

bayes

Lu Jiawei

1/30/2020

1)実データ、例えば

<https://cran.r-project.org/web/packages/Ecdat/Ecdat.pdf> (<https://cran.r-project.org/web/packages/Ecdat/Ecdat.pdf>)

にあるデータなどを用い、

・プロビット回帰 ・トービット回帰

プロビット回帰モデルで、分析するのですが、同じく出力変数（被説明変数）のロジスティックモデルとの違いについては、判別したい2値について、説明変数のとりうる値の分布が別々の場合はロジットを使う場合が良いのに対して、説明変数のとりうる値の分布が共通の場合プロビット・モデルを使った方が良い。

この考え方のもとで、データを選ぶ。

Panel Datas of Individual Wages

というデータの正社員年数、既婚者と教育年数を選択する。

既婚者か否かを非説明変数、正社員年数と教育年数を説明変数にする。

まず正社員年数と教育年数をそれぞれ正規分布にしたがっているか検定する。

```
require(Ecdat)
```

```
## Loading required package: Ecdat
```

```
## Loading required package: Ecfun
```

```
##  
## Attaching package: 'Ecfun'
```

```
## The following object is masked from 'package:base':  
##  
##      sign
```

```
##  
## Attaching package: 'Ecdat'
```

```
## The following object is masked from 'package:datasets':  
##  
##      Orange
```

```
data(Wages)  
shapiro.test(Wages$ed)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  Wages$ed  
## W = 0.92426, p-value < 2.2e-16
```

```
shapiro.test(Wages$exp)
```

```
##  
##  Shapiro-Wilk normality test  
##  
## data:  Wages$exp  
## W = 0.95232, p-value < 2.2e-16
```

両方ともp-value < 2.2e-16なので、二つの変数共に正規分布に仕上がっていることがわかった。

```
df <- data.frame("ed" = Wages$ed,  
                  "exp" = Wages$exp, "married" = Wages$married)  
head(df)
```

```
##      ed exp married  
## 1    9   3      yes  
## 2    9   4      yes  
## 3    9   5      yes  
## 4    9   6      yes  
## 5    9   7      yes  
## 6    9   8      yes
```

```
myprobit <- glm(married ~ ed+exp, family = binomial(link = "probit"),
               data = df)
summary(myprobit)
```

```
##
## Call:
## glm(formula = married ~ ed + exp, family = binomial(link = "probit"),
##      data = df)
##
## Deviance Residuals:
##      Min       1Q   Median       3Q      Max
## -2.3827   0.4476   0.5736   0.7106   0.8887
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  0.272744   0.127076   2.146   0.0318 *
## ed           0.014563   0.008520   1.709   0.0874 .
## exp          0.023273   0.002255  10.319  <2e-16 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for binomial family taken to be 1)
##
##      Null deviance: 3996.5  on 4164  degrees of freedom
## Residual deviance: 3882.2  on 4162  degrees of freedom
## AIC: 3888.2
##
## Number of Fisher Scoring iterations: 4
```

考察:

正社員の年数だけ有意(5%)、という結果になった。

プロビット回帰の結果によると、教育程度と、結婚率の関係は有意ではない(5%)ことがわかった。結婚率は、教育程度よりも、正社員の年数に強く影響されていることがわかった。