

Rによる検定と 推定と実践（後半）

内田 治

西澤 英子

福井 克法

3つ以上のグループの母平均に関する検定

- ・一元配置分散分析

- ・二元配置分散分析

(二つの因子を分析できる)

一元配置分散分析 例題

次のデータは3箇所の異なる水源（変数 x ）から収集した上水道水に含まれる塩素の量（変数 y ）を記録したものである。水源地の違いによって含まれる塩素の量に違いがあるといえるだろうか。

x	y
1	1.75
1	0.63
1	0.44
1	1.74
1	1.27
1	1.2
1	1.05

x	y
2	1.12
2	0.77
2	0.3
2	1.16
2	0.29
2	0.11
2	0.06

x	y
3	0.64
3	0.47
3	0.42
3	0.74
3	0.56
3	1.03
3	0.21

一元配置分散分析 検定の手順

1. 仮説を立てる

第1グループの母平均と第2グループと第3グループの母平均が等しいか？

帰無仮説 $H_0 : \mu_1 = \mu_2 = \mu_3$

(第1グループと第2グループ、第3グループの母平均は等しい)

対立仮説 H_1 : 少なくとも1つの母平均が他の母平均と異なる

一元配置分散分析 検定の手順

2. 有意水準の設定

有意水準 $\alpha = 0.05$

3. 使用する統計的手法の決定

「一元配置分散分析」を実施する。関数`oneway.test()`を使用

一元配置分散分析 Rの操作

1. Rへのデータの読み込み

```
> setwd("C:/R")  
> library(xlsx)  
> data3_6 <- read.xlsx("data3-6.xlsx", sheet="sheet1")
```

2. グラフの作成

グループ別の平均値プロットを作成する

パッケージggplotsのダウンロードが必要となる

グラフの作成には関数plotmeans()を利用する

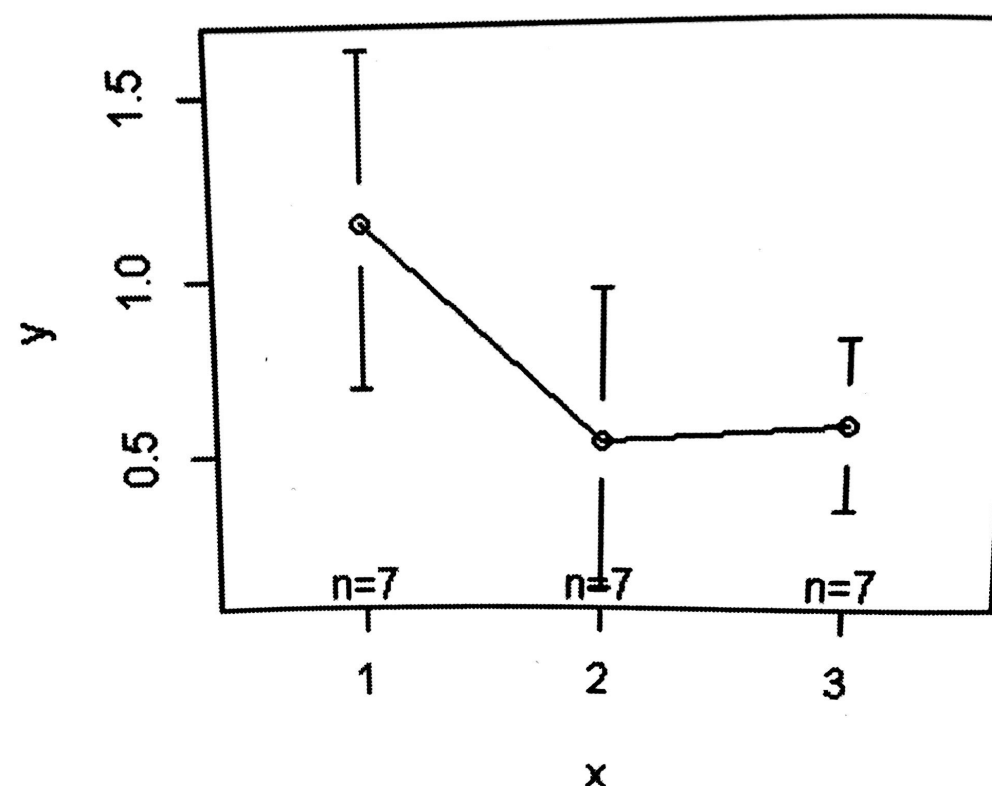
一元配置分散分析 Rの操作 (グラフの作成)

- Rの操作

- > library(gplots)

- > plotmeans(y ~ x, data = data3_6)

plotmeans(測定変数 ~ グループ変数, data=データフレーム名)



平均値プロットの例

一元配置分散分析 Rの操作（関数の実施）

3. 関数oneway.test()の実施

```
> oneway.test(y~x, data=data3_6, var.equal=T)
```

初期設定では、「ウェルチの検定」を実施

グループの分散が等しいと考えられる場合 `var.equal = T`

一元配置分散分析 検定の結果

- 結果

One-way analysis of means

data: y and x

$F = 4.5688, \text{num df} = 2, \text{denom df} = 18, \text{p-value} = 0.02485$

p-valueと有意水準を比較

$\text{p-value} = 0.02485 < \text{有意水準} = 0.05$

帰無仮説 H_0 は棄却され、3つのグループの塩素の量は等しくない（少なくとも1グループは異なる）と判断される

二次元分散分析（繰り返しあり）

例題

化粧品の品質を高めるための条件を探すために、原料の一種であるヒアルロン酸の濃度の種類と、コラーゲンの分子構造の種類を因子として実験を行なった結果。ヒアルロン酸の濃度（A）3水準、コラーゲンの分子構造(B)4水準。繰り返し2回の実験を行った。このデータを解析する。

	B1	B2	B3	B4
A1	26	28	32	28
	27	30	31	29
A2	28	30	32	29
	27	30	31	29
A3	30	30	30	30
	29	31	31	30

二次元分散分析（繰り返しあり） 検定の手順

1. 仮説を立てる

- ・ 因子Aについて

帰無仮説H0 : 因子Aの水準間に差はない

対立仮説H1 : 因子Aの水準間に差はある

- ・ 交互作用について

帰無仮説H0 : 因子Aと因子Bの交互作用はない

帰無仮説H1 : 因子Aと因子 Bの交互作用はある

- ・ 因子Bについて

帰無仮説H0 : 因子Bの水準間に差はない

対立仮説H1 : 因子Bの水準間に差はある

二次元分散分析（繰り返しあり） 検定の手順

2. 有意水準の設定

有意水準 $\alpha = 0.05$

3. 使用する統計的手法の決定

「二元配置分散分散」を実施する。関数aov()を使用

二次元分散分析（繰り返しあり）

Rの操作

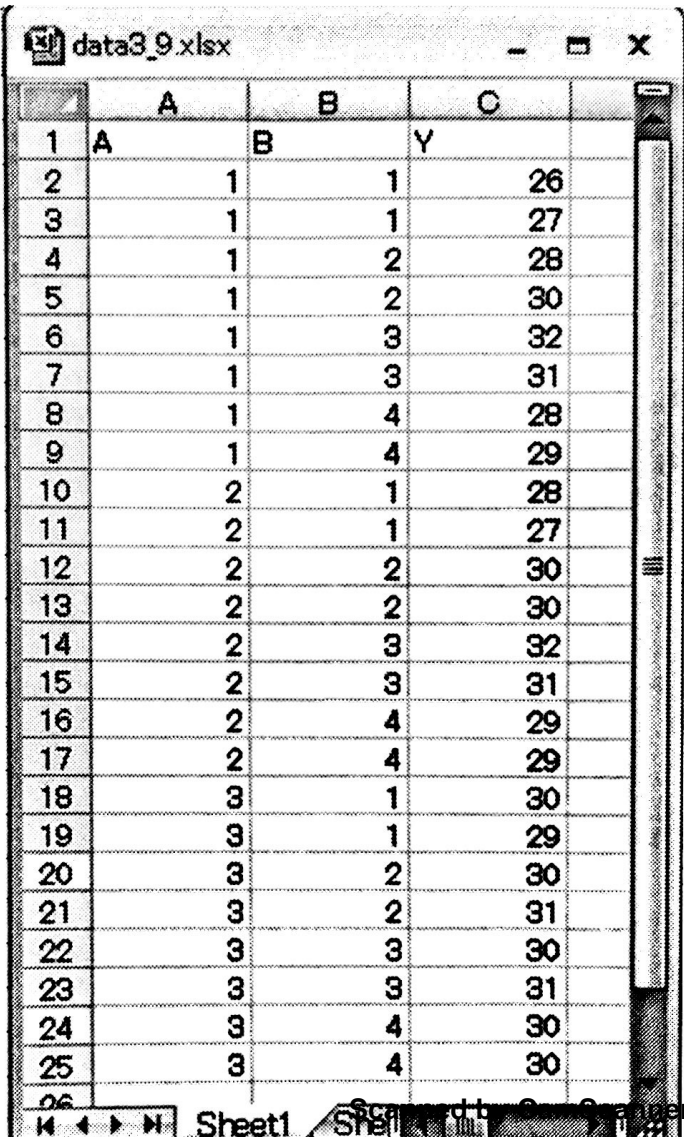
1. Rへのデータの読み込み

```
>data3_9 <- read.xlsx("data3_9.xlsx", sheetName = "Sheet1")
```

2. 因子の定義

```
>data3_9$A <- factor(data3_9$A)
```

```
>data3_9$B <- factor(data3_9$B)
```



	A	B	C
1	A	B	Y
2	1	1	26
3	1	1	27
4	1	2	28
5	1	2	30
6	1	3	32
7	1	3	31
8	1	4	28
9	1	4	29
10	2	1	28
11	2	1	27
12	2	2	30
13	2	2	30
14	2	3	32
15	2	3	31
16	2	4	29
17	2	4	29
18	3	1	30
19	3	1	29
20	3	2	30
21	3	2	31
22	3	3	30
23	3	3	31
24	3	4	30
25	3	4	30

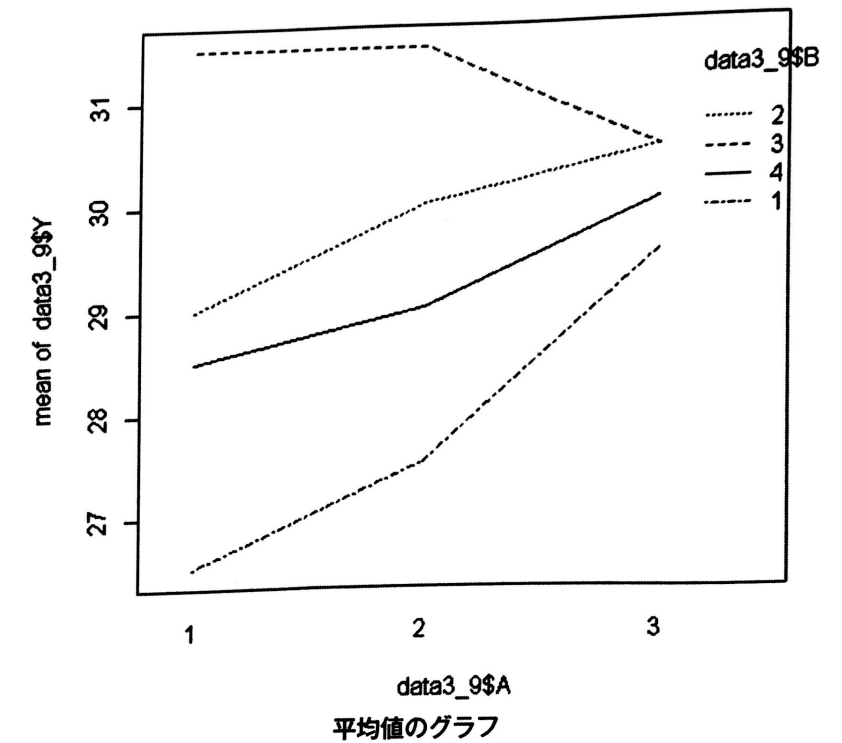
二次元分散分析（繰り返しあり）

Rの操作

3. グラフの作成

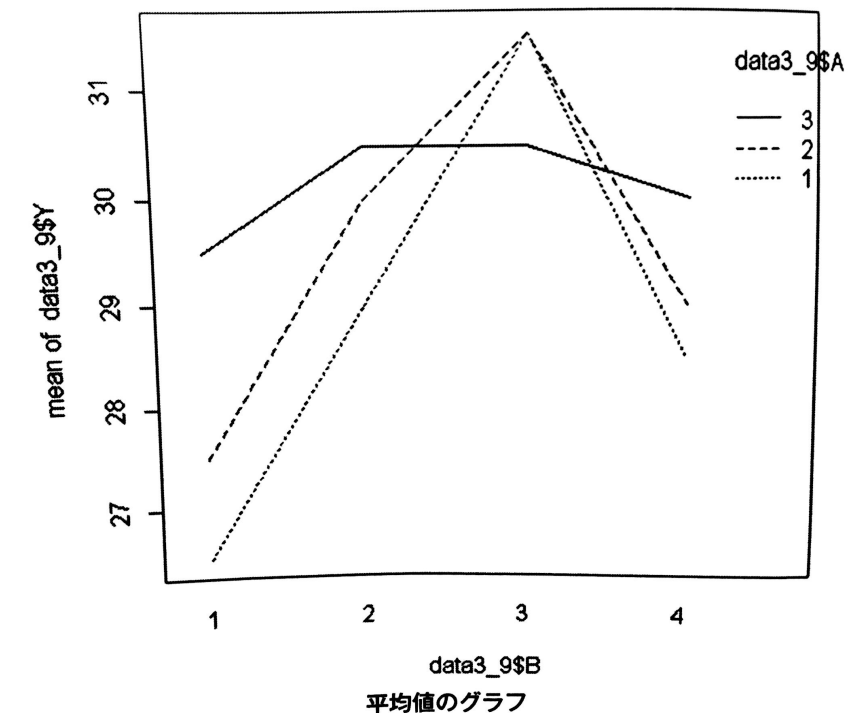
因子Aをx軸に指定した折れ線グラフを作成する

```
>interaction.plot(data3_9$A, data3_9$B, data3_9$Y)
```



因子Bをx軸に指定した折れ線グラフを作成する

```
>interaction.plot(data3_9$B,data3_9$A,data3_9$Y)
```



二次元分散分析（繰り返しあり） 関数の実施 検定の結果

```
>summary(aov(Y ~ A*B, data = data3_9))
```

検定の結果

	Df	Sum Sq	Mean Sq	F value	pr(>F)
A	2	6.25	3.125	1.81	0.0138 *
B	3	34.67	11.557	6.41	2.83e-05 ***
A:B	6	9.08	1.513	0.85	0.0485 *
Residuals	12	6.00	0.500		

Signif. codes: 0 '***' 0.01 '**' 0.05 '*' 0.1 '.' 1

1つのアスタリスク表示 0.05以下の水準で有意である

2つ 0.001以下の水準で有意である

因子A,Bの主効果、ならびに因子AとBの主効果、ならびに因子Aと因子Bの交互作用共に品質に影響を与えることがわかる

相関分析

- ・ 関係性の強さを強さを数値で表現する

相関係数の計算法

- ・ スピアマンの順位相関係数
- ・ ケードルマンの順位相関係数
- ・ ピアソンの相関係数

相関係数は-1~1までの値をとる。

$0.7 \leq |r|$ 強い相関がある

$0.5 \leq |r| < 0.7$ 相関がある

$0.3 \leq |r| < 0.5$ 弱い相関あり

$|r| < 0.3$ ほとんど相関なし

無相関の検定

データ同士に相関があるかを分析する

- ・ 仮説を立てる

帰無仮説 $H_0 : \rho = 0$ (母相関係数は0である)

対立仮説 $H_1 : \rho \neq 0$ (母相関係数は0ではない)

有意水準と比較することにより相関があるか判断できる。

クロス集計表 例題

男性と女性のチョコレート好み

	A.少な	B.中程度	C.多め	合計
男性	9	13	18	40
女性	11	19	20	50
合計	20	32	38	90

クロス集計表 検定の手順

- ・ 仮説を立てる

帰無仮説 H_0 : 男性と女性のチョコレートへの好みには違いがない

対立仮説 H_1 : 男性と女性のチョコレートへの好みには違いがある

- ・ 有意水準 $\alpha = 0.05$
- ・ 使用する統計的手法の決定

「独立性の検定（クロス集計表によるカイ 2 乗検定）」を実施する。`chisq.test()`を使って実施する。

クロス集計表

検定の結果

- ・ 結果

Pearson's Chi-squared test

data: data3_13

X-squared = 2.2747, df = 2, p-value = 0.3207

p-valueを有意水準と比較する。

p-value = 0.3207 > 有意水準 = 0.05

この結果より、帰無仮説H0は棄却されず、男性とチョコレートの好みには違いがあるとはいえない。