

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
4 相関・回帰
4.4 誤差を考慮した推定

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

●表示4.4.3 区間推定

表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル

Green1-4-4.R

利用した関数

lm

方法（全体）

Excel ファイルからデータを読み込、df に付値

lm 関数で回帰分析

結果を lm_out に付値

```
df <- read_excel("Green1-4.xlsx",  
                 sheet = "4-soukan3")  
df <- data.frame(df)  
lm_out <- lm(y ~ x, data = df)
```

回帰分析の結果
を収めたオブジェクト

Excel ファイルのデータ

x	y
1	5
3	5
4	7
5	6
7	9
10	10

単回帰分析：グラフ（信頼区間、予測区間）

p.245

●表示4.4.3 区間推定

表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル：Green1-4-4.R

利用した関数：lm、predict

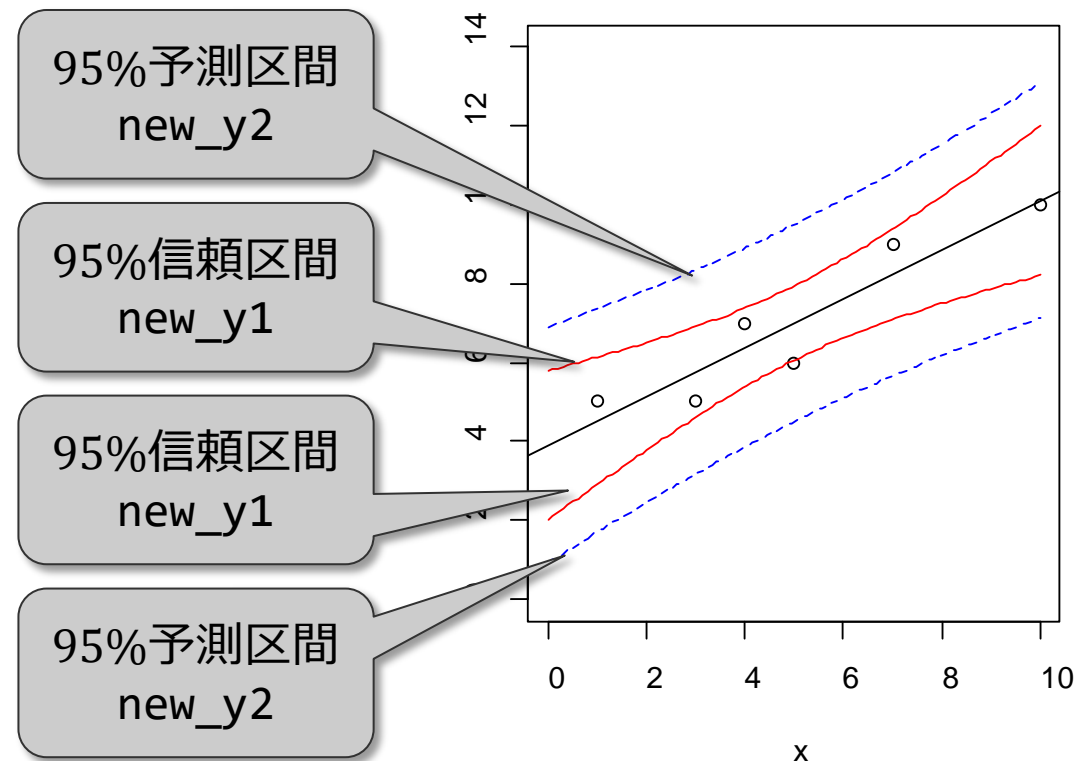
plot、abline、lines

方法：lm 関数の結果からpredict 関数で

95%信頼区間、95%予測区間を推定

(信頼区間：応答の期待値 η に対する信頼区間)

(予測区間：個別の応答 y に対する信頼区間)



```
new_x <- data.frame(x = seq(0, 10, length = 100))
new_y1 <- predict(lm_out, newdata = new_x,
                  interval = "confidence", level = 0.95)
new_y2 <- predict(lm_out, newdata = new_x,
                  interval = "prediction", level = 0.95)
```

単回帰分析：グラフ（信頼区間、予測区間）

p.245

●表示4.4.3 区間推定

表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル：Green1-4-4.R

利用した関数：lm、predict

plot、abline、lines

方法：lm 関数の結果からpredict 関数で

95%信頼区間、95%予測区間を推定

(信頼区間：応答の期待値 η に対する信頼区間)

(予測区間：個別の応答 y に対する信頼区間)

new_x\$x

```
new_x <- data.frame(x = seq(0, 10, length = 100))
new_y1 <- predict(lm_out, newdata = new_x,
                  interval = "confidence", level = 0.95)
new_y2 <- predict(lm_out, newdata = new_x,
                  interval = "prediction", level = 0.95)
```

データフレーム

95%信頼区間
new_y1

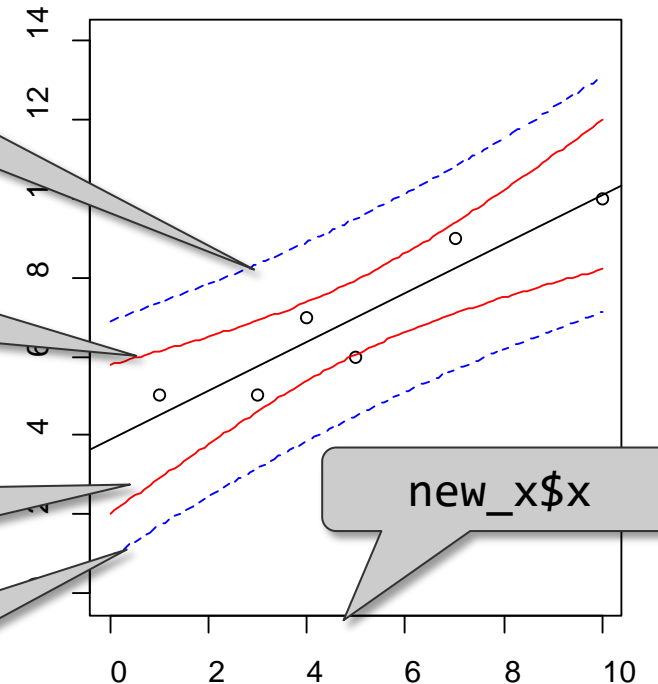
95%予測区間
new_y2

95%予測区間
new_y2

95%信頼区間
new_y1

95%信頼区間
new_y1

95%予測区間
new_y2



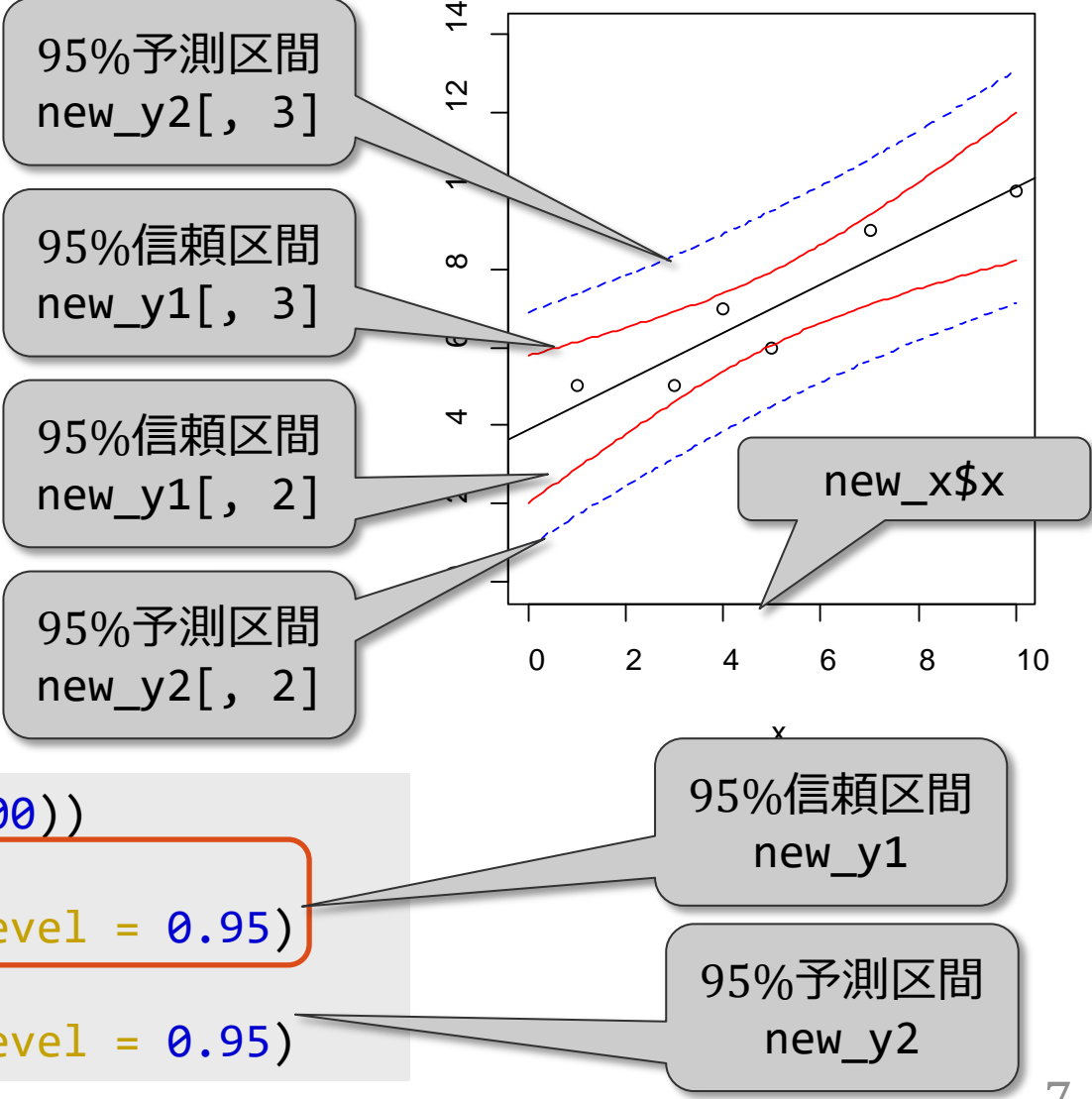
new_x\$x

単回帰分析：グラフ（信頼区間、予測区間）

```
> new_y1
##           fit           lwr           upr
## 1    3.900000  2.010113  5.789887
## 2    3.962626  2.101300  5.823952
## 3    4.025253  2.192336  5.858169
## 4    4.087879  2.283214  5.892544
## ...
## 98   9.974747  8.141831 11.807664
## 99  10.037374  8.176048 11.898700
## 100 10.100000  8.210113 11.989887
```

new_y1[, 1] new_y1[, 2] new_y1[, 3]

```
new_x <- data.frame(x = seq(0, 10, length = 100))
new_y1 <- predict(lm_out, newdata = new_x,
                  interval = "confidence", level = 0.95)
new_y2 <- predict(lm_out, newdata = new_x,
                  interval = "prediction", level = 0.95)
```

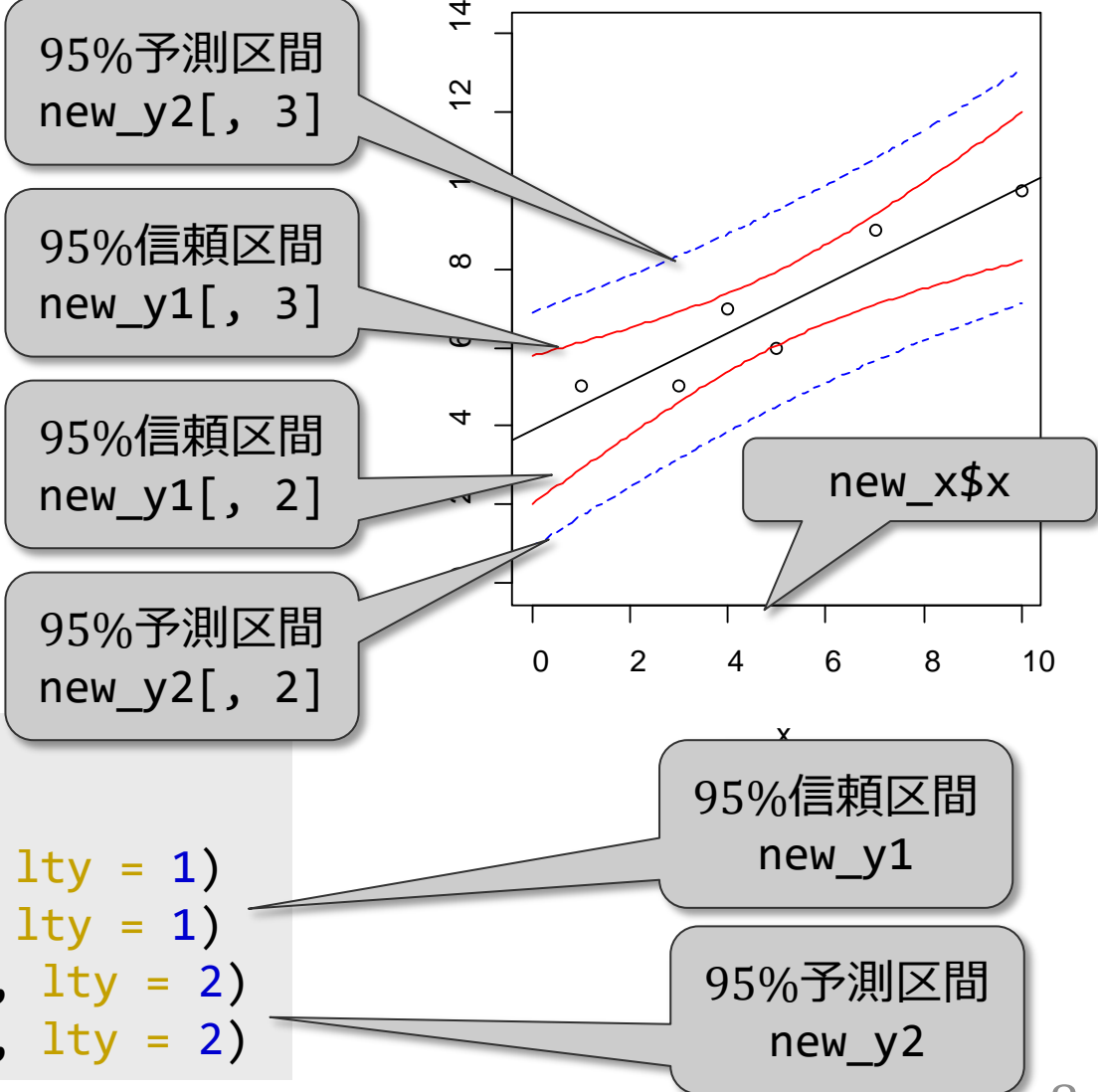


単回帰分析：グラフ（信頼区間、予測区間）

```
> new_y1
##          fit          lwr          upr
## 1    3.900000  2.010113  5.789887
## 2    3.962626  2.101300  5.823952
## 3    4.025253  2.192336  5.858169
## 4    4.087879  2.283214  5.892544
##      . . . . .
## 98   9.974747  8.141831 11.807664
## 99  10.037374  8.176048 11.898700
## 100 10.100000  8.210113 11.989887
```

new_y1[, 1] new_y1[, 2] new_y1[, 3]

```
plot(df)
abline(lm_out) # 回帰直線
lines(new_x$x, new_y1[, 2], col = "red", lty = 1)
lines(new_x$x, new_y1[, 3], col = "red", lty = 1)
lines(new_x$x, new_y2[, 2], col = "blue", lty = 2)
lines(new_x$x, new_y2[, 3], col = "blue", lty = 2)
```



単回帰分析：グラフ（信頼区間、予測区間）

p.245

●表示4.4.3 区間推定

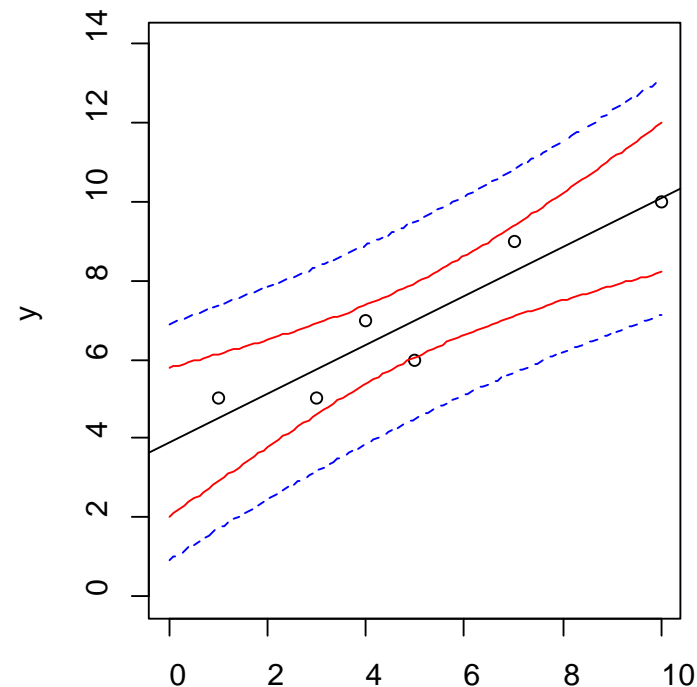
表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル：Green1-4-4.R

```
lm_out <- lm(y ~ x, data = df)

new_x <- data.frame(x = seq(0, 10, length = 100))
new_y1 <- predict(lm_out, newdata = new_x,
                  interval = "confidence", level = 0.95)
new_y2 <- predict(lm_out, newdata = new_x,
                  interval = "prediction", level = 0.95)

plot(df, xlim = c(0, 10), ylim = c(0, 14), pch = 21)
abline(lm_out)
lines(new_x$x, new_y1[, 2], col = "red", lty = 1)
lines(new_x$x, new_y1[, 3], col = "red", lty = 1)
lines(new_x$x, new_y2[, 2], col = "blue", lty = 2)
lines(new_x$x, new_y2[, 3], col = "blue", lty = 2)
```



95%信頼区間
new_y1

95%予測区間
new_y2

- 表示4.4.3 区間推定

表示4.4.7 JMP「二変量の関係」
による回帰式

スクリプトファイル

Green1-4-4.R

利用した関数（全体）

lm、summary、avova

方法（全体）

lm 関数で回帰分析

summary 関数で結果を出力

```
summary(lm_out)
```

```
##
```

```
## Call:
```

```
## lm(formula = y ~ x)
```

```
##
```

```
## Residuals:
```

```
##      1      2      3      4      5      6  
## 0.48 -0.76  0.62 -1.00  0.76 -0.10
```

```
##
```

```
## Coefficients:
```

```
##              Estimate Std. Error t value Pr(>|t|)  
## (Intercept)   3.9000     0.6807   5.730  0.00459 **  
## x              0.6200     0.1179   5.259  0.00626 **
```

```
## ---
```

```
## Residual standard error: 0.8337 on 4 degrees of freedom
```

```
## Multiple R-squared:  0.8736, Adjusted R-squared:  0.842
```

```
## F-statistic: 27.65 on 1 and 4 DF,  p-value: 0.00626
```

モデル式

残渣 e

- 表示4.4.3 区間推定

表示4.4.7 JMP 「二変量の関係」

による回帰式

スクリプトファイル

Green1-4-4.R

利用した関数（全体）

lm、summary、avova

方法（全体）

lm 関数で回帰分析

summary 関数で結果を出力

```
summary(lm_out)
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      1      2      3      4      5      6
##  0.48 -0.76  0.62 -1.00  0.76 -0.10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)   3.9000     0.6807   5.730  0.00459 **
## x              0.6200     0.1179   5.259  0.00626 **
## ---
## Residual standard error: 0.8337 on 4 degrees of freedom
## Multiple R-squared:  0.8736, Adjusted R-squared:  0.842
## F-statistic: 27.65 on 1 and 4 DF,  p-value: 0.00626
```

回帰係数と切片の
推定値、標準誤差、t 値、p 値

Coefficients: table with 5 columns: Estimate, Std. Error, t value, Pr(>|t|), and significance markers (**).

- 表示4.4.3 区間推定

表示4.4.7 JMP 「二変量の関係」
による回帰式

スクリプトファイル

Green1-4-4.R

利用した関数（全体）

lm、summary、avova

方法（全体）

lm 関数で回帰分析

summary 関数で結果を出力

```
summary(lm_out)
##
## Call:
## lm(formula = y ~ x)
##
## Residuals:
##      1      2      3      4      5      6
##  0.48 -0.76  0.62 -1.00  0.76 -0.10
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  3.9000      0.6807   5.7298  0.00459 **
## x            0.6200      0.1179   5.259  0.00626 **
## ---
## Residual standard error: 0.8337 on 4 degrees of freedom
## Multiple R-squared:  0.8736, Adjusted R-squared:  0.842
## F-statistic: 27.65 on 1 and 4 DF,  p-value: 0.00626
```

R2値、自由度調整R2値

モデルの F 値と p 値

残差標準偏差

残差の自由度

- 表示4.4.3 区間推定

表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル

Green1-4-4.R

利用した関数（全体）

lm、summary、anova

方法（全体）

lm 関数で回帰分析

anova 関数で結果を出力

```
anova(lm_out)
```

```
## Analysis of Variance Table
```

```
##
```

```
## Response: y
```

```
##          Df Sum Sq Mean Sq F value Pr(>F)
```

```
## x          1  19.22  19.220  27.655 0.00626 **
```

```
## Residuals  4   2.78   0.695
```

残差の
平方和、自由度、
平均平方

回帰の
平方和、自由度、
平均平方、F値、p値

- 表示4.4.7 JMP 「二変量の関係」
による回帰式
スクリプトファイル
Green1-4-4.R
利用した関数
lm
stargazer::stargazer
方法
lm 関数で回帰分析
stargazer 関数で結果を出力

```
stargazer(lm_out, type = "text")
=====
##                               Dependent variable:
##                               -----
##                               y
## -----
## x                               0.620***
##                               (0.118)
##
## Constant                         3.900***
##                               (0.681)
## -----
## Observations                       6
## R2                                 0.874
## Adjusted R2                        0.842
## Residual Std. Error                0.834 (df = 4)
## F Statistic                        27.655*** (df = 1; 4)
## =====
```



単回帰分析：パラメータの区間推定

- 表示4.4.7 JMP「二変量の関係」による回帰式

スクリプトファイル：Green1-4-4.R

利用した関数：lm、avova、coef、(coefficients)、confint

方法：lm 関数で回帰分析

coef 関数で係数と切片を抽出

confint 関数で係数と切片の区間推定

```
coe <- coef(lm_out)
con <- confint(lm_out)
cbind(Estimate = coe, con)

##           Estimate      2.5 %      97.5 %
## (Intercept)      3.90 2.0101129 5.7898871
## x                0.62 0.2926619 0.9473381
```

●表示4.4.3 区間推定

スクリプトファイル

Green1-4-4.R

利用した関数

lm、predict

方法

lm 関数で回帰分析

新しい x の値を `new_xx` に付値

`predict` 関数で `new_xx` に対する x の
推定と区間推定を行う

信頼区間：応答の期待値 η に対する信頼区間

予測区間：個別の応答 y に対する信頼区間

```
new_xx <- data.frame(x = c(2.0, 6.0, 8.0, 9.0))
```

```
predict(lm_out, newdata = new_xx, level = 0.95,  
        interval = "confidence") # 信頼区間
```

```
##      fit      lwr      upr  
## 1 5.14 3.777183 6.502817  
## 2 7.62 6.619966 8.620034  
## 3 8.86 7.497183 10.222817  
## 4 9.48 7.865280 11.094720
```

```
predict(lm_out, newdata = new_xx, level = 0.95,  
        interval = "prediction") # 予測区間
```

```
##      fit      lwr      upr  
## 1 5.14 2.453966 7.826034  
## 2 7.62 5.098576 10.141424  
## 3 8.86 6.173966 11.546034  
## 4 9.48 6.657797 12.302203
```


● 表示4.4.5 逆推定

表示4.4.8 JMP 「モデルのあてはめ」

による逆推定(1)

スクリプトファイル

Green1-4-4.R

利用した関数

lm、investr::calibrate、investr::plotFit

plot、abline

方法

lm 関数で回帰分析

$y = 8$ に対する x の値を

calibrate 関数で逆推定

信頼区間：応答の期待値 η に対する

信頼区間

```
new_yy <- 8.0
cal_out1 <- calibrate(lm_out, y0 = new_yy,
                     interval = "inversion",
                     mean.response = TRUE,
                     level = 0.95)

cal_out1

## estimate      lower      upper
## 6.612903 5.088091 9.384410

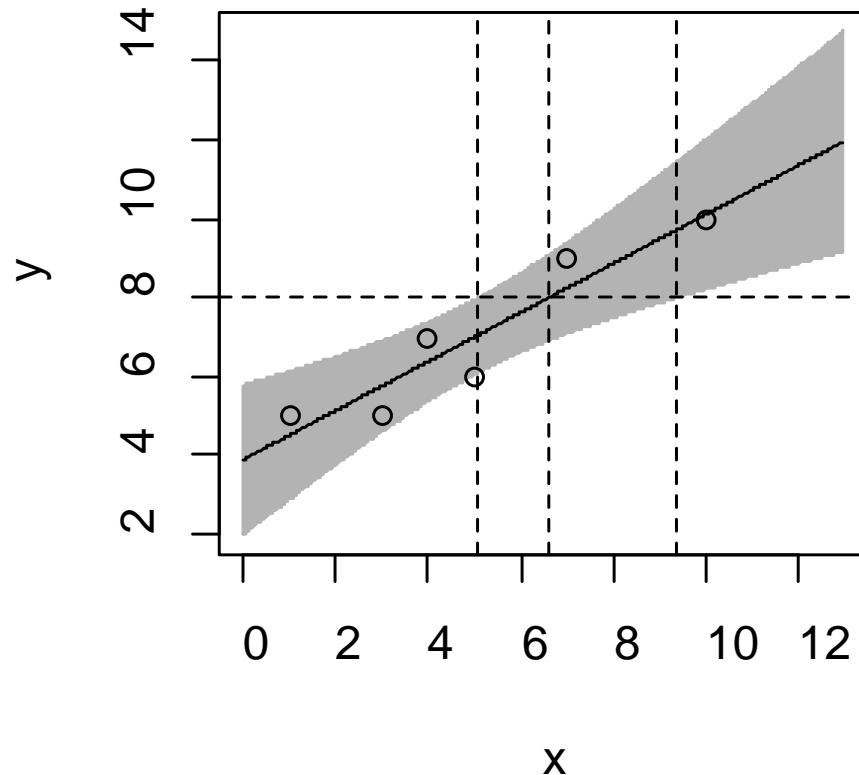
plotFit(lm_out, xlim = c(0, 13), level = 0.95,
        interval = "confidence", shade = TRUE)
abline(h = new_yy,
       v = c(cal_out1$lower,
            cal_out1$estimate,
            cal_out1$upper),
       lty = 2)
```

単回帰分析：逆推定（信頼区間）

p.246

●表示4.4.5 逆推定

表示4.4.8 JMP 「モデルのあてはめ」
による逆推定(1)



```
new_yy <- 8.0
cal_out1 <- calibrate(lm_out, y0 = new_yy,
                    interval = "inversion",
                    mean.response = TRUE,
                    level = 0.95)
```

```
cal_out1
## estimate      lower      upper
## 6.612903 5.088091 9.384410
```

```
plotFit(lm_out, xlim = c(0, 13), level = 0.95,
        interval = "confidence", shade = TRUE)
abline(h = new_yy,
       v = c(cal_out1$lower,
             cal_out1$estimate,
             cal_out1$upper),
       lty = 2)
```

- 表示4.4.5 逆推定

表示4.4.9 JMP「モデルのあてはめ」

による逆推定(2)

スクリプトファイル

Green1-4-4.R

利用した関数

lm、investr::calibrate、investr::plotFit

plot、abline

方法

lm 関数で回帰分析

calibrate 関数で逆推定

予測区間：個別の応答 y に対する
信頼区間

```
cal_out2 <- calibrate(lm_out, y0 = 8.0,  
                     interval = "inversion",  
                     mean.response = FALSE,  
                     level = 0.95)
```

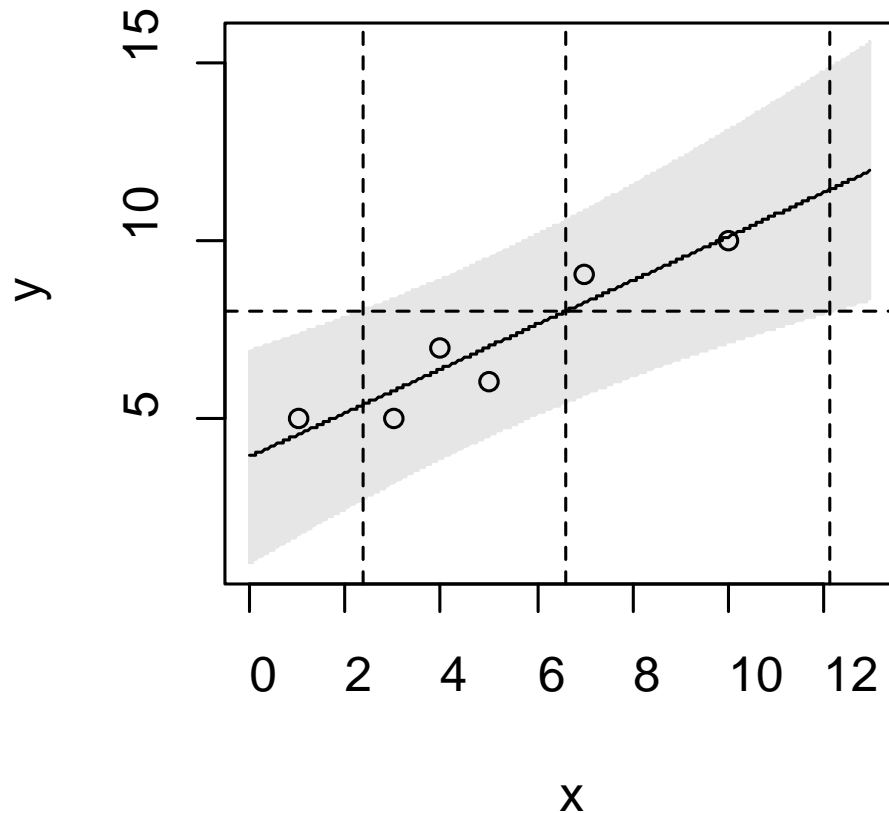
```
cal_out2
```

```
## estimate      lower      upper  
## 6.612903  2.343567 12.128934
```

```
plotFit(lm_out, xlim = c(0, 13), level = 0.95,  
        interval = "prediction", shade = TRUE)  
abline(h = new_yy,  
       v = c(cal_out2$lower,  
             cal_out2$estimate,  
             cal_out2$upper),  
       lty = 2)
```

- 表示4.4.5 逆推定

表示4.4.9 JMP「モデルのあてはめ」
による逆推定(2)



```
cal_out2 <- calibrate(lm_out, y0 = 8.0,  
                    interval = "inversion",  
                    mean.response = FALSE,  
                    level = 0.95)
```

```
cal_out2
```

```
## estimate      lower      upper  
## 6.612903  2.343567 12.128934
```

```
plotFit(lm_out, xlim = c(0, 13), level = 0.95,  
        interval = "prediction", shade = TRUE)  
abline(h = new_yy,  
       v = c(cal_out2$lower,  
             cal_out2$estimate,  
             cal_out2$upper),  
       lty = 2)
```



- 作成 片瀬雅彦
- 作成時期 2021年8月14日