

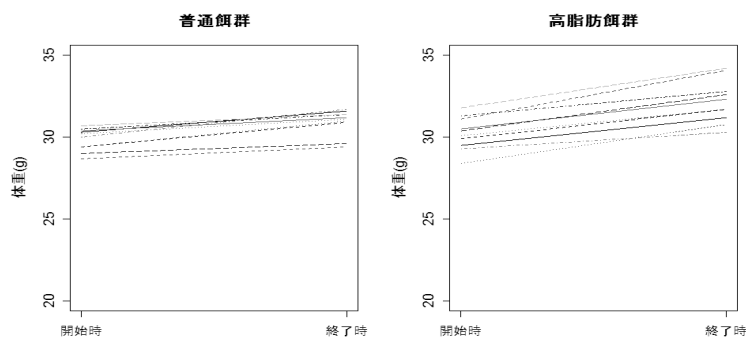
医学情報処理演習第7回「一元配置分散分析と多重比較」*1

2006年11月20日 中澤 港 (nminato@med.gunma-u.ac.jp)

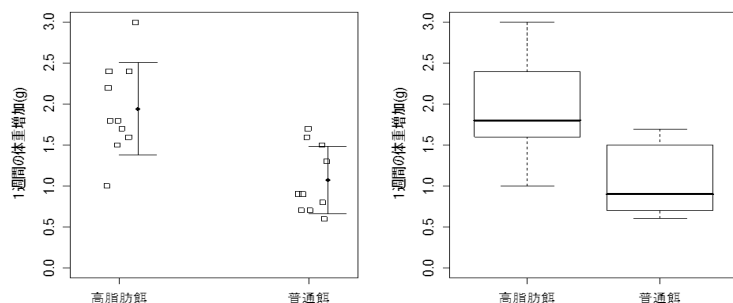
前回の課題の回答例

一見、対応のある検定かと思わせるが、2群にランダムに割り付けたマウスの体重増加を比較するというデザインであり、普通餌摂取群の1匹目と高脂肪餌摂取群の1匹目には何も関係はないので、これは独立2標本の平均値の差の検定(資料 p.3)になる。

図示は、2群別々に体重変化の折れ線グラフを作って並べると、情報量損失を最小限に抑えることができる。このグラフだと、開始時に、体重そのものに2群間で差があった可能性^aがわかる一方、体重変化量を2群間で比較するには向いていない。



むしろ、まず2群別々に1週間の体重増加量を求め、下図のように、増加量を2群間で比較するストリップチャートや箱ヒゲ図を描く方が、この課題で検証すべき「普通餌摂取群と高脂肪餌摂取群で体重変化に差があるかどうか」を見るには適している。



図では差がありそうに見える。続いて「普通餌摂取群と高脂肪餌摂取群で1週間の体重増加に差がない」という帰無仮説を検定してみる。資料 p.3 の母分散が未知の場合になるので、まず「2群の分散に差が無い」という帰無仮説を F 検定する。 $F = 0.5306$, $\text{num df} = 9$, $\text{denom df} = 9$, $p\text{-value} = 0.3591$ という結果から、有意水準 5% で帰無仮説は棄却されない(分散には有意差がない)ので、続いて資料 p.3 分散に差がない場合の通り、通常の t 検定を実施する。結果として $t = -3.9148$, $\text{df} = 18$, $p\text{-value} = 0.001015$ が得られるので、有意水準 5% で帰無仮説は棄却される。以上より、普通餌群と高脂肪餌群では体重増加に有意な差があると判断できる。以上の作図と検定を実施するプログラムを次ページに示す。

^a もしそうだとすると、ランダム割付がうまく行っていなかったということになる。その場合、データを救うのは難しいが、体重増加量ではなく体重増加率で検討するのは一案である。

*1 本資料は <http://phi.med.gunma-u.ac.jp/medstat/it07-2006.pdf> としてダウンロード可能である。

it06-ans-2006.R

```
dat <- read.delim("http://phi.med.gunma-u.ac.jp/medstat/it06-k-2006.txt")
attach(dat)
layout(t(1:2))
matplot(rbind(rep(1,10),rep(2,10)),rbind(NDS,NDE),type="l",
  ylim=c(20,35),ylab="体重 (g)",main="普通餌群",xaxt="n",xlab="")
axis(1,1:2,c("開始時","終了時"))
matplot(rbind(rep(1,10),rep(2,10)),rbind(HFDS,HFDE),type="l",
  ylim=c(20,35),ylab="体重 (g)",main="高脂肪餌群",xaxt="n",xlab="")
axis(1,1:2,c("開始時","終了時"))
NDD <- NDE-NDS
HFDD <- HFDE-HFDS
WeightGain <- c(NDD,HFDD)
Diet <- as.factor(c(rep("普通餌",10),rep("高脂肪餌",10)))
IX <- c(1.1,2.1)
MWG <- tapply(WeightGain,Diet,mean)
SDWG <- tapply(WeightGain,Diet,sd)
layout(t(1:2))
stripchart(WeightGain~Diet,method="jitter",vert=T,ylim=c(0,3),
  ylab="1週間の体重増加 (g)")
points(IX,MWG,pch=18)
arrows(IX,MWG-SDWG,IX,MWG+SDWG,angle=90,code=3)
boxplot(WeightGain~Diet,ylim=c(0,3),ylab="1週間の体重増加 (g)")
print(res<-var.test(NDD,HFDD))
VAREQ <- ifelse(res$p.value < 0.05, FALSE, TRUE)
t.test(NDD,HFDD,var.equal=VAREQ)
detach(dat)
```

多群の平均値を比較する2つの思想

前は2群の平均値を比較したが、今回は3群以上の多群の平均値を比較する方法を説明する。3群以上を比較するために、単純に2群間の差の検定を繰り返すことは誤りである。なぜなら、 n 群から2群を抽出するやりかたは ${}_nC_2$ 通りあって、1回あたりの第1種の過誤(第5回に触れたとおり、本当は差がないのに、誤って差があると判定してしまう確率)を5%未満にしたとしても、3群以上の比較全体として「少なくとも1組の差のある群がある」というと、全体としての第1種の過誤が5%よりずっと大きくなってしまふからである。

この問題を解消するには、(1)多群間の比較という捉え方をやめて、群分け変数が注目している量の変数に与える効果があるかどうかという捉え方にするのが一つの方法であり、具体的には、一元配置分散分析やクラスカル=ウォリス(Kruskal-Wallis)の検定がこれに当たる。

別のアプローチとして、(2)有意水準5%の2群間の検定を繰り返すことによって全体としては大きくなってしまふ第1種の過誤を調整することによって、全体としての検定の有意水準を5%に抑える方法もある。このやり方は「多重比較」と呼ばれる。

これら2つのアプローチは別々に行うというよりも、段階を踏んで行うものと考えるのが一般的である*2。

*2 ただし、永田、吉田(1997)が指摘するように、段階を踏んで実行すると、ここにまた検定の多重性の問題が生じるので、両方はやるべきではない、という考え方も一理ある【典拠：永田靖、吉田道弘『統計的多重比較法の基礎』、サイエンティスト社、1997年】。つまり、厳密に考えれば、群分け変数が量的変数に与える効果があるかどうかを調べたいのか、群間で量的変数に差があるかどうかを調べたいのかによって、これら2つのアプローチを使い分けるべきかもしれない。

一元配置分散分析やクラスカル=ウォリスの検定によって群間に何らかの差があると結論されてから、初めてどの群とどの群の差があるのかを調べるために多重比較を使うというわけである。その意味で、多重比較は *post hoc* な解析と呼ばれることがある。仮に多重比較で有意な結果が出たとしても、一元配置分散分析の結果が有意でなければ、偶然のばらつきの効果が群間の差よりも大きいということなので、特定群間の差に意味があると結論することはできない。

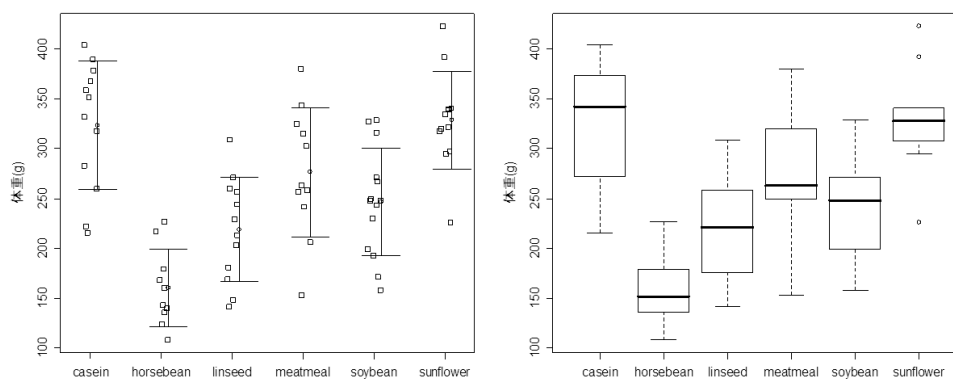
一元配置分散分析

一元配置分散分析の思想は、データのばらつき（変動）を、群間の違いという意味のはっきりしているばらつき（群間変動）と、各データが群ごとの平均からどれくらいばらついているか（誤差）のすべての群についての合計（誤差変動）とに分解し、前者が後者よりもどれくらい大きいかを検討することによって、群分け変数がデータの変数に与える効果が誤差に比べて有意に大きいかどうかを調べるということである。帰無仮説は、「群分け変数がデータの変数に与える効果が誤差の効果に比べて大きくない」ということになる。言い換えると「すべての群の母平均が等しい」が帰無仮説である。

では、具体的に、R に含まれているデータ `chickwts` で説明しよう。これは、既に一部使ったが、71羽の鶏を孵化直後にランダムに6群に分けて、それぞれ異なる餌を与え、6週間後に何グラムになったかを示すデータである（R Console で `chickwts` と入力してヘルプをみると、出典は、Anonymous (1948) *Biometrika*, 35: 214. である）。すべての値を下表に示す。

餌	その餌を食べて6週間育った鶏の体重 (g)
カゼイン (casein)	368, 390, 379, 260, 404, 318, 352, 359, 216, 222, 283, 332
ソラマメ (horsebean)	179, 160, 136, 227, 217, 168, 108, 124, 143, 140
アマニの種 (linseed)	309, 229, 181, 141, 260, 203, 148, 169, 213, 257, 244, 271
肉の配合餌 (meatmeal)	325, 257, 303, 315, 380, 153, 263, 242, 206, 344, 258
大豆 (soybean)	243, 230, 248, 327, 329, 250, 193, 271, 316, 267, 199, 171, 158, 248
ヒマワリの種 (sunflower)	423, 340, 392, 339, 341, 226, 320, 295, 334, 322, 297, 318

`chickwts` はデータフレームであり、体重を示す数値型変数 `weight` と、餌の種類を示す因子型変数 `feed` という形でデータが入っている。変数が2つで、オブザーベーションが71個という形になっていることは、`str(chickwts)` とすれば確認できる。餌の種類によって鶏の体重に差が出るかをみるためには、まずグラフ表示をしてみると、何となく差がありそうに見える。



そこで、群間で体重に差がないという帰無仮説を検定するためには、weight という量的変数に対して、feed という群分け変数の効果を見る形で一元配置分散分析することになる。R Console に入力するコマンドは、summary(aov(weight~feed)) または anova(lm(weight~feed)) である。どちらでも同じ結果が下枠内の通りに得られる。後者は、一元配置分散分析が線形モデルの一種であることを利用した書き方だが、詳しくは後日説明することとし、今回は前者の方法を用いる。

```

              Df Sum Sq Mean Sq F value    Pr(>F)
feed           5 231129   46226  15.365 5.936e-10 ***
Residuals     65 195556    3009
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

このような結果の表を分散分析表という。右端の*の数には有意性を示す目安だが、有意確率そのもの(Pr(>F)の下の数字)に注目してみるほうがよい。

Sum Sq は、平方和 (sum of squares) の略である。

feed の Sum Sq の値 231129 は、餌の種類が異なる群ごとの平均値から総平均を引いて二乗した値を、餌の種類が異なる群ごとの鶏の個体数で重み付けした和である。群間変動または級間変動と呼ばれ、feed 間でのばらつきの程度を意味する。

Residuals の Sum Sq の値 195556 は各鶏の体重から、その鶏が属する餌群の鶏の平均体重を引いて二乗したものの総和であり、誤差変動と呼ばれ、餌群によらない(それ以外の要因がないとすれば偶然の)ばらつきの程度を意味する。

Mean Sq は平均平方和 (mean square) の略であり、平方和を自由度 (Df) で割ったものである。平均平方和は分散なので、feed の Mean Sq の値 46226 は群間分散または級間分散と呼ばれることがあり、Residuals の Mean Sq の値 3009 は誤差分散と呼ばれることがある。

F value は分散比と呼ばれ、群間分散の誤差分散に対する比である。この場合の分散比は第 1 自由度 5、第 2 自由度 65 の F 分布に従う。一般に、一元配置分散分析の場合は、対立仮説の下では $F > 1$ であることが期待されるため右片側検定すればよく、分散比がこの実現値よりも偶然大きくなる確率は、 $1 - \text{pf}(15.365, 5, 65)$ で得られる。Pr(>F) の下の数字は、まさにその値を示すものである。この例では 5.936×10^{-10} と*³、ほとんどゼロといえるくらい小さいので、feed の効果は 5%水準で有意であり、帰無仮説は棄却される。つまり、鶏の体重は、生後 6 週間に与えた餌の種類によって差があることになる。

ただし、一元配置分散分析は、各群が等しい母分散をもつ正規分布に従うことを仮定して行っているパラメトリックな分析法なので、データの分布がひどく歪んでいる場合は結果がおかしくなる場合がある。各群の母分散が等しいかどうかを調べる検定法として、パートレット (Bartlett) の検定と呼ばれる方法がある。R では bartlett.test(量的変数~群分け変数) で実行できる。帰無仮説「各群の母分散が等しい」が棄却された場合は、後日説明するクラスカル=ウォリスの検定のような、ノンパラメトリックな方法を使った方がいいかもしれない。

この例では、bartlett.test(weight~feed) と入力して得られる結果の p-value をみると、0.66 であり、5%よりずっと大きいので帰無仮説は棄却されず、一元配置分散分析を実行しても問題ないことになる。以上の作図と検定を実行する R のコードは次の枠内の通りである*⁴。

*³ 注: 5.936×10^{-10} の意味。 $1 - \text{pf}(15.365, 5, 65)$ の結果とは 10^{-13} の位で 1 違うのは F value の丸め誤差による。

*⁴ ただし、きれいな作図のため、グラフィックデバイスとして Windows 拡張メタファイルを指定し、図の大きさとフォントも指定するため、layout(t(1:2)) の前に win.metafile("./it07-1-2006.emf", width=13, height=6, pointsize=12) と、par(family="sans") を実行し、作図終了後、つまり boxplot(...) の次の行に dev.off() を実行した。できた拡張メタファイルを IrfanView で開いて、Adobe のポストスクリプトプリンタドライバである Acrobat Distiller J に、用紙サイズを幅 13cm、

it07-1-2006.R

```
attach(chickwts)
mw <- tapply(weight,feed,mean)
sw <- tapply(weight,feed,sd)
ix <- 1:length(table(feed))+0.1
layout(t(1:2))
stripchart(weight~feed,vert=T,method="jitter",ylab="体重 (g)")
points(ix,mw)
arrows(ix,mw-sw,ix,mw+sw,angle=90,code=3)
boxplot(weight~feed,ylab="体重 (g)")
print(res.bt <- bartlett.test(weight~feed))
ifelse(res.bt$p.value<0.05,
cat("不等分散！ Bartlett の検定で p=",res.bt$p.value,"\n"),
summary(aov(weight~feed)))
detach(chickwts)
```

検定の多重性を調整する「多重比較」

この鶏の体重の例では、一元配置分散分析の結果、餌群の効果が有意だったので、次に調べたいことは、具体的にどの餌とどの餌の間で差がでてくるかであろう。

単純に考えると、2種類の餌ずつ、

it07-2-2006.R

```
attach(chickwts)
kf <- names(table(feed))
k <- length(kf)
for (i in 1:(k-1)) { for (j in (i+1):k) {
  cat("** Compare ",kf[i]," and ",kf[j]," **\n")
  print(RV <- var.test(weight[feed==kf[i]],weight[feed==kf[j]]))
  ifelse(RV$p.value<0.05,VRES<-FALSE,VRES<-TRUE)
  print(t.test(weight[feed==kf[i]],weight[feed==kf[j]],var.equal=VRES)) }}
detach(chickwts)
```

と t 検定を繰り返せば良さそうであり、この方法が使われている本や論文もないわけではない。しかし、6種類の餌についてこれをやると6つから2つを取り出す全ての組み合わせについて検定するため、15回の比較をすることになり、個々の検定について有意水準を5%にすると、全体としての第1種の過誤は明らかに5%より大きくなる。したがって、先に述べた通り、 t 検定の繰り返しは不都合である。これに似た方法として無制約LSD（最小有意差）法やFisherの制約つきLSD法（一元配置分散分析を行って有意だった場合にのみLSD法を行うという方法）があるが、これらも第1種の過誤を適切に調整できない（ただし制約つきの場合は3群なら大丈夫）ことがわかっているので、使ってはいけない。現在では、この問題は広く知られているので、 t 検定の繰り返しやLSD法で分析しても論文はacceptされない。

多重比較の方法にはいろいろあるが^{*5}、ボンフェローニ (Bonferroni) の方法、シェフェ (Scheffé) の方法、ダンカン (Duncan) の方法、テューキー (Tukey) のHSD、ダネット (Dunnett) の方法、ウィリアムズ

高さ6cm, ポストスクリプトオプションでEPSと指定して出力し、直接 encapsulated postscript ファイルを作成して pLaTeX に取り込んでいる。前回の回答例のグラフでも同様のテクニックを用いた。

^{*5} 以下、一般論は、「Rによる統計解析の基礎」第10章の内容と基本的に同じである。

(Williams)の方法がよく使われている。このうち、ダンカンの方法は、数学的に間違っていることがわかってるので使ってはいけない。ボンフェローニの方法とシェフェの方法も検出力が悪いので、特別な場合を除いては使わない方がよい。データが正規分布に近ければ、チューキーのHSDを使うべきである。ダネットの方法は対照群が存在する場合に対照群と他の群との比較に使われるので、適用場面が限定されている*6。ウィリアムズの方法は対照群があって他の群にも一定の傾向が仮定される場合には最高の検出力を発揮する。

上記いくつかの方法がよく使われている理由は、限定された用途で高い検出力をもつダネットとウィリアムズを除けば、たんにそれらが歴史的に古く考案され、昔の統計学の教科書にも説明され、多くのソフトウェアに実装されているからに過ぎない。現在では、かなり広い用途をもち、ノンパラメトリックな分析にも適応可能なホルム (Holm) の方法 (ボンフェローニの方法を改良して開発された方法) が第一に考慮されるべきである。その上で、全ての群間の比較をしたい場合はペリ (Peritz) の方法、対照群との比較をしたいならダネットの逐次棄却型検定 (これはステップダウン法と呼ばれる方法の1つであり、既に触れたダネットの方法とは別) も考慮すればよい。

多重比較においては、帰無仮説が単純ではない。例えば、3群間の差を調べるとしよう。一元配置分散分析での帰無仮説は、 $\mu_1 = \mu_2 = \mu_3$ である。これを包括的帰無仮説と呼び、 $H_{\{1,2,3\}}$ と書くことにする。さて第1群から第3群までの母平均 $\mu_1 \sim \mu_3$ の間で等号関係が成り立つ場合をすべて書き上げると、 $H_{\{1,2,3\}} : \mu_1 = \mu_2 = \mu_3$, $H_{\{1,2\}} : \mu_1 = \mu_2$, $H_{\{1,3\}} : \mu_1 = \mu_3$, $H_{\{2,3\}} : \mu_2 = \mu_3$ の4通りである。このうち、 $H_{\{1,2,3\}}$ 以外のものを部分帰無仮説と呼ぶ。

すべての2つの群の組み合わせについて差を調べるということは、 $\{H_{\{1,2\}}, H_{\{1,3\}}, H_{\{2,3\}}\}$ が、考慮すべき部分帰無仮説の集合となる。第1群が対照群で他の群のそれぞれが第1群と差があるかどうかを調べたい場合は、考慮すべき部分帰無仮説の集合は $\{H_{\{1,2\}}, H_{\{1,3\}}\}$ となって、「すべての2つの群の組み合わせについて調べる」場合とは異なる。これらの集合をその多重比較における「帰無仮説族」と呼ぶ。

ここで多重比較の目的を「帰無仮説族」というコトバを使って言い換えてみる。個々の帰無仮説で有意水準を5%にしてしまうと、帰無仮説族に含まれる帰無仮説のどれか1つが誤って棄却されてしまう確率が5%より大きくなってしまう。それではまずいので、その確率が5%以下になるようにするために、何らかの調整を必要とするわけで、この調整をする方法が多重比較なのである。つまり、帰無仮説族の有意水準を定める (例えば5%にする) ことが、多重比較の目的である。このことからわかるように、差のなさそうな群をわざと入れておいて帰無仮説族を棄却されにくくしたり、事後的に帰無仮説を追加したりすることは、統計を悪用していることになり、やってはいけない。

計算方法については、ボンフェローニとホルム、チューキーのHSDだけを簡単に紹介する。より詳しく知りたい場合には、永田、吉田 (1997) を参照されたい。

ボンフェローニの方法

ボンフェローニの方法とは、ボンフェローニの不等式に基づく多重比較法である。きわめて単純な考え方に基づいているために、適用可能な範囲が広いが、検出力が落ちてしまいがちなので、ベストな方法ではない。ボンフェローニの不等式とは、 k 個の事象 E_i ($i = 1, 2, \dots, k$) に対して成り立つ、

$$Pr(\cup_{i=1}^k E_i) \leq \sum_{i=1}^k Pr(E_i)$$

*6 ただし、対照群が他の群との比較のすべての場合において差があるといいたい場合は、多重比較をするのではなく、 t 検定を繰り返して使うのが正しいので注意が必要である。

をいう。左辺は k 個の事象 E_i のうち少なくとも1つが成り立つ確率を示し、右辺は各事象 E_i が成り立つ確率を加え合わせたものなので、この式が成り立つことは自明であろう（個々の事象がすべて独立な場合にのみ等号が成立する）。

次に、この不等式を多重比較にどうやって応用するかを示す。まず、帰無仮説族を $\{H_{01}, H_{02}, \dots, H_{0k}\}$ とする。 E_i を「正しい帰無仮説 H_{0i} が誤って棄却される事象」と考える。この表現をボンフェローニの不等式にあてはめれば、

$$\begin{aligned} & Pr(\text{正しい帰無仮説のうちの少なくとも1つが誤って棄却される}) \\ & \leq \sum_{i=1}^k Pr(\text{正しい帰無仮説 } H_{0i} \text{ が誤って棄却される}) \end{aligned}$$

右辺が α 以下になるためには、もっとも単純に考えれば、足しあわされる各項が α/k に等しいかより小さければよい。つまり、ボンフェローニの方法とは、有意水準 α で帰無仮説族を検定するために、個々の帰無仮説の有意水準を α/k にするものである*7。手順としてまとめると、以下の通りである。

1. 帰無仮説族を明示し、そこに含まれる帰無仮説の個数 k を求める。
2. 帰無仮説族についての有意水準 α を定める。 $\alpha = 0.05$ または $\alpha = 0.01$ と定めることが多い。
3. 帰無仮説族に含まれているそれぞれの帰無仮説に対して検定統計量 T_i ($i = 1, 2, \dots, k$) を選定する。
4. データを取り、検定統計量 T_i を計算する。
5. 各検定統計量 T_i について有意水準 α/k に対応する棄却限界値（通常は分布関数の $(1 - \alpha/k) \times 100\%$ 点を c_i とするとき、 $T_i \geq c_i$ ならば H_{0i} を棄却し、 $T_i < c_i$ なら H_{0i} を保留する（採択ではない））

なお、R では、各々の帰無仮説の有意水準を α/k とする代わりに、各々の帰無仮説に対して得られる有意確率が k 倍されて（ただし1を超えるときは1として）表示されるので、各々の比較に対して表示される有意確率と帰無仮説族について設定したい有意水準との大小によって仮説の棄却/保留を判断してよい。

ホルムの方法

ボンフェローニの方法では、すべての H_{0i} について有意水準を α/k としたのが良くなかったので、ホルムの方法は、そこを改良したものである。以下、ホルムの方法の手順をまとめる。

1. 帰無仮説族を明示し、そこに含まれる帰無仮説の個数 k を求める。
2. 帰無仮説族についての有意水準 α を定める。 $\alpha = 0.05$ または $\alpha = 0.01$ と定めることが多い。ここまではボンフェローニと同じ。
3. $\alpha_1 = \alpha/k, \alpha_2 = \alpha/(k-1), \dots, \alpha_k = \alpha$ を計算する。
4. 帰無仮説族に含まれているそれぞれの帰無仮説に対して検定統計量 T_i ($i = 1, 2, \dots, k$) を選定する。
5. データを取り、検定統計量 T_i を計算する。
6. 各検定統計量 T_i について有意確率 P_i を求め、小さい順に並べ換える。
7. P_i の小さいほうから順に α_i と P_i の大小を比べる。
8. $P_i > \alpha_i$ ならばそれよりも有意確率が大きい場合の帰無仮説をすべて保留して終了する。 $P_i \leq \alpha_i$ なら H_{0i} を棄却して、次に小さい P_i について比較する。 $i = k$ となるまで繰り返す。

ホルムの方法についても、R では、手順7. で P_i と α_i の大小を比べる代わりに $P'_i = P_i \times (k - i + 1)$ が表示されるので、値そのものを有意水準と比較すればよい。

*7 ここで注意しなければいけないことは、検定すべき帰無仮説族に含まれる個々の帰無仮説は、データをとるまえに定められていなければならないことである。データをとった後で有意になりそうな帰無仮説を k 個とってきて帰無仮説族を構成するのでは、帰無仮説族に対しての第1種の過誤をコントロールできないのでダメである。

テューキーの HSD

テューキーの HSD では、母集団の分布は正規分布とし、すべての群を通して母分散は等しいと仮定する。

データが第 1 群から第 a 群まであって、各々が n_i 個 ($i = 1, 2, \dots, a$) のデータからなるものとする。第 i 群の j 番目のデータを x_{ij} と書くことにすると、第 i 群の平均 \bar{x}_i と分散 V_i は、

$$\begin{aligned}\bar{x}_i &= \sum_{j=1}^{n_i} x_{ij} / n_i \\ V_i &= \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / (n_i - 1)\end{aligned}$$

となり、誤差自由度 P_E と誤差分散 V_E は、

$$\begin{aligned}P_E &= N - a = n_1 + n_2 + \dots + n_a - a \\ V_E &= \sum_{i=1}^a \sum_{j=1}^{n_i} (x_{ij} - \bar{x}_i)^2 / P_E = \sum_{i=1}^a (n_i - 1) V_i / P_E\end{aligned}$$

で得られる。

簡単にいえば、テューキーの HSD は、すべての群間の比較について誤差分散を使った t_0 統計量を計算し、 t 分布ではなくて、ステューデント化された範囲の分布 (Studentized range distribution) と呼ばれる分布の $(1 - \alpha) \times 100\%$ 点を $\sqrt{2}$ で割った値との大小で有意水準 α の検定をする方法である。以下手順としてまとめると。

1. 帰無仮説族を明示する。テューキーの HSD の場合は、通常、

$$\{H_{\{1,2\}}, H_{\{1,3\}}, \dots, H_{\{1,a\}}, H_{\{2,3\}}, \dots, H_{\{a-1,a\}}\}$$

2. 有意水準 α を定める。 $\alpha = 0.05$ または $\alpha = 0.01$ と定めることが多い。
3. データを取り、すべての群について \bar{x}_i, V_i を計算し、 P_E, V_E を計算する。
4. すべての 2 群間の組み合わせについて、検定統計量 t_{ij} を

$$t_{ij} = (\bar{x}_i - \bar{x}_j) / \sqrt{V_E(1/n_i + 1/n_j)}$$

により計算する ($i, j = 1, 2, \dots, a; i < j$)

5. $|t_{ij}| \geq q(a, P_E; \alpha) / \sqrt{2}$ なら $H_{\{i,j\}}$ を棄却し、 i 群と j 群の平均値には差があると判断する (比較の形からわかるように、これは両側検定である)。 $|t_{ij}| < q(a, P_E; \alpha) / \sqrt{2}$ なら $H_{\{i,j\}}$ を保留する。ここで $q(a, P_E; \alpha)$ は、群数 a 、自由度 P_E のステューデント化された範囲の分布の $(1 - \alpha) \times 100\%$ 点である。つまり、 $\alpha = 0.05$ ならば、 $q(a, P_E, 0.05)$ は、群数 a 、自由度 P_E のステューデント化された範囲の分布の 95% 点である。R では、この値を与える分位点関数は、群数 a 、自由度 df として、`qtukey(0.95, a, df)` だが、すべての群間比較を手計算するのは面倒なので、`TukeyHSD()` 関数を使って自動的に実行させるのが普通である。

上の鶏の体重の例について、実際に R で多重比較を試みよう。

```
attach(chickwts)
pairwise.t.test(weight, feed, p.adjust.method="bonferroni")
detach(chickwts)
```

とすれば、ボンフェローニの方法で有意水準を調整した、すべての餌群間の体重の差を t 検定した結果の有意確率を、下三角行列の形で出してくれる*8。もし `p.adjust.method` を指定しなければホルムの方

*8 ただし、 t 検定とは言っても、`pool.sd=F` というオプションをつけない限りは、 t_0 を計算するときに全体の誤差分散を使うので、ただの t 検定の繰り返しとは違う。

法になるが、明示したければ、`p.adjust.method="holm"`とすればよい。データが正規分布していれば、`TukeyHSD(aov(weight~feed))`としてテューキーのHSDを行ってもよい。実行結果は、2群ずつの組み合わせのそれぞれについて、テューキーのHSDで調整した、差の95%信頼区間という形で得られる。`lwr`と`upr`の数値の間にゼロが含まれない組み合わせには有意水準5%で有意差があるといえる。

なお、CRANから`multcomp`パッケージをインストールしておけば他の多重比較も可能である*⁹。

多群の平均値を比べたいときの手順をまとめる。

- 群ごとの分布の正規性をチェックする。正規分布と大きくずれていたらノンパラを考える。そうでなければ次へ。
- Bartlettの検定で「群間で分散に差がない」帰無仮説を検定する。帰無仮説が棄却されたら差がないとはいえないのでノンパラとか別の方法を考える（原則として）。棄却されなければ次へ。
- 一元配置分散分析で群分け変数がデータに有意な効果を与えているか検討する。有意でなければ群間で平均値に差がないと判断される。有意なら次へ。
- どの群とどの群に差があるか、検定の多重性を調整しながら検定する。TukeyのHSDかHolmの方法が普通。

課題

<http://phi.med.gunma-u.ac.jp/medstat/p07.txt> は、パプアニューギニアのある地方の4つの村に居住する成人男性を対象に、約20年前に行われた血液検査の結果得られた、ヘモグロビン濃度のデータを一部加工した（演習のため架空のデータを付け加えたり削除したりした）ものである。VILという変数が村の番号を示し（その地方にある13の村のうち、異なる生態学的条件を代表する典型的な4つの村を選んだ）、HBがヘモグロビン濃度(g/dL)を示す。村によってヘモグロビン濃度に差があるかどうか検討せよ。

配布するレポート用紙に検定方法、検定結果（検定統計量、自由度、有意確率）とその解釈を文章として記入し、学籍番号と氏名も記入して提出すること。結果の提出をもって出席確認とする。

*⁹ 管理者権限があれば、`install.packages("multcomp", CRAN="http://cran.md.tsukuba.ac.jp/")`でインストールできる。`library(multcomp)`と打ってから、`simtest(weight~feed, type="Dunnett")`とするとダネットの方法での多重比較が実行できる。群分け変数の最初のカテゴリが対照群であるとみなされるので、このデータでは`casein`を飼育されて育った鶏が対照群となる。もっともこれは、このデータに相応しい解析ではない。