

# R と RStudio の使い方

芳賀敏郎 (2011) 医薬品開発のための統計解析 第1部 基礎  
2 1組のデータの解析  
2.1 データの特徴の記述

# テキストと利用上の注意

---

## ●テキスト

芳賀敏郎（2011）医薬品開発のための統計解析

第1部 基礎 改訂版、サイエンティスト社、p.275

（サイトへアップすることに対して、サイエンティスト社の了解を得ています）

## ●Rによる解析事例を紹介

R スクリプトの出力結果を紹介します（tidyverse 系には次期バージョンで対応します）

R スクリプト（文字コードUTF-8に設定）を、このサイトから[ダウンロード](#)できます

R スクリプトを [Compile Report] することにより、Word または HTML で見ることができます

R と RStudio の設定と基本的な使い方は「[R と RStudio の使い方](#)」を参照してください

R の出力結果の見方は、テキストとそれを解説した [PDF ファイル](#) を参照してください

グラフ表示は、解析手段として、必要最小限の表現に止めています

## ●自己責任で利用

上記のことを理解した上で、自己責任により利用してください

# 第1部 基礎

---

- 1. 統計の基礎 . . . . .
  - 1.1 宝くじの期待値と分散、1.2 サイコロの目の数の期待値と分散
  - 1.3 分散の加法性・中心極限定理・正規分布、1.4 統計的推測、1.5 モデル
- 2. **1組のデータの解析** **2.1 データの特徴の記述**、2.2 データのグラフ表示と外れ値
  - 2.3 対数変換と対数正規分布、2.4 平均に関する推測（母標準偏差  $\sigma$  既知）
  - 2.5 分散に関する推測、2.6 平均に関する推測（母標準偏差  $\sigma$  未知）
- 3. **2組のデータの解析** **3.1 データのグラフ化**、3.2 平均値の差の  $t$  検定、3.3 分散の違いの検定
  - 3.4 分散が異なる場合の平均値の差の比較
  - 3.5 対応のある場合の平均値の差の  $t$  検定、3.6 検出力と  $n$  の決め方
  - 3.7 ノンパラメトリック検定
- 4. 相関・回帰 . . . . .
  - 4.1 散布図、4.2 相関係数、4.3 回帰モデルとモデルの推定
  - 4.4 誤差を考慮した推定、4.5 回帰分析適用上の諸問題

## ●表示2.1.2 平方和、平均平方、標準偏差の計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt

cbind、apply、function

方法

オブジェクト  $x$  にデータをベクトルとして付値

残差  $e$  とその2乗値を計算し、行列にまとめた

```
>mx
##      x      e      e2
## [1,]  1 -5.25 27.5625
## [2,]  2 -4.25 18.0625
## [3,]  3 -3.25 10.5625
## [4,]  4 -2.25  5.0625
## [5,]  5 -1.25  1.5625
## [6,]  7  0.75  0.5625
## [7,] 10  3.75 14.0625
## [8,] 18 11.75 138.0625
```

```
x <- c(1, 2, 3, 4, 5, 7, 10, 18)
e  <- x - mean(x)
e2 <- e^2
mx <- cbind(x = x, e = e, e2 = e2)
```

ベクトル演算  
 $e$  はベクトル  
 $e2$  はベクトル

## ●表示2.1.2 平方和、平均平方、標準偏差の計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt

cbind、apply、function

方法

オブジェクト  $x$  にデータをベクトルとして付値

残差  $e$  とその2乗値を計算し、行列にまとめた

```
ave_x <- mean(mx[, 1]) # x 列の平均
ave_e <- mean(mx[, 2]) # e 列の平均
ss_x <- sum((mx[, 1] - mean(mx[, 1]))^2) # x 列の偏差平方和
ss_e <- sum((mx[, 2] - mean(mx[, 2]))^2) # e 列の偏差平方和
ss_e2 <- sum((mx[, 3] - mean(mx[, 3]))^2) # e2列の偏差平方和
```

```
>mx
##           x           e           e2
## [1,]      1      -5.25      27.5625
## [2,]      2      -4.25      18.0625
## [3,]      3      -3.25      10.5625
## [4,]      4      -2.25       5.0625
## [5,]      5      -1.25       1.5625
## [6,]      7       0.75       0.5625
## [7,]     10       3.75      14.0625
## [8,]     18     11.75     138.0625
```

```
##           x           e           e2
## ave      6.250      0.00           NA
## ss     215.500     215.50     215.50
```

## ●表示2.1.2 平方和、平均平方、標準偏差の計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt

cbind、apply、function

方法

オブジェクト x にデータをベクトルとして付値

残差 e とその 2 乗値を計算し、行列にまとめた

```
>mx
##           x      e      e2
## [1,]    1 -5.25  27.5625
## [2,]    2 -4.25  18.0625
## [3,]    3 -3.25  10.5625
## [4,]    4 -2.25   5.0625
## [5,]    5 -1.25   1.5625
## [6,]    7  0.75   0.5625
## [7,]   10  3.75  14.0625
## [8,]   18 11.75 138.0625
```

```
ave_x <- mean(mx[, 1]) # x 列の平均
ave_e <- mean(mx[, 2]) # e 列の平均
ss_x  <- sum((mx[, 1] - mean(mx[, 1]))^2) # x 列の偏差平方和
ss_e  <- sum((mx[, 2] - mean(mx[, 2]))^2) # e 列の偏差平方和
ss_e2 <- sum((mx[, 3] - mean(mx[, 3]))^2) # e2 列の偏差平方和
```

apply 関数を使い  
1 行で記述

```
##           x      e      e2
## ave      6.250  0.00      NA
## ss     215.500 215.50 215.50
```

## ●表示2.1.2 平方和、平均平方、標準偏差の計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt

cbind、apply、function

方法

オブジェクト x にデータをベクトルとして付値

残差 e とその 2 乗値を計算し、行列にまとめた

第 1 引数

第 2 引数

第 3 引数

```
ave <- apply(mx, 2, mean) # mx の 3 列の平均
ave[3] <- NA
ss <- apply(mx, 2, function(x) {sum((x - mean(x))^2)})
```

第 1 引数

第 2 引数

第 3 引数

2 か所に x がある

```
>mx
##           x           e           e2
## [1,]      1      -5.25      27.5625
## [2,]      2      -4.25      18.0625
## [3,]      3      -3.25      10.5625
## [4,]      4      -2.25       5.0625
## [5,]      5      -1.25       1.5625
## [6,]      7       0.75       0.5625
## [7,]     10       3.75      14.0625
## [8,]     18     11.75     138.0625
```

```
##           x           e           e2
## ave      6.250      0.00           NA
## ss     215.500     215.50     215.50
```

## ●表示2.1.2 平方和、平均平方、標準偏差の計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt

cbind、apply、function

方法

オブジェクト  $x$  にデータをベクトルとして付値  
残差  $e$  とその2乗値を計算し、行列にまとめた

```
ave <- apply(mx, 2, mean)
ave[3] <- NA
ss <- apply(mx, 2,
  function(x) {sum((x - mean(x))^2)})
```

```
##           x           e           e2
##  1.000000 -5.250000  27.5625
##  2.000000 -4.250000  18.0625
##  3.000000 -3.250000  10.5625
##  4.000000 -2.250000   5.0625
##  5.000000 -1.250000   1.5625
##  7.000000  0.750000   0.5625
## 10.000000  3.750000  14.0625
## 18.000000 11.750000 138.0625
##           NA           NA           NA
## ave    6.250000  0.000000           NA
## ss  215.500000 215.500000 215.5000
## n    8.000000  8.000000   8.0000
## dof  7.000000  7.000000           NA
## ms  30.785714 30.785714           NA
## std  5.548488  5.548488           NA
## cv   0.887758           NA           NA
```

区切り



● 表示2.1.3 四分位値、ひずみ、とがりの計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt、quantile

e1071::skewness、e1071::kurtosis

cbind、apply、function

方法

```
x1 <- c(1, 2, 3, 4, 5, 7, 10, 18)
x2 <- c(1, 3, 5, 8, 9, 11, 13, 18)
mxx <- cbind(x1, x2)
```

```
##                x1                x2
## 1.000000      1.000000
## 2.000000      3.000000
## 3.000000      5.000000
## . . . . .
## 10.000000     13.000000
## 18.000000     18.000000
##                NA                NA
## nn           8.000000      8.000000
## min          1.000000      1.000000
## q25          2.750000      4.500000
## med          4.500000      8.500000
## q75          7.750000     11.500000
## max          18.000000     18.000000
## qrange       5.000000      7.000000
## q135         3.703704      5.1851852
## skew         1.568120      0.3700423
## kurt         2.510521     -0.2824846
```

## ● 表示2.1.3 四分位値、ひずみ、とがりの計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt、quantile

e1071::skewness、e1071::kurtosis

cbind、apply、function

方法

```
q25 <- apply(mxx, 2, quantile, probs = 0.25)
med  <- apply(mxx, 2, median)
```

第1引数    第2引数    第3引数    第4引数

第1引数    第2引数    第3引数

##	x1	x2
##	1.000000	1.000000
##	2.000000	3.000000
##	3.000000	5.000000
	.....	
##	10.000000	13.000000
##	18.000000	18.000000
##	NA	NA
## nn	8.000000	8.000000
## min	1.000000	1.000000
## q25	2.750000	4.500000
## med	4.500000	8.500000
## q75	7.750000	11.500000
## max	18.000000	18.000000
## qrange	5.000000	7.000000
## q135	3.703704	5.1851852
## skew	1.568120	0.3700423
## kurt	2.510521	-0.2824846

## ● 表示2.1.3 四分位値、ひずみ、とがりの計算

スクリプトファイル：Green1-2-1a.R

利用した関数

mean、sum、length、sqrt、quantile

e1071::skewness、e1071::kurtosis

cbind、apply、function

方法

ひずみ（歪度）を skewness 関数で計算

とがり（尖度）を kurtosis 関数で計算

Joanes and Gill（1998）による3種類の方法を指定可

ここでは、タイプ2で計算（JMP、SASで採用）

```
skew <- apply(mxx, 2, skewness, type = 2)
kurt <- apply(mxx, 2, kurtosis, type = 2)
```

##	x1	x2
##	1.000000	1.000000
##	2.000000	3.000000
##	3.000000	5.000000
	. . . . .	. . . . .
##	10.000000	13.000000
##	18.000000	18.000000
##	NA	NA
## nn	8.000000	8.000000
## min	1.000000	1.000000
## q25	2.750000	4.500000
## med	4.500000	8.500000
## q75	7.750000	11.500000
## max	18.000000	18.000000
## qrange	5.000000	7.000000
## q135	3.703704	5.1851852
## skew	1.568120	0.3700423
## kurt	2.510521	-0.2824846

- 表示2.1.6 JMP による基本統計量の計算

スクリプトファイル：Green1-2-1b.R

利用した関数：summary、psych::describe

方法：データを付値したデータフレームから

必要なデータ（10列と11列）を抽出して関数で計算

転置  
行と列の入れ替え

```
summary(df[, 10:11])  
t(describe(df[, 10:11]))
```

summaryの出力

##	TN	BS
## Min.	:6.500	Min. : 73.00
## 1st Qu.	:7.200	1st Qu.: 82.00
## Median	:7.400	Median : 85.00
## Mean	:7.372	Mean : 88.08
## 3rd Qu.	:7.600	3rd Qu.: 93.25
## Max.	:8.300	Max. :117.00

skewness(vt, type = 3)、kurtosis(vt, type = 3) に相当

psych::describe の出力

##	TN	BS
## vars	1.00	2.00
## n	40.00	40.00
## mean	7.37	88.08
## sd	0.36	9.53
## median	7.40	85.00
## trimmed	7.37	86.62
## mad	0.30	7.41
## min	6.50	73.00
## max	8.30	117.00
## range	1.80	44.00
## skew	0.09	1.28
## kurtosis	0.24	1.24
## se	0.06	1.51

## ● 表示2.1.6 JMP による基本統計量の計算

スクリプトファイル：Green1-2-1b.R

利用した関数：pasteecs::stat.desc

方法：データを付値したデータフレームから  
必要なデータを抽出して関数で計算

```
stat_out <- stat.desc(df[, 10:11],  
                      basic = TRUE,  
                      desc = TRUE,  
                      norm = TRUE,  
                      p = 0.95)  
lucid(stat_out, dig = 6, na.value = "")
```

skewness(vt, type = 3)、kurtosis(vt, type = 3) に相当

##		TN	BS
##	nbr.val	40	40
##	nbr.null	0	0
##	nbr.na	0	0
##	min	6.5	73
##	max	8.3	117
##	range	1.8	44
##	sum	294.9	3523
##	median	7.4	85
##	mean	7.3725	88.075
##	SE.mean	0.0565	1.507
##	CI.mean.0.95	0.11428	3.047
##	var	0.12769	90.789
##	std.dev	0.35733	9.528
##	coef.var	0.04847	0.108
##	skewness	0.08803	1.28
##	skew.2SE	0.11776	1.712
##	kurtosis	0.24153	1.244
##	kurt.2SE	0.16484	0.849

- 表示2.1.6 JMP による基本統計量の計算

スクリプトファイル：Green1-2-1b.R

利用した関数

mean、sd、length、qt、sem、var、sum

e1071::skewness、e1071::kurtosis

c、data.frame

方法

表示2.1.3 と同様に、

統計量を個々に計算する関数を使って、

個別にBS（血糖値）の統計量を算出

ひずみ（歪度）を skewness 関数で計算

とがり（尖度）を kurtosis 関数で計算

Joanes and Gill（1998）による3種類の方法を指定可

ここでは、タイプ2で計算（JMP、SASで採用）

BSの統計量

##	Statistics	Value
## 1	mean	88.075
## 2	s.d.	9.528
## 3	se	1.507
## 4	lower95CI	85.028
## 5	upper95CI	91.122
## 6	size	40
## 7	total	3523
## 8	var	90.789
## 9	skewness	1.382
## 10	kurtosis	1.832
## 11	Coef var	10.818

- 表示2.1.7 四分位値の JM と Excel による違い

スクリプトファイル : Green1-2-1b.R

利用した関数

quantile

function, vapply

方法

quantile 関数を使って BS (血糖値) の四分位値を計算

9 種類の方法を順に選択して出力

BS の四分位値 (9 種類)

##		25%	50%	75%
##	[1,]	82	85	93.00000
##	[2,]	82	85	93.50000
##	[3,]	82	85	93.00000
##	[4,]	82	85	93.00000
##	[5,]	82	85	93.50000
##	[6,]	82	85	93.75000
##	[7,]	82	85	93.25000
##	[8,]	82	85	93.58333
##	[9,]	82	85	93.56250

## ● 補足 基本統計量の計算

スクリプトファイル：Green1-2-1b.R

利用した関数：psych::describe

方法：データを付値したデータフレームから 3列～13列を抽出して関数で計算

```
describe(df[, 3:13])
```

##	vars	n	mean	sd	median	trimmed	mad	min	max	range	skew	kurtosis	se
## height	1	40	164.32	5.00	164.5	164.44	5.19	154.0	177.0	23.0	-0.01	-0.34	0.79
## weight	2	40	62.58	6.83	62.5	62.59	8.15	50.0	75.0	25.0	-0.01	-1.00	1.08
## SBP	3	40	121.20	14.63	118.0	119.94	11.86	98.0	152.0	54.0	0.65	-0.43	2.31
## DBP	4	40	77.85	11.04	75.0	76.81	10.38	60.0	108.0	48.0	0.77	-0.26	1.75
## TC	5	40	192.90	28.78	194.5	189.84	22.98	153.0	300.0	147.0	1.55	3.71	4.55
## HDL	6	40	51.83	17.13	50.0	49.59	10.38	28.0	110.0	82.0	1.74	3.89	2.71
## TG	7	40	142.50	100.14	117.5	120.88	34.84	50.0	580.0	530.0	2.95	8.93	15.83
## TN	8	40	7.37	0.36	7.4	7.37	0.30	6.5	8.3	1.8	0.09	0.24	0.06
## BS	9	40	88.08	9.53	85.0	86.62	7.41	73.0	117.0	44.0	1.28	1.24	1.51
## BMI	10	40	0.00	0.00	0.0	0.00	0.00	0.0	0.0	0.0	0.15	-1.04	0.00
## deltaBP	11	40	43.35	9.42	44.0	42.62	10.38	28.0	68.0	40.0	0.56	-0.13	1.49





- 作成 片瀬雅彦
- 作成時期 2021年8月25日