

R でチューキー・クレーマーによる多重検定

清水顕史

例データの解析

チューキー・クレーマー検定は、複数種類間の平均の有意差を計算するのに、信頼性のある多重比較検定方法です。同じ目的でダンカンの多重範囲検定や LSD 法などが使用される場合がありますが、これらは多重性の問題を考慮しておらず不適切とされています（詳しくは永田靖・吉田道弘 1997『統計的多重比較法の基礎』など）。

例データとして、或るストレス条件で調べたイネ 10 品種の形質値 [rice10var.txt](#) を利用します。Web ページのリンクからダウンロードしたファイルを、R の作業用フォルダに移動しておいてください。このデータはエクセルなどを利用して作成されたタブ区切りのテキストファイルで、列 1 は品種名、列 2 が数値になっており、一行目はそのラベル名が記述してあります。各品種 8 個体反復のデータが記録されています。

	A	B	
1	treat	data	
2	V13	0.6	
3	V13	0.8	
4	V13	0.9	
5	V13	0.9	
6	V13	0.8	
7	V13	0.7	
8	V13	0.9	
9	V13	0.5	
10	V17	3.1	
11	V17	1.8	
⋮	⋮	⋮	

このデータの、品種間差をチューキー・クレーマー検定してみましょう。

以下で使用しているコマンドは、Web ページのリンク [tukey code.txt](#) に記載しており、これをコピー&ペースすれば操作を再現できます。コマンド入力するときは、タイプミスに注意しましょう。レジユメの通りに結果が表示されずエラー・メッセージがでた場合は、慌てずにエラー内容を注意深く確認する必要があります。

自前のデータを同様に解析する場合は、以降のコードの **緑文字** 部分を適宜書き換えます。まずデータの読み込みに `read.table` 関数を使用し、中身を行列 `d` に代入します。その際、入力ファイルの一行目をヘッダーとみなします（オプション `header=T` で設定）。

```
d <- read.table("tukey_test.txt", header=T)
```

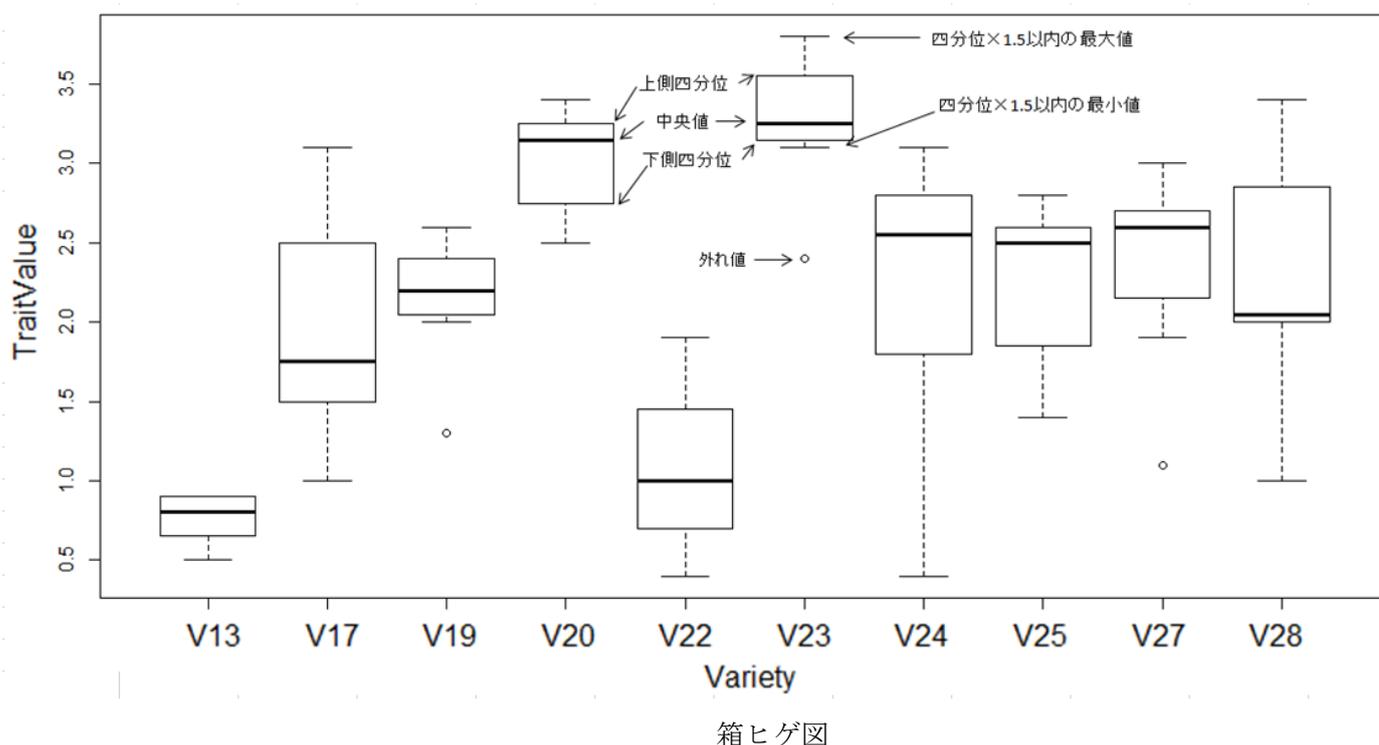
次に、行列 `d` の `treat` 列と、`data` 列をそれぞれ `Variety`(品種名)と `TraitValue`(形質値)に割り当て、作成したデータフレームを新たに `Pd` とします。

```
Pd <- data.frame(Variety=factor(d$treat), TraitValue=d$data)
```

以降の解析には、`Pd` のデータに一次線形モデル $A \sim B$ を当てはめます。 $A \sim B$ は $A = B +$ 残差の意味で、`A` に応答変数、`B` に独立変数を割り当てます。例データの場合、品種間の根長を比較したいので、根長を応答変数、品種を独立変数とします。

この線形モデルは箱ヒゲ図で図示できます。箱ヒゲ図は、各群のデータの中央値とバラツキを効率よく表せる優れた図で、例データのように複数種類の平均の比較をしたい場合に適しています。棒グラフよりも複雑ですが、Rを使うと簡単に作図できます。

```
plot(TraitValue~Variety, data=Pd)
```



データフレーム `Pd` を、根長~品種という線形モデルに基づき一元分散分析します。

```
res1 <- aov(TraitValue~Variety, data=Pd) #一元分散分析  
summary(res1) #分散分析表の出力
```

```

          Df Sum Sq Mean Sq F value    Pr(>F)
Variety     9  41.65   4.628   14.19 1.02e-12 ***
Residuals  70   22.83   0.326
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

チューキー・クレーマー検定を行います。

```
library(multcomp) #パッケージのインストール
```

```

要求されたパッケージ mvtnorm をロード中です
要求されたパッケージ survival をロード中です
要求されたパッケージ splines をロード中です

```

glht 関数を用い、線形モデル根長~品種の多重検定を行います。手法には、チューキー・クレーマー検定を選びました。

```
res2 <- glht(res1, linfct = mcp(Variety = "Tukey")) #チューキー・クレーマー検定
summary(res2) #検定結果を表出します
```

Simultaneous Tests for General Linear Hypotheses

Multiple Comparisons of Means: Tukey Contrasts

```
Fit: aov(formula = TraitValue ~ Variety, data = Pd)
```

Linear Hypotheses:

	Estimate	Std. Error	t value	Pr(> t)	
V17 - V13 == 0	1.1875	0.2856	4.159	0.00339	**
V19 - V13 == 0	1.3875	0.2856	4.859	< 0.001	***
V20 - V13 == 0	2.2625	0.2856	7.923	< 0.001	***
V22 - V13 == 0	0.3125	0.2856	1.094	0.98373	
V23 - V13 == 0	2.5000	0.2856	8.755	< 0.001	***
V24 - V13 == 0	1.4625	0.2856	5.122	< 0.001	***
V25 - V13 == 0	1.5000	0.2856	5.253	< 0.001	***
V27 - V13 == 0	1.6125	0.2856	5.647	< 0.001	***
V28 - V13 == 0	1.5125	0.2856	5.297	< 0.001	***
V19 - V17 == 0	0.2000	0.2856	0.700	0.99943	
V20 - V17 == 0	1.0750	0.2856	3.765	0.01186	*
V22 - V17 == 0	-0.8750	0.2856	-3.064	0.08470	.
V23 - V17 == 0	1.3125	0.2856	4.596	< 0.001	***
:	:	:	:	:	:
V28 - V25 == 0	0.0125	0.2856	0.044	1.00000	
V28 - V27 == 0	-0.1000	0.2856	-0.350	1.00000	

```

---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Adjusted p values reported -- single-step method)

```


最後に、箱ヒゲ図によるチューキー・クレーマー検定結果の描出です。図の上端のアルファベットの昇順が平均の昇順に対応し、平均間に有意差がない場合に同じアルファベットが付与されます。

```
boxp_res2 <- cld(res2) #有意差を付与  
opar <- par(mai=c(1, 1, 1.5, 1)) #余白(底辺、左、上、右)の変更  
plot(boxp_res2) #結果の表示
```

