

R と RStudio の使い方

芳賀敏郎 (2014) 医薬品開発のための統計解析 第2部 実験計画法
4 共分散分析
4.2 解析手順

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2部 実験計画法 改訂版、サイエンティスト社、p.294

（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2部 実験計画法

- 1 因子実験・・・質的因子
 - 1.1 繰り返し数が等しい場合、1.2 繰り返し数が異なる場合
 - 1.3 多重比較、1.4 ばらつきを特性値とする実験
 - 1.5 ノンパラメトリック検定
- 量的因子
 - 2.1 直線関係の場合、2.2 非直線関係の場合
 - 2.3 ダミー変数による質的因子の効果の推定
- 乱塊法・・・3.1 質的因子の乱塊法、3.2 量的因子の乱塊法、3.3 欠測値のある場合
- 共分散分析**・・・4.1 共分散分析の目的、**4.2 解析手順**、4.3 医薬品開発における共分散分析の例
- 2 因子実験・・・5.1 2 因子実験の基礎、5.2 質的因子×質的因子、5.3 質的因子×量的因子
- 5.4 質的因子×量的因子（変形）、5.5 量的因子×量的因子
- 多因子実験・・・6.1 多因子実験の基礎、6.2 スクリーニング計画、6.3 応答局面計画
- 変量モデルほか・・・7.1 1 因子実験、7.2 枝分れ実験、7.3 乱塊法の拡張、7.4 経時データ、7.5 交差試験

- 表示4.1.1 年収の比較調査の結果

スクリプトファイル：Green2-4-1.R

利用した関数：by、summary

3社（A1, A2, A3）の年収と年齢

```
df
##      group age  income
## 1      A1  34    684
## 2      A1  33    788
## 3      A1  34    764
## 4      A1  37    836
## 5      A1  29    606
## 6      A1  26    696
## 7      A1  37    766
## 8      A1  38    862
## . . . . .
```

by 関数により 3社ごとの年収の基本統計量を比較

```
by(df$incom, df$group, summary)
## df$group: A1
##  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
## 606.0 687.0 736.0 731.6 782.5 862.0
## -----
## df$group: A2
##  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
## 580.0 689.8 706.0 727.3 790.8 843.0
## -----
## df$group: A3
##  Min. 1st Qu. Median  Mean 3rd Qu.  Max.
## 678.0 732.8 811.5 800.8 871.0 905.0
```

● 表示4.2.1 年収の分散分析表と平均値の差の検定

スクリプトファイル

Green2-4-2a.R

利用した関数

lm、anova、pairwise.t.test、
emmeans::emmeans、pairs

方法

emmeans 関数を使って

個々の平均値を比較

多重性を考慮するには

「adjust = tukey」を

指定する

分散分析表

```
lm_out <- lm(income ~ group, data = df)
anova(lm_out)

## Analysis of Variance Table
##
## Response: income
##           Df Sum Sq Mean Sq F value Pr(>F)
## group      2  34031 17015.6   2.3877  0.111
## Residuals 27 192410   7126.3
```

平均値の比較
(多重性の
考慮なし)

```
emmeans_out1 <- emmeans(lm_out, specs = "group")
pairs(emmeans_out1, adjust = "none")

## contrast estimate SE df t.ratio p.value
## A1 - A2          4.3 37.8 27  0.114  0.9102
## A1 - A3        -69.2 37.8 27 -1.833  0.0779
## A2 - A3        -73.5 37.8 27 -1.947  0.0620
```

年齢と年収の関係

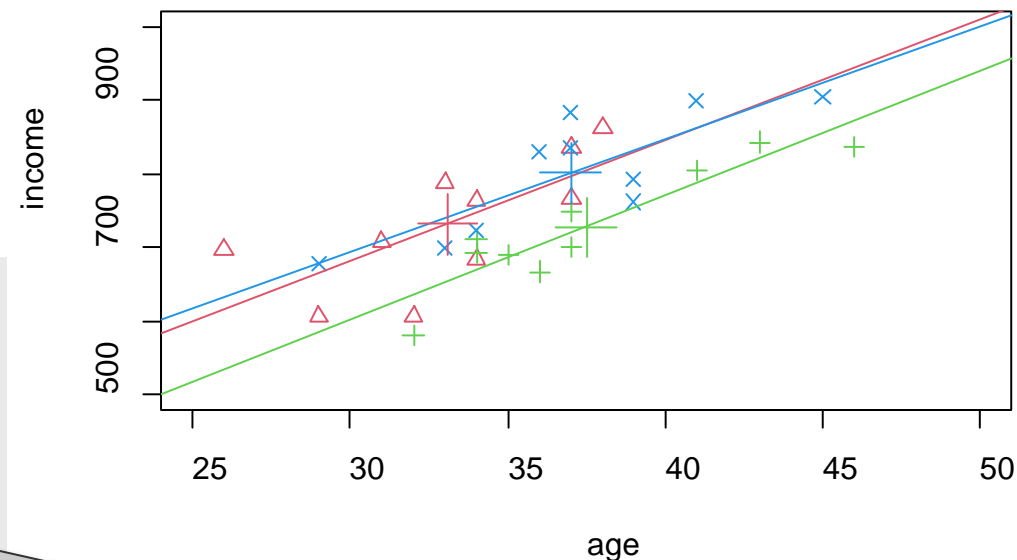
p.137

● 表示4.2.2 年齢と年収の関係

スクリプトファイル：Green2-4-2a.R

利用した関数：plot、abline、points

```
g <- 2:4
plot(x = df$age, y = df$income, type = "p",
     col = g[unclass(df$group)],
     pch = g[unclass(df$group)],
     xlim = c(25, 50),
     ylim = c(500, 1000),
     xlab = "age", ylab = "income")
abline(lm_out1, col = 2)
abline(lm_out2, col = 3)
abline(lm_out3, col = 4)
points(mean(df1$age), mean(df1$income), pch = 3, cex = 3, col = 2)
points(mean(df2$age), mean(df2$income), pch = 3, cex = 3, col = 3)
points(mean(df3$age), mean(df3$income), pch = 3, cex = 3, col = 4)
```



group は因子型(1, 2, 3)



ダミー変数による解析

- 表示4.2.5 LiNEST 関数による解析・・・ダミー変数に変換してLINEST 関数で解析
2種類のダミー変数 (dA2, dA3) (DA1, DA2) 参照：[§2.3](#) (p.97)

group	age	income	dA2	dA3	DA1	DA2
A1	34	684	0	0	1	0
A1	33	788	0	0	1	0
A1	34	764	0	0	1	0
A1	37	836	0	0	1	0
A1	29	606	0	0	1	0
A1	26	696	0	0	1	0
A1	37	766	0	0	1	0
A1	38	862	0	0	1	0
A1	32	606	0	0	1	0
A1	31	708	0	0	1	0
A2	34	692	1	0	0	1
A2	34	712	1	0	0	1
...

group	age	income	dA2	dA3	DA1	DA2
...
A2	41	805	1	0	0	1
A2	35	689	1	0	0	1
A3	34	723	0	1	-1	-1
A3	39	762	0	1	-1	-1
A3	37	883	0	1	-1	-1
A3	29	678	0	1	-1	-1
A3	33	699	0	1	-1	-1
A3	36	830	0	1	-1	-1
A3	41	900	0	1	-1	-1
A3	37	835	0	1	-1	-1
A3	45	905	0	1	-1	-1
A3	39	793	0	1	-1	-1

ダミー変数による解析

● 表示4.2.5 LiNEST 関数による解析

2種類のダミー変数 (dA2, dA3) (DA1, DA2)

参照 : §2.3 (p.97)

表示4.2.5の
ダミー変数

group	age	income	dA2	dA3	DA1	DA2
A1	34	684	0	0	1	0
A1	33	788	0	0	1	0
A1	34	764	0	0	1	0
A1	37	836	0	0	1	0
A1	29	606	0	0	1	0
A1	26	696	0	0	1	0
A1	37	766	0	0	1	0
A1	38	862	0	0	1	0
A1	32	606	0	0	1	0
A1	31	708	0	0	1	0
A2	34	692	1	0	0	1
A2	34	712	1	0	0	1
...

group	age	income	dA2	dA3	DA1	DA2
...
A2	41	805	1	0	0	1
A2	35	689	1	0	0	1
A3	34	723	0	1	-1	-1
A3	39	762	0	1	-1	-1
A3	37	883	0	1	-1	-1
A3	29	678	0	1	-1	-1
A3	33	699	0	1	-1	-1
A3	36	830	0	1	-1	-1
A3	41	900	0	1	-1	-1
A3	37	835	0	1	-1	-1
A3	45	905	0	1	-1	-1
A3	39	793	0	1	-1	-1

- 表示4.2.5 LiNEST 関数による解析

スクリプトファイル

Green2-4-2b.R

利用した関数

lm、summary

方法

lm 関数で回帰分析

上段はダミー変数 1 による

表示4.2.5 と一致

下段はダミー変数 2 による

表示4.2.8 の

JMP の結果と一致

ダミー変数 1 による解析

```
lm_out1 <- lm(income ~ age + dA2 + dA3 , data = df)
summary(lm_out1)
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	195.283	76.750	2.544	0.01723	*
##	age	16.203	2.269	7.141	1.39e-07	***
##	dA2	-75.593	24.484	-3.087	0.00476	**
##	dA3	6.009	24.044	0.250	0.80463	

ダミー変数 2 による解析

```
lm_out2 <- lm(income ~ age + DA1 + DA2 , data = df)
summary(lm_out2)
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	lm(formula = income ~ age + DA1 + DA2, data = df)					
##	(Intercept)	172.088	81.893	2.101	0.045449	*
##	age	16.203	2.269	7.141	1.39e-07	***
##	DA1	23.195	14.353	1.616	0.118164	
##	DA2	-52.398	13.429	-3.902	0.000604	***

● 表示4.2.5 LiNEST 関数による解析

$$y = 195.28 + \begin{pmatrix} 0 \\ -75.59 \\ 6.01 \end{pmatrix} + 16.20x$$
$$= \begin{pmatrix} 195.28 \\ 119.69 \\ 201.29 \end{pmatrix} + 16.20x \quad \begin{pmatrix} A1 \\ A2 \\ A3 \end{pmatrix}$$

(p.140)

$$y = 172.09 + \begin{pmatrix} 23.20 \\ -52.40 \\ 29.20 \end{pmatrix} + 16.20x$$
$$= \begin{pmatrix} 195.29 \\ 119.69 \\ 201.29 \end{pmatrix} + 16.20x \quad \begin{pmatrix} A1 \\ A2 \\ A3 \end{pmatrix}$$

$23.20 + (-52.40) + 29.20 = 0$
(3社の効果の和が0)

ダミー変数1による解析

```
lm_out1 <- lm(income ~ age + dA2 + dA3 , data = df)
summary(lm_out1)
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	195.283	76.750	2.544	0.01723	*
##	age	16.203	2.269	7.141	1.39e-07	***
##	dA2	-75.593	24.484	-3.087	0.00476	**
##	dA3	6.009	24.044	0.250	0.80463	

ダミー変数2による解析

```
lm_out2 <- lm(income ~ age + DA1 + DA2 , data = df)
summary(lm_out2)
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	172.088	81.893	2.101	0.045449	*
##	age	16.203	2.269	7.141	1.39e-07	***
##	DA1	23.195	14.353	1.616	0.118164	
##	DA2	-52.398	13.429	-3.902	0.000604	***

● 表示4.2.8 傾きの等しい直線のおてはめ

効果の和がゼロになるように設定

スクリプトファイル

Green2-4-2c.R

利用した関数

lm、summary

car::Anova

方法

欠測値のある

乱塊法と同じ考え方

参照：p.144 [注意]

§3.3 (p.123)

```
lm_out <- lm(income ~ group + age, data = df,
             contrasts = list(group = "contr.sum"))
summary(lm_out)
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	172.088	81.893	2.101	0.045449	*
##	group1	23.195	14.353	1.616	0.118164	
##	group2	-52.398	13.429	-3.902	0.000604	***
##	age	16.203	2.269	7.141	1.39e-07	***

```
Anova(lm_out, type = 3)
```

##		Sum Sq	Df	F value	Pr(>F)	
##	(Intercept)	11036	1	4.4159	0.045449	*
##	group	39101	2	7.8232	0.002189	**
##	age	127434	1	50.9928	1.39e-07	***
##	Residuals	64976	26			

平方和 タイプIII を指定

傾きの等しい直線のアてはめ：共分散分析

- 表示4.2.8 傾きの等しい直線のアてはめ
スクリプトファイル：Green2-4-2c.R

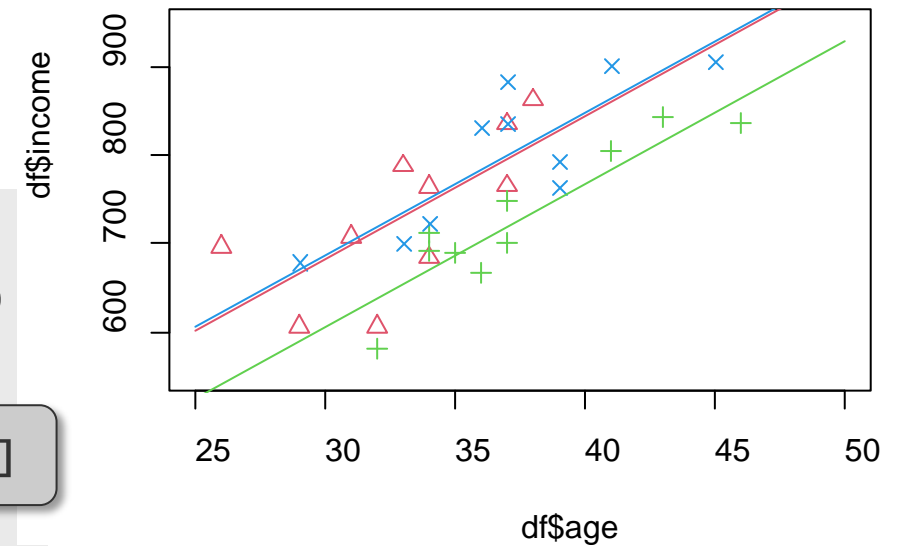
```
lm_out <- lm(income ~ group + age, data = df,  
             contrasts = list(group = "contr.sum"))  
  
coef_out <- coef(lm_out)  
est_a <- coef_out[1] # 172.09  
est_g1 <- coef_out[2] # 23.20  
est_g2 <- coef_out[3] # -52.40  
est_g3 <- -(est_g1 + est_g2) # 効果の和が 0  
est_b <- coef_out[4] # 172.09  
fn1 <- function(x) est_a + est_g1 + est_b * x  
fn2 <- function(x) est_a + est_g2 + est_b * x  
fn3 <- function(x) est_a + est_g3 + est_b * x
```

効果の和がゼロ

$$23.20 + (-52.40) + 29.20 = 0$$

$$y = 172.09 + \begin{pmatrix} 23.20 \\ -52.40 \\ 29.20 \end{pmatrix} + 16.20x \quad \begin{pmatrix} A1 \\ A2 \\ A3 \end{pmatrix}$$

表示4.2.8 傾きの等しい直線のアてはめ

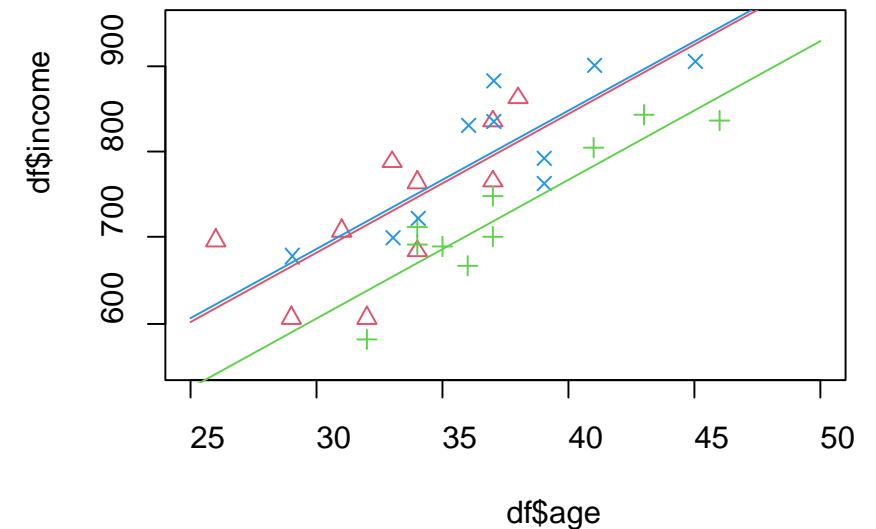


傾きの等しい直線のアてはめ：共分散分析

p.145

- 表示4.2.8 傾きの等しい直線のアてはめ
スクリプトファイル：Green2-4-2c.R

表示4.2.8 傾きの等しい直線のアてはめ



```
g <- 2:4
plot(x = df$age, y = df$income, type = "p",
     pch = g[unclass(df$group)], col = g[unclass(df$group)],
     xlim = c(25, 50), ylim = c(550, 950))
curve(fn1, xlim = c(25, 50), ylim = c(550, 950), col = 2, add = TRUE)
curve(fn2, xlim = c(25, 50), ylim = c(550, 950), col = 3, add = TRUE)
curve(fn3, xlim = c(25, 50), ylim = c(550, 950), col = 4, add = TRUE)
```

- 表示4.2.9 A2-A1、A3-A1 の推定と検定

スクリプトファイル

Green2-4-2c.R

利用した関数

lm、emmeans::emmeans

pairs

方法

多重性を考慮しない比較

```
emmeans_out1 <- emmeans(lm_out, specs = "group")
emmeans_out1
##   group emmean   SE df lower.CL upper.CL
##   A1      776 17.0 26     741     811
##   A2      701 16.2 26     667     734
##   A3      782 16.0 26     750     815
##
## Confidence level used: 0.95
pairs(emmeans_out1, adjust = "none")
##   contrast estimate   SE df  t.ratio    p.value
##   A1 - A2      75.59 24.5 26   3.087  0.0048
##   A1 - A3     -6.01 24.0 26  -0.250  0.8046
##   A2 - A3    -81.60 22.4 26  -3.645  0.0012
```

- 表示4.2.9 A2-A1、A3-A1 の推定と検定

対照群の指定

スクリプトファイル

Green2-4-2c.R

利用した関数

lm、emmeans::emmeans

pairs

方法

多重性を考慮した比較
(DunnettXの方法)

```
emmeans(lm_out, specs = trt.vs.ctrl ~ group, ref = 1)
## $emmeans
##   group emmean   SE df lower.CL upper.CL
##   A1      776 17.0 26     741     811
##   A2      701 16.2 26     667     734
##   A3      782 16.0 26     750     815
##
## Confidence level used: 0.95
##
## $contrasts
##   contrast estimate   SE df t.ratio p.value
##   A2 - A1     -75.59 24.5 26  -3.087  0.0092
##   A3 - A1      6.01 24.0 26   0.250  0.9442
##
## P value adjustment: dunnettX method for 2 tests
```

● 表示4.2.11 傾きの違いの検定

スクリプトファイル

Green2-4-2c.R

利用した関数

lm、

car::Anova

方法

中心化をして、
交互作用を含めた
回帰分析を実施

主効果と交互作用を指定

中心化を指定

```
center = mean(df$age)
lm_out1 <- lm(income ~ group * I(age - center), data = df,
              contrasts = list(group = "contr.sum"))
Anova(lm_out1, type = 3)

## Anova Table (Type III tests)
##
## Response: income
##
```

	Sum Sq	Df	F value	Pr(>F)	
## (Intercept)	13388144	1	4962.3968	< 2.2e-16	***
## group	36938	2	6.8456	0.004443	**
## I(age - center)	124797	1	46.2567	4.921e-07	***
## group:I(age - center)	226	2	0.0418	0.959111	
## Residuals	64750	24			



- 作成 片瀬雅彦
- 作成時期 2021年7月4日