

# 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  
2.5 分散  $\sigma^2$  に関する推定

# 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 で見ることが出来ます

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2.4.1 100組のサンプルの観測値（一部のみ）と基本統計量

スクリプトファイル：Green1-2-5a.R

利用した関数

rnorm、matrix、mean、sd、cbind

signif、rownames、colnames

方法

rnorm 関数で 900個の正規乱数を発生させ、

ベクトル vt に付値、

vt から行列 mx (100×9) を生成 (前節 §2.4)

```
Mu <- 50 # 母平均
Sigma <- 10 # 母標準偏差
Nt <- 900 # データ総数
Nr <- 100 # 行数 100×9 の行列
Nc <- 9 # 列数 (サンプルサイズ)
Xmin <- 30 # グラフの最小値
Xmax <- 70 # グラフの最大値

vt <- rnorm(Nt, mean = Mu, sd = Sigma)
mx <- matrix(vt, nrow = Nr, ncol = Nc)
```

n = 9のサンプルが  
100組

標準偏差、平方和  
カイ2乗値 など

```
##           [,1] [,2] [,3] [,4] [,5] [,6] [,7] [,8] [,9]
## [1,] 44.4 42.9 72.0 42.8 49.3 44.0 60.7 42.7 53.6
## [2,] 47.7 52.6 63.1 42.5 38.3 40.1 49.7 34.6 43.4
## [3,] 65.6 47.5 47.3 40.6 43.7 60.3 49.7 43.1 58.6
##           . . . . .
## [99,] 47.6 43.9 50.2 37.4 53.6 61.8 54.2 52.3 53.7
## [100,] 39.7 38.1 62.5 46.5 55.5 68.6 33.9 56.5 52.4
```

● 表示2.5.1 100組のサンプルの標準偏差など

スクリプトファイル: Green1-2-5a.R

利用した関数: rnorm、matrix、mean、sd、rownames、colnames

方法: 正規乱数 900 個を発生、行列 mx (100×9) に付値、n=9 のサンプル100組 (前節 [§2.4](#))  
標準偏差 (s.d.)、母平均を使った平和和 (S0)、通常の平和和 (S)、  
カイ2乗値 (chi2)、H0: 母標準偏差 = 10 の p 値、母標準偏差の95%信頼区間計算

##	s.d.	S0	S	chi2	p	lowerCI	upperCI
## 1	10.209	834.358	833.721	8.337	0.802	6.895	19.557
## 2	8.633	756.781	596.191	5.962	0.697	5.831	16.538
## 3	8.709	611.164	606.766	6.068	0.721	5.883	16.684
## 4	8.459	573.537	572.481	5.725	0.644	5.714	16.206
		. . . . .		. . . . .			
## 99	6.972	391.466	388.889	3.889	0.266	4.709	13.357
## 100	11.715	1099.474	1097.932	10.979	0.406	7.913	22.443
## mean	9.670	879.632	797.476	7.975	NA	NA	NA
## sd	2.498	432.957	407.333	4.073	NA	NA	NA

# カイ2乗分布（シミュレーション）

- 表示2.5.1 100組のサンプルの標準偏差など

スクリプトファイル：Green1-2-5a.R

利用した関数：rnorm、matrix、mean、sd、rownames、colnames

方法：正規乱数 900 個を発生、行列 mx (100×9) に付値、n=9 のサンプル100組（前節 [§2.4](#)）

標準偏差 (s.d.)、母平均を使った平和和 (S0)、通常の平和和 (S)、

カイ2乗値 (chi2)、H0：母標準偏差 = 10 の p 値、母標準偏差の95%信頼区間計算

##	s.d.	S0	S	chi2	p	lowerCI	upperCI
## 1	10.209	834.358	833.721	8.337	0.802	6.895	19.557
## 2	8.633	756.781	596.191	5.962	0.697	5.831	16.538
## 3	8.709	611.164	606.766	6.068	0.721	5.883	16.684
## 4	8.459	573.537	572.481	5.725	0.644	5.714	16.206
		. . . . .		. . . . .	. . . . .		
## 99	6.972	391.466	388.889	3.889	0.266	4.709	13.357
## 100	11.715	1099.474	1097.932	10.979	0.406	7.913	22.443
## mean	9.670	879.632	797.476	7.975	NA	NA	NA
## sd	2.498	432.957	407.333	4.073	NA	NA	NA



# カイ 2 乗分布 (シミュレーション)

- 表示2.5.2  $\chi^2$  の分布

  - スクリプトファイル

    - Green1-2-5a.R

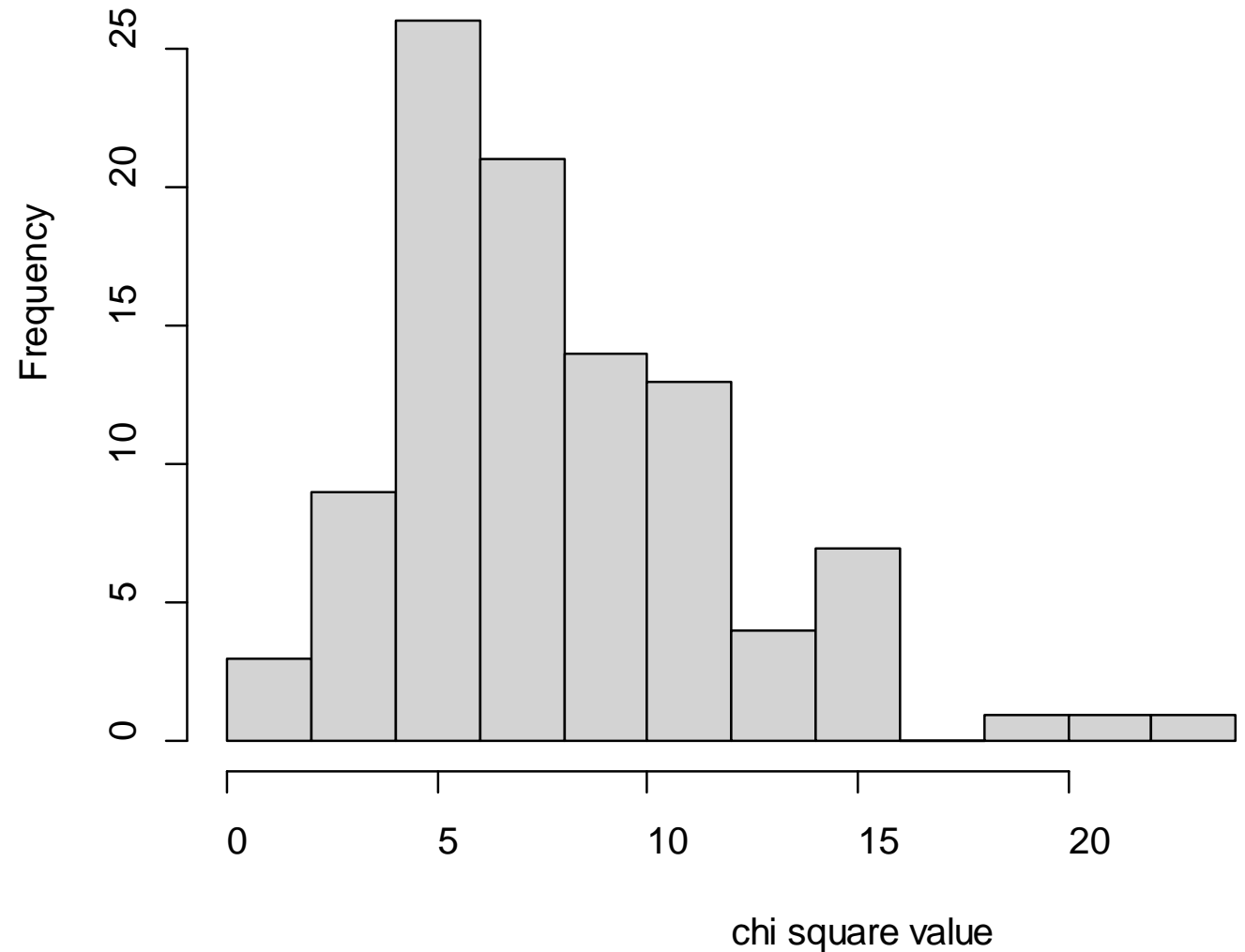
  - 利用した関数

    - hist

  - 方法

    - 行列 mx のカイ 2 乗値の列を

    - hist 関数でヒストグラムを作成





# カイ 2 乗分布

## ●表示2.5.3 自由度によるカイ 2 乗分布の変化

スクリプトファイル

Green1-2-5b.R

利用した関数

curve、dchisq、abline、legend

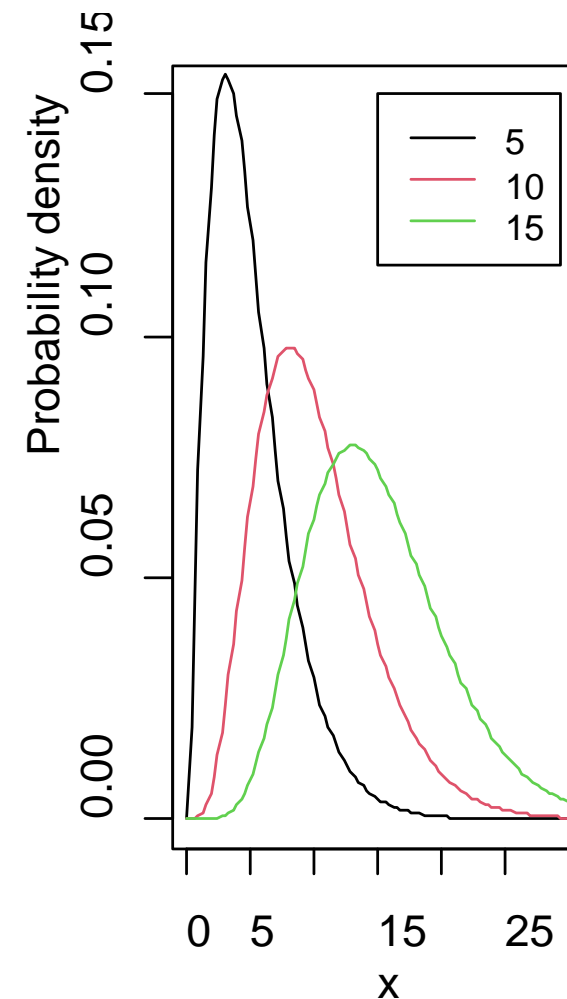
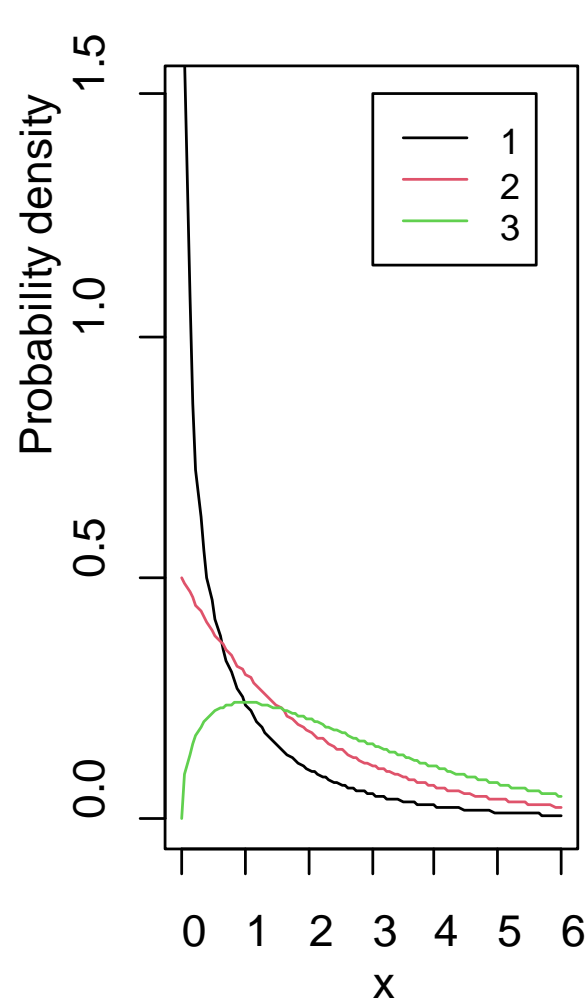
par、plot.new

方法

curve 関数と dchisq関数で

カイ 2 乗分布を描画

legend 関数で凡例を描画





# カイ2乗分布 (シミュレーション)

p.103

## ● 表示2.5.4 標準偏差の点推定と区間推定

スクリプトファイル

Green1-2-5a.R

利用した関数

plot、segments、abline、points

par、plot.new、points

方法

表示2.5.1 のデータを基に、

plot 関数で枠を描画

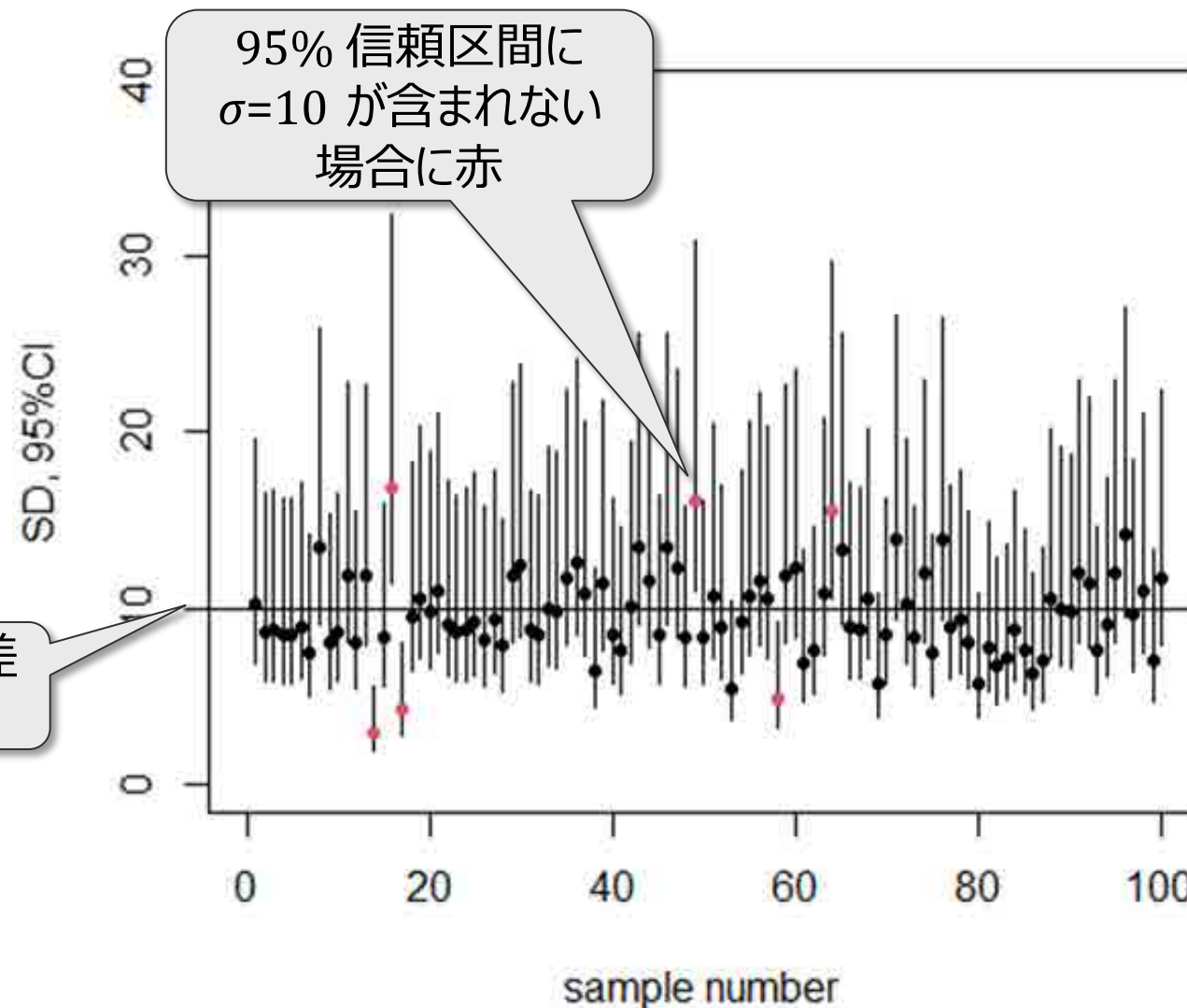
segments 関数で

95%信頼区間を表示

points 関数で点推定値を表示

信頼区間に  $\sigma$  が含まれない場合は赤

(前節 [§2.4](#) 表示2.4.3 参照)



- 表示2.5.5 分散に関する推測

スクリプトファイル

Green1-2-5b.R

利用した関数

pchisq、qchisq、length、  
sum、mean、sd、sqrt

方法

ベクトルにデータを入力し、  
各統計量を求める関数で計算

帰無仮説  $H_0: \sigma = 10$

区間推定

```
x <- c(44, 39, 41, 60, 72, 56, 47, 59, 69)

## n                9.0000000
## sum of squares   1156.8888889
## s.d.             12.0254360
## degree of freedom 8.0000000
## H0 sigma         10.0000000
## chi square       11.5688889
## upper p value    0.1715001
## lower p value    0.8284999
## both sided p value 0.3430002
## alpha            0.0500000
## lower var 95%CI  65.9776922
## upper var 95%CI  530.7485295
## lower SD 95%CI   8.1226653
## upper SD 95%CI   23.0379802
```

- 表示2.5.5 分散に関する推測  
スクリプトファイル  
Green1-2-5.R  
利用した関数  
EnvStats::varTest

```
x <- c(44, 39, 41, 60, 72, 56, 47, 59, 69)
varTest_out <- varTest(x, alternative = "two.sided",
                        sigma.squared = sigma^2,
                        conf.level = 0.95)

## Chi-Squared Test on Variance
##
## data: x
## Chi-Squared = 11.569, df = 8, p-value = 0.343
## alternative hypothesis: true variance is not equal to 100
## 95 percent confidence interval:
## 65.97769 530.74853
## sample estimates:
## variance
## 144.6111

## sigma 95%CI = 8.122665
## sigma 95%CI = 23.03798
```



- 作成 片瀬雅彦
- 作成時期 2021年8月25日