

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
2 量的因子の1因子実験
2.3 ダミー変数による質的因子の効果の推定

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

本PDF ファイルをダウンロードし、Adobe Acrobat Reader DC で注釈のメモを表示してください

●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



テキストの内容を R で実行

ダミー変数

ダミー変数 1 とダミー変数 2

y 列を省略

● 表示2.3.3 元データとダミー変数

スクリプトファイル：Green2-2-3a.R

利用した関数：read_excel

方法

df0：元データのデータフレーム

df1：ダミー変数 1 のデータフレーム

df2：ダミー変数 2 のデータフレーム

(df1、df2の y 列の表示は省略)

df0 の質的変数 group を

ダミー変数 1 とダミー変数 2 に変換し、

lm 関数で解析してその結果を比較

df0：元データ

y	group
43	A1
45	A1
42	A1
47	A1
49	A1
50	A1
47	A2
51	A2
49	A2
54	A3
48	A3
57	A3
55	A4
58	A4
61	A4
52	A5
48	A5
53	A5

43	A1
45	A1
42	A1
47	A1
49	A1
50	A1
47	A2
51	A2
49	A2
54	A3
48	A3
57	A3
55	A4
58	A4
61	A4
52	A5
48	A5
53	A5

df1：ダミー変数 1

dA2	dA3	dA4	dA5
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
0	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	0	1	0
0	0	0	1
0	0	0	1
0	0	0	1

df2：ダミー変数 2

DA1	DA2	DA3	DA4
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
1	0	0	0
0	1	0	0
0	1	0	0
0	1	0	0
0	0	1	0
0	0	1	0
0	0	1	0
0	0	0	1
0	0	0	1
0	0	0	1
-1	-1	-1	-1
-1	-1	-1	-1
-1	-1	-1	-1

ダミー変数 1 を使った回帰分析

● 表示2.3.4 LINEST 関数の解

スクリプトファイル

Green2-2-3a.R

利用した関数

lm、summary

ダミー変数 1

デフォルトなので省略可

通常の lm 関数の出力

表示2.3.4 の LINEST 関数の
結果（上段）と一致

```
# 表示2.3.3、表示2.3.4 (ダミー変数 1 による解) (p.100)
```

```
lm_out1 <- lm(y ~ dA2 + dA3 + dA4 + dA5, data = df1)
```

```
##
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	46.000	1.311	35.096	2.89e-14	***
##	dA2	3.000	2.270	1.321	0.209132	
##	dA3	7.000	2.270	3.083	0.008721	**
##	dA4	12.000	2.270	5.286	0.000147	***
##	dA5	5.000	2.270	2.202	0.046290	*

```
lm_out2 <- lm(y ~ group, data = df0,
```

```
contrasts = list(group = "contr.treatment"))
```

```
##
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	46.000	1.311	35.096	2.89e-14	***
##	groupA2	3.000	2.270	1.321	0.209132	
##	groupA3	7.000	2.270	3.083	0.008721	**
##	groupA4	12.000	2.270	5.286	0.000147	***
##	groupA5	5.000	2.270	2.202	0.046290	*

ダミー変数 1 を使った回帰分析

● 表示2.3.4 LINEST 関数の解

スクリプトファイル

Green2-2-3a.R

利用した関数

lm、summary

ダミー変数 1

group A1 は 0

通常の lm 関数の出力

表示2.3.4 の LINEST 関数の
結果（上段）と一致

```
# 表示2.3.3、表示2.3.4 (ダミー変数 1 による解) (p.100)
```

```
lm_out1 <- lm(y ~ dA2 + dA3 + dA4 + dA5, data = df1)
```

```
##
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	46.000	1.311	35.096	2.89e-14	***
##	dA2	3.000	2.270	1.321	0.209132	
##	dA3	7.000	2.270	3.083	0.008721	**
##	dA4	12.000	2.270	5.286	0.000147	***
##	dA5	5.000	2.270	2.202	0.046290	*

```
lm_out2 <- lm(y ~ group, data = df0,  
              contrasts = list(group = "contr.treatment"))
```

```
##
```

##		Estimate	Std. Error	t value	Pr(> t)	
##	(Intercept)	46.000	1.311	35.096	2.89e-14	***
##	groupA2	3.000	2.270	1.321	0.209132	
##	groupA3	7.000	2.270	3.083	0.008721	**
##	groupA4	12.000	2.270	5.286	0.000147	***
##	groupA5	5.000	2.270	2.202	0.046290	*

ダミー変数 2 を使った回帰分析

● 表示2.3.4 LINEST 関数の解

スクリプトファイル

Green2-2-3a.R

利用した関数

lm、summary

ダミー変数 2

設定

表示2.3.4 の LINEST 関数の

結果 (下段) と一致

JMP の出力 (表示1.2.3、 p.44)

と一致

```
# 表示2.3.3、表示2.3.4 (ダミー変数 2 による解) (p.100)
```

```
lm_out3 <- lm(y ~ DA1 + DA2 + DA3 + DA4, data = df2)
```

```
##
```

```
##          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  51.4000    0.7864  65.359 < 2e-16 ***
```

```
## DA1          -5.4000    1.2842  -4.205  0.00103 **
```

```
## DA2          -2.4000    1.6371  -1.466  0.16641
```

```
## DA3           1.6000    1.6371   0.977  0.34623
```

```
## DA4           6.6000    1.6371   4.032  0.00142 **
```

```
lm_out4 <- lm(y ~ group, data = df0,
```

```
  contrasts = list(group = "contr.sum"))
```

```
##
```

```
##          Estimate Std. Error t value Pr(>|t|)
```

```
## (Intercept)  51.4000    0.7864  65.359 < 2e-16 ***
```

```
## group1       -5.4000    1.2842  -4.205  0.00103 **
```

```
## group2       -2.4000    1.6371  -1.466  0.16641
```

```
## group3        1.6000    1.6371   0.977  0.34623
```

```
## group4        6.6000    1.6371   4.032  0.00142 **
```

ダミー変数 2 を使った回帰分析

● 表示2.3.4 LINEST 関数の解

スクリプトファイル

Green2-2-3a.R

利用した関数

lm、summary

表示2.3.3、表示2.3.4 (ダミー変数 2 による解) (p.100)

```
lm_out3 <- lm(y ~ DA1 + DA2 + DA3 + DA4, data = df2)
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  51.4000    0.7864  65.359 < 2e-16 ***
## DA1          -5.4000    1.2842  -4.205  0.00103 **
## DA2          -2.4000    1.6371  -1.466  0.16641
## DA3           1.6000    1.6371   0.977  0.34623
## DA4           6.6000    1.6371   4.032  0.00142 **
```

$group1 + group2 + \dots + group5 = 0$
 $-(-5.4 - 2.4 + 1.6 + 6.6) = -0.4$

```
lm_out4 <- lm(y ~ group, data = df0,
               contrasts = list(group = "contr.sum"))
##
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept)  51.4000    0.7864  65.359 < 2e-16 ***
## group1      -5.4000    1.2842  -4.205  0.00103 **
## group2      -2.4000    1.6371  -1.466  0.16641
## group3       1.6000    1.6371   0.977  0.34623
## group4       6.6000    1.6371   4.032  0.00142 **
```

表示2.3.4 の LINEST 関数の

下段の結果と一致

JMP の出力 (表示1.2.3、p.44)

と一致



R における contrasts の設定

補足



R における contrasts の種類と設定

● contrasts の種類

contrasts の設定で質的因子（カテゴリカル因子）の処理方法を設定

contrasts の種類

"contr.treatment" "contr.sum" "contr.poly" "contr.SAS" "contr.helmert"

"contr.treatment" : テキストにおけるダミー変数 1 (0, 1) が用いられる、R のデフォルト

"contr.sum" : テキストにおけるダミー変数 2 (-1, 0, 1) が用いられる

● contrasts の設定

(1) 実行する関数の引数で設定（`contrasts` という引数を利用）

その関数の実行時において、設定した contrasts が用いられる

(2) データフレームの一部として contrasts を設定（`contrasts` 関数を利用）

データフレームが任意のモデルで呼び出されるたびに、設定した contrasts が用いられる

(3) 一連の作業（セッション）で設定（`options` 関数を利用）

contrasts の設定が影響するすべての関数に、設定した contrasts が反映する

Rにおけるcontrastsの設定方法

●Rにおける contrasts の設定：contr.treatment（デフォルト）

スクリプトファイル：Green2-3-3b.R

方法：lm 関数による回帰分析をデフォルトで実行（contrasts は contr.treatment に設定）

```
df <- read_excel("Green2-2.xlsx", sheet = "original")
df <- data.frame(df)
```

Excelからの
読み込み

```
df
##      y group
## 1  43   A1
## 2  45   A1
.....
## 17 48   A5
## 18 53   A5
```

質的因子

因子型に変換

```
df$group <- factor(df$group, levels = c("A1", "A2", "A3", "A4", "A5"))
lm_out1 <- lm(y ~ group, data = df)
```

水準の位置

Rにおけるcontrastsの設定方法

●Rにおける contrasts の設定：contr.treatment（デフォルト）

スクリプトファイル：Green2-3-3b.R

方法：デフォルトで、lm 関数による回帰分析をで実行（contrasts は contr.treatment に設定）

```
df$group <- factor(df$group, levels = c("A1", "A2", "A3", "A4", "A5"))  
lm_out1 <- lm(y ~ group, data = df)
```

先頭の水準

contrast の表示

```
contrasts(df$group)
```

##		A2	A3	A4	A5
##	A1	0	0	0	0
##	A2	1	0	0	0
##	A3	0	1	0	0
##	A4	0	0	1	0
##	A5	0	0	0	1

基準は
先頭の水準

ダミー変数 1

R における contrasts の設定方法

●R における contrasts の設定 : `contr.sum` (1)

スクリプトファイル : `Green2-3-3b.R`

方法 : `contrasts` を `contr.sum` に設定して、`lm` 関数による回帰分析を実行

`lm` 関数の `contrasts` 引数で設定 (この `lm` 関数の実行時のみ)

```
# (1)
df$group <- factor(df$group, levels = c("A1", "A2", "A3", "A4", "A5"))
lm_out1 <- lm(y ~ group, data = df,
              contrasts = list(group = "contr.sum"))
```

`contr.sum` 引数

```
contrasts(df$group)
```

contrast の表示

```
##      A2 A3 A4 A5
## A1   0  0  0  0
## A2   1  0  0  0
## A3   0  1  0  0
## A4   0  0  1  0
## A5   0  0  0  1
```

ダミー変数 1

R における contrasts の設定方法

●R における contrasts の設定 : `contr.sum` (2)

スクリプトファイル : `Green2-3-3b.R`

方法 : `contrasts` を `contr.sum` に設定して、`lm` 関数による回帰分析を実行

`contrasts` 関数で設定 (`df1` が関数に呼び出されるたびに、`contr.sum` で処理される)

dfのコピー

```
# (2)
df1 <- df
contrasts(df1$group) <- contr.sum
lm_out2 <- lm(y ~ group, data = df1)
```

`contr.sum` に設定

```
contrasts(df1$group)
```

contrast の表示

```
##      [,1] [,2] [,3] [,4]
## A1      1      0      0      0
## A2      0      1      0      0
## A3      0      0      1      0
## A4      0      0      0      1
## A5     -1     -1     -1     -1
```

ダミー変数 2

-1 は末尾

R における contrasts の設定方法

●R における contrasts の設定 : contr.sum (3)

スクリプトファイル : Green2-3-3b.R

方法 : contrasts を contr.sum に設定して、lm 関数による回帰分析を実行、options 関数で設定

options(contrasts = c(順序のない質的変数の設定, 順序のある質的変数の設定))

現状の設定を
一時退避

全体の設定を
変更する

有効

設定を
元に戻す

```
# (3)
constr_def <- options("contrasts")
options(contrasts = c("contr.sum", "contr.sum"))
lm_out3 <- lm(y ~ group, data = df)
summary(lm_out3)
options(contrasts = constr_def$contrasts)
```

options 設定前の contrasts (デフォルト)

```
options("contrasts")
## $contrasts
## unordered      ordered
## "contr.treatment" "contr.poly"
```

options 設定後の contrasts

```
options("contrasts")
## $contrasts
## unordered      ordered
## "contr.sum"    "contr.sum"
```



R における contrasts の設定

2 水準の事例

4 水準の事例

ダミー変数の基準（ベースライン）にする水準の決め方

Rにおけるcontrastsの設定：2水準の事例

●contrastsの設定

スクリプトファイル：Green2-3-3c.R
方法

factor関数で因子型に設定
levels引数で水準の位置を設定

contrasts関数で
contrastsの種類を設定
現在のcontrastsを表示

ダミー変数1の基準は先頭的水準
(0)
ダミー変数2の基準は末尾的水準
(-1)

```
vt2 <- c("g1", "g2", "g1", "g1", "g2")  
# ダミー変数1、基準："g1"  
ft1 <- factor(vt2, levels = c("g1", "g2"))  
# contrasts(ft1) <- contr.treatment  
contrasts(ft1) # 現在の contrasts を表示  
##      g2  
## g1  0  
## g2  1  
  
# ダミー変数2、基準："g2"  
ft2 <- factor(vt2, levels = c("g1", "g2"))  
contrasts(ft2) <- contr.sum  
contrasts(ft2)  
##      1  
## g1  1  
## g2 -1
```

先頭的水準

デフォルト
なので不要

先頭的水準

末尾的水準

末尾の水準

Rにおけるcontrastsの設定：2水準の事例

●contrastsの設定

スクリプトファイル：Green2-3-3c.R
方法

factor関数で因子型に設定
levels引数で水準の位置を設定

contrasts関数で
contrastsの種類を設定
現在のcontrastsを表示

ダミー変数1の基準は**先頭**の水準
(0)
ダミー変数2の基準は**末尾**の水準
(-1)

```
vt2 <- c("g1", "g2", "g1", "g1", "g2")
# ダミー変数1、基準："g1"
ft1 <- factor(vt2, levels = c("g1", "g2"))
# contrasts(ft1) <- contr.treatment
contrasts(ft1) # 現在の contrasts を表示

##      g2
## g1  0
## g2  1

# ダミー変数2、基準："g2"
ft2 <- factor(vt2, levels = c("g1", "g2"))
contrasts(ft2) <- contr.sum
contrasts(ft2)

##      1
## g1  1
## g2 -1
```

先頭の水準

デフォルト
なので不要

先頭の水準

末尾の水準

末尾の水準

Rにおけるcontrastsの設定：4水準の事例

●contr.treatmentの設定

スクリプトファイル

Green2-3-3c.R

方法

factor関数で因子型に設定

levels引数で水準の位置を
設定

contrasts関数で

contrastsの種類を設定

現在のcontrastsを表示

ダミー変数1の基準は

先頭の水準

```
vt4 <- c("g1", "g3", "g3", "g4", "g2", "g2", "g4", "g1")
```

```
# 基準: "g1"
```

```
ft3 <- factor(vt4, levels = c("g1", "g2", "g3", "g4"))
```

```
contrasts(ft3)
```

```
##      g2 g3 g4
```

```
## g1  0  0  0
```

```
## g2  1  0  0
```

```
## g3  0  1  0
```

```
## g4  0  0  1
```

```
# 基準: "g2"に変更
```

```
ft4 <- factor(vt4, levels = c("g2", "g3", "g4", "g1"))
```

```
contrasts(ft4)
```

```
##      g3 g4 g1
```

```
## g2  0  0  0
```

```
## g3  1  0  0
```

```
## g4  0  1  0
```

```
## g1  0  0  1
```

先頭の水準

先頭の水準

先頭の水準

先頭の水準

Rにおけるcontrastsの設定：4水準の事例

●contr.treatmentの設定

スクリプトファイル

Green2-3-3c.R

方法

factor関数で因子型に設定

levels引数で水準の位置を

設定

contr.treatment関数の

base引数により、

基準になる水準を設定

relevel関数のref引数により、

基準になる水準を設定

```
# 基準："g2"に変更
ft5 <- factor(vt4, levels = c("g1", "g2", "g3", "g4"))
contrasts(ft5) <- contr.treatment(4, base = 2)
contrasts(ft5)
##      1 3 4
## g1 1 0 0
## g2 0 0 0
## g3 0 1 0
## g4 0 0 1

# 基準："g2"に変更
ft6 <- factor(vt4, levels = c("g1", "g2", "g3", "g4"))
ft6 <- relevel(ft6, ref = 2)
contrasts(ft6)
##      g1 g3 g4
## g2 0 0 0
## g1 1 0 0
## g3 0 1 0
## g4 0 0 1
```

2 番目の水準

2 番目の水準

2 番目の水準

2 番目の水準

2 番 → 1 番

Rにおけるcontrastsの設定：4水準の事例

●contr.sumの設定

スクリプトファイル

Green2-3-3c.R

方法

factor関数で因子型に設定

levels引数で水準の位置を
設定

contrasts関数で、

ダミー変数を表示

ダミー変数2の基準は

末尾の水準

```
# 基準 : "g4"  
ft6 <- factor(vt4, levels = c("g1", "g2", "g3", "g4"))  
contrasts(ft6) <- contr.sum  
contrasts(ft6)  
##      1  2  3  
## g1  1  0  0  
## g2  0  1  0  
## g3  0  0  1  
## g4 -1 -1 -1
```

末尾の水準

末尾の水準

```
# 基準 : "g2"に変更  
ft7 <- factor(vt4, levels = c("g1", "g3", "g4", "g2"))  
contrasts(ft7) <- contr.sum  
contrasts(ft7)  
##      1  2  3  
## g1  1  0  0  
## g3  0  1  0  
## g4  0  0  1  
## g2 -1 -1 -1
```

末尾の水準

末尾の水準

Rにおけるcontrastsの設定：4水準の事例

●contr.sumの設定

スクリプトファイル：Green2-3-3c.R

方法

factor関数で因子型に設定、levels引数で水準の位置を設定

memisc::contr.sum関数で、基準になる水準（-1）を、base引数で設定

```
library("memisc")  
# 基準："g2"に変更  
ft8 <- factor(vt4, levels = c("g1", "g2", "g3", "g4"))  
contrasts(ft8) <- memisc::contr.sum(4, base = 2)  
contrasts(ft8)
```

##		1	3	4
##	g1	1	0	0
##	g2	-1	-1	-1
##	g3	0	1	0
##	g4	0	0	1

ロード

2番目の水準

2番目の水準



- 作成 片瀬雅彦
- 作成時期 2021年5月26日
- 改定 2021年11月8日、11月26日