

R ハンズオン2016年10月29日

[Code ▼](#)

注意

データは架空のものです。

準備

rmsパッケージ, Matchingパッケージをインストールする。 分析に必要なデータを取り込む(dat.csvというファイルです)

データの概要

strokeは脳卒中のあるなし。0がなし, 1があり。 genderは性別。0が男性, 1が女性とする。 sakeは飲酒量, smokingは喫煙量, hbpは最高血圧。

やりたいこと

性別による脳卒中への因果効果を調べたい。

```
summary(dat)
```

[Hide](#)

```
  id      stroke      sake      gender      smoking
Min.   :1.00 Min.   :0.00 Min.   :1.100 Min.   :0.0 Min.   :0.00
1st Qu.: 5.75 1st Qu.:0.00 1st Qu.:2.000 1st Qu.:0.0 1st Qu.:0.75
Median :10.50 Median :0.00 Median :3.250 Median :1.0 Median :1.00
Mean   :10.50 Mean   :0.25 Mean   :3.260 Mean   :0.7 Mean   :1.30
3rd Qu.:15.25 3rd Qu.:0.25 3rd Qu.:4.525 3rd Qu.:1.0 3rd Qu.:2.00
Max.   :20.00 Max.   :1.00 Max.   :6.400 Max.   :1.0 Max.   :3.00
      hbp
Min.   :133.0
1st Qu.:139.8
Median :147.0
Mean   :148.0
3rd Qu.:155.0
Max.   :176.0
```

傾向スコアを求める。カンタン。

```
kekka = glm(gender ~ sake + smoking + hbp, data = dat, family = binomial)
```

[Hide](#)

```
kekka
```

[Hide](#)

```
Call: glm(formula = gender ~ sake + smoking + hbp, family = binomial,
data = dat)
```

Coefficients:

```
(Intercept)      sake      smoking      hbp
    9.18941    3.22120   -2.65238   -0.09206
```

Degrees of Freedom: 19 Total (i.e. Null); 16 Residual

Null Deviance: 24.43

Residual Deviance: 11.09 AIC: 19.09

c統計量はどうかだろうか？

```
library(rms)
c.stat = lrm(gender ~ sake + smoking + hbp, data = dat)
c.stat
```

Hide

Logistic Regression Model

```
lrm(formula = gender ~ sake + smoking + hbp, data = dat)
      Model Likelihood Discrimination Rank Discrim.
      Ratio Test      Indexes      Indexes
Obs      20 LR chi2  13.35 R2    0.690 C    0.940
0        6 d.f.      3 g    5.619 Dxy  0.881
1        14 Pr(> chi2) 0.0039 gr 275.661 gamma 0.881
max lderivl 0.002          gp  0.389 tau-a 0.389
      Brier 0.096
```

```
      Coef S.E. Wald Z Pr(>|Z|)
Intercept 9.1837 15.3782 0.60 0.5504
sake      3.2205 1.8151 1.77 0.0760
smoking   -2.6519 1.6973 -1.56 0.1182
hbp       -0.0920 0.1184 -0.78 0.4371
```

傾向スコアによるマッチング分析

分析に必要なものを取り出す。傾向スコア, 性別, 脳卒中の有無

```
ps = kekka$fitted.values
gender = dat$gender
stroke = dat$stroke
```

Hide

Trが割当。Yが結果変数。Xが傾向スコア。

```
library(Matching)
inga = Match(Tr=gender, Y=stroke, X=ps, estimand = "ATE")
summary(inga)
```

Hide

Estimate... -0.55
AI SE..... 0.26325
T-stat..... -2.0893
p.val..... 0.036683

Original number of observations..... 20
Original number of treated obs..... 14
Matched number of observations..... 20
Matched number of observations (unweighted). 20

逆確率重み付け法 (IPW) の場合

```
ivec1 = dat$gender  
ivec0 = rep(1, nrow(dat))-ivec1  
ivec = cbind(ivec1, ivec0)  
iestp1 = ivec1/kekka$fitted  
iestp0 = ivec0/(1-kekka$fitted)  
iestp = iestp1 + iestp0  
ipw = lm(dat$stroke ~ 0 + ivec, weights = iestp)  
summary(ipw)
```

Hide

Call:
lm(formula = dat\$stroke ~ 0 + ivec, weights = iestp)

Weighted Residuals:
Min 1Q Median 3Q Max
-0.73977 -0.25782 -0.16597 0.03737 0.99218

Coefficients:
Estimate Std. Error t value Pr(>|t|)
ivecivec1 0.1640 0.1248 1.314 0.20535
ivecivec0 0.4820 0.1648 2.925 0.00904 **

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

Residual standard error: 0.5413 on 18 degrees of freedom
Multiple R-squared: 0.3636, Adjusted R-squared: 0.2929
F-statistic: 5.142 on 2 and 18 DF, p-value: 0.01713

```
ipw$coefficients[1] - ipw$coefficients[2]
```

Hide

```
ivecivec1  
-0.3179901
```