

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

2.4 平均 μ に関する推定 (母標準偏差 σ 既知)

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 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

●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：Green1-2-4a.R

利用した関数

rnorm、matrix、mean、sd、cbind

round、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 # 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

```
##           [,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2.4.1 100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：Green1-2-4a.R

利用した関数

mean、sd、sqrt

signif、rownames、colnames

方法

apply 関数で行列の各行の
平均値を計算（[§2.1](#) 参照）

```
s_mean <- apply(mx, 1, mean) # サンプル (Nc個) の平均
t_mean <- mean(mx)          # 全体の平均
t_sd    <- sd(mx)           # 全体の標準偏差
m_mean  <- mean(s_mean)     # 平均の平均
m_sd    <- sd(s_mean)       # 平均の標準誤差 (推定値)
sem     <- Sigma / sqrt(Nc) # 平均の標準誤差 (理論値)
```

```
##
## [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
```



平均に関する検定と推定 (シミュレーション)

● 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「*」

表示2.4.1 で得られた
n=9 のサンプル 100組

##											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	43.1	51.8	54.2	52.3	53.7	50.5			
##	[100,]	39.7	38.6	58.6	33.9	56.5	52.4	50.4			

母標準偏差が既知
(母分散)

##		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	

平均に関する検定と推定 (シミュレーション)

- 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

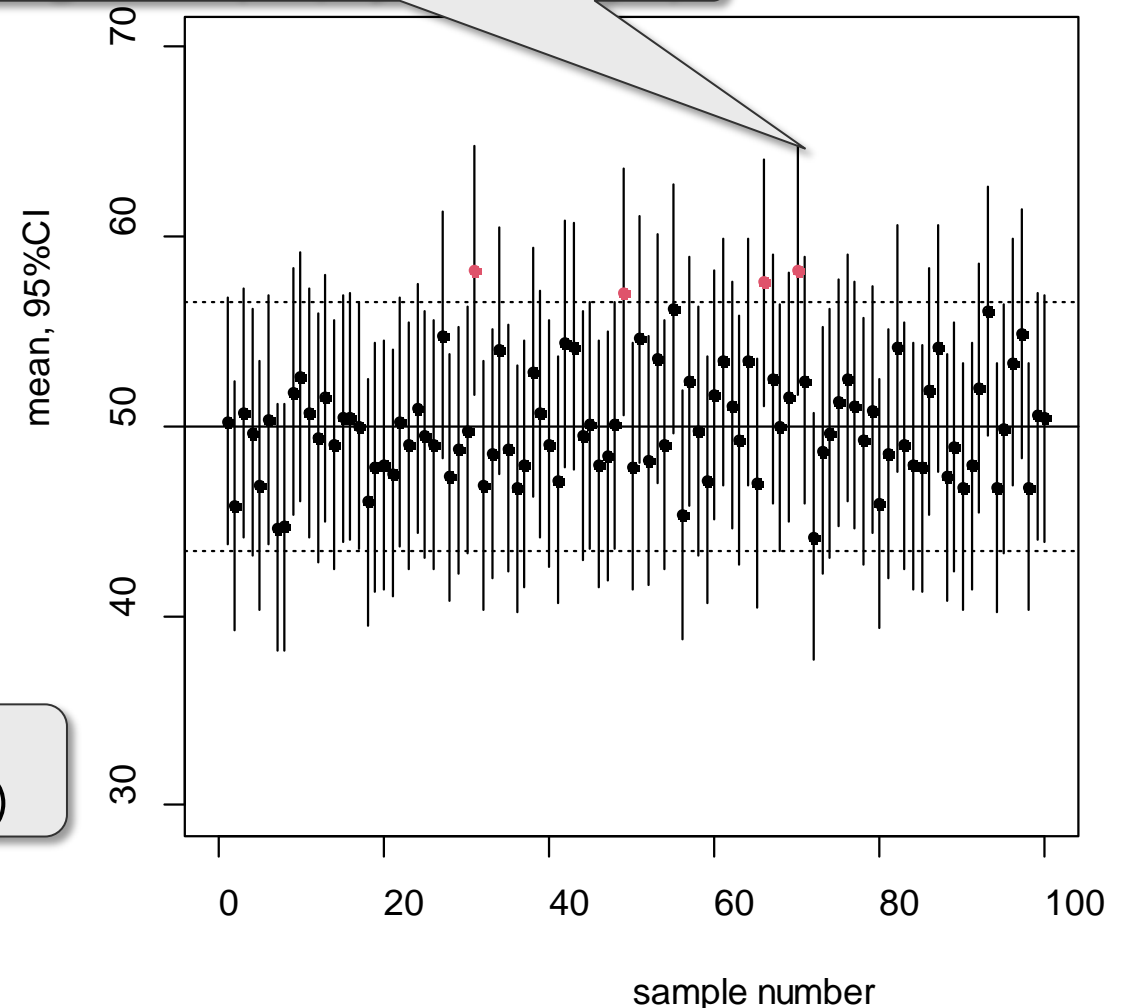
利用した関数: plot.new、par、plot、
segments、abline、points

方法

```
plot(NA,  
     xlim = c(0, Nr), ylim = c(Xmin, Xmax),  
     xlab = "sample number",  
     ylab = "mean, 95%CI")  
segments(1:Nr, s_ciL, 1:Nr, s_ciU)  
abline(h = Mu)  
abline(h = critU, lty = 3)  
abline(h = critL, lty = 3)  
color <- 2 -  
      (critL < s_mean) * (s_mean < critU)  
points(1:Nr, s_mean, col = color, pch = 20)
```

x と y はベクトル
(要素数が100個)

segments(x1, y1, x2, y2)



平均に関する検定と推定 (シミュレーション)

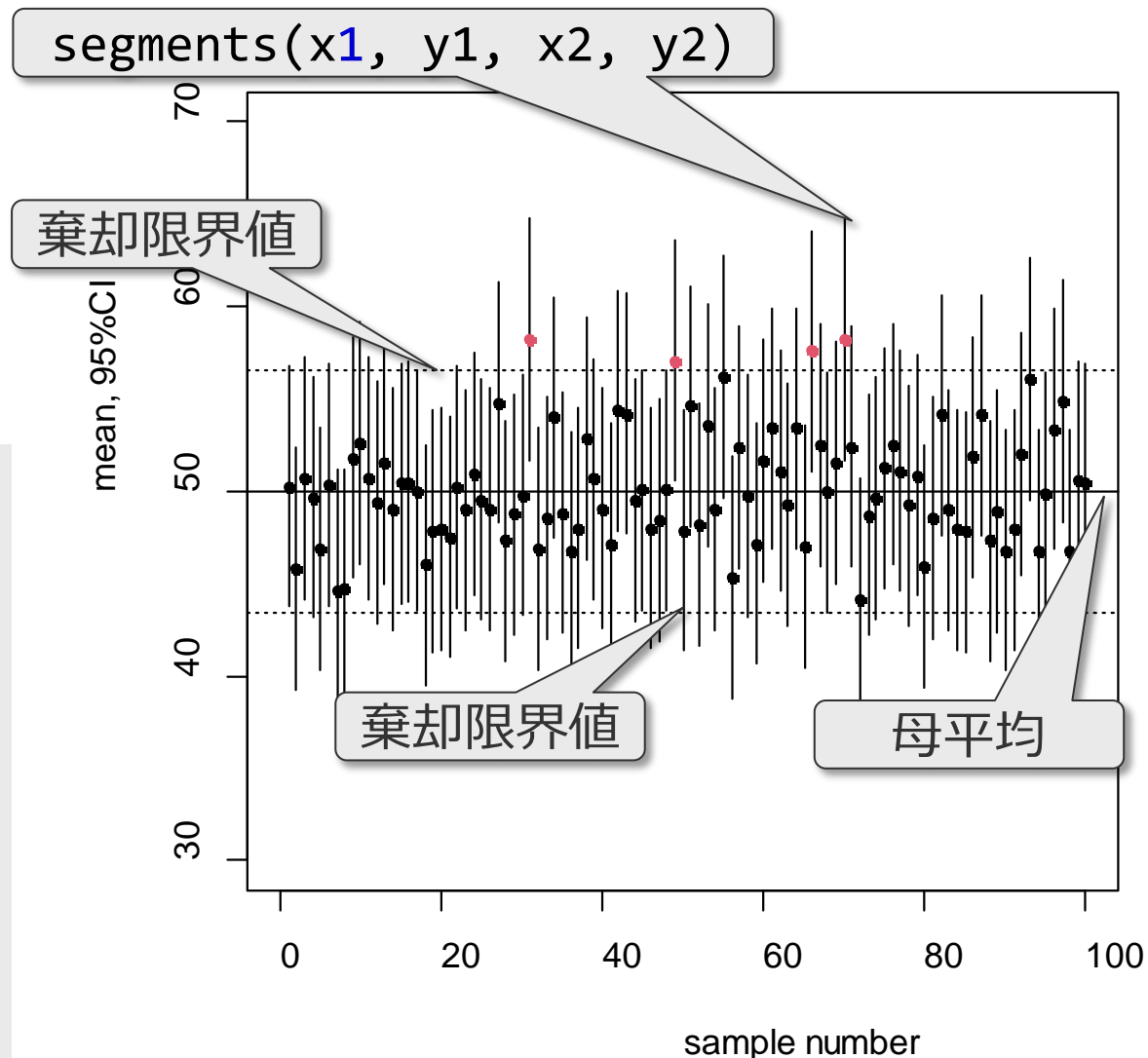
- 表示2.4.3 仮説検定と区間推定の関係 (σ 既知)

スクリプトファイル: Green1-2-4a.R

利用した関数: plot.new、par、plot、
segments、abline、points

方法

```
plot(NA,  
     xlim = c(0, Nr), ylim = c(Xmin, Xmax),  
     xlab = "sample number",  
     ylab = "mean, 95%CI")  
segments(1:Nr, s_ciL, 1:Nr, s_ciU)  
abline(h = Mu)  
abline(h = critU, lty = 3)  
abline(h = critL, lty = 3)  
color <- 2 -  
      (critL < s_mean) * (s_mean < critU)  
points(1:Nr, s_mean, col = color, pch = 20)
```



平均に関する検定と推定 (シミュレーション)

- 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

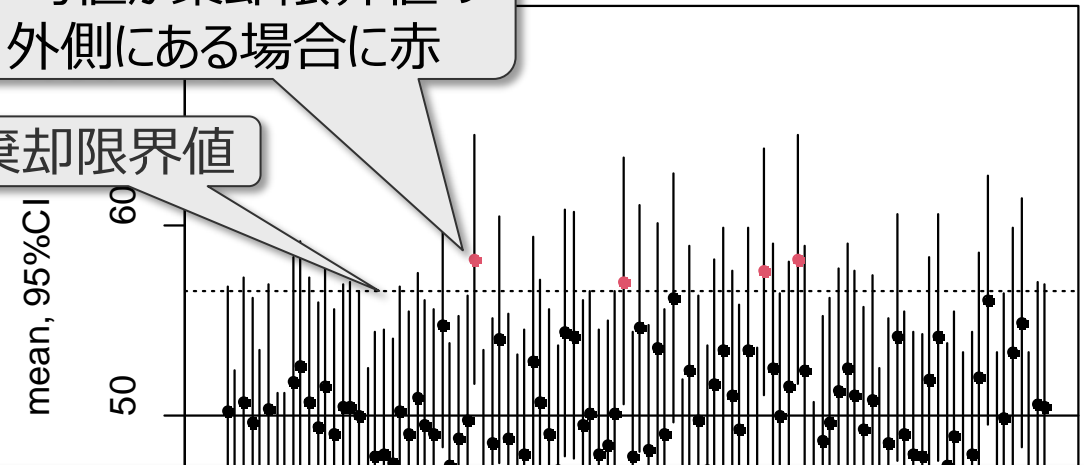
利用した関数: plot.new、par、plot、segments、abline、points

方法

```
plot(NA,
     xlim = c(0, Nr), ylim = c(Xmin, Xmax),
     xlab = "sample number",
     ylab = "mean, 95%CI")
segments(1:Nr, s_ciL, 1:Nr, s_ciU)
abline(h = Mu)
abline(h = critU, lty = 3)
abline(h = critL, lty = 3)
color <- 2 -
      (critL < s_mean) * (s_mean < critU)
points(1:Nr, s_mean, col = color, pch = 20)
```

平均値が棄却限界値の外側にある場合に赤

棄却限界値



平均値が棄却限界値の上限の外側の場合

$$2 - (\text{critL} < s_mean) * (s_mean < \text{critU}) = 2 - (1) * (0) = 2 - 0 = 2$$

平均値が棄却限界値の上限と下限の内側の場合

$$2 - (\text{critL} < s_mean) * (s_mean < \text{critU}) = 2 - (1) * (1) = 2 - 1 = 1$$

sample number

x と \bar{x} の分布の関係 (母標準偏差 σ 既知)

p.90

● 表示2.4.2 x と \bar{x} の分布の関係

スクリプトファイル: Green1-2-4b.R

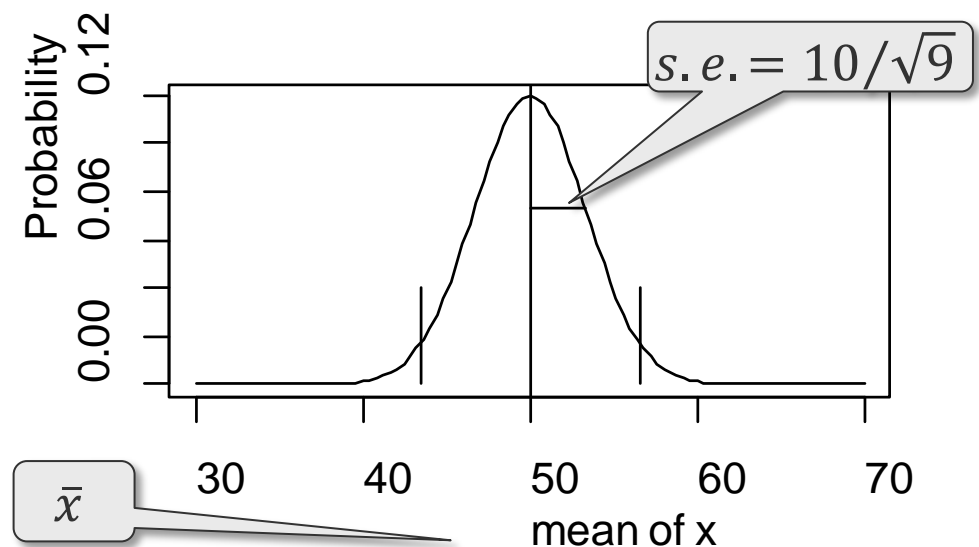
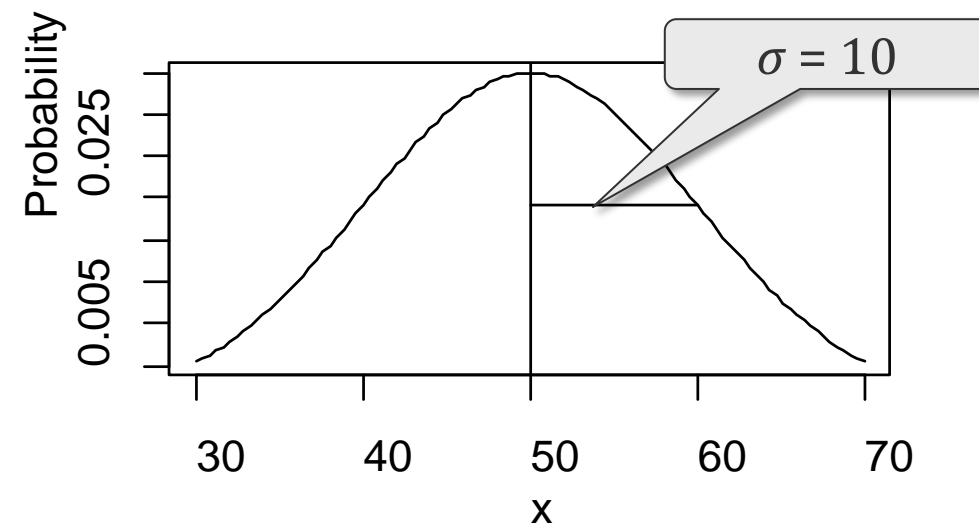
利用した関数: curve、dnorm、pnorm、qnorm

abline、segments、pars、rownames、matrix

方法: curve 関数とdnorm 関数で正規分布を描画

segments 関数と abline 関数で直線を追加

```
##           [,1]
## mu       50.000000
## sigma    10.000000
## n        9.000000
## alpha    0.050000
## u        1.959964
## lower critical x 43.466787
## upper critical x 56.533213
```



x と \bar{x} の分布の関係 (母標準偏差 σ 既知)

●表示2.4.4 x と \bar{x} の分布の関係

スクリプトファイル: Green1-2-4b.R

利用した関数: curve、dnorm、qnorm、pnorm

abline、segments、pars、polygon、text、seq

方法: curve 関数とdnorm 関数で

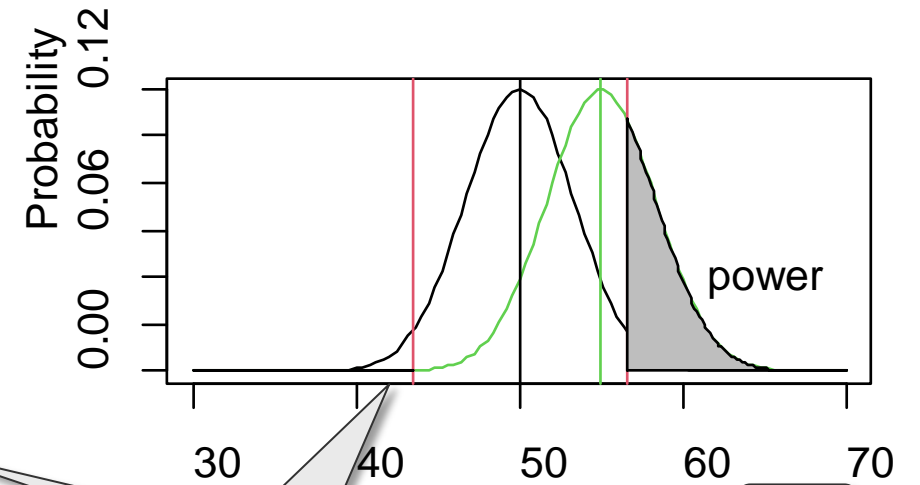
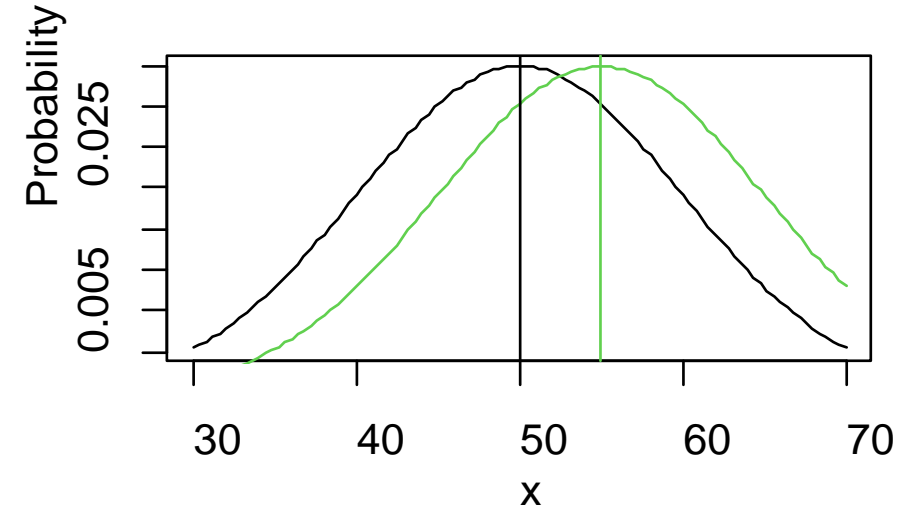
正規分布を描画、segments 関数

abline 関数で直線を追加

polygon 関数で塗りつぶし

```
##           [,1]
## Sigma 10.00000
## n      9.00000
## sem   3.33333
## Alpha 0.05000
## u     1.959964
```

```
##           H0      H1
## Mu       50.00000 55.00000
## lower critical x 43.46679 NA
## upper critical x 56.53321 NA
## lower significance p 0.02500 0.00027
## upper significnace p 0.02500 0.32277
## total significance p 0.05000 0.32304
```



左 ごく僅か

\bar{x}

- 表示2.4.5 n と μ による検出力の変化

スクリプトファイル：Green1-2-4c.R

利用した関数：pnorm、qnorm、ifelse、fuction、matrix

方法：pnorm 関数により棄却限界値の外側の確率を計算する関数を、function 関数で定義

対立仮説の母平均：48~64、サンプルサイズ：2~80

n 、 μ による検出力の変化（両側検定）

サンプル
サイズ

対立仮説の母平均

##	48	49	50	51	52	53	54	56	58	60	62	64
## 80	0.432	0.145	0.05	0.145	0.432	0.765	0.947	1.000	1.000	1.000	1.000	1.000
## 40	0.244	0.097	0.05	0.097	0.244	0.475	0.716	0.967	0.999	1.000	1.000	1.000
## 20	0.145	0.073	0.05	0.073	0.145	0.269	0.432	0.765	0.947	0.994	1.000	1.000
## 10	0.097	0.062	0.05	0.062	0.097	0.158	0.244	0.475	0.716	0.885	0.967	0.993
## 5	0.073	0.056	0.05	0.056	0.073	0.103	0.145	0.269	0.432	0.609	0.765	0.879
## 3	0.064	0.053	0.05	0.053	0.064	0.081	0.107	0.180	0.283	0.410	0.547	0.679
## 2	0.059	0.052	0.05	0.052	0.059	0.071	0.087	0.136	0.205	0.293	0.396	0.508

- 表示2.4.5 n と μ による検出力の変化（補足）

スクリプトファイル：Green1-2-4c.R

利用した関数：pnorm、qnorm、ifelse、fuction、matrix

方法：pnorm 関数により棄却限界値の外側の確率を計算する関数を、function 関数で定義

対立仮説の母平均：48~64、サンプルサイズ：2~80

n 、 μ による検出力の変化（片側検定）

対立仮説の母平均

サンプル
サイズ

##	48	49	50	51	52	53	54	56	58	60	62	64
## 80	0.000	0.006	0.05	0.226	0.557	0.850	0.973	1.000	1.000	1.000	1.000	1.000
## 40	0.002	0.011	0.05	0.156	0.352	0.600	0.812	0.984	1.000	1.000	1.000	1.000
## 20	0.006	0.018	0.05	0.116	0.226	0.381	0.557	0.850	0.973	0.998	1.000	1.000
## 10	0.011	0.025	0.05	0.092	0.156	0.243	0.352	0.600	0.812	0.935	0.984	0.997
## 5	0.018	0.031	0.05	0.078	0.116	0.165	0.226	0.381	0.557	0.723	0.850	0.931
## 3	0.023	0.035	0.05	0.071	0.097	0.130	0.171	0.272	0.398	0.535	0.668	0.782
## 2	0.027	0.037	0.05	0.066	0.087	0.111	0.140	0.213	0.304	0.409	0.521	0.631

検出力 (母標準偏差 σ 既知)

- 表示2.4.6 検出力曲線

スクリプトファイル

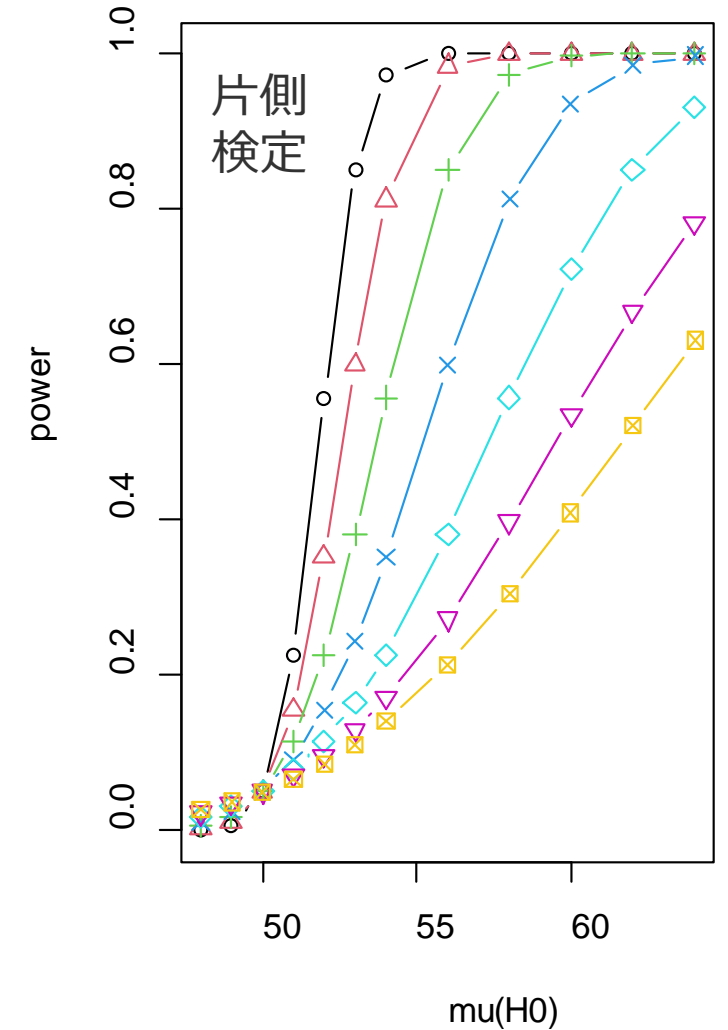
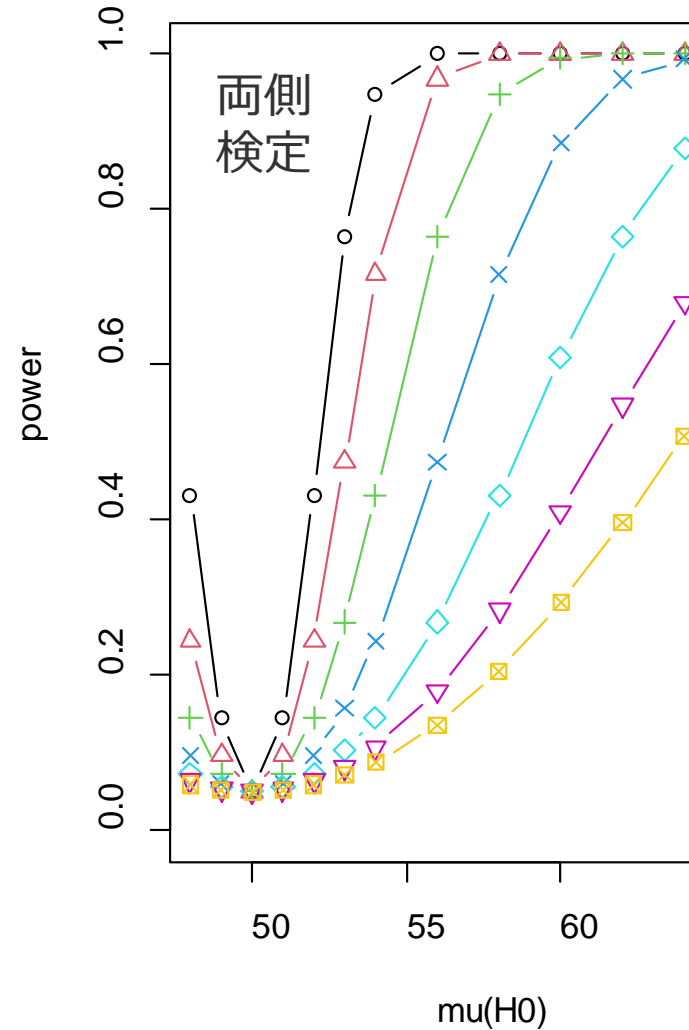
Green1-2-4c.R

利用した関数

plot、lines、for

方法

plot 関数で大枠を描画し、
lines 関数により個々の折れ線を描画



平均値に関する検定と推定 (母標準偏差 σ 既知)

p.98

● 表示2.4.7 平均値に関する推測 (σ 既知)

スクリプトファイル: Green1-2-4d.R

利用した関数: pnorm、qnorm、length、mean、c、cbind

方法

```
mu_H0 <- 50      # 帰無仮説
sigma <- 10      # 母標準偏差
alpha <- 0.05    # 有意水準
x      <- c(44, 39, 41, 60, 72, 56, 47, 59, 69)

n      <- length(x)
s_mean <- mean(x)
sem    <- sigma / sqrt(n)
u      <- (s_mean - mu_H0) / sem
p      <- pnorm(abs(u), lower.tail = FALSE) * 2

temp   <- qnorm(1 - alpha / 2) * sem
ciL    <- s_mean - temp      # 95%信頼区間 上限
ciU    <- s_mean + temp      # 95%信頼区間 下限
```

個々の観測値

母標準偏差が既知
(母分散)

```
##           [,1]
## n           9.000
## mean        54.111
## H0: mu      50.000
## sigma       10.000
## s.e.         3.333
## u value     1.233
## p value     0.217
## alpha(two sided) 0.050
## lower CI    47.578
## upper CI    60.644
```



- 作成 片瀬雅彦
- 作成時期 2021年8月26日