5月20日 15:30-16:30/ 23日 16:30-17:30

統計用言語Rの使い方





今日の話のメインターゲット

- そもそも使ったことが無い人
- Rの存在を知らなかった人



R/Rstudioを**インストール**して 簡単な計算を**実行**する. (Rで何ができるか)把握する.



1. R言語とは 何ができるのか? T. 2. インストールから実行まで Rのダウンロード/インストール I. II. Rstudioのダウンロード/インストール III. コンソールとスクリプト 3. 簡単な統計解析 I. 平均,分散を求める+a 4. グラフを描いてみよう 余った時間: I. データのプロット R-番外編 線形回帰 II.



[本節の内容] -

- ✔ Rって何
- ✓ Rを使って何ができるのか
- ✓ Rstudioの紹介



● 統計解析用のプログラム言語





何ができるのか

• だいたいなんでもできる

- 通常の四則演算+行列やベクトルの演算
- 様々な統計分析(回帰や分散分析etc)

データのグラフ表示





超優秀なツール RStudio

Rのための統合開発環境(便利なツール)
もちろん



- Rを見やすくしたようなもの.
- データを分析している感が出て恰好良い

● 導入も簡単

RStudio



RからRstudioへ

RStudioをインストールする前に, **先にRをインストールする**必要があります!

インストールから実行まで

[本節の内容] -

- ✓ Rをインストール
- ✓ Rstudioをインストール
- ✓ Rstudioを実行する

Rをダウンロード(1)

CRANからプログラムをダウンロード http://cran.r-project.org/index.html



$\lceil \mathsf{Download} \ \mathsf{R} \ \mathsf{for} \ (\mathsf{Windows}) \rfloor \rightarrow \lceil \mathsf{base} \rfloor$

Rをダウンロード(2)





Rをインストール(3) ダウンロードしたファイルを実行. R-***-win.exe 「次へ」で進む. 起動オプションなど設定する必要は無い 実行ファイルがインストールできました.



これでRが使用可能に.

Rstudioをダウンロード(1)

公式サイトからプログラムをダウンロードする. https://www.rstudio.com/



Rstudioをダウンロード(2)



Rstudioをダウンロード(3)



Rstudioをインストール(4)

特に設定は必要ない.「次へ」で進めばよい.



RStudioが使用可能に.

Rstudioを実行する(コンソール)

● 起動(初期画面).







電卓的利用.

もう少しプログラミングっぽい使い方



コンソールとスクリプト コンソール = 1行ずつ実行 スクリプト = プログラムを**まとめて**実行





出来るだけスクリプトを使いましょう.

● 直接コンソールに打つと, 間違った時が大変.



統計っぽい解析

簡単な統計解析

[本節の内容]

- ✓ "ベクトル"という考え方
- ✓ 平均,分散を計算する
- ✓ (偏差値を計算する)
- ✓ (ヒストグラムを描く)

"ベクトル"という考え方

- 配列みたいなもの.
- 数字の列ぐらいに思ってください.

	≥ 1		Source on Save
1	a	<-	c(1,2,3,4,5)

aに(1,2,3,4,5)という数字の列を入れる

aに5つの数字が入っている.

ベクトルに対して、様々な関数が用意されています.

> max(a)	> mean(a)	> var(a)
[1] 5	[1] 3	[1] 2.5
最大値	平均值	分散



例) あるテストの平均点を求める.

	Aさん	Bさん	Cさん	Dさん	Eさん	Fさん	Gさん	Hさん
点数	55	45	65	30	85	90	95	50

1) 点数のデータをベクトルで表現

pts <- c(55,45,65,30,85,90,95,50)

2) 平均点を計算 (meanを使う) > mean(pts) [1] 64.375

3) 最高点等も計算できる!

> max(pts) [1] 95

ベクトルに名前を付ける

- ① > pts
 [1] 55 45 65 30 85 90 95 50 ←誰が何点なのか分からない
- (2) pts.names <- c("A","B","C","D","E","F","G","H")</pre>
 - 名前を入れたベクトルを作成. (点数の順番と対応)



もっと巨大なデータの処理

200人分の点数のデータがあったとする.

> pts																									
[1]	70	60	65	56	62	86	73	54	36	64	50	61	58	22	49	63	44	65	59	62	64	64	65	68	71
[26]	47	56	53	51	51	57	20	39	67	63	76	62	95	49	58	38	40	73	74	37	81	74	90	48	53
[51]	66	32	65	47	64	62	56	74	46	68	71	37	32	61	65	80	57	48	42	48	56	40	59	75	63
[76]	49	80	67	92	48	75	36	59	69	43	68	58	78	84	72	56	40	56	67	35	66	47	61	62	51
[101]	51	50	51	50	57	62	73	43	63	71	48	66	68	68	58	60	73	54	38	40	49	58	69	51	87
[126]	54	57	52	64	57	48	64	62	64	61	80	62	93	71	34	63	62	55	75	39	53	57	42	34	45
[151]	33	25	50	52	70	76	65	85	71	57	56	81	69	32	53	53	90	47	71	69	45	73	41	75	67
[176]	68	49	56	58	74	39	71	77	73	68	40	77	47	50	39	52	82	65	56	71	59	45	64	66	47

mean, max, minを調べる







さっきの点数のヒストグラムを描く

たった一言命令するだけ

> hist(pts)



まとめ

- R/Rstudioのダウンロード/インストール
- 計算はベクトルが基本
- コマンドいろいろあって便利です

グラフを描いてみよう

[本節の内容] -

✓ データのプロット

✓ 線形回帰

データのプロット

事例)最高気温とアイスコーヒーの注文個数

最高気温 (℃)	アイスコーヒー注文数 (個数)
22	300
23	310
23	320
24	:330:
24	320 -
25	330:
25	310
26	:320:
26	310
27	340:
27	360
28	:350:
29	360
32	:400:
28	370
24	ៈ310ៈ
31	360
31.	390:
32	390
33	400
33	410
34	450
34	460
35	440
35	480

「アイスコーヒーの注文数を予測しよう」 http://markezine.jp/article/detail/16164

・タを入力

●頑張って入力します.

temperature <- c(22,23,23,24,24,25,25,26,26,27 coffee <- c(300,310,320,330,320,330,310,320,31

表の上から順に入力 (順番を変えてはいけません)









plot() だと図は自動で調整される が, 手動で修正できる.

 :x軸,y軸の範囲 :グラフのタイトル



色々なsettingを試してみましょう. (点の間を線で結んでくれる設定等)

図を修正する(2)

pchとか変えてみる



http://cse.naro.affrc.go.jp/takezawa/r-tips/r/53.html



相関係数を計算

$$\operatorname{Cor}(X,Y) := \frac{\operatorname{Cov}(X,Y)}{\sqrt{\operatorname{V}(X)\operatorname{V}(Y)}}$$

面倒な計算も一発

> cor(temperature,coffee)
[1] 0.9431283

かなり強い相関がある (>0.8) グラフから見ても線形に近似できそう





 $coffee = a \times temperature + b$

計算の方法について, 興味のある人は「線形回帰」で調べましょう 省略!

> lm(coffee ~ temperature)





 $coffee = 11.76 \times temperature + 30.21$

たった一言で式が求まった!

これをグラフに書き込んでみる



回帰直線を図に書き込む

prd <- lm(coffee~temperature) abline(prd)</pre>



この章のまとめ

データとplot()で図を書ける 手動調整用のコマンドも豊富 線形回帰も楽々

今日やったことのおさらい

- R/RStudioのDL/Install
- ●簡単な統計解析
 - 平均や分散の計算
 - 偏差値とヒストグラム
- グラフを描いてみる
 - plot() 関数
 - 手動でのグラフの調整
 - 線形回帰



●講習会終了後の質問は,

- 火曜日15:00-18:00
- 金曜日15:00-18:00
 @利用支援カウンター

● または, E-mail:

sogo-ta52@library.osaka-u.ac.jp

まで.



[本節の内容] —

- ✓ 今日使ったソースコード
- ✓ 参考文献

今日使ったコード(1)

```
#analyze points
pts <- c(55,45,65,30,85,90,95,50)
mean(pts)
max(pts)</pre>
```

```
#big data
pts<-ceiling(rnorm(200,mean=60,sd=15))
min(pts)
max(pts)
mean(pts)</pre>
```

```
#standard score
normalized.pts <- scale(pts,center=TRUE,scale=TRUE)
50 + 10*normalized.pts</pre>
```

```
#coffee ~ temperature
temperature <-
    c(22,23,23,24,24,25,25,26,26,27,27,28,29,32,28,24,31,31,32,33,33,34,34,35,35)
coffee <- c(300,310,320,330,320,330,310,320,310,340,360,350,360,400,370,310,</pre>
```

360,390,390,400,410,450,460,440,480)

今日使ったコード(2)

#simple plot
plot(temperature,coffee)

```
#manual
plot(temperature,coffee,xlim=c(22,35),
    ylim=c(300,480),
    main="coffee~temperature",
    pch=17)
```

```
#linear regression
lm(coffee~temperature)
```

```
#plot predicted line
prd <- Im(coffee~temperature)
abline(prd)</pre>
```



定番の本

The R tips データ解析環境Rの基本技・グラフィックス活用集 / 舟尾暢男著

言語: Japanese

出版物情報:東京オーム社 2009.11.

版:第2版

出版時期: 2009

同じ内容が無料で公開(web)



R は有名な統計言語『S 言語』をオープンソースとして実装し直した統計解析ソフトです。さまざまなブラットフォーム(0: ロードすることができます。それにも関わらず、世界中の専門家が開発に携わっており、日々新しい手法・アルゴリズム が速い上にグラフィングも充実しているので数値計算などにも持ってこいです。このドキュメントは Windows 版 R と Mac ンドを調べた足跡です。

ちなみに、この頁の内容を新しくした書籍は こちら、電子書籍版は こちらで販売されております。

●入門篇●

リンク	The R Project	リンク	RipWiki
リンク	PDF版 R-Tips(200頁*3Mb)	索引	<u>この頁の索引</u>
第01節	R のセットアップ+参考文献	第02節	<u>R の起動と終了</u>
第03節	<u>簡単な計算</u>	第04節	<u>R 用エディタ</u>
第05節	<u>オブジェクトと代入(付値)</u>	第06節	作業ディレクトリの3
第07節	<u>ヘルブを見る</u>	第08節	バッケーシ ライフラ
441.00.642	ニ、カの刑	答案の答案	チーデジー クレ ヘキテ

http://cse.naro.affrc.go.jp/takezawa/rtips/r.html

R-Tips



[本節の内容]

✓ ヘルプを読む(?mean)✓ プログラミングっぽい書き方

✓ ベクトルの応用



● コマンド(mean等)のヘルプを読む.

例) mean() の使い方を調べる									
	> ?mean	"? + コマンド名"							

	Files Plots Packages Help Viewer								
	😓 🄿 🏠 🚍 🗊 🤤								
E	R: Arithmetic Mean 👻 Find in Topic								
	mean {base}								
	Arithmetic Mean								
	Description								
-	Generic function for the (trimmed) arithmetic mean.								
ot ¢	Usage								
	mean(x,)								
*	<pre>## Default S3 method: mean(x, trim = 0, na.rm = FALSE,)</pre>								
	Arguments								
	x An R object. Currently there are methods for numeric/logical vect time and time interval objects. Complex vectors are allowed for to	ors and <u>da</u> rim = 0,							
	trim the fraction (0 to 0 5) of abaanations to be trimmed from each or	d of a bofe							

RStudioの右下のウインドウに ヘルプが表示される.

mean()Ohelp Usage Arguments(引数) mean(x, ...) オプションみたいなもの ## Default S3 method: mean(x trim = 0, na)FALSE. ...) Arguments An R object. Currently there are methods for numeric/logical vectors and date, dateх time and time interval objects. Complex vectors are allowed for trim = 0, only. the fraction (0 to 0.5) of observations to be trimmed from each end of x before the trim mean is computed. Values of trim outside that range are taken as the nearest endpoint. na. Im a logical value indicating whether NA values should be stripped before the meanではtrimという引数を入力できる. (入力しなければデフォルト値=0)



meanの引数-trim

ヘルプの下の方に、Exampleが載っている.

Examples

```
x <- c(0:10, 50)
xm <- mean(x)
c(xm, mean(x, trim = 0.10))</pre>
```



trimの続き

 the fraction (0 to 0.5) of observations to be trimmed from each end of x before the mean is computed. Values of trim outside that range are taken as the nearest endpoint.



他にも色々調べてみましょう

• ?plot, ?hist, ?lm, ?abline, …等々

けど分かりにくい場合もある. (英語を読むのが面倒臭い)

メジャーな関数なら…

Web検索. またはR-tipsの方が分かりやすいかも.

加) 平均です。 X> 0 の場合しか意味を持ちません。

少なくするた<u>めに、 点数を大きさの順に並べて、 両側から同数ずつ削除してから平均を求めること</u>がありま |り込み平均<mark>(trimmed mean)などと呼びます。 Rでは mean() に trim=... というオブションを与えて</mark>計算しま E確には, 個数×0.2 を切り捨てて整数にした個数ずつ両側から外します)。

プログラミングっぽい書き方

1+2+3+…+100を求めたい.

方法2) forを使う方法.

```
> s <- 0
> for(i in 1:100) s <- s + i
> cat(s)
5050
```

少しプログラミング風になった.

プログラミングっぽい書き方

• for文について.

- Rでfor文を使うと遅い.
- forではなく**ベクトルを**使う方が良い.



通常の利用では,単なる好みの問題





ベクトルで計算する(発展)

$$\sum_{k=1}^{100} (3k^2 + 2k)$$
を求めるにはどうするか?

k <- seq(1,100,1)
sum(3*k^2 + 2*k)</pre>

> sum(3*k^2 + 2*k)
[1] 1025150

行列で計算する(発展)

 $\sum_{i,j=1}^{10} \frac{1}{ij}$ をどうやって計算するか?

a <- seq(1,10,1) b <- a %*% t(a) sum(1/b)

> b										
	[,1]	[,2]	[,3]	[,4]	[,5]	[,6]	[,7]	[,8]	[,9]	[,10]
[1,]	1	2	3	4	5	6	7	8	9	10
[2,]	2	4	6	8	10	12	14	16	18	20
[3,]	3	6	9	12	15	18	21	24	27	30
[4,]	4	8	12	16	20	24	28	32	36	40
[5,]	5	10	15	20	25	30	35	40	45	50
[6,]	6	12	18	24	30	36	42	48	54	60
[7,]	7	14	21	28	35	42	49	56	63	70
[8,]	8	16	24	32	40	48	56	64	72	80
[9,]	9	18	27	36	45	54	63	72	81	90
[10,]	10	20	30	40	50	60	70	80	90	100

(i,j)要素=ij

> sum(1/b) [1] 8.578855

R-番外編, まとめ

- ●困ったらヘルプを読みましょう.
 - Web検索, R-tips,
 - またはRから呼び出せるヘルプへ.
- プログラムっぽい書き方もできます.
 - for, while, repeat系も利用可.
 - ただしベクトルを利用した方が早い.
 - (発展)行列を使って書くのがRの醍醐味.
 - 関数も作れます.活用しましょう.