

## 11/13 の講義内容追記・途中で終わってしまった Tukey-Kramer's HSD test について

R コンソール上で Tukey-Kramer's HSD test を行うモジュールを実行する際の、データの編集方法に誤りがありました。

群馬大の青木氏が公開している `tukey` コマンドは、R コンソール上で

```
> tukey(x, y)
```

という形で実行しますが、`x` には検定に用いるデータを、`y` にはデータが所属する実験区の名称を、1 対 1 の関係で与える必要があったようです。

よって、もとの Excel データ("20091113.xls"のシート"test2")は、次の手順で加工するのみでよいようです。参考として、私が作成した csv ファイルを、<http://shushu.me/archives/test23.csv> に置いておきます。

- 1) 3 行目(`captured_number` の右隣)に、`captured_number` の値を対数変換したものを並べる。  
1 列目は `log_captured_number` などと項目名を記入する。
- 2) そのまま csv ファイルとして保存する。

ファイルを保存後、R コンソールから次の順でコマンドを実行します。

```
> z<-read.csv("test23.csv", header=TRUE, sep=",")
> x<-data$log_captured_number
> y<-data$group
> tukey(x, y)
```

`z` には先の csv ファイルの中身がそのまま代入されます。`x` には `z` のうち "`log_captured_number`" の中身(対数変換した後の数値)が、`y` には各値がどの実験区なのかがそれぞれ代入されます ("`captured_number`" の行は、`x` にも `y` にも代入しない(統計に利用されない)ため、csv ファイルを作るときにあってもなくても構いません)。その上で `tukey(x, y)` を実行すると、`x` と `y` に既に 1:1 の関係で値が代入されているので、Tukey-Kramer's HSD test が実行され、R コンソール上に結果が返されます。

操作が正しくできていれば、以下のような結果が帰ってくるはずです。

```
$result1
      n      Mean  Variance
Group1 15 1.0332134 0.02085730
Group2 16 1.3353327 0.04294126
Group3 16 1.6301354 0.01601811
Group4 16 0.0000000 0.00000000
Group5 16 0.3152327 0.03619281
```

```
$Tukey
      t          p
1:2  5.514995 4.777807e-06
1:3 10.896431 0.000000e+00
1:4 18.860654 0.000000e+00
1:5 13.106281 0.000000e+00
2:3  5.470391 5.711558e-06
2:4 24.778580 0.000000e+00
2:5 18.929087 0.000000e+00
3:4 30.248971 0.000000e+00
3:5 24.399478 0.000000e+00
4:5  5.849493 1.231487e-06
```

```
$phi
[1] 74
```

```
$v
[1] 0.02323358
```

これらの出力されたデータのうち、実際に必要なのは各群を比較した p 値のみになります。  
(先程のデータを `tukey(x, y)` にかけて、なぜか control と D が入れ替わってしまい、A が Group1、  
B が Group2、C が Group3、D が Group5、Control が Group4 と表示されてしまいます)

例えば、"1:2"(Group1 と Group2 の比較)の p 値は"4.777807e-06"(0.000004777807)であり、有意水準 5% の基準である " $p < 0.05$ " よりも非常に小さい値を示しています。よって、Tukey-Kramer's HSD test は、A と B の間には有意差があると判断したことになります。

他の値も一様に  $p < 0.05$  を示している (" $p = 0.000000e+00$ " のところは、p 値が小さすぎて四捨五入したら 0 になってしまったことを示します)ため、Control を含む全ての区は、全て互いに有意差があると判断されます。このような場合、表およびグラフは下のようになります。

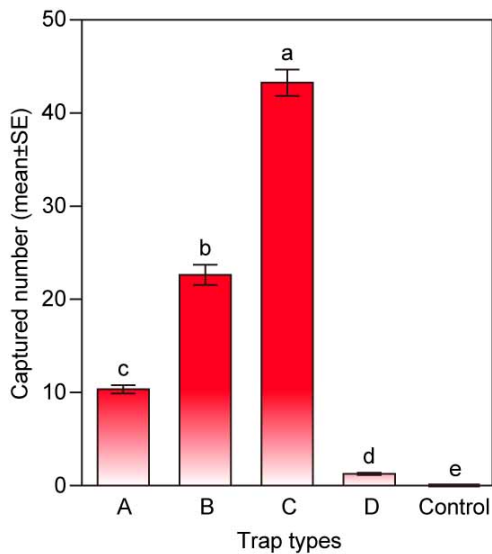


Table 1. (insert title)

Trap types	Captured number (mean±SE)	
A	10.33 ± 0.91	c
B	22.63 ± 2.18	b
C	43.25 ± 2.83	a
D	1.25 ± 0.23	d
Control	0	e

Values followed by a different letter are significantly different at the 5% level by Tukey-Kramer test.

Fig. 1. (insert title)  
Values followed by a different letter are significantly different at the 5% level by Tukey-Kramer test.

今回の例は、全ての実験区間に有意差があるサンプルでしたが、参考として、そうではないサンプルファイルも用意しました。<http://shushu.me/archives/test24.csv> をダウンロードして、先程と同じように `tukey` コマンドを実行してみてください。

```
> z<-read.csv("test24.csv", header=TRUE, sep=",")
> x<-z$log_captured_number
> y<-z$group
> tukey(x, y)
$result1
      n      Mean  Variance
Group1 10 1.454474 0.01318039
Group2 10 1.512200 0.03820642
Group3 10 1.619124 0.02267830
Group4 10 0.000000 0.00000000
Group5 10 1.268503 0.03900162

$Tukey
      t      p
1:2 0.858367 9.104169e-01
1:3 2.448295 1.212610e-01
1:4 21.627609 2.362555e-13
1:5 2.765330 5.991225e-02
2:3 1.589928 5.114320e-01
2:4 22.485976 2.362555e-13
2:5 3.623697 6.292632e-03
3:4 24.075904 2.362555e-13
3:5 5.213624 4.288784e-05
4:5 18.862279 2.362555e-13

$phi
[1] 45

$v
[1] 0.02261335
```

今度は、"1:2"(p=0.9104) "1:3"(p=0.1213) "1:5"(p=0.5599) "2:3"(p=0.5114)の間には有意差が存在しない、つまり「AはB, C, Dと有意差がない」「BとCには有意差がない」「DはAと有意差がないが、B, Cとは有意差がある」「ControlはA, B, C, Dのいずれとも有意差がある」という結果になりました。これを図あるいは表にすると、次のようになります。

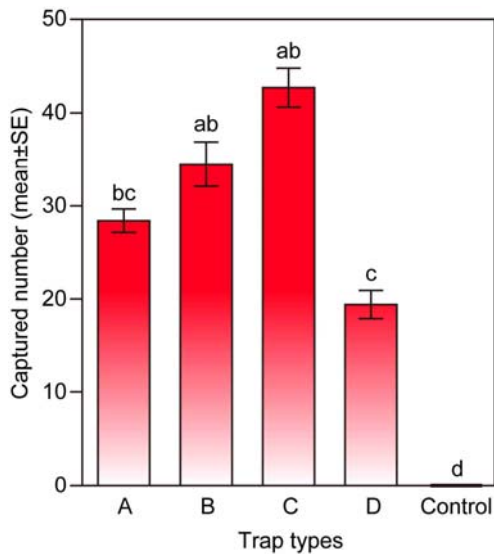


Table 2. (insert title)

Trap types	Captured number (mean±SE)	
A	28.40 ± 2.52	b c
B	34.50 ± 4.78	a b
C	42.70 ± 4.16	a b
D	19.40 ± 3.03	c
Control	0	d

Values followed by a different letter are significantly different at the 5% level by Tukey–Kramer test.

Fig. 2. (insert title)

Values followed by a different letter are significantly different at the 5% level by Tukey–Kramer test.

一見すると符号の関係がわかりにくいですが、「同じ符号の群間には有意差がない」ことを念頭に、ゆっくり考えてみてください。符号と統計結果との関係が見えてくると思います。

レポート課題は、Excel ファイル"20091113.xls"のシート"subject"に含まれるデータに適切な変数変換と統計処理を施し、その結果を「論文に掲載するための英語の表」にすることです。また、レポートには、氏名・学籍番号・課題の表とともに、その表を作成するに至った経緯や思考、必要な知識などをなるべく詳しく書いてください。レポートを作成する際、参考にした資料やサイトなどがあれば、その一覧(文献ならページ数など、サイトならサイト名とURI)を必ず明記してください。講義の出席如何に関わらず、レポート提出者は、レポートの内容に従って A~C の 3 段階で評価します。

シート"subject"の脚注に書き忘れてしまいましたが、表中"Cn-Cc"は *Chrysoperla nipponensis* ♀ × *Chrysoperla carnea* ♂ の交配を示します。他も同様です。

レポートの提出期限は、11月27日(金)の午後5時までとします。

講義内容、レポート課題などについて質問がある場合は、農学部4180号室まで来ていただくか、あるいは [chun@muses.tottori-u.ac.jp](mailto:chun@muses.tottori-u.ac.jp) まで連絡してください。