

# 「Rによるデータ分析入門」正誤表

2024/09/23

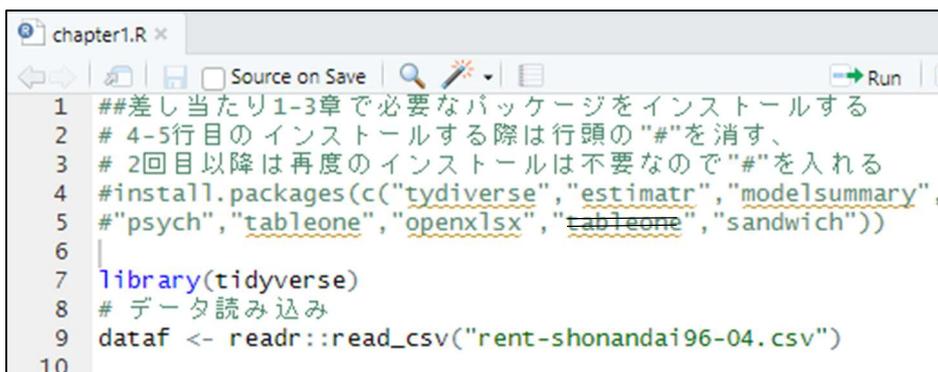
P. iv 9 行目

誤) 高野祐介

正) 高野佳佑

## 第1章 RとRStudioの基本動作

P.27 6 行目 "tableone"は一つ不要 (動作上は問題ありません)



```
chapter1.R x
Source on Save
Run

1 ##差し当たり1-3章に必要なパッケージをインストールする
2 # 4-5行目のインストールする際は行頭の"#"を消す、
3 # 2回目以降は再度のインストールは不要なので"#"を入れる
4 #install.packages(c("tidyverse", "estimatr", "modelsummary",
5 # "psych", "tableone", "openxlsx", "tableone", "sandwich"))
6
7 library(tidyverse)
8 # データ読み込み
9 dataf <- readr::read_csv("rent-shonandai96-04.csv")
10
```

## 第2章 統計表の作成

P.65 下から 11 行目

誤) 1um 千円

正) 1 千円

## 第3章 回帰分析

P.73 上から 7 行目

誤) 傾き :  $\hat{b}$

正) 傾き :  $\hat{\beta}$

P.78 5 行目 t 値の定義式の右辺の分母

誤)  $t = \frac{\text{係数}}{\text{標準偏差}}$

正)  $t = \frac{\text{係数}}{\text{標準誤差}}$

P.82 下から 10 行目

誤) 傾き :  $b_1$

正) 傾き :  $\beta_1$

P.84 1 行目

誤) たとえば age を 1/12 して (age\_rev) 築何か月かを示す指標にすると

正) たとえば age を 1/12 (age\_rev) にすると

P.85 囲み、自由度調整済決定係数の最後の項

誤)  $\frac{n-1}{n-k}$

正)  $\frac{n-1}{n-k-1}$

P.88 上から 6-7 行目

誤) `select(rent_total, y_hat, floor, age)`

正) `select(rent_total, y_hat, res, floor, age)`

P.91 下から 5 行目

誤) `results.docx`

正) `results.docx`

P.92 上から 2 行目

誤) `msummary(results, ...)`

正) `msummary(regs, ...)`

P.95 下から 11 行目

誤) 0 の値をとる

正) 0 か 1 の値をとる

P.102 上から 3 行目

誤) `wage_census`

正) `wage-census`

P.107 下から 2 行目

誤) `dataf <- dataf %>% dplyr::mutate(age_male = age * male)`

正) `dataf <- dataf %>% dplyr::mutate(age_male = age * male)`

P.108 下から 7 行目

誤) 性の賃金と年齢の関係

正) **男性**の賃金と年齢の関係

P.109 下から 4 行目

誤) 使用するデータ : wage-census-ca.**R**

誤) 使用するデータ : wage-census-ca.csv

P.117 下から 14 行目

誤) 男性ダミーの係数が**マイナス**

正) 男性ダミーの係数が**プラス**

P.120 上から 3-4 行目

誤) 係数  $b$  の意味

正) 係数  $\beta$  の意味

P.131 13 行目 「これで回帰分析を実施します。」の後ろに以下の一文を挿入

“なお、Chapter3-4.R では「city が"kunitachi\_city"なら 1」をとるダミー変数

kunitachi\_city を作成し、下記の回帰分析に導入しています。”

P.137 6 行目

誤) library("estimatr")

正) library(estimatr)

#### 第 4 章 離散選択モデル

P.160 文章部上から 2 行目

誤) ほぼ数値が

正) ほぼ同じ数値が

P.161 上から 1 行目

誤) output="mrozframe"

正) output="data.frame"

P.162 上から 3 行目

誤) left cancered

正) left censored

P.162 下から 6 行目

誤) 0 か 0 以上か

正) 0 より大ききか、0 以下か

P.170 上から 10 行目

誤) 自民党を支持している

正) 自民党を支持**し**ている

## 第 5 章 差の差の分析とパネルデータ分析

P. 179 下から 9 行目、P.180 下から 5 行目

誤) ITC

正) ICT

P. 181 2 行目

(誤)  $Y_{Tit} - Y_{Tit}$

(正)  $Y_{Tit} - Y_{Tit-1}$

P. 181 7 行目

(誤)  $Y_{Cit} - Y_{Cit}$

(正)  $Y_{Cit} - Y_{Cit-1}$

P. 181 9 行目

(誤)  $(Y_{Tit} - Y_{Tit}) - (Y_{Cit} - Y_{Cit})$

(正)  $(Y_{Tit} - Y_{Tit-1}) - (Y_{Cit} - Y_{Cit-1})$

P. 183 7 行目

(誤)  $Treat$

(正)  $Treat_i$

P.184 ボックス内の数式

誤)  $a$

正)  $\alpha$

P.184 13 行目 (ボックス内の 3 つ目の数式)

誤)  $Y_{T,2004} = +\beta_1 Treat + \dots$

正)  $Y_{T,2004} = \alpha + \beta_1 Treat + \dots$

P.185 4 行目、7 行目

誤) `station=="shonandai"`

正) `station=="Shonandai"`

P.185-186 chapter5-1.R で `msummary` で推計結果を出力すると前後比較と差

の差の分析のサンプル数 (N of obs) が同一になってしまいます。以下のように

(chapter5-1\_rev.R を参照) `lm_robust` を使わないで、代わりに `msummary` のオ

プションに `vcov` 修正すると、サンプル数が正しく表示されます。

```
===== chapter5-1.R =====  
results <-  
  list(  
    "model1"
```

```

=estimatr::lm_robust(rent_total~floor+age+dist+treat2004, data=data,
station=="Shonandai"),
  "model2"
=estimatr::lm_robust(rent_total~floor+age+dist+treat+year2004+treat200
4, data=datasf)
)
modelsummary::msummary(results, stars=TRUE,
gof_omit='RMSE|AIC|BIC|Log.Lik.|F')

```

---

```

===== chapter5-1_rev.R =====
results <-
  list(
    "model1"
=lm(rent_total~floor+age+dist+treat2004, data=datasf, station=="Shonandai
"),
    "model2"
=lm(rent_total~floor+age+dist+treat+year2004+treat2004, data=datasf))
modelsummary::msummary(results,
stars=TRUE, gof_omit='RMSE|AIC|BIC|Log.Lik.|F', vcoc="HC2")

```

---

以下は、chapter5-1\_rev.Rの実行結果です。

	model1	model2
(Intercept)	3.742***	3.752***
	(0.455)	(0.291)
floor	0.112***	0.109***
	(0.006)	(0.004)
age	-0.092***	-0.090***
	(0.016)	(0.010)
dist	-0.010	-0.045**
	(0.026)	(0.016)
treat2004	0.699**	0.938**
	(0.264)	(0.327)
treat		0.522*
		(0.258)
year2004		-0.204
		(0.234)
Num.Obs.	124	228
R2	0.785	0.804
R2 Adj.	0.778	0.798

+ p < 0.1, \* p < 0.05, \*\* p < 0.01, \*\*\* p < 0.001

P.201 下から 4 行目

誤) Fixe-Effects

正) Fixed-Effects

P.203 表の第 3 列 1 行目

誤)  $b \rightarrow$  正)  $\beta$

P.206 3 行目

誤)  $+\alpha + b_1 X_{i2} + \gamma_1 D_1 \dots$

正)  $+\alpha + \beta_1 X_{i2} + \gamma_1 D_1 \dots$

P.231 下から 4 行目

誤) t 期も非参加なら 1

正) t 期も非参加なら 0

P.232 下から 6 行目～、「ここでは～回帰」を削除

最後に回帰分析結果を見ておきましょう。ここでは t 時点で労働組合に入っていない人が、t+1 期にかけて組合に参加すると賃金に変化するかを見ようとしていますので、以下のように `dplyr::filter()` 関数で t-1 時点で労働組合に入っていない人に限定したオブジェクトを作成します。

```
pdataf2 <- pdataf %>% dplyr::filter(plm::lag(pdataf$union)==0)
```

この新しいオブジェクトを使って実施した回帰分析の結果が以下です。説明変数は lag 関数を使って 1 期前の値を用いています。

P.237 下から 3 行目

誤) `tidyer ::pivot_longer()`関数を使います

正) `tidyr ::pivot_longer()`関数を使います

P.237 下から 1 行目

誤) `tidier ::pivot_longer(~`

正) `tidyr ::pivot_longer(~`

P.238 11 行目

誤) `caprate <-tidier ::pivot_longer(~`

正) `caprate <-tidyr ::pivot_longer(~`

P.247 13 行目

誤) ヒスパニックダミー (hisssp)

正) ヒスパニックダミー (hisp)

P.248 3-5 行目

被説明変数に犯罪発**声**率→被説明変数に犯罪発生率

逮捕される確率(l**probarr**)→逮捕される確率(lprbarr)

逮捕されたのち有罪になる確率(l**probconv**)→逮捕される確率(lprbconv)

収監される確率(l**probpris**)→収監される確率(lprbpris)

## 第 6 章 操作変数法

P.262 2 行目

誤) fitstat=~**ivwald+ivwald.p+sargan+sargan.p**,se.below=TRUE)

正) fitstat=~ivf+ivf.p,se.below=TRUE)

※WEB サポートで配布されている Chapter6-1.R には変更ありません。

P.267 12 行目

誤) Cigarret**2**

正) Cigarettes**SW**

P.273 下から 1 行目

逮捕される確率 (l**probarr**) →逮捕される確率 (lprbarr)

P.274

逮捕されたのち有罪になる確率(lprobconv)→逮捕される確率(lprbconv)

収監される確率(lprobpris)→収監される確率(lprbpris)

## 第7章 傾向スコア法

P.280 ボックス内

誤) 使用するデータ：nswcps\_pamatch.csv

正) 使用するデータ：nswcps\_psmatch.csv

P.292 13 行目, 16 行目 educ2 と married を追加

※P.291 下から 11 行目の予測確率の計算で educ2 と married が含まれているの

に対して、P.292 の処置効果の計算では含まれていないため、これを追加

誤)  $\text{lm}(\text{re78} \sim \text{treated} + \text{age} + \text{age}^2 + \text{educ} + \text{nodegree})$

正)  $\text{lm}(\text{re78} \sim \text{treated} + \text{age} + \text{age}^2 + \text{educ} + \mathbf{\text{educ2}} + \mathbf{\text{married}} + \text{nodegree})$

これにより表 7.1 の結果が以下のように変わり、P.293 の数値も変更になります。

	OLS (unweighted)	PS-reg (weighted)
treated	839.579 (548.675)	1298.100*** (107.153)
その他の変数	Yes	Yes
Num. Obs.	16177	16177
R2	0.476	0.075
R2 Adj.	0.476	0.074

注) カッコ内は標準誤差. \*\*\*は1%水準で統計的に有意であることを示す.

P.293 2 行目, 3 行目

誤) 最小二乗法 (OLS(unweighted)) では、treated の係数は 739 と小さく

正) 最小二乗法 (OLS(unweighted)) では、treated の係数は 840 と小さく

誤) 傾向スコア回帰 (PS-reg(weighted)) では、係数は 1332 と

正) 傾向スコア回帰 (PS-reg(weighted)) では、係数は 1298 と

P.302 12 行目

誤) 雇用の状況(empster\_cat)

正) 雇用の状況(empstat\_edt)

## 補論

P.324 上から 4 行目

誤) 背景を城に

正) 背景を白に

P.354 下から 3 行目、P.355 上から 2 行目

rent\_totla→rent\_total

### **参考文献**

P.357 5 行目

野口春子→野口晴子

P.357 9 行目

久保克之→久保克行