医学情報処理演習第６回「２群の平均値の差の検定」[*1]

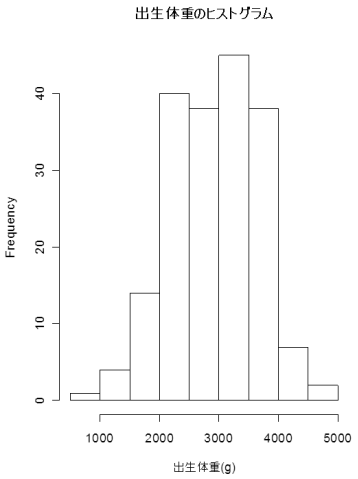
2006 年 11 月 13 日 中澤　港 (nminato@med.gunma-u.ac.jp)

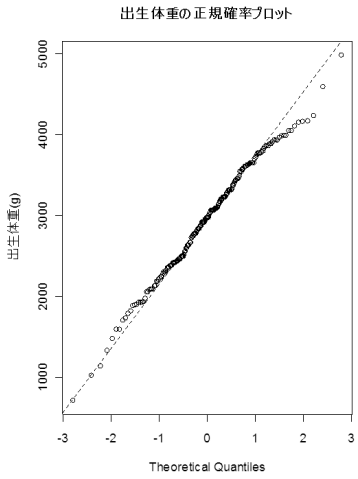
前回の課題の回答例

it05-ans-2006.R

```
library(MASS)
attach(birthwt)
layout(t(1:2))
hist(bwt,main="出生体重のヒストグラム",xlab="出生体重 (g)")
qqnorm(bwt,main="出生体重の正規確率プロット",ylab="出生体重 (g)")
qqline(bwt,lty=2)
source("http://phi.med.gunma-u.ac.jp/medstat/it05-3-2006.R")
shapiro.test(bwt)
geary.test(bwt)
detach(birthwt)
```

結果は次のグラフと枠内の通り（情報量のない行は削除済み)。グラフを見るだけでも正規分布に近そうだと見当はつく。何度も繰り返すが，検定よりも先に作図をすることは非常に重要である。






```
    Shapiro-Wilk normality test
data:  bwt
W = 0.9924, p-value = 0.4354

Geary's test for normality:
 G= 0.8126568  / p= 0.1693836
```

「出生体重データは正規分布にしたがう」という帰無仮説を，Shapiro-Wilk の検定と Geary の検定により，有意水準 5%で検定した結果，有意確率 (p 値) が 0.05 よりずっと大きいので，有意水準 5%で帰無仮説は棄却できない。よって，とりあえず正規分布にしたがっているとみなしてよい。

[*1] 本資料は http://phi.med.gunma-u.ac.jp/medstat/it06-2006.pdf としてダウンロード可能である。

母平均値と標本平均の差の検定

第 4 回で，標本データから母平均とその 95%信頼区間を推定する方法を説明したが，その復習も兼ねて，まずは標本平均と母平均の差の検定を扱ってみる。なお，ここでは検定だけを説明するが，R の出力には 95%信頼区間も表示されるので，統計的な結果としては，そちらの方が実は情報量が多い[*2]。ただ，疫学の専門誌以外は仮説検定の方が通りがいい場合が多いので，この講義では検定を中心に説明する。

サイズ n の標本 X について, 標本平均 $E(X) = \sum_{i=1}^{n} X_i/n$ と既知の母平均値 μ_X の差の検定は, 母分散 V_X が既知のとき, $z_0 = |E(X) - \mu_X|/\sqrt{V_X/n}$ が標準正規分布に従うことを使って検定できる[*3]。V_X が未知のときは，標本の不偏分散 $S_X = \sum_{i=1}^{n} (X_i - E(X))^2/(n-1) = \mathrm{var}(X)$ を使って，$t_0 = |E(X) - \mu_X|/\sqrt{S_X/n}$ が自由度 $n-1$ の t 分布に従うことを使って検定できる（暗黙の仮定として，ランダムサンプルで，母集団の分布が正規分布であることが必要)。つまり，t_0 が自由度 $n-1$ の t 分布の 2.5%点より小さいか 97.5%点より大きかったら，有意水準 5%で有意差があるとみなす。前回の資料にもあるように，この場合，帰無仮説が「差がない」であり，対立仮説は「大きいか小さい」なので，このような**両側検定**になる。実用上，両側検定の場合は，t 分布はゼロに対して左右対称なので，有意確率は，t_0 に対する確率母関数の値を 1 から引いた上側確率を 2 倍すれば得られる[*4]。

第 4 回に説明した，未知の母平均値の信頼区間の推定はこの裏返しである。 つまり，母平均値の 95%信頼区間の下限は，不偏分散を標本数 n で割ったものの平方根に自由度 $n-1$ の t 分布の 97.5%点を掛けた値を標本平均から引いた値になり，上限は，同じ値を標本平均に足した値になる。

R では，第 4 回に触れたように，`t.test()` 関数がこれらを両方やってくれる。例えば，量的変数 `X` が母平均 120 の母集団からのランダムサンプルであるという帰無仮説を検定するには，`t.test(X,mu=120)` とする。`X <- rnorm(100,120,10)` の場合と，`X <- rnorm(100,110,10)` の場合について結果を比べてみるとよい。

例題

平成 10 年の国民栄養調査によれば，50-59 歳男性の平均 BMI は 23.6 であった。同じ年にA社の職員健診を受診した 50-59 歳男性 248 人の平均 BMI が 24.6 で，その不偏分散が 8.6 であったとき，A社の 50-59 歳男性の BMI の平均値は全国平均と差があるといえるかどうか検定せよ。

母分散が未知なので，標本の不偏分散で代用すれば，$t_0 = |24.6 - 23.6|/\sqrt{8.6/248} = 5.37$ より，自由度 247 の t 分布で 5.37 よりも大きい値をとる確率はほぼ 0 なので，両側検定のために 2 倍しても有意に異なるといえる。R のコマンド入力を以下のようにすると，有意確率が得られる。

```
t0 <- (24.6-23.6)/sqrt(8.6/248)
2*(1-pt(t0,247))
```

[*2] 「平均値の差がない」という帰無仮説を有意水準 5%で検定するよりも，平均値の差の 95%信頼区間を推定する方が情報量が多い。平均値の差の 95%信頼区間が 0 を跨いでいれば，「差がない」という帰無仮説が棄却されないことがわかるので，信頼区間を表示すれば仮説検定の結果も同時に分かる。けれども，区間推定をしたということは，区間そのものの方が，有意差の有無を判別するという意思決定よりも重要だとみなしているということなので，その結果について検定的な解釈をすべきではない。

[*3] つまり，$E(X)$ が，平均 μ_X，標準偏差 $\sqrt{V_X/n}$ の正規分布に従うということ。これは中心極限定理そのものである。

[*4] データ以外の情報によって予め X が母平均より小さくなることはないとわかっているときは，小さい側を考えなくてよくなるので，有意確率は t_0 に対する t 分布の確率母関数の値を 1 から引いた上側確率そのものとなるし，95%信頼区間も 95%点を考えればよい。このような場合を片側検定とよぶ。

独立２標本の平均値の差の検定

次に，標本調査によって得られた独立した２つの量的変数 X と Y（サンプル数が各々 n_X と n_Y とする）について，平均値に差があるかどうかを検定することを考える。

母分散が既知で等しい V である場合（稀）

この場合は，言い換えると，これらの独立 2 標本が同じ母集団からのサンプルであるというのが帰無仮説になる。$z_0 = |E(X) - E(Y)|/\sqrt{V/n_X + V/n_Y}$ が標準正規分布に従うことを使って検定する。

母分散が未知の場合（通常はこちら）

1. F **検定（2 群の分散に差が無いという帰無仮説の検定）**：２つの量的変数 `X` と `Y` の不偏分散 `SX<-var(X)` と `SY<-var(Y)` の大きい方を小さい方で（以下の説明では `SX>SY` だったとする）割った `F0<-SX/SY` が第 1 自由度 `DFX<-length(X)-1`，第 2 自由度 `DFY<-length(Y)-1` の F 分布に従うことを使って検定する（一般に，互いに独立な分散の比は F 分布に従うと考えてよい）。有意確率は `1-pf(F0,DFX,DFY)` で得られる。しかし，`F0` を手計算しなくても，`var.test(X,Y)` で等分散かどうかの検定が実行できる[*5]。また，１つの量的変数 `X` と１つの群分け変数 `C` があって，`C` の２群間で `X` の分散が等しいかどうか検定するというスタイルでデータを入力してある場合は，`var.test(X~C)` とすればよい。
2. 分散に差があるか差がないかによって，平均値が等しいかどうかの検定法は異なる（以下に詳述）。分散に差があるときは，その事実をもって別の母集団からとられた標本であると判断し，平均値が等しいかどうかを検定する意味はないとする考え方もあるが，一般には Welch の方法を使うか，ノンパラメトリックな方法[*6]を使って検定する。

分散に差がない場合

まず母分散 `S` を `S<-(DFX*SX+DFY*SY)/(DFX+DFY)` として推定する（２つの分散の自由度で重み付けした平均をとる）。

`t0<-abs(mean(X)-mean(Y))/sqrt(S/length(X)+S/length(Y))` が自由度 `DFX+DFY` の t 分布に従うことから，帰無仮説「X と Y の平均値には差がない」を検定すると，`(1-pt(t0,DFX+DFY))*2` が有意確率となる。両側検定なので上側確率を出して 2 倍している。

R では，`t.test(X,Y,var.equal=T)` とする。また，F 検定のところで触れた量的変数と群分け変数という入力の仕方の場合は，`t.test(X~C,var.equal=T)` とする。ただしこれだと両側検定なので，片側検定したい場合は，`t.test(X,Y,var.equal=T,alternative="less")` などとする（`alternative="less"`は対立仮説が `X<Y` という意味なので，帰無仮説が `X>=Y` であることを意味する）。

[*5] 「R による統計解析の基礎」では第 3 刷まで，「この場合は，R が勝手に入れ替えてくれるので，X の不偏分散の方が Y の不偏分散より大きいかどうか気にしなくてもよい。」と書いていたが，実は，古川・丹後「医学への統計学」（朝倉書店）で２つの方法の１つとして触れられている，「帰無仮説：SX=SY，対立仮説：SX≠SY」で大小を区別せず F 比を算出して両側検定するのがデフォルトになっているので注意されたい。

[*6] 例えば Mann-Whitney の U 検定（Wilcoxon の順位和検定と数学的に同値）が良く用いられる。その場合は，代表値としても平均値と標準偏差でなく，中央値と四分位範囲または四分位偏差を表示するのが相応しい。ノンパラメトリックな方法は，分散が異なる場合というよりも，分布が歪んでいたり外れ値がある場合によく用いられる。

分散に差がある場合（Welch の方法）

$t_0 = |E(X) - E(Y)|/\sqrt{S_X/n_X + S_Y/n_Y}$ が自由度 ϕ の t 分布に従うことを使って検定する。但し，ϕ は下式による。

$$\phi = \frac{(S_X/n_X + S_Y/n_Y)^2}{\{(S_X/n_X)^2/(n_X-1) + (S_Y/n_Y)^2/(n_Y-1)\}}$$

R では，`t.test(X,Y,var.equal=F)` だが，`var.equal` の指定を省略した時は等分散でないと仮定して Welch の検定がなされるので省略して `t.test(X,Y)` でいい。量的変数と群分け変数という入力の仕方の場合は，`t.test(X~C)` とする。

実は，`t.test(X,Y,var.equal=(var.test(X,Y)$p.value>=0.05))` とすれば，自動的にこれら 2 つの場合に応じた分析が行われるが，ただ機械的にそう実行するのではなく，`stripchart()` や `boxplot()` などで 2 群をプロットし，生データのばらつきと位置の様子を確認した上で，`var.test()` の結果もみて，それに応じて `t.test()` を実行する方がよい。

例題

2001 年に，厚生科学研究で「少子化の見通しに関する専門家調査」が行われた。この調査は，「人口学，経済学，家族社会学，公衆衛生学を中心とした専門家を対象として少子化研究会のメンバーが対象候補者を抽出し，回答者の偏りや不足等について検討を加えた上で，748 名を対象として調査を実施した」もので，回収率は 44%であった。この調査では，2025 年の合計出生率がいくつになるかという予測値があるが，出生率がそのうち回復するとみるか，低下し続けるとみるかという質問項目もあり，この答えの違いによって，2025 年の予測値には違いがあると考えられる。
回復するとみる人たちの 2025 年の合計出生率の予測値は，`1.40 1.40 1.56 1.50 1.40 ...` **（後略）** となっており（サンプルサイズ 58，平均 1.487，不偏分散 0.0275），低下し続けるとみる人たちの予測値は，`1.38 1.30 1.15 1.31 1.37 ...` **（後略）** となっていた（サンプルサイズ 221，平均 1.356，不偏分散 0.0211）。2 群の平均値に有意な差があるといえるか，有意水準 5%で検定せよ。

R で計算するには，まず下枠内のように入力する。

```
F0 <- 0.0275/0.0211
1-pf(F0,57,220)
```

とすれば，0.0915... という結果が得られるので分散に有意差はないといえる。従って，Welch の方法でなく，通常の t 検定を行う。

```
S <- ((58-1)*0.0275+(221-1)*0.0211)/(58+221-2)
t0 <- abs(1.487-1.356)/sqrt(S/58+S/221)
2*(1-pt(t0,58+221-2))
```

の結果として `8.97506e-09` が得られ（8.97506×10^{-9} という意味），5%より遙かに小さいので，出生率の見通しの異なる専門家集団間で，2025 年合計出生率の予測値の平均は有意に異なるといえる。

なお，このように，既に平均値と不偏標準偏差が計算されている場合の図示は，エラーバー付きの棒グラフを使うことが多い（誤解を生む場合があるので，必ずしもいい図示ではないのだが，伝統的に良く使われている）。R では，下枠内のように入力すれば作図できる。

```
X <- c(1.487,1.356)
SX <- c(sqrt(0.0275),sqrt(0.0211))
IX <- barplot(X,ylim=c(0,max(X+SX)*1.5))
arrows(IX,X,IX,X+SX,angle=90)
```

一方，生データがあるときの図示には，`stripchart()` か `boxplot()` を用いる。そのためには，量的変数と群別変数という形にしなくてはいけない。例えば，平均 10，標準偏差 2 の正規乱数 100 個からなる変数 `V` と，平均 12，標準偏差 3 の正規乱数 60 個からなる変数 `W` を比較して図示するためのコードは次の枠内の通り。なお，最後の行は，これら 2 つの変数の平均に有意差があるかどうかを検定するためのコードである。

it06-1-2006.R

```
RNGkind("Mersenne-Twister")
set.seed(1)
V <- rnorm(100,10,2)
W <- rnorm(60,12,3)
X <- c(V,W)
C <- as.factor(c(rep("V",100),rep("W",60)))
stripchart(X~C,method="jitter",vert=T,ylim=c(0,20))
MX <- tapply(X,C,mean)
SX <- tapply(X,C,sd)
IX <- c(1.1,2.1)
points(IX,MX,pch=18)
arrows(IX,MX-SX,IX,MX+SX,angle=90,code=3)
t.test(V,W,var.equal=(var.test(V,W)$p.value>=0.05))
```

対応のある 2 標本の平均値の差の検定

先の例題と同じ専門家調査の結果で，2005 年の予測値と 2025 年の予測値に差があるかないかという問題を考えよう。この場合は同じ人について両方の値があるので，全体の平均に差があるかないかだけをみるのではなく，個人ごとの違いを見るほうが情報量が失われない。このような場合は，独立 2 標本の平均値の差の検定をするよりも，対応のある 2 標本として分析する方が切れ味がよい（差の検出力が高い）[*7]。対応のある 2 標本の差の検定は，paired-t 検定と呼ばれ，意味合いとしてはペア間の値の差を計算して値の差の母平均が 0 であるかどうかを調べることになる。R で対応のある変数 X と Y の paired-t 検定をするには，`t.test(X,Y,paired=T)` で実行できるし，それは `t.test(X-Y,mu=0)` と等価である。

2025 年の予測値は，`1.38 1.50 1.30 ...`**（後略）**であり（回答数は 311，平均値は 1.385，不偏分散は 0.0252），2005 年の予測値は，`1.30 1.35 1.34 ...`**（後略）**であった（回答数は 311，平均値は 1.334，不偏分散は 0.00259）。これを普通に t 検定するなら，明らかに分散が異なるので，Welch の検定によって $t_0 = 5.37$，自由度が 373.1 より両側検定の有意確率は 1.37×10^{-7} となるが，対応のある t 検定をすると，2025 年と 2005 年の予測値の差が，`-0.08 -0.15 0.04 ...`**（後略）**となりサンプル数 311，平均 -0.0508，不偏分散 0.0192 より，$t_0 = 6.46$ となり自由度 310 の t 分布で上側確率を求めて 2 倍すれば，$p = 3.942 \times 10^{-10}$ となり，こちらの方が有意確率は小さくなる。いずれにせよ 5%よりずっと小さいので，2025 年の予測値と 2005 年の予測値は 5%水準で有意に異なるといえる。

[*7] 分布が歪んでいる場合や，分布が仮定できない場合の対応のある 2 標本の分布の位置の差があるかどうか検定するには，ウィルコクソンの符号順位検定を用いる。R では `wilcox.test(`**変数 1**`,`**変数 2**`,paired=T)` で実行できる。詳細は後日。

例題

10 人の健康な日本人成人男性ボランティアを募り，同じ日の 9:00 と 21:00 に採血をして血清鉄濃度 (mg/L) を測定した結果が下表のように得られたとする（注：架空のデータである)。9:00 と 21:00 の血清鉄濃度に有意差があるといえるか？　有意水準 5%で検定せよ。

時刻＼対象者	1	2	3	4	5	6	7	8	9	10
9:00	0.98	0.87	1.12	1.34	0.88	0.91	1.04	1.21	1.17	1.09
21:00	1.03	0.78	1.04	1.52	0.97	0.84	1.32	1.12	1.09	1.32

回答を得るには，下枠内のように R で入力すればよい（対応がある場合の図示は，このように一組づつ線で結ぶことが多い)。

it06-2-2006.R

```
BX <- c(0.98,0.87,1.12,1.34,0.88,0.91,1.04,1.21,1.17,1.09)
AX <- c(1.03,0.78,1.04,1.52,0.97,0.84,1.32,1.12,1.09,1.32)
t.test(BX,AX,paired=T)
plot(c(1,2),c(BX[1],AX[1]),type="l",ylim=c(0,2),xaxt="n",xlab="",
 ylab="血清鉄濃度 (mg/L)",col=1)
axis(1,1:2,c("9:00","21:00"))
for (j in 2:length(BX)) { lines(c(1,2),c(BX[j],AX[j]),col=j) }
```

課題

20 匹の 8 週齢の ICR マウスをランダムに 10 匹ずつ 2 群にわけて，片方には普通の餌を自由に食べさせ，もう片方には高脂肪の餌を自由に食べさせ，飲水，運動などもとくに制限せずに 1 週間飼育したとする。この 1 週間の前後でのマウスの体重 (g) が，以下の表のように得られたとき，高脂肪餌の摂取は普通餌摂取に比べてマウスの体重を有意に増加させる効果があると言えるかどうか検定せよ。群別の体重変化を図示した上で，帰無仮説を明示し，手順も書くこと。

なお，下記のデータは，タブ区切りテキスト形式でダウンロードできる[*8]。

普通餌		高脂肪餌	
開始時	終了時	開始時	終了時
30.3	31.6	29.5	31.2
28.7	29.4	31.1	34.1
30.2	31.1	30.1	31.7
30.5	31.4	31.3	32.8
30.7	31.4	31.8	34.2
30.4	31.2	30.5	32.3
29.4	30.9	29.9	31.7
29.4	31.0	28.4	30.8
30.0	31.7	29.3	30.3
29.0	29.6	30.4	32.6

結果は図を Powerpoint に貼り付け，検定結果とその解釈を，学籍番号，氏名とともにテキスト枠に記入した上で印刷し，氏名を自筆して提出すること。結果の提出をもって出席確認とする。

[*8] URL は `http://phi.med.gunma-u.ac.jp/medstat/it06-k-2006.txt` であり，変数名は，普通餌開始時が NDS，普通餌終了時が NDE，高脂肪餌開始時が HFDS，高脂肪餌終了時が HFDE となっている。