統計モデリング入門 筑波大 (大塚) 集中講義 [03]

R の練習: 次の時間の例題データ

久保拓弥 kubo@ees.hokudai.ac.jp, @KuboBook

筑波大集中講義 http://goo.gl/HvRhXn

2015–02–28

ファイル更新時刻: 2015–02–28 07:58

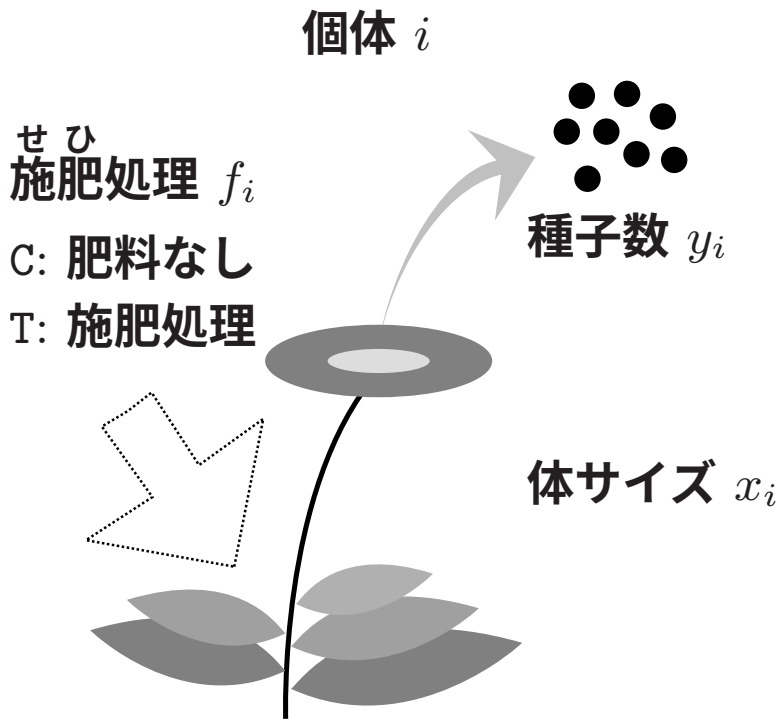
1. ポアソン回帰の例題: 架空植物の種子数データ

植物個体の属性，あるいは実験処理が種子数に影響?

まずはデータの概要を調べる

個体サイズと実験処理の効果を調べる例題

- **応答変数: 種子数** $\{y_i\}$
- **説明変数:**
  - **体サイズ** $\{x_i\}$
  - **施肥処理** $\{f_i\}$



標本数

- **無処理** ($f_i = \mathtt{C}$): 50 sample ($i \in \{1, 2, \cdots 50\}$)
- **施肥処理** ($f_i = \mathtt{T}$): 50 sample ($i \in \{51, 52, \cdots 100\}$)

データファイルを読みこむ

data3a.csv **は** CSV (comma separated value) format file **なので，** R **で読みこむには以下のようにする**:

```
> d <- read.csv("data3a.csv")
```

データは d **と名付けられた** data frame (**表みたいなもの**) **に格納される**

とりあえず

data frame d **を表示**

```
> d
     y      x  f
1    6   8.31  C
2    6   9.44  C
3    6   9.50  C
... (中略) ...
99   7 10.86  T
100  9   9.97  T
```

data frame d を調べる: dx, dy

```
> d$x
  [1]  8.31  9.44  9.50  9.07 10.16  8.32 10.61 10.06
  [9]  9.93 10.43 10.36 10.15 10.92  8.85  9.42 11.11
... (中略) ...
 [97]  8.52 10.24 10.86  9.97

> d$y
  [1]  6  6  6 12 10  4  9  9  9 11  6 10  6 10 11  8
 [17]  3  8  5  5  4 11  5 10  6  6  7  9  3 10  2  9
... (中略) ...
 [97]  6  8  7  9
```

data frame d を調べる: d$f — factor type!

施肥処理の有無をあらわす f 列はちょっと様子がちがう

```
> d$f
  [1] C C C C C C C C C C C C C C C C C C C C C C C C C
 [26] C C C C C C C C C C C C C C C C C C C C C C C C C
 [51] T T T T T T T T T T T T T T T T T T T T T T T T T
 [76] T T T T T T T T T T T T T T T T T T T T T T T T T
Levels: C T
```

因子型データ: いくつかの水準をもつデータ
ここでは C と T の 2 水準

R のデータのクラスとタイプ

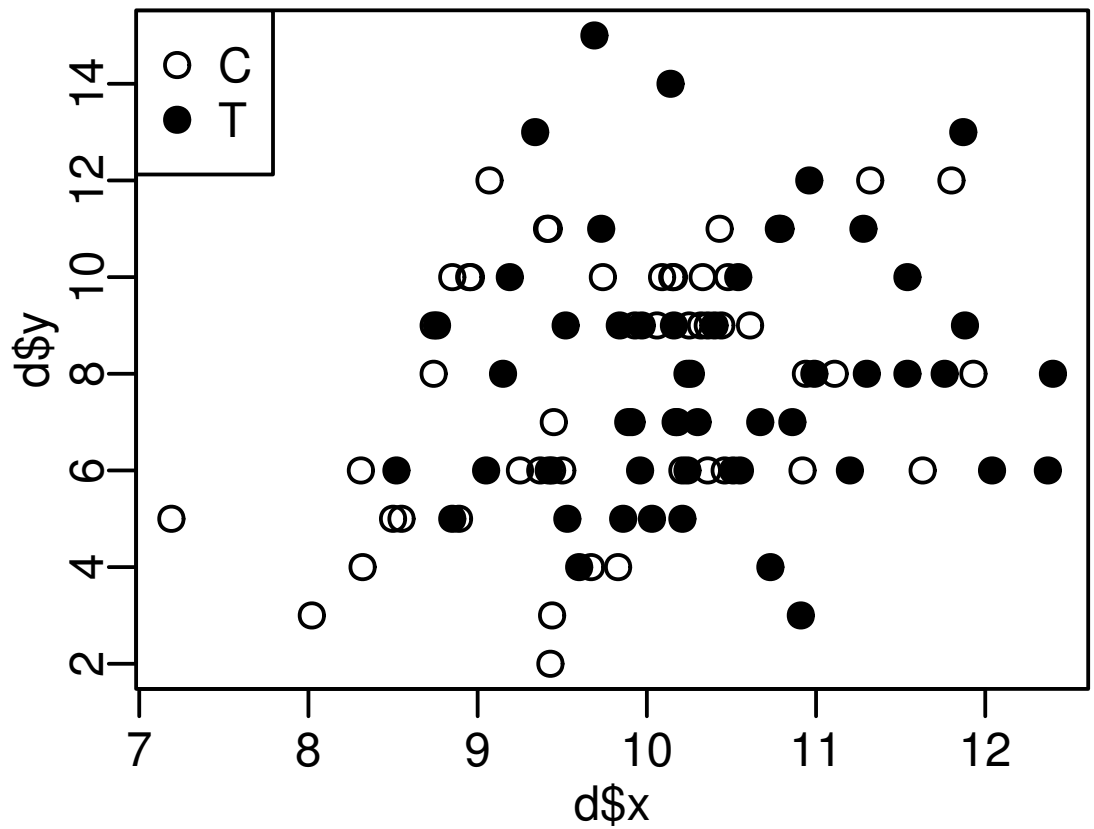
```
> class(d) # d は data.frame クラス
[1] "data.frame"
> class(d$y) # y 列は整数だけの integer クラス
[1] "integer"
> class(d$x) # x 列は実数も含むので numeric クラス
[1] "numeric"
> class(d$f) # そして f 列は factor クラス
[1] "factor"
```

data frame の summary()

```
> summary(d)
       y                x               f
 Min.   : 2.00   Min.   : 7.190   C:50
 1st Qu.: 6.00   1st Qu.: 9.428   T:50
 Median : 8.00   Median :10.155
 Mean   : 7.83   Mean   :10.089
 3rd Qu.:10.00   3rd Qu.:10.685
 Max.   :15.00   Max.   :12.400
```

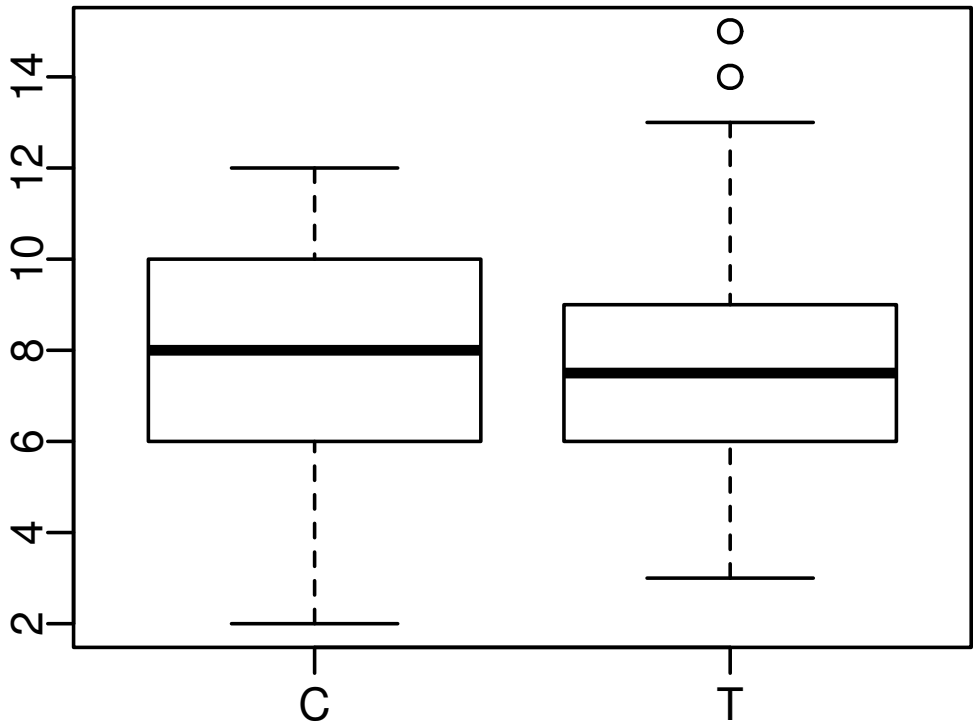
データはとにかく図示する!

```
> plot(d$x, d$y, pch = c(21, 19)[d$f])
> legend("topleft", legend = c("C", "T"), pch = c(21, 19))
```



施肥処理 f を横軸とした図

```
> plot(d$f, d$y)
```



2. ちょっと R 実習

このデータを R であつかう

RStudio 使ってみますかね?

File Edit Code View Plots Session Build Debug Tools Help

Console ~/public_html/stat/2015/

```
> load("Fig/distribution/data.RData")
> summary(data)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   0.00    2.00    3.00    3.56    4.75    7.00
> hist(data, breaks = 0:9 - 0.5)
> lines(0:9, 50 * dpois(0:9, lambda = 3.56), type = "b", col = "#ff400080",
lwd = 4)
>
```

Environment History

Import Dataset Clear List

Global Environment

Data	
d	100 obs. of 3 variables
Values	
data	num [1:50] 2 2 4 6 4 5 2 3 1 2 ...
Functions	
lsfont	function ()
lsp	function (package, all.names = TRUE, pattern…
passwd	function (length = 8, cset = c(letters, LETT…
sw	function (w)

Files Plots Packages Help Viewer

Zoom Export Clear All