

統計学 2

3. 記述統計とデータの可視化

矢内 勇生

2019年4月18日

高知工科大学 経済・マネジメント学群

Rと誕生日問題

- 問題「40人の中に少なくとも1組同じ誕生日のペアがいる確率は？」
- R の **pbirthday()** で簡単に解ける

`pbirthday(40)`

- 他の人数のときは？

全体の人数を変える

`pbirthday(0)`

`pbirthday(1)`

`pbirthday(2)`

`...`

`pbirthday(366)`

Rを使いこなせるようになると

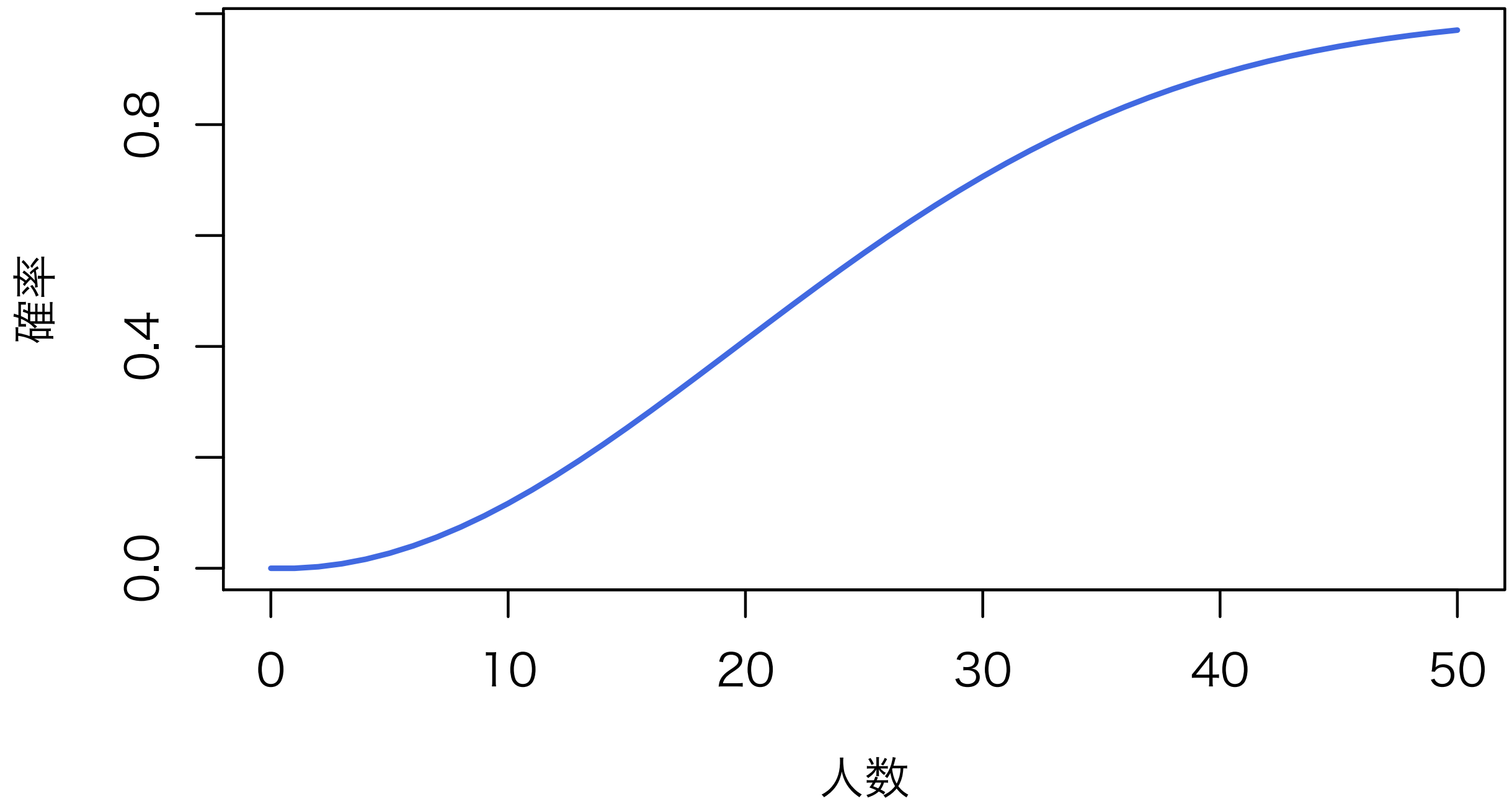
```
sapply(0:366, pbirthday)
```

- 1行で済む

- 追加で以下を書くと

```
prob_pair <- sapply(0:50, pbirthday)
plot(0:50 prob_pair, type = "l", lwd = 2,
     col = "royalblue",
     main = "誕生日問題：少なくとも1組は同じ誕生日",
     xlab = "人数", ylab = "確率")
```

誕生日問題：少なくとも1組は同じ誕生日



今日の目標

- RStudioの基本操作を覚える！
- Rで基本的な統計量を計算する！
- Rで図を作ってみる！

RStudioを利用する



- R用の統合開発環境 (IDE)
- RStudio も無料！ <http://www.rstudio.com/>
- Rをより効率的に使える
 - 画面を4分割して必要な情報を表示
 - Project 機能で研究の管理を容易に
 - セッション中に作った図を遡って確認できる

RStudioのスクリーンショット

The screenshot displays the RStudio interface with the following components:

- Source Editor:** Contains R code for generating data and fitting a model.
- Environment:** Shows the current data environment.
- Files:** Lists files in the project directory.
- Console:** Displays the output of the executed code, including model fit statistics.

```
20 myd1$x5 <- rnorm(500, mean = 10, sd = 0.01)
21
22 ## set parameter values
23 b0 <- 1      #y intercept
24 b1 <- 4
25 b2 <- -0.1
26 b3 <- 0.8
27 b4 <- 10^(-6)
28 b5 <- 8
29
30 myd1$y <- b0 + b1*myd1$x1 + b2*myd1$x2 + b3*myd1$x3 + b4*myd1$x4 + b5*myd1$x5 + rnorm(500)
31
32
33 OLS
34 {r}
35 fit1 <- lm(y ~ x1, data = myd1)
36 summary(fit1)
37
38
39 Compare the fit to the true model.
40 {r}
41 true.model <- lm(y ~ x1 + x2 + x3 + x4 + x5, data = myd1)
42 summary(true.model)
43
```

Environment:

Object	Value
myd1	500 obs. of 6 variables
b0	1
b1	4
b2	-0.1
b3	0.8
b4	1e-06
b5	8
fit1	List of 12
model1	List of 12
true.model	List of 12

Files:

Name	Size	Modified
..		
.Rhistory	2 KB	Sep 19, 2014, 3:33 PM
00_syllabus		
01_イントロダクション		
beer-around-kobe.html	2.3 KB	Sep 17, 2014, 4:51 PM
graphics-R-figure		
graphics-R.md	794 B	Sep 17, 2014, 4:17 PM
graphics-R.Rpres	537 B	Sep 17, 2014, 4:17 PM
graphics.Rmd	933 B	Sep 23, 2014, 11:18 PM
intro-R-graphics.Rmd	1003 B	Sep 17, 2014, 4:58 PM
introduction-to-R.html	622.3 KB	Sep 17, 2014, 3:59 PM
introduction-to-R.Rmd	8.4 KB	Sep 17, 2014, 4:00 PM
rm1_slides_Template.tex	835 B	Sep 22, 2014, 8:39 PM
rm1_slides_Template.toc	153 B	Sep 22, 2014, 8:39 PM
ToDo4pm1.pages		
政治学方法論.Rproj	229 B	Sep 13, 2014, 7:36 AM

Console:

```
Residuals:
    Min       1Q   Median       3Q      Max
-2.55522 -0.64459 -0.00711  0.61403  3.16160

Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept) 41.3259847  43.2662835   0.955  0.340
x1          3.9906856   0.0144468  276.232 <2e-16
x2         -0.0997436   0.0007321 -136.249 <2e-16
x3          0.8010602   0.0036837  217.462 <2e-16
x4          0.0013757   0.0014239   0.966  0.334
x5          3.9696815   4.3263289   0.918  0.359

Residual standard error: 0.9423 on 494 degrees of freedom
Multiple R-squared:  0.9966,    Adjusted R-squared:  0.9965
F-statistic: 2.856e+04 on 5 and 494 DF,  p-value: < 2.2e-16
```

困ったときは

- Googleする
- Rjp Wiki を見る：
 - <http://www.okadajp.org/RWiki/?RjpWiki>
- コンピュータに詳しくそうな友人に訊く
- メールで質問する
- オフィスアワーに来る

RStudio で R を使ってみよう！

- RStudioを開く
- 授業のWebページ
 - 「[RStudio入門と記述統計](#)」を一緒に進める
 - 授業後の自習課題：「[R入門](#)」を参考に、いろいろ試してみよう！

今週の課題

- 授業のウェブページの「RStudio入門と記述統計」にある課題すべて
 - ▶ A4用紙に (1) Rのコード, (2) 結果 (課題の答え) を印刷して提出
 - ▶ 学籍番号と氏名 (ふりがな付) を必ず書くこと
- **締め切り** : 4月25日 (木) の授業開始時