

Rによるミクロ計量経済分析

5 ポアソン回帰モデル

この章では R によるポアソン回帰モデルの推定の仕方について説明します。

5.1 データの読み込み

推定に用いるデータとして、ゼミのページにある “fertil.csv” というファイルのデータを用います。これは、1972年の156人のアメリカの女性のデータで、Wooldridge (2010)で使われたものです。¹ このファイルを開くと、以下のようなデータがあります。

```
# Data on 156 women in U.S.
# kids: number of children ever born
# educ: years of education of the woman
# age: age of the woman
# agesq: square of age
# black: it is 1 if the woman is a black, and 0 otherwise
kids  edu  age  agesq  black
  4   12   48   2304    0
  3   17   46   2116    0
....
```

ここで分析したいのは、子供の数と女性の教育年数との関係です。この分析のために、ここではポアソン回帰分析を行います。このポアソン回帰分析では、子供の数を被説明変数、このファイルにおけるその他の変数を説明変数とします(各説明変数についてはデータのファイルを参照して下さい)。作業ファイル(Working directory)を “fertile.csv” が保存されているフォルダ(ディレクトリー)に変更した後、以下のように打ち込んで、ファイルのデータをRに読み込ませてください。

```
> fertil = read.csv("fertil.csv", header=T, skip=6)
```

データの最初の7行を見るために、以下のように打ち込んでください。

```
> head(fertil, 7)
```

出力結果は以下のようになります。

```
   kids educ age agesq black
1     4   12  48   2304     0
2     3   17  46   2116     0
3     2   12  53   2809     0
4     2   12  42   1764     0
5     2   12  51   2601     0
6     4    8  50   2500     0
7     0   12  47   2209     0
```

¹ Wooldridge, J.M. (2010) *Econometric Analysis of Cross Section and Panel Data*, 2nd edition. MIT Press, Cambridge, MA.

5.2 ポアソン回帰モデルの推定

ポアソン回帰モデルは以下のようなモデルです。

$$y_i | \mathbf{x}_i \sim Po(m(\mathbf{x}_i, \boldsymbol{\beta})), \quad m(\mathbf{x}_i, \boldsymbol{\beta}) = \exp(\mathbf{x}_i' \boldsymbol{\beta}),$$

ここで y_i は(スカラーの)被説明変数、 $\mathbf{x}_i$ は説明変数ベクトル、 $Po(\lambda)$ は期待値 λ のポアソン分布を表します。ここでの分析では、 $y_i = \text{kids}_i$ 、 $\mathbf{x}_i = [\text{educ}_i, \text{age}_i, \text{agesq}_i, \text{black}_i]'$ となります。ここで、例えば kids_i は変数“kids”の i 番目のデータのことです。このモデルにおいて $E(y_i | \mathbf{x}_i) = \exp(\mathbf{x}_i' \boldsymbol{\beta})$ が成り立ちます。

このモデルを推定するために、R の `glm()` 関数を用います。ここで、ひとまずは、モデルの定式化は正しいと仮定します。つまり、 y_i は実際に上記のモデルに従うと仮定します。まずはこの仮定の下で、モデルに含まれるパラメーターを最尤推定法によって推定します(後ほどこの仮定を緩め、疑似最尤推定法による推定を考えます)。以下のように打ち込んでください。

```
> Poiresult=glm(kids~educ+age+agesq+black,family=poisson(link="log"),data=fertil)
```

推定結果を見るために、`summary()` 関数を用います。以下のように打ち込んでください。

```
> summary(Poiresult)
```

出力結果は以下のようになります。

Call:

```
glm(formula=kids ~ educ + age + agesq + black, family = poisson(link = "log"),
     data = fertil)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-2.8131	-0.7750	-0.1384	0.6210	2.0926

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)
(Intercept)	-2.958415	3.223340	-0.918	0.359
educ	- 0.014821	0.020585	-0.720	0.472
age	0.213878	0.145489	1.470	0.142
agesq	-0.002632	0.001643	-1.602	0.109
black	0.172169	0.157581	1.093	0.275

(Dispersion parameter for poisson family taken to be 1)

Null deviance: 210.70 on 155 degrees of freedom
Residual deviance: 201.92 on 151 degrees of freedom
AIC: 630.2

Number of Fisher Scoring iterations: 5

この出力結果より、このモデルの推定において、p 値は全て 0.1 より大きいことがわかります。これは、このモデルにおいて、10%有意となる説明変数は一つもないということです。つまりすべての説明変数が説明力がほとんどないことになります。

5.3 当てはめ値に計算

Next, we calculate the fitted values, or the expected values of y_i 's with estimated β . 次に当てはめ値、すなわち β の推定値を用いて計算した y_i の期待値、を計算します。次のように打ち込んでください。

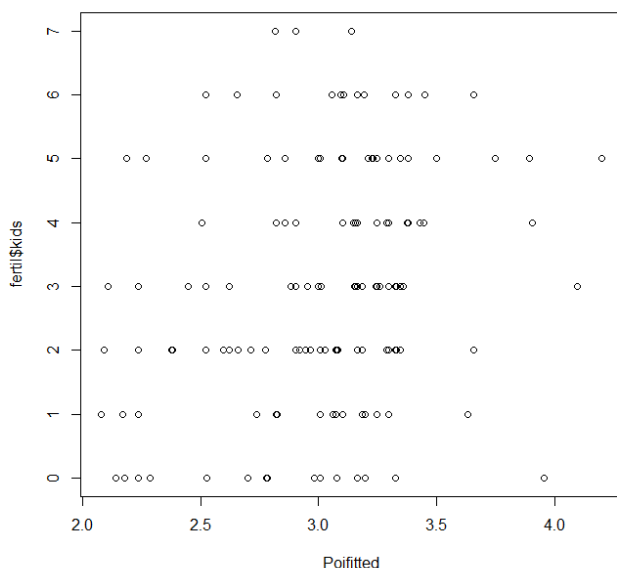
```
> Poifitted=Poiresult$fitted
> Poifitted
      1      2      3      4      5      6      7      8      9     10
2.901136 2.880920 2.237127 3.330972 2.521742 2.818531 3.008046 3.195860 3.346333 3.164206
     11     12     13     14     15     16     17     18     19     20
3.071789 3.373985 2.783334 2.270530 3.007639 3.227254 3.894965 3.153716 3.346333 3.078326
      :      :      :      :      :      :      :      :      :      :
    141    142    143    144    145    146    147    148    149    150
2.902673 2.901136 2.521742 3.431186 3.380707 2.999933 2.982072 3.346333 3.324348 3.056876
    151    152    153    154    155    156
2.944454 3.287347 3.746401 3.008046 2.623123 3.330972
```

この出力結果より、例えば $E(y_{148} | \mathbf{x}_{148})$ の推定値は 3.346333 であることがわかります。

実際の観測値と当てはめ値をプロットするために、以下のように打ち込みます。

```
> plot(Poifitted,fertil$kids)
```

これより、以下のような散布図を得ます。



この散布図より、このモデルの当てはめ値による予測はあまりうまくいっていないことが示唆されます(ほとんどの点が45度線の上に載っていません)。

5.3 疑似最尤法による推定

先ほどは、 y_i が実際にポアソン回帰モデルに従っていると仮定しました(特に y_i がポアソン分布に従うと仮定しました)。実際にはこの仮定は正しくないかもしれません。その場合でも、期待値関数の定式化さえ正しければ、その期待値関数に含まれる未知パラメーターベクトルは、ポアソン回帰モデルの推定量によって、一致推定することができます。この場合、ポアソン回帰モデルの最尤推定量は疑似最尤推定量と呼ばれます。これは同じ推定量(ポアソン回帰推定量)に対して、その仮定が満たされないとして、呼び方を変えただけで、推定量自体は全く同じです(つまり推定値自体は全く同じです)。では何が変わるのかと言うと標準誤差の計算の仕方が変わります。

正しい標準誤差を得るためには、(ポアソン回帰モデルの仮定が正しいと想定したもとで得た最尤法の標準誤差ではなく)、疑似最尤推定法のもとでの漸近分散から計算した標準誤差を計算する必要があります。

これは、先ほどのコードを若干変更した、以下のコードによって得ることができます。

```
> Poiquasi=glm(kids~educ+age+agesq+black,family=quasipoisson(link="log"),data=fertil)
> summary(Poiquasi)

Call:
glm(formula = kids ~ educ + age + agesq + black, family = quasipoisson(link = "log"),
    data = fertil)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.8131  -0.7750  -0.1384   0.6210   2.0926

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.958415   3.392328  -0.872   0.385
educ         -0.014821   0.021664  -0.684   0.495
age           0.213878   0.153116   1.397   0.165
agesq        -0.002632   0.001729  -1.522   0.130
black         0.172169   0.165842   1.038   0.301

(Dispersion parameter for quasipoisson family taken to be 1.107601)

Null deviance: 210.70  on 155  degrees of freedom
Residual deviance: 201.92  on 151  degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 5
```

この出力結果を見ると、推定値自体は先ほどと全く同じですが、その標準誤差(Std. Error)が先ほどと異なっていることが見て取れます。

6 フラクショナル回帰モデル

次に、フラクショナル回帰モデルの推定を行います。フラクショナル回帰モデルは、モデルにおける被説明変数 y_i の取りうる範囲が $[0, 1]$ であるようなモデルです（この場合、一般的には y_i は 0 と 1 を取るような二値変数ではないですが、二値変数の場合もフラクショナル回帰モデルの特別な場合として含まれます）。

フラクショナル回帰モデルは期待値関数 $E(y_i | \mathbf{x}_i)$ を定式化します。よく使われる定式化として、以下のフラクショナルロジットモデルとフラクショナルプロビットモデルがあります。

$$(\text{フラクショナルロジット}) \quad E(y_i | \mathbf{x}_i) = \frac{1}{1 + \exp(-z_i)}, \quad z_i = \mathbf{x}_i' \boldsymbol{\beta},$$

$$(\text{フラクショナルプロビット}) \quad E(y_i | \mathbf{x}_i) = \Phi(z_i), \quad z_i = \mathbf{x}_i' \boldsymbol{\beta},$$

ここで、 $\mathbf{x}_i$ は説明変数ベクトル、 $\boldsymbol{\beta}$ はその係数ベクトル、 $\Phi(\cdot)$ は標準正規分布の分布関数です。この係数ベクトルは、以下の疑似対数尤度関数を最大化することによって、疑似最尤推定できます。

$$L_N^{(F)}(\boldsymbol{\beta}) = \sum_{i=1}^N \ell_i^{(F)}(\boldsymbol{\beta}),$$

ここで $\ell_i^{(F)}(\boldsymbol{\beta}) = (1 - y_i) \log[1 - G(\mathbf{x}_i' \boldsymbol{\beta})] + y_i \log[G(\mathbf{x}_i' \boldsymbol{\beta})]$ および $G(\mathbf{x}_i' \boldsymbol{\beta}) = E(y_i | \mathbf{x}_i)$ です。フラクショナル回帰モデルについてより詳しくは、講義スライドを参照して下さい。

以下では、フラクショナルロジットおよびフラクショナルプロビットモデルの推定の前に、まずは講義でも紹介したフラクショナル変数 y_i に対する慣習的なモデルおよびその推定方法について紹介します（ただしここでは y_i の取りうる値の範囲は $(0, 1)$ とします（つまり 0 と 1 はとらないとします））。

6.1 フラクショナル変数への慣習的な方法

被説明変数であるフラクショナル変数 y_i が 0 より大きく 1 より小さい値しか取らない場合には、慣習的な方法として y_i を以下のようにモデル化するという方法があります。

$$y_i = \frac{1}{1 + \exp(-w_i)}, \quad w_i = \mathbf{x}_i' \boldsymbol{\beta} + u_i,$$

ここで、 u_i はある確率変数、 $\mathbf{x}_i$ は説明変数ベクトル、 $\boldsymbol{\beta}$ はその係数ベクトルです。このように定式化した場合、 w_i が正の ∞ に近づいていくと、 y_i はどんどん 1 に近づいていき、また、 w_i が負の ∞ に近づいていくと、 y_i はどんどん 0 に近づいていきます。この変数 w_i は以下のように y_i の関数として表せます。

$$\begin{aligned} y_i + y_i \exp(-w_i) &= 1 \Leftrightarrow \exp(-w_i) = (1 - y_i)/y_i \\ &\Leftrightarrow w_i = \log(y_i) - \log(1 - y_i). \end{aligned}$$

よって、 w_i は観測される y_i 変数から計算でき、係数ベクトル $\boldsymbol{\beta}$ は、 w_i を $\mathbf{x}_i$ に回帰することによって、通常の最小二乗推定法によって推定することができます。

このモデルにおいては

$$E(y_i | \mathbf{x}_i) = E \left[\frac{1}{1 + \exp(-\mathbf{x}_i' \boldsymbol{\beta} - u_i)} \middle| \mathbf{x}_i \right] = E \left[\frac{1}{1 + \exp(-\mathbf{x}_i' \boldsymbol{\beta}) \exp(-u_i)} \middle| \mathbf{x}_i \right],$$

ですので、これは一般的には $1/(1+\exp(-\mathbf{x}_i' \boldsymbol{\beta}))$ にも $\Phi(\mathbf{x}_i' \boldsymbol{\beta})$ にも等しくないことに注意して下さい(ただし u_i の分布によっては等しくなることもあります)。よって、慣習的な方法におけるこのモデルは一般的にはフラクショナルロジットやフラクショナルプロビットとは同じモデルではないことに注意して下さい。

以下ではまず、慣習的なモデルを推定します。

ここでは“401k.csv”というファイルにある 401(k) Pension Plan(アメリカの年金制度です)についてのデータを用います。このデータは以下の変数を含みます。

prate: 各企業の従業員のこの年金制度への参加率(フラクショナル変数),
mrate: 401k plan match rate
age: age of 401k plan
totemp: total number of firm employees
sole: =1 if 401k is firm's sole plan
ltotemp: log of totemp

このデータについてより詳しくは L.E. Papke (1995), "Participation in and Contributions to 401(k) Pension Plans: Evidence from Plan Data," *Journal of Human Resources*, 30, 311-325 を参照して下さい。以下では、“prate”を被説明変数とし、“mrate”, “age”, “sole”, および “ltotemp” を説明変数します。

データを見るとわかる通り、被説明変数の“prate”は 1 を取る場合がたくさんあります。よって慣習的な方法はこのままでは直接適用できません。よって、かなり恣意的なやり方ですが、“prate”に 0.999 をかけることによって、新しいデータを作成し、そのデータに対して、慣習的な方法を適用することにします。これはかなり恣意的なやり方ですので、実際の分析ではよくない方法かと思いますが、ここでは単に R の練習として行います。

まずデータを読み込みます。作業ディレクトリを“401k.csv”ファイルが保存されているフォルダに変更した後に、以下を打ち込みます。

```
> data401k=read.csv("401k.csv",header=T,skip=7)
> head(data401k,10)
  prate mrate age totemp sole  ltotemp
1 0.261  0.21  8   8709    0 9.072112
2 1.000  1.42  6    315    1 5.752573
```

```

3 0.976 0.91 10 275 1 5.616771
4 1.000 0.42 7 500 0 6.214608
5 0.825 0.53 28 933 1 6.838405
6 1.000 1.82 7 143 1 4.962845
7 1.000 0.53 31 628 1 6.442540
8 0.925 0.34 13 823 0 6.712956
9 1.000 0.22 21 279 1 5.631212
10 0.968 0.60 10 141 1 4.948760

```

次に、変数 “ w_i ” を作ります。

```
> w=log(0.999*data401k$prate)-log(1-0.999*data401k$prate)
```

最後にこの w_i を説明変数に線形回帰し、最小二乗法で β を推定します。

```
> OLSresult=lm(w~mrate+age+sole+ltotemp,data=data401k)
> summary(OLSresult)
```

Call:

```
lm(formula = w ~ mrate + age + sole + ltotemp, data = data401k)
```

Residuals:

```

      Min       1Q   Median       3Q      Max
-6.7428 -2.1590 -0.4271  2.3502  5.2129

```

Coefficients:

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept)  4.753340    0.365164   13.017 < 2e-16 ***
mrate        0.964971    0.085561   11.278 < 2e-16 ***
age          0.035891    0.007285    4.927 9.27e-07 ***
sole         0.854230    0.140674    6.072 1.59e-09 ***
ltotemp     -0.362414    0.048585   -7.459 1.45e-13 ***
---

```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 2.56 on 1529 degrees of freedom
```

```
Multiple R-squared:  0.1798,    Adjusted R-squared:  0.1777
```

```
F-statistic: 83.82 on 4 and 1529 DF,  p-value: < 2.2e-16
```

出力結果より、これらの説明変数の係数は全て有意に 0 と異なっていることが分かります。

6.2 Estimation of the fractional logit model

次に、`glm(.)` 関数を用いて、フラクショナルロジットモデルを推定します。

Here, for a comparison to the OLS result, we use `0.999*prate` as the dependent variable (later we estimate the model with `prate` too).

ここではまず、先ほどの最小二乗法による推定結果との対比のために、`0.999*prate` を被説明変数として推定を行います(のちほど `prate` を被説明変数として推定も行います)。以下のように打ち込みます。

```

> Flogit=glm(0.999*prate~mrate+age+sole+ltotemp,data=data401k,
(ここまで打ち込んだらエンターキーを押してください。行が変わって “+” が付きますが、そのままコードが続けて打ち込めます)
+ family=quasibinomial('logit'))
> summary(Flogit)

Call:
glm(formula = 0.999 * prate ~ mrate + age + sole + ltotemp, family =
quasibinomial("logit"),
    data = data401k)

Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.7834  -0.2017   0.1882   0.4249   0.8599

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)  2.363123    0.206573  11.440 < 2e-16 ***
mrate        0.901063    0.098874   9.113 < 2e-16 ***
age          0.031985    0.004961   6.447 1.52e-10 ***
sole         0.166899    0.083141   2.007 0.0449 *
ltotemp      -0.206631    0.026715  -7.735 1.87e-14 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.2362364)

Null deviance: 376.52 on 1533 degrees of freedom
Residual deviance: 307.32 on 1529 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6

```

この推定結果においてもやはり、すべての説明変数の係数は5%で有意になっています。ただし、推定値の値は最小二乗法の推定値ともちろん異なります。

次に、同じモデルを被説明変数を(“0.999*prate”ではなく)“prate”にして推定します。

```

> Flogit2=glm(prate~mrate+age+sole+ltotemp,data=data401k,
(ここでエンターキーを押してください)
+ family=quasibinomial('logit'))

```

推定結果は

```

> summary(Flogit2)

Call:
glm(formula = prate ~ mrate + age + sole + ltotemp, family =
quasibinomial("logit"),
    data = data401k)

```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7889	-0.2028	0.1961	0.4334	0.8690

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2.370495	0.209172	11.333	< 2e-16 ***
mrates	0.916716	0.101053	9.072	< 2e-16 ***
age	0.032236	0.005032	6.406	1.98e-10 ***
sole	0.167686	0.084204	1.991	0.0466 *
ltotemp	-0.208002	0.027042	-7.692	2.58e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.2406718)

Null deviance: 384.49 on 1533 degrees of freedom
Residual deviance: 314.53 on 1529 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6

となります。予想できたことですが、推定結果は被説明変数として“0.999*prate”を用いた場合とほぼ同じになります。

6.3 フラクショナルプロビットモデルの推定

同様に、フラクショナルプロビットモデルを推定します。これは先ほどのコードにおいて“logit”を“probit”に置き換えるだけです。

```
> Fprobit=glm(0.999*prate~mrates+age+sole+ltotemp,data=data401k,  
+ family=quasibinomial('probit'))  
> summary(Fprobit)
```

Call:

```
glm(formula = 0.999 * prate ~ mrates + age + sole + ltotemp, family =  
quasibinomial("probit"),  
data = data401k)
```

Deviance Residuals:

Min	1Q	Median	3Q	Max
-1.7729	-0.2061	0.1935	0.4254	0.8267

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	1.422534	0.111932	12.709	< 2e-16 ***
mrates	0.398019	0.044539	8.936	< 2e-16 ***
age	0.016726	0.002544	6.574	6.72e-11 ***
sole	0.109694	0.044535	2.463	0.0139 *
ltotemp	-0.113904	0.014672	-7.763	1.51e-14 ***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for quasibinomial family taken to be 0.2374011)

Null deviance: 376.52 on 1533 degrees of freedom
Residual deviance: 309.21 on 1529 degrees of freedom
AIC: NA

Number of Fisher Scoring iterations: 6

推定結果はフラクショナルロジットモデルの場合とほとんど同じになります。

7 二項回帰モデル

ここでは、二項回帰モデルの推定について説明します。まず二項回帰モデルについて簡単に復習しておきましょう。二項回帰モデルにおいては、試行回数 n_i と説明変数ベクトル $\mathbf{x}_i$ で条件付けした場合に(ここでは共に観測されると仮定します)、被説明変数 y_i は二項分布に従うと仮定します。より具体的には二項回帰モデルは以下のようなモデルです。

$$y_i | n_i, \mathbf{x}_i \sim B(p_i, n_i) \text{ with } p_i = F(\mathbf{x}_i' \boldsymbol{\beta}),$$

ここで試行回数 n_i は 0 より大きい整数²、 $F(\cdot)$ は定義域が $(-\infty, \infty)$ 、値域が $(0, 1)$ である関数とします。ここでは、上記のモデルにおいて、未知パラメータ $\boldsymbol{\beta}$ を推定することが目的です。

この関数 $F(\cdot)$ としては、標準正規分布の分布関数

$$\Phi(x) = \int_{-\infty}^x \phi(z) dz, \text{ where } \phi(\cdot) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right),$$

やまたは(標準)ロジスティック分布の分布関数

$$\Lambda(x) = \frac{\exp(x)}{1 + \exp(x)}.$$

がよく用いられます。

7.1 データ “fertil2.csv” の読み込み

データをダウンロードして作業ディレクトリに保存して下さい。This data is about the women's fertility in Botswana. このデータはボツワナの女性についての子供の数のデータです。以下のように打ち込んでデータを読み込んでください。

² 試行回数 n_i の値は 0 を取ることもできます。しかしながら、対数尤度関数を見てみると、 $n_i = 0$ の場合は、そのデータは $\boldsymbol{\beta}$ の推定に全く影響を及ぼさないことがわかります(なぜならその場合そのデータの対数尤度関数は $\boldsymbol{\beta}$ に依存しない定数になるからです)。つまりそのデータを入れて推定しても入れないで推定しても推定結果は同じになるので、入れないとしても $\boldsymbol{\beta}$ についての分析結果には影響しないことになります。

```
> fertil2=read.csv("fertil2.csv",header=T,skip=12)
```

最初の 6 行を見るために以下のように打ち込んでください。

```
> head(fertil2,6)
  children ceb educ age agesq electric radio tv bicycle urban evermarr
1         0   0  12  24   576         1     1  1         1     1         0
2         3   3  13  32  1024         1     1  1         1     1         1
3         1   1   5  30   900         1     0  0         0     1         1
4         2   3   4  42  1764         1     0  1         0     1         1
5         2   2  11  43  1849         1     1  1         1     1         1
6         1   1   7  36  1296         1     0  0         0     1         1
```

これを見るとこのデータは “children”, “ceb”, “educ”, “age”, “agesq”, “electric”, “radio”, “tv”, “bicycle”, “urban”, および “evermarr” の 11 個の変数を含んでいることがわかります。ここで Here “children” は生存している子供の数、“ceb” は今までに生まれた子供の数です。そのほかの変数についての説明はファイルを参照して下さい。以下では被説明変数 y_i を children_i とし、 $n_i = \text{ceb}_i$ および、説明変数ベクトルを $\mathbf{x}_i = [\text{educ}_i, \text{age}_i, \text{agesq}_i, \text{electric}_i, \text{radio}_i, \text{tv}_i, \text{bicycle}_i, \text{urban}_i, \text{evermarr}_i]'$ とします。ここで下付き文字 “ i ” はそのデータが i 番目の個人のものを意味しています。ここでは被説明変数 y_i は二項回帰モデルに従うとモデル化し、そこに含まれる未知パラメータベクトル β を推定することとします。このモデルにおいて $F(\mathbf{x}_i'\beta)$ は i 番目の個人の子供の生存率を表します。

7.2 R 関数“glm(.)”による推定

まず、 $F(.) = \Lambda(.)$ として、 β を推定しましょう。この場合、 β を推定するには

```
> binlogit=glm(cbind(children,ceb-children)~educ+age+agesq+electric+radio+tv+
  (ここでエンターキーを押して新しい行から入力するとよいでしょう)
+ bicycle+urban+evermarr,family=binomial(link=logit),data=fertil2)
```

と打ち込みます。このコードにおいて、一般的には、最初の入変数は “ $\text{cbind}(y, n-y)$ ” の形をとります。ここで y は成功回数に相当する変数、 n は試行回数に相当する変数です(よってこの場合 $n-y$ は失敗の回数を表しているので、ここに失敗の回数に相当するデータを入力しても構いません)。推定結果を見るには以下のように打ち込んでください。

```
> summary(binlogit)

Call:
glm(formula = cbind(children, ceb - children) ~ educ + age +
    agesq + electric + radio + tv + bicycle + urban + evermarr,
    family = binomial(link = logit), data = fertil2)

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
```

```
(Intercept) 3.8183714 0.8295870 4.603 4.17e-06 ***
educ 0.0348046 0.0132145 2.634 0.00844 **
age -0.0489337 0.0478772 -1.022 0.30675
agesq 0.0002054 0.0006696 0.307 0.75902
electric 0.1715742 0.1679760 1.021 0.30706
radio 0.1368960 0.0939249 1.458 0.14498
tv 0.2525566 0.2149227 1.175 0.23995
bicycle 0.1261075 0.0985633 1.279 0.20074
urban -0.1448674 0.0899742 -1.610 0.10738
evermarr -0.0318957 0.1037203 -0.308 0.75845
---
Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 2009.9 on 2756 degrees of freedom
Residual deviance: 1930.1 on 2747 degrees of freedom
AIC: 2764.7
```

Number of Fisher Scoring iterations: 5

この推定結果より、“educ”と言う変数のみが(1%の有意水準で)有意に0と異なっていることが分かります。また、当てはめ値 $\Lambda(\mathbf{x}; \hat{\beta})$ を計算するには以下のように打ち込んでください。

```
> binlogit$fitted
      1      2      3      4      5      6      7      8
0.9763901 0.9684969 0.9372657 0.9249842 0.9519943 0.9283361 0.9663095 0.9800511
      9     10     11     12     13     14     15     16
0.9259603 0.8988694 0.9592201 0.9501395 0.8969706 0.9014761 0.9639224 0.9565692
      :
```

次に、 $F(.) = \Phi(.)$ の場合に β を推定してみましょう。以下のように打ち込んでください。

```
> binprobit=glm(cbind(children,ceb-children)~educ+age+agesq+electric+radio+tv+
+ bicycle+urban+evermarr,family=binomial(link=probit),data=fertil2)
```

このコードと先ほどのコードを比べると “link=logit” の部分が “link=probit” に置き換わっているだけであることがわかります。推定結果を見るには以下のように打ち込んでください。

```
> summary(binprobit)

Call:
glm(formula = cbind(children, ceb - children) ~ educ + age +
    agesq + electric + radio + tv + bicycle + urban + evermarr,
    family = binomial(link = probit), data = fertil2)

robit
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept)  2.005e+00  3.899e-01   5.141 2.73e-07 ***
educ         1.638e-02  6.341e-03   2.583 0.00979 **
age         -1.982e-02  2.277e-02  -0.870 0.38408
```

```

agesq      4.099e-05  3.218e-04  0.127  0.89865
electric   8.553e-02  7.966e-02  1.074  0.28298
radio      6.815e-02  4.625e-02  1.474  0.14061
tv         1.228e-01  9.988e-02  1.229  0.21889
bicycle    6.268e-02  4.751e-02  1.319  0.18707
urban     -6.791e-02  4.403e-02 -1.543  0.12293
evermarr  -1.406e-02  5.015e-02 -0.280  0.77927
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 2009.9  on 2756  degrees of freedom
Residual deviance: 1929.7  on 2747  degrees of freedom
AIC: 2764.3

Number of Fisher Scoring iterations: 5

```

この場合もやはり、“*educ*” のみが有意になっています。