

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
2.6 平均 μ に関する推測 (母標準偏差 σ 未知)

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 σ 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 σ 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

正規乱数によるシミュレーション

● 表示2.4.1 100組のサンプルの観測値と基本統計量 (§2.4)

母標準偏差が既知
(母分散)

スクリプトファイル: Green1-2-4a.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

```
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 # 癖ラフの最大値

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

n = 9 のサンプルが
100組

← 平均値、u 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2.4.3 仮説検定と区間推定の関係 (σ 既知)

スクリプトファイル

Green1-2-4a.R

利用した関数

mean、sd、qnorm

signif、nownames、colnan

方法

行ごとの平均

u 値、p 値、95%信頼区間
(母標準偏差が既知)

$H_0 : \mu=50$ に対するp 値

p < 0.05 の行にマーク「*」

```
##                                     mean
## [1,] 44.4 42.9 72.0 42.8 49.3 44.0 60.7 42.7 53.6 50.3
## [2,] 47.7 52.6 63.1 42.5 38.3 40.1 49.7 34.6 43.4 45.8
## [3,] 65.6 47.5 47.3 40.6 43.7 60.3 49.7 43.1 58.6 50.7
## . . . . .
## [99,] 47.6 43.9 50.2 37.4 53.6 61.8 54.2 52.3 53.7 50.5
## [100,] 39.7 38.1 62.5 46.5 55.5 68.6 33.9 56.5 52.4 50.4
```

```
##      mean  u.value p.value lower.95CI upper.95CI sig
## 1  50.27  0.07977  0.9364      43.73      56.8
## 2  45.78 -1.267    0.2051      39.24      52.31
## 3  50.7   0.2097   0.8339      44.17      57.23
## . . . . .
## 31 58.19  2.458    0.01398      51.66      64.73  *
## . . . . .
## 99 50.54  0.1605   0.8725      44         57.07
## 100 50.41  0.1242   0.9012      43.88      56.95
```

標準偏差に関する検定と推定（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（[§2.5](#)）

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（前節 [§2.4](#)）、n=9 のサンプル100組
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

母標準偏差は未知
推定値として
サンプルの標準偏差を
使う

平均に関する検定と推定 (シミュレーション)

● 表示2.6.1 仮説検定と区間推定の関係 (σ 未知)

スクリプトファイル: Green1-2-6a.R

利用した関数: c、cbind、rownames、colnames、colMeans、apply、signif

方法: 正規乱数 900 個を発生させ、マトリックス (9 × 100) に格納

9 個ずつのサンプルの平均値、標準誤差、t 値、 $H_0: \mu=50$ の p 値、 μ の 95%CI を算出

##		mean	s.e.	t value	p value	lower 95%CI	upper 95%CI	t*s.e.		
##	1	50.266	3.403	0.078	0.940	42.419	58.113	7.847		
##	2	45.776	2.878	-1.468	0.180	39.140	52.412	6.636		
##	3	50.699	2.903	0.241	0.816	44.005	57.393	6.694		
##	4	49.657	2.820	-0.122	0.906	43.155	56.160	6.502		
##	5	46.877	2.813	-1.110	0.299	40.390	53.365	6.487		
									
##	99	50.535	2.324	0.230	0.824	45.176	55.894	5.359		
##	100	50.414	3.905	0.106	0.918	41.409	59.419	9.005		
##	mean	50.201	3.223	0.069	NA	NA	NA	NA		
##	sd	3.030	0.833	1.007	0.286	3.564	NA	NA		

n = 9 のサンプルが
100組

平均に関する検定と推定 (シミュレーション)

● 表示2.6.1 仮説検定と区間推定の関係 (σ 未知)

スクリプトファイル: Green1-2-6a.R

利用した関数: c、cbind、rownames、colnames、colMeans、apply、signif

方法: 正規乱数 900 個を発生させ、マトリックス (9 × 100) に格納

9 個ずつのサンプルの平均値、標準誤差、t 値、 $H_0: \mu=50$ の p 値、 μ の 95%CI を算出

```
s_sd <- apply(mx, 1, sd)
s_sem <- s_sd / sqrt(Nc)
s_t <- (s_mean - Mu) / s_sem
```

##	mean	s.e.	t value	p value	lower 95%CI	upper 95%CI	t*s.e.
## 1	50.266	3.403	0.078	0.940	42.419	58.113	7.847
## 2	49.766	3.678	-1.468	0.180	39.140	52.412	6.636
## 3	49.299	2.903	0.241	0.816	44.005	57.393	6.694
## 4	49.577	2.820	-0.122	0.906	43.155	56.160	6.502
## 5	49.777	2.813	-1.110	0.299	40.390	53.365	6.487
.....							
## 99	50.535	2.324	0.230	0.824	45.176	55.894	5.359
## 100	50.414	3.905	0.106	0.918	41.409	59.419	9.005
## mean	50.201	3.223	0.069	NA	NA	NA	NA
## sd	3.030	0.833	1.007	0.286	3.564	NA	NA

平均に関する検定と推定（シミュレーション）

- 表示2.6.1 仮説検定と区間推定の関係（ σ 未知）

スクリプトファイル：Green1-2-6a.R

利用した関数：c、cbind、rownames、colnames、colMeans、apply、signif

方法：正規乱数 900 個を発生させ、マトリックス（ 9×100 ）に格納

9 個ずつのサンプルの平均値、標準誤差、t 値、 $H_0: \mu=50$ の p 値、 μ の 95%CI を算出

##	mean	s.e.	t value	p value	lower 95%CI	upper 95%CI	t*s.e.
## 1	50.266	3.403	0.078	0.940	42.419	58.113	7.847
## 2	45.776	2.878	-1.468	0.180	39.140	52.412	6.636
## 3	50.699	2.903	0.241	0.816	44.005	57.393	6.694
## 4	49.657	2.820	-0.122	0.906	43.155	56.160	6.502
## 5	46.877	2.813	-1.110	0.299	40.390	53.365	6.487
						
## 99	50.535	2.324	0.230	0.824	45.176	55.894	5.359
## 100	50.414	3.905	0.106	0.918	41.409	59.419	9.005
## mean	50.201	3.223	0.069	NA	NA	NA	NA
## sd	3.030	0.833	1.007	0.286	3.564	NA	NA

- 表示2.6.1 仮説検定と区間推定の関係 (σ 未知)

スクリプトファイル

Green1-2-6a.R

利用した関数

plot、segments、abline、points

ifelse、par

方法

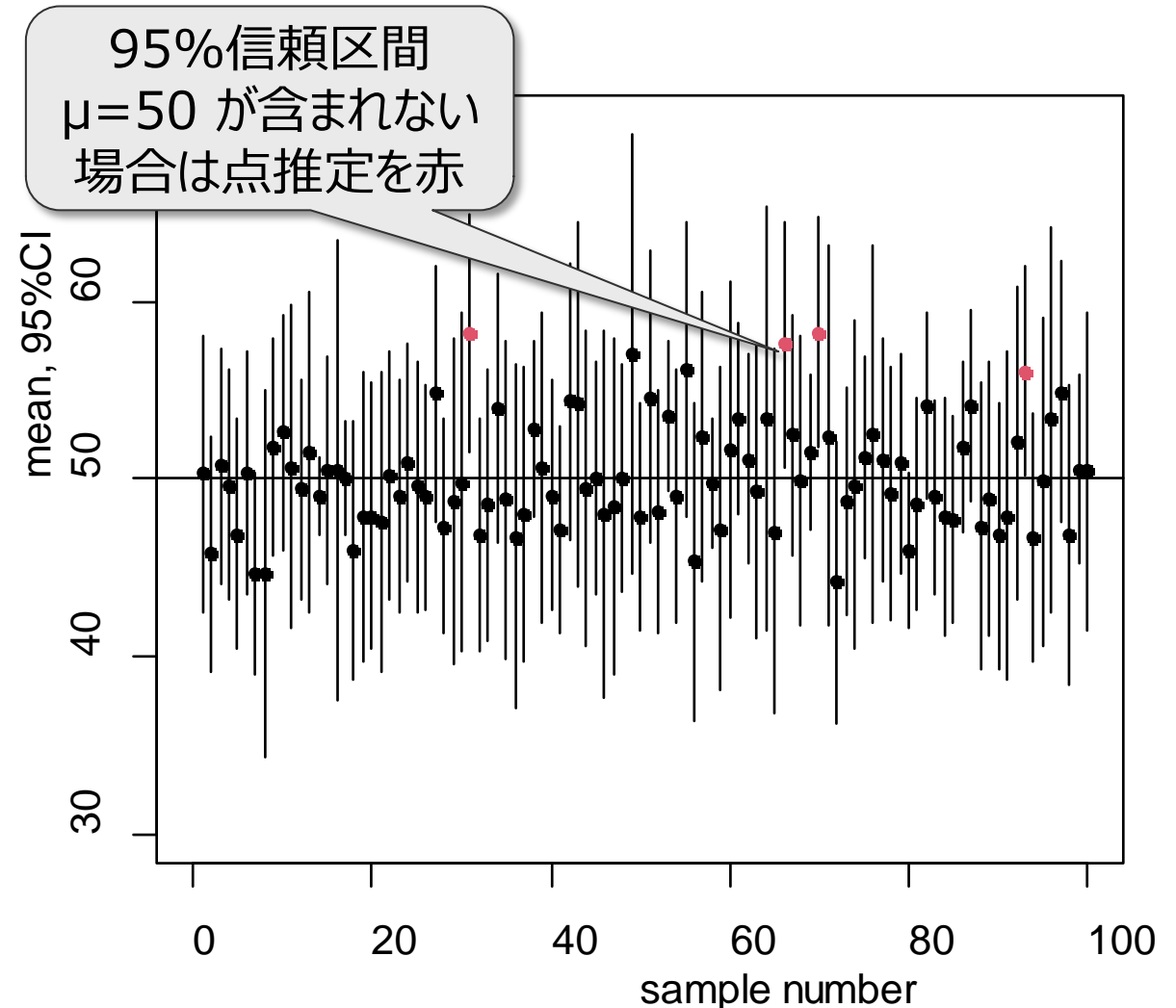
plot 関数で枠組みを表示

segments 関数で母平均の95%信頼区間を表示

points 関数で母平均の点推定値を表示

95%信頼区間に 50 が含まれない場合は赤

(前々節 [§2.4](#) 表示2.4.3 参照)



●表示2.6.2 標準正規分布と t 分布

スクリプトファイル

Green1-2-6c.R

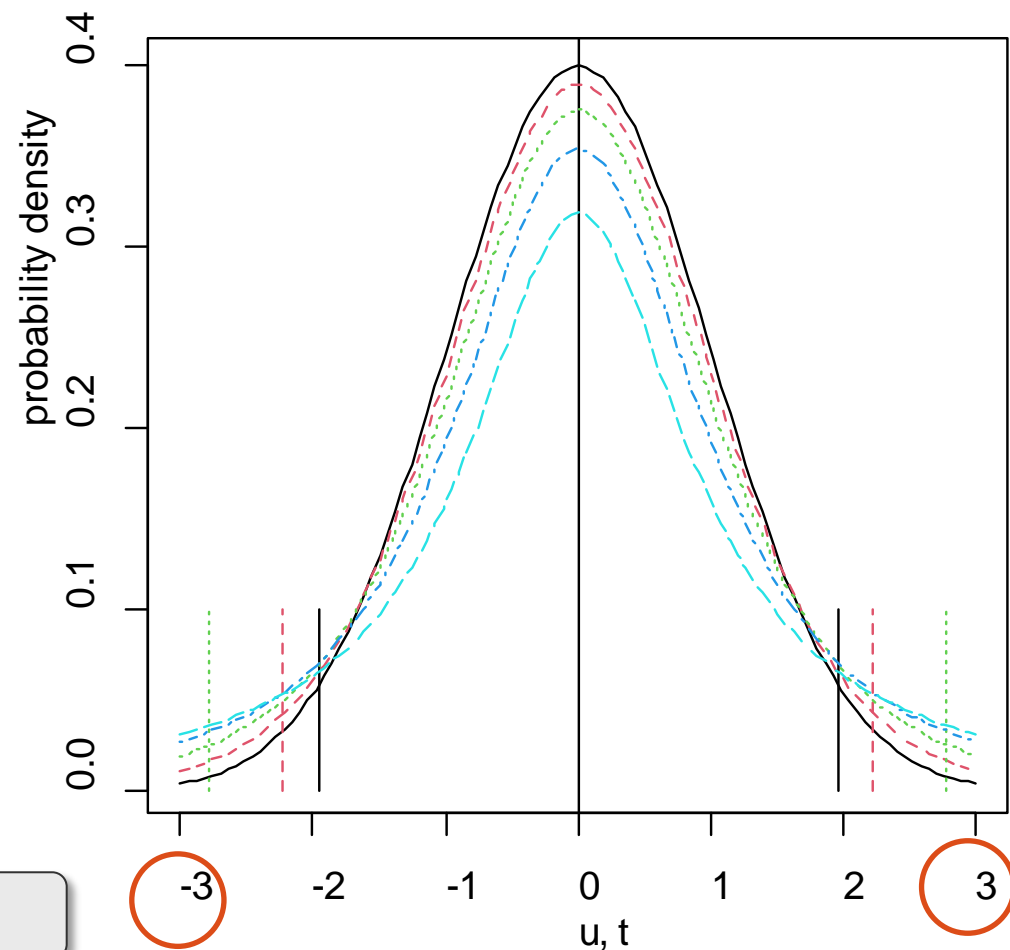
利用した関数

curve、dnorm、dt、legend

方法

```
curve(dnorm(x, mean = 0, sd = 1), -3, 3)
curve(dt(x, df = 10), -3, 3, add = TRUE)
curve(dt(x, df = 4), -3, 3, add = TRUE)
curve(dt(x, df = 2), -3, 3, add = TRUE)
curve(dt(x, df = 1), -3, 3, add = TRUE)
```

x 軸の最大値と最小値



● 表示2.6.3 標準正規分布と t 分布

スクリプトファイル：Green1-2-6c.R

利用した関数：curve、dnorm、dt、function、sqrt、segments、abline、par
方法

```
mu0 <- 50 # 帰無仮説
mu1 <- 60 # 対立仮説
sigma <- 10 # 母標準偏差
s <- 10 # サンプルの標準偏差
n <- 9 # サンプルサイズ
alpha <- 0.05 # 有意水準
dof <- n - 1 # 自由度
conf <- 1 - alpha / 2 # 信頼率
```

シグマ既知の場合の平均値の分布

```
u0 <- (mu0 - mu0) / (sigma / sqrt(n)) # = 0
u1 <- (mu1 - mu0) / (sigma / sqrt(n))
f_u0 <- function(u, u0) dnorm(u, u0, 1) # 帰無仮説 H0
f_u1 <- function(u, u1) dnorm(u, u1, 1) # 対立仮説 H1
limN <- qnorm(conf, 0, 1) # 棄却限界値
```

シグマ未知の場合の平均値の分布

```
t0 <- (mu0 - mu0) / (s / sqrt(n))
t1 <- (mu1 - mu0) / (s / sqrt(n))
f_t0 <- function(t, dof, t0) dt(t, dof, t0) # H0
f_t1 <- function(t, dof, t1) dt(t, dof, t1) # H1
limT <- qt(conf, dof, 0)
```

t 分布と非心 t 分布

●表示2.6.3 標準正規分布と t 分布

スクリプトファイル：Green1-2-6c.R

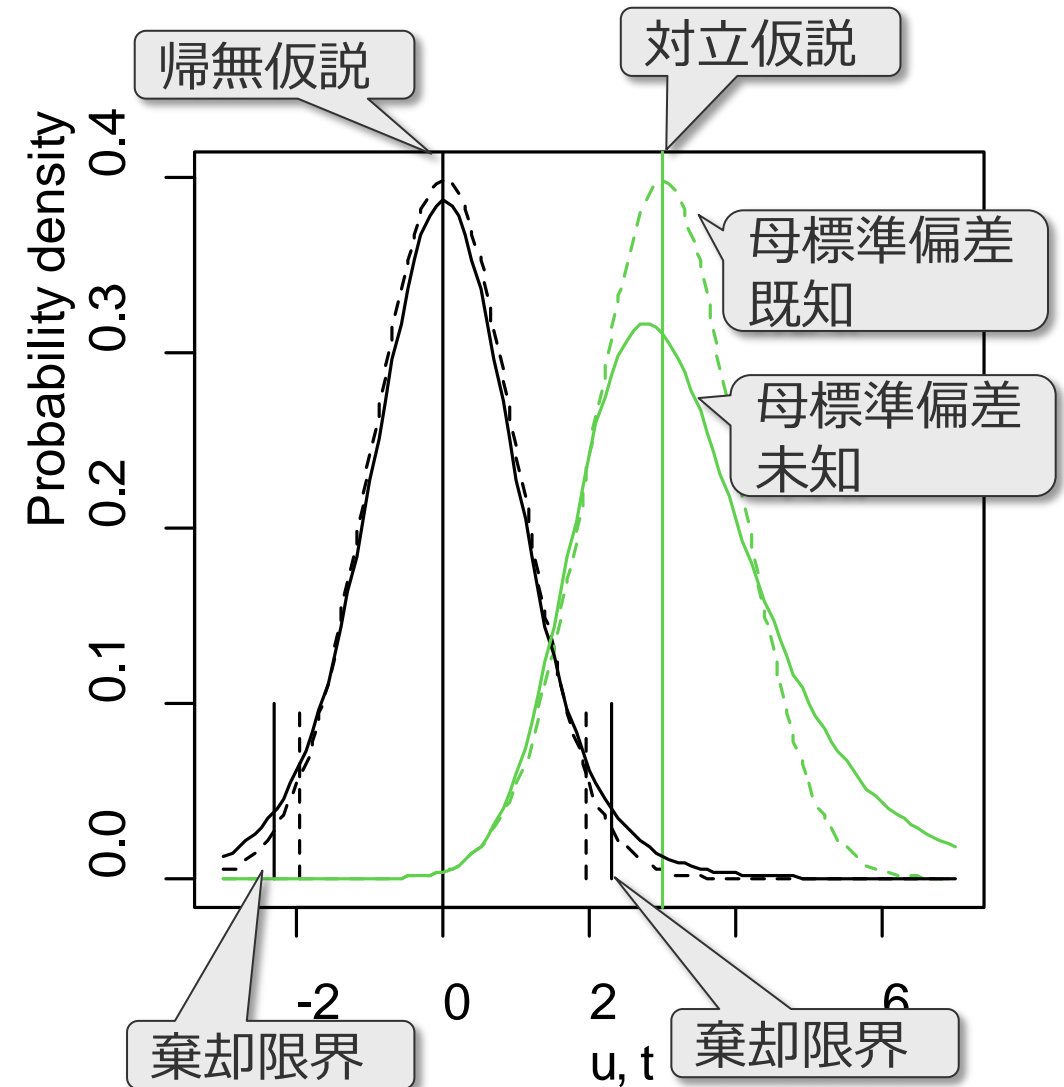
利用した関数：curve、dnorm、dt など

方法

```
xMin <- -3
xMax <- 7
curve(f_u0(u = x, u0), xMin, xMax)
segments(limN, 0, limN, 0.1, lty = 2)
segments(-limN, 0, -limN, 0.1, lty = 2)
curve(f_u1(u = x, u1), xMin, xMax,
      add = TRUE)
abline(v = u0, col = 1)
abline(v = u1, col = 3)

curve(f_t0(t = x, dof, t0), xMin, xMax,
      add = TRUE)
curve(f_t1(t = x, dof, t1), xMin, xMax,
      add = TRUE)
```

非心度



平均に関する検定と推定（母標準偏差 σ 未知）

p.109

● 表示2.6.4 平均値に関する推測（ σ 未知）

スクリプトファイル：Green1-2-6d.R

```
mu_H0 <- 50          # 帰無仮説
alpha <- 0.05        # 有意水準
x      <- c(44, 39, 41, 60, 72, 56, 47, 59, 69)

n      <- length(x)  # サンプルサイズ
s_mean <- mean(x)    # サンプルの平均値
s_sd   <- sd(x)      # サンプルの標準偏差
dof    <- n - 1      # サンプルの自由度
sem    <- s_sd / sqrt(n) # 平均の標準誤差
t      <- (s_mean - mu_H0) / sem
p      <- pt(abs(t), df = dof, lower.tail = FALSE) * 2

temp   <- qt(1 - alpha / 2, dof) * sem
ciL    <- s_mean - temp   # 母平均の95%信頼区間
ciU    <- s_mean + temp   # 母平均の95%信頼区間
```

```
## n          9.000
## mean       54.111
## s.d.       12.025
## H0: mu     50.000
## s.e.       4.008
## t value    1.026
## p value    0.335
## alpha(two sided) 0.050
## lower CI   44.868
## upper CI   63.355
```

一変量の分布 (母標準偏差 σ 未知)

●表示2.6.7 JMP による解析結果(1)

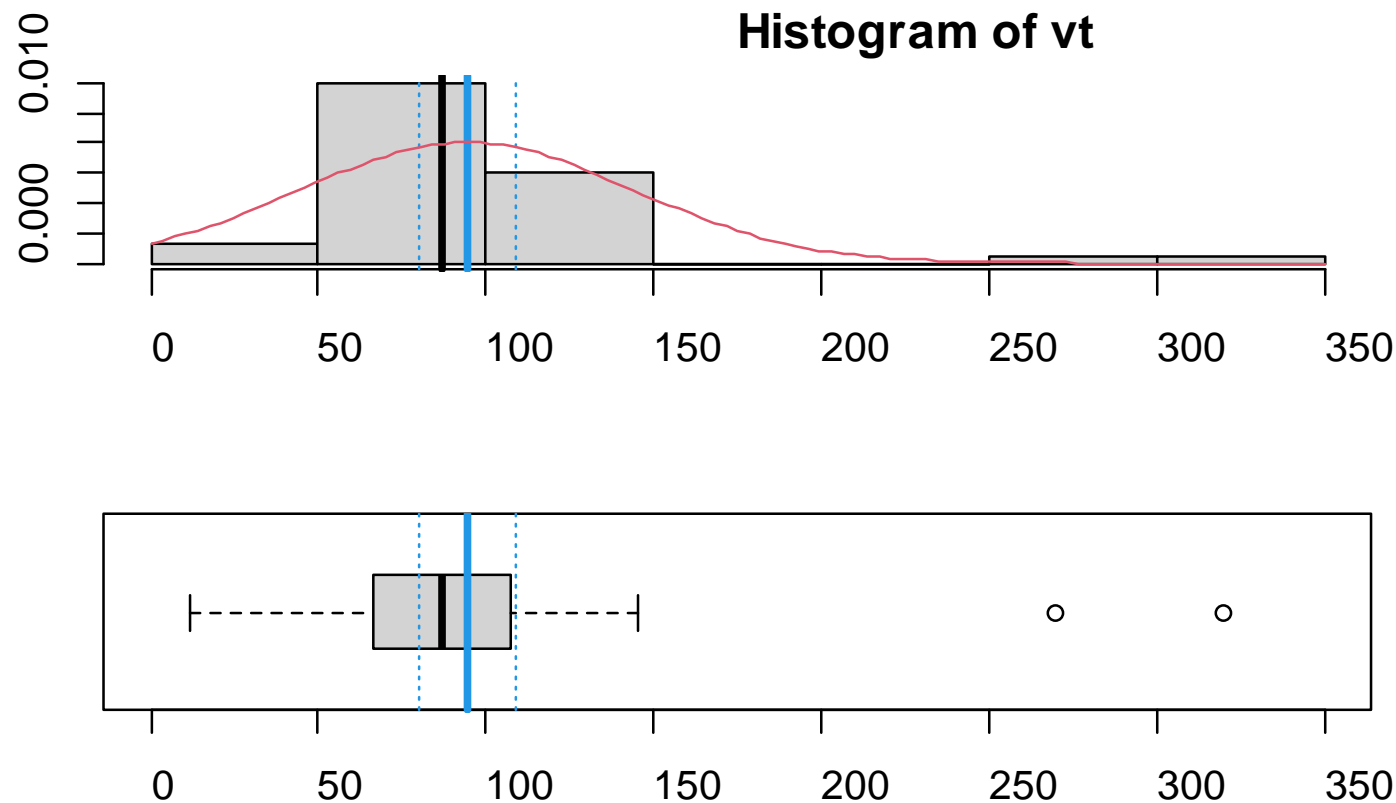
スクリプトファイル：Green1-2-6b.R

利用した関数：length、mean、sd、median、e1071::skewness、e1071::kurtosis、hist、boxplot

方法：表に関して [§2.1](#) 参照

グラフに関して [§2.2](#) 参照

```
## mean      94.340000
## sd        49.735057
## se         7.033599
## median    86.500000
## lower 95%CI 80.205453
## upper 95%CI 108.474547
## N         50.000000
## variance  2473.575918
## CV         52.718950
## skewness   2.729118
## kurtosis  10.544280
```



- 表示2.6.8 JMP による解析結果(2)

表示2.6.9 JMP による解析結果(3)

スクリプトファイル

Green1-2-6b.R

利用した関数

BSDA::z.test、readxl::read_excel

t.test、

方法

z.test 関数により、z 検定を実施
(母標準偏差が既知)

t.test 関数により、t 検定を実施
(母標準偏差が未知)

```
## One-sample z-Test
## data: vt
## z = -0.80044, p-value = 0.4235
## alternative hypothesis: true mean
## is not equal to 100
## 95 percent confidence interval:
## 80.48096 108.19904
## sample estimates:
## mean of x 94.34

## One Sample t-test
## data: vt
## t = -0.80471, df = 49, p-value = 0.4249
## alternative hypothesis: true mean
## is not equal to 100
## 95 percent confidence interval:
## 80.20545 108.47455
## sample estimates:
## mean of x 94.34 (一部表示を改変)
```

- 表示2.6.8 JMP による解析結果(2)

- 表示2.6.9 JMP による解析結果(3)

- スクリプトファイル：Green1-2-6b.R

- 利用した関数：EnvStats::varTest

- 方法：varTest 関数により、標準偏差の検定（母分散に関するカイ 2 乗検定）を実施

```
## Chi-Squared Test on Variance
##
## data: vt
## Chi-Squared = 33.668, df = 49, p-value = 0.09318
## alternative hypothesis: true variance is not equal to 3600
## 95 percent confidence interval:
## 1726.019 3841.088  $\leftarrow \sqrt{1726.019} = 41.545 \quad \sqrt{3841.088} = 61.977$ 
## sample estimates:
## variance
## 2473.576  $\leftarrow \sqrt{2473.576} = 49.735$  (一部表示を改変)
```



- 作成 片瀬雅彦
- 作成時期 2021年8月25日