時起動ライセンス

ご利用環境により下記3通りの運用方法からご選択いただけます。

PC毎にスタンドアロンインストールをして運用
例:2ライセンスの場合



クライアント＆サーバ環境によるネットワーク運用
例:2ライセンスの場合



1台の仮想マシンにユーザがアクセスする運用
例:2ライセンスの場合

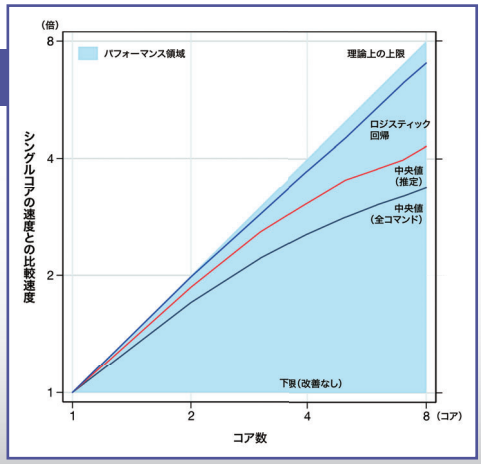
＼データ処理時間の削減ならMP／

Stata／MPの演算能力

Stata MPはPCの持つマルチコアの特性を活かして、処理を分散・並列する機能を備えます。およそ85%以上のコマンドで処理速度が向上し、コア数に応じた計算時間の短縮が期待できます。コマンドごとの処理速度の向上の度合いについては以下のページや資料をご覧ください。

コア数	全てのコマンド	推定コマンド	ロジスティック回帰
2	1.7倍	1.8倍	1.9倍
4	2.6倍	3.1倍	3.8倍
8	3.3倍	4.2倍	6.8倍

弊社Webページ内の『Stata／MP』
https://www.lightstone.co.jp/stata/statamp.html



株式会社 ライトストーン
〒101−0031
東京都千代田区東神田 2−5−12　龍角散ビル 7F
TEL：03−3864−5211　e-mail：sales@lightstone.co.jp
FAX：03−3865−0050　WEB：https://www.lightstone.co.jp/