

# 医学情報処理演習第6回「2群の平均値の差：検定と推定」

中澤 港 (nminato@med.gunma-u.ac.jp)

2004年11月17日

## 質問・意見への回答

- プリントの字は小さくなりすぎないように気をつけます。
- Rの日本語マニュアルは、有志が訳したバージョン1.7.0のものが下のURLにあります。  
<http://phi.med.gunma-u.ac.jp/swtips/R-jp-docs/>
- 資料に目を通して全部わかれば、課題の回答を書いて提出して先に退出してかまいません。
- 3Dグラフも作れます。library(scatterplot3d)としてから、example(scatterplot3d)と打ってみてください。鳥瞰図もpersp()関数で作れます。

## 1 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。

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  $\mu_X$  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。 $V_X$  が未知のときは、標本の不偏分散  $S_X = \sum_{i=1}^n (X_i - E(X))^2/(n-1) = \text{var}(X)$  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）。

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。使用法を下枠内に示す。

```
X <- rnorm(100,120,10)
t.test(X,mu=120)
```

### 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歳男性は全国平均に比べてBMIに差があると言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 独立 2 標本の平均値の差の検定

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。

### 2.1 母分散が既知で等しい $V$ である場合

$z_0 = |E(X) - E(Y)| / \sqrt{V/n_X + V/n_Y}$  が標準正規分布に従うことを使って検定する\*1。

### 2.2 母分散が未知の場合

調査データを分析する場合は母分散が既知であることはほとんどなく, こちらが普通である。手順は以下の通り。

1.  $F$  検定 (分散が等しいかどうか): 2 つの量的変数  $X$  と  $Y$  の不偏分散  $SX \leftarrow \text{var}(X)$  と  $SY \leftarrow \text{var}(Y)$  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  $F0 \leftarrow SX/SY$  が第 1 自由度  $DFX \leftarrow \text{length}(X) - 1$ , 第 2 自由度  $DFY \leftarrow \text{length}(Y) - 1$  の  $F$  分布に従うことを使って検定する。有意確率は  $1 - \text{pf}(F0, DFX, DFY)$  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\*2。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を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を使って検定する。

### 2.3 分散に差がない場合

母分散  $S$  を  $S \leftarrow (DFX * SX + DFY * SY) / (DFX + DFY)$  として推定し,  $t_0 \leftarrow \text{abs}(\text{mean}(X) - \text{mean}(Y)) / \sqrt{S / \text{length}(X) + S / \text{length}(Y)}$  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,  $(1 - \text{pt}(t_0, DFX + DFY)) * 2$  が有意確率となる。

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  $X \geq Y$  であることを意味する)。

### 2.4 分散が差がある場合 (Welch の方法)

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

\*1 分布がひどく歪んでいる場合には, Mann-Whitney の  $U$  検定 (Wilcoxon の順位和検定と数学的に同値) を行う。詳細は後日説明するが, 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。

\*2 「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が気にしなくてもよい。」と書いていたが, 少なくとも現在のバージョン 1.9.1 とか 2.0.0 とか 2.0.1 ではそうではなくて, 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:  $SX \neq SY$ 」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。

#### 例題

前回触れた出生率の専門家調査には、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りそうである。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(サンプル数 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(サンプル数 221, 平均 1.356, 不偏分散 0.0211)。2群間に違いがあると言っていいか?

Rで計算すると,

```
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 にしなくていい。続けて R で計算すると,

```
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 ( $8.97506 \times 10^{-9}$  という意味), ほぼゼロに近いので, 平均には有意差があるといえる。

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のが常道である\*3。R では, 下枠内のように入力すればよい。

```
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)
```

### 3 対応のある2標本の平均値の差の検定

例えば, 先に触れた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見るのではなく, 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)\*4。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),

\*3 生データの図示では `stripchart()` 関数を用いる。そのためには, 量的変数と群別変数という形にしておく必要がある。2つの量的変数 `V <- rnorm(100,10,2)` と `W <- rnorm(60,12,3)` があつたら, 予め

```
X <- c(V,W)
C <- as.factor(c(rep("V",length(V)),rep("W",length(W))))
```

のように変換しておく必要がある。プロットには次のように入力すればよい。

```
stripchart(X~C,method="jitter",vert=T)
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)
```

\*4 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。

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  $t$  検定するならば、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  $1.37 \times 10^{-7}$  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。

#### 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, 例えば下枠内のように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)。確かめよ。

```
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))
for (j in 2:length(BX)) { lines(c(1,2),c(BX[j],AX[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。

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

結果は配布する紙に学籍番号, 氏名と共に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。