

# 医学情報処理演習第 11 回「ノンパラメトリックな分析」

中澤 港 (nminato@med.gunma-u.ac.jp)

2005 年 1 月 17 日

(注) 2005 年 1 月 17 日 6:00 頃から 16:30 頃まで公開されていたもの、及び講義時に配布したものは致命的なミス 2 つを含む誤植があるので、差し替えられたい。

前回の課題の回答例

ある国で炉心溶融事故を起こした原子力発電所を含む郡の人口 40000 人のうち、事故後 5 年間で 400 人が白血病を発症したとする。事故の 10 年前には人口 35000 人で、事故の 5 年前までの 5 年間では、100 人が白血病を発症したとする。この炉心溶融事故は白血病リスクをどれくらい高めたか？ リスク比とその 95%信頼区間を計算して答えよ（注：この数値例はまったく架空の話である）。

この研究デザインは、非曝露群として対照にとるのが過去の同一集団なので、疾病統計が取られているという意味でこの郡の人口全体をコホートと考えれば、ケースコホート研究（事故の 10 年前の人口 35000 人の中には事故後に発症する人も含まれるため）とみなすことができる。

曝露群の 5 年間の白血病発症リスクは、 $400/40000$  であり、非曝露群の 5 年間の白血病発症リスクは  $100/35000$  なので、リスク比の点推定量は、 $(400/40000)/(100/35000)$  より、3.5 となる。この数値は、炉心溶融事故の曝露が、白血病罹患リスクを 3.5 倍に高めたことを意味する。95%信頼区間は近似式通りに計算すればよいので、

$$3.5 * \exp(\text{qnorm}(0.025) * \sqrt{1/400 - 1/40000 + 1/100 - 1/35000})$$

$$3.5 * \exp(\text{qnorm}(0.975) * \sqrt{1/400 - 1/40000 + 1/100 - 1/35000})$$

より、下限が 2.81、上限が 4.35（または 4.36）である。

## 2 群の分布の位置の差に関するノンパラメトリックな検定

### ノンパラメトリックな検定とは？

パラメータ (parameter) とは母数という意味である。これまで説明してきた検定法の多くは、母数、つまり母集団の分布に関する何らかの仮定をおいていた。その意味で、 $t$  検定も  $F$  検定もパラメトリックな分析法といえる。パラメトリックな分析法は、原則としては、その分析法が前提とする、母数に関する仮定を満たすようなデータに対して使うべきである。一方、フィッシャーの正確な確率は母数を仮定しないのでパラメトリックでない。ノンパラメトリックな分析とは、パラメトリックでない分析、つまり母数を仮定しない分析をさし、分布がひどく歪んでいたり、上限や下限があったりするようなデータに対しても使うことができるとい

う利点がある反面、理想的な場合のパラメトリックな分析に比べると検出力は高くない\*1。

2群の分布の位置の差に関して問題を定式化すると、次のようになる。

1. 標本データ  $X_1, X_2, \dots, X_n$  が互いに独立に分布  $F(x)$  に従い、別の標本データ  $Y_1, Y_2, \dots, Y_n$  が互いに独立かつ  $X$  とも独立で分布  $G(y)$  に従う。
2.  $F$  と  $G$  には連続分布であるという以外には制約をおかない。
3. このとき、「2つの分布に差はない」という帰無仮説 ( $H_0: F(x) \equiv G(x)$ ) を検定する。

つまり、2群の分布の位置の差についてのノンパラメトリックな検定では、「母数を仮定しない」とは言っても、連続分布であることだけは仮定する。もっとも理想的には分布の形が同じで位置だけがずれているという、「ズレのモデル」が仮定できると話は簡単である。具体的な方法としては、Wilcoxon の順位和検定、符号付順位和検定、符号検定などがある。得られたデータがある種の経験分布関数に一致するかどうかを調べるために良く使われる検定法としてはコルモゴロフ＝スミルノフ検定 (KS 検定) がある\*2。

## Wilcoxon の順位和検定

Wilcoxon の順位和検定は、パラメトリックな検定でいえば、 $t$  検定を使うような状況、つまり、独立2標本の分布の位置に差がないかどうかを調べるために用いられる。Mann-Whitney の U 検定と (これら2つほど有名ではないが、Kendall の S 検定とも) 数学的に等価である。

データがもつ情報の中で、単調変換に対して頑健なのは順位なので、これを使って検定しようという発想である。以下、Wilcoxon の順位和検定の手順を箇条書きする。

1. 変数  $X$  のデータを  $x_1, x_2, \dots, x_m$  とし、変数  $Y$  のデータを  $y_1, y_2, \dots, y_n$  とする。
2. まず、これらをませこぜにして小さい方から順に番号をつける\*3。例えば、 $x_8[1], y_2[2], y_{17}[3], \dots, x_4[N]$  のようになる (但し  $N = m + n$ )。
3. ここで問題にしたいのは、それぞれの変数の順位の合計がいくつになるかということである。ただし、順位の大計は  $(N + 1)N/2$  に決まっているので、片方の変数だけ考えれば残りは引き算でわかる。そこで、変数  $X$  だけ考えることにする。
4.  $X$  に属する  $x_i$  ( $i = 1, 2, \dots, m$ ) の順位を  $R_i$  と書くと、 $X$  の順位の大計は

$$R_X = \sum_{i=1}^m R_i$$

となる。 $R_X$  があまり大きすぎたり小さすぎたりすると、 $X$  の分布と  $Y$  の分布に差がないという帰無仮説  $H_0$  が疑わしいと判断されるわけである。では、帰無仮説が成り立つ場合に、 $R_X$  はどのくらいの値になるのだろうか\*4

\*1 ただし、厳密に考えると区分はそれほど明確ではない。例えば、カイ二乗検定では母集団の分布には特定の仮定は置いていないので、定義からすると、実はノンパラメトリックな分析になる。ただし、カイ二乗統計量がカイ二乗分布に従うためにはデータ数が十分に多いことが必要である。もっとも、そう言ってしまうば正規近似する場合の順位和検定もデータ数が多いために必要なので、問題は何を検定の本質と見なすかという話になってくる。一般には、量的な変数を分析するのに、量の情報を使わずに大小関係、即ち順位の情報だけを使う分析をノンパラメトリックな解析と呼ぶことが多い。なお、相関については、既にスピアマンとケンドールの順位相関を説明済みである。

\*2 R では `ks.test`(変数 1, 変数 2) で実行可能である。

\*3 同順位がある場合の扱いは後述する。

\*4 以下説明するように、順位和  $R$  をそのまま検定統計量として用いるのが Wilcoxon の順位和検定であり、 $R_X, R_Y$  の代わりに、 $U_X = mn + n(n+1)/2 - R_Y, U_Y = mn + m(m+1)/2 - R_X$  として、 $U_X$  と  $U_Y$  の小さいほうを  $U$  として検定統計量として用いるのが、Mann-Whitney の U 検定である。また、 $U_X - U_Y$  を検定統計量とするのが Kendall の S 検定である。有意確率を求めるために参照する表が異なる (つまり帰無仮説の下で検定統計量が従う分布の平均と分散は、これら3つですべて異なる) が、数学的には等価な検定である。R では、Wilcoxon の順位和統計量の分布関数が提供されているので、例えばここで得られた順位和を  $RS$  と書くことにすると、 $2*(1-pwilcox(RS, m, n))$  で両側検定の正確な有意確率が得られる。

5. もし  $X$  と  $Y$  に差がなければ,  $X$  は  $N$  個のサンプルから偶然によって  $m$  個取り出したものであり,  $Y$  がその残りである, と考えることができる。順位についてみると,  $1, 2, 3, \dots, N$  の順位から  $m$  個の数値を取り出すことになる。同順位がなければ, ありうる組み合わせは,  ${}_N C_m$  通りある\*5。
6.  $X > Y$  の場合には,  ${}_N C_m$  通りのうち, 合計順位が  $R_X$  と等しいかより大きい場合の数を  $k$  とする ( $X < Y$  の場合は, 合計順位が  $R_X$  と等しいかより小さい場合の数を  $k$  とする)。
7.  $k/{}_N C_m$  が有意水準  $\alpha$  より小さいときに  $H_0$  を疑う。  $N$  が小さいときは有意になりにくい,  $N$  が大きすぎると計算が大変面倒である\*6。そこで, 正規近似を行う (つまり, 期待値と分散を求めて, 統計量から期待値を引いて分散の平方根で割った値が標準正規分布に近似的に従うという関係を用いて検定する)。
8. 帰無仮説  $H_0$  のもとでは, 期待値は

$$E(R) = \sum_{i=1}^m E(R_i) = m(1 + 2 + \dots + N)/N = m(N + 1)/2$$

(1 から  $N$  までの値を等確率  $1/N$  でとるから) 分散はちょっと面倒で,

$$\text{var}(R) = E(R^2) - (E(R))^2$$

から,

$$E(R^2) = E\left(\left(\sum_{i=1}^m R_i\right)^2\right) = \sum_{i=1}^m E(R_i^2) + 2 \sum_{i < j} E(R_i R_j)$$

となるので\*7,

$$E(R_i^2) = (1^2 + 2^2 + \dots + N^2)/N = (N + 1)(2N + 1)/6$$

と

$$\begin{aligned} E(R_i R_j) &= \frac{1}{N(N-1)} \left\{ \left(\sum_{k=1}^N k\right)^2 - \sum_{k=1}^N k \right\} \\ &= \frac{1}{N(N-1)} \left( \frac{N^2(N+1)^2}{4} - \frac{N(N+1)(2N+1)}{6} \right) \\ &= \frac{(N+1)(3N+2)}{12} \end{aligned}$$

を代入して整理すると, 結局,  $\text{var}(R_X) = m(N+1)(N-m)/12 = mn(N+1)/12$  となる。

9. 標準化\*8して連続修正\*9し,  $z_0 = \{|R_X - E(R_X)| - 1/2\} / \sqrt{\text{var}(R_X)}$  を求める。  $m$  と  $n$  が共に大きければこの値が標準正規分布に従うので, 例えば  $z_0 > 1.96$  ならば, 両側検定で有意水準 5% で有意である。 R で有意確率を求めるには,  $z_0$  を  $z_0$  と書けば,  $2 * (1 - \text{pnorm}(z_0, 0, 1))$  とすればよい。

\*5 R では `choose(N,m)` によって得られる。

\*6 もっとも, 今ではコンピュータにやらせればよい。例えば R であれば, `wilcox.test(X,Y,exact=T)` とすれば, サンプル数の合計が 50 未満で同順位の値がなければ, 総当たりして正確な確率を計算してくれる。が, つい 15 年くらいまではコンピュータは誰もが使える道具ではなかったし, 総当たりをするには計算時間がかかりすぎた。今のコンピュータでもサンプルサイズが大きいと, 総当たりでは計算時間がかかりすぎて実用的でない。

\*7 第 1 項が対角成分, 第 2 項がそれ以外に相当する。  $m = 2$  の場合を考えてやればわかるが,

$$E\left(\sum_{i=1}^2 R_i^2\right) = E((R_1 + R_2)^2) = E(R_1^2 + R_2^2 + 2R_1 R_2) = \sum_{i=1}^2 E(R_i^2) + 2 \sum_{i < j} E(R_i R_j)$$

となる。

\*8 何度も出てくるが, 平均 (期待値) を引いて分散の平方根で割る操作である。

\*9 これも何度も出てくるが, 連続分布に近づけるために  $1/2$  を引く操作である。

10. ただし、同順位があった場合は、ステップ2の「小さい方から順に番号をつける」ところで困ってしまう。例えば、変数  $X$  が  $\{2, 6, 3, 5\}$ 、変数  $Y$  が  $\{4, 7, 3, 1\}$  であるような場合には、 $X$  にも  $Y$  にも 3 という値が含まれる。こういう場合は、下表のように平均順位を両方に与えることで、とりあえず解決で

属する変数	$Y$	$X$	$X$	$Y$	$Y$	$X$	$X$	$Y$
値	1	2	3	3	4	5	6	7
順位	1	2	3.5	3.5	5	6	7	8

11. ただし、このやり方では、正規近似をする場合に分散が変わる<sup>\*10</sup>。帰無仮説の下で、 $E(R_X) = m(N+1)/2$  はステップ8と同じだが、分散が

$$var(R_X) = mn(N+1)/12 - mn/\{12N(N-1)\} \cdot \sum_{t=1}^T (d_t^3 - d_t)$$

となる。ここで  $T$  は同順位が存在する値の総数であり、 $d_t$  は  $t$  番目の同順位のところにいくつのデータが重なっているかを示す。上の例では、 $T = 1$ 、 $d_1 = 2$  となる。なお、あまりに同順位のものが多い場合は、この程度の補正では追いつかないので、値の大小があるクロス集計表として分析することも考慮すべきである（例えば Cochran-Armitage 検定などが考えられる）。

#### 例題 1

以下は、外来と入院患者に対して、それぞれ3ヶ月治療したときの胃潰瘍の面積変化割合の表である<sup>a</sup>。

	人数	面積変化割合（肩数字はその割合を示した人数を意味する）
X: 入院患者	32	-100 <sup>(12)</sup> , -93, -92, -91 <sup>(2)</sup> , -90, -85, -83, -81, -80, -78, -46, -40, -34, 0, 29, 62, 75, 106, 147, 1321
Y: 外来患者	32	-100 <sup>(5)</sup> , -93, -89, -80, -78, -75, -74, -72, -71, -66, -59, -41, -30, -29, -26, -20, -15, 20, 25, 37, 55, 68, 73, 75, 145, 146, 220, 1044

入院患者と外来患者の間で、面積変化割合に差があるといえるだろうか？ Wilcoxon の順位和検定を実行してみよう。

<sup>a</sup> 出典：Doll R, Pygott F: Factors influencing the rate of healing of gastric ulcers: admission to hospital, phenobarbitone, and ascorbic acid. *Lancet*, i: 171-175, 1952. から Armitage P, Berry G, Matthews JNS: *Statistical Methods in Medical Research* 4th ed. Blackwell Publishing, 2002, pp.281. に引用されているもの

まず、2群を併せて順位を計算する。下枠内のように直接データ入力してもいいが、Excel で入院患者と外来患者を1行ずつ使って、横に32カラム使って面積変化割合を入力し、まず入院患者の全データを選択してコピーしてから、Rに移って `X <- scan("clipboard")` として、次に再びExcelに戻って外来患者の全データを選択してコピーしてからRに移って `Y <- scan("clipboard")` としてデータ入力の部分を済ませる方が便利だろう。

<sup>\*10</sup> 正確な確率を求めることができれば問題ないけれども、同順位がある場合には、Rでは正確な確率は求められない。

```

X <- c(rep(-100,12), -93, -92, rep(-91,2), -90, -85, -83, -81, -80, -78,
      -46, -40, -34, 0, 29, 62, 75, 106, 147, 1321)
Y <- c(rep(-100,5), -93, -89, -80, -78, -75, -74, -72, -71, -66, -59, -41,
      -30, -29, -26, -20, -15, 20, 25, 37, 55, 68, 73, 75, 145, 146, 220, 1044)
d <- data.frame(
  gr=factor(c(rep('X',length(X)),rep('Y',length(Y)))),
  val=c(X,Y))
rnk <- rank(d$val)
dd <- data.frame(d,rk=rnk)
RX <- dd$rk[dd$gr=='X']
RY <- dd$rk[dd$gr=='Y']

```

RX に X の各値の順位が得られ、RY に Y の各値の順位が得られる。-100 はすべて 9 位となる (1 位から 17 位までが -100 なので平均順位の 9 位が与えられる) ので、RX や RY を表示させても最初の方の値は 9.0 ばかりである。

RX と RY それぞれの合計は、 $\text{sum}(RX)$  より  $R_X = 858$ 、 $\text{sum}(RY)$  より  $R_Y = 1222$  となる。RY の合計の方が大きいのでそちらを検定統計量として採用し、同順位がある場合の正規近似から、 $R_Y$  の期待値  $E(R_Y)$  は、 $E(R_Y) = 32 \times (32 + 32 + 1) / 2 = 1040$  となり、分散  $V(R_Y)$  は、 $V(R_Y) = 32 \times 32 \times (32 + 32 + 1) / 12 - 32 \times 32 / \{12 \times (32 + 32 + 1) \times (32 + 32)\} \times \{17^3 - 17 + (2^3 - 2) \times 4\} = 5445.744$  となるので、

$$\frac{R_Y - E(R_Y) - 1/2}{\sqrt{V(R_Y)}} = 2.459507$$

より<sup>\*11</sup>、 $2 \times (1 - \text{pnorm}(2.459507, 0, 1)) = 0.01391280$  となるので<sup>\*12</sup>、有意水準 5% で帰無仮説は棄却され、2 群には有意な差があるといえる。この計算部分の R のプログラムは下枠内の通り。

```

print(sum(RX))
print(sum(RY))
# 見比べると sum(RY) が大きいとわかるので E(RY) と V(RY) を計算する。
N <- 32+32+1
ERY <- 32*N/2
VRV <- 32*32*N/12-32*32/(12*N*(N-1))*(17^3-17+(2^3-2)*5)
2*(1-pnorm((sum(RY)-ERY-0.5)/sqrt(VRV),0,1))

```

## 例題 2

R の組み込みデータ `sleep` は、20 人の患者を 10 人ずつ 2 群に分けて、それぞれ異なる催眠剤を与えたときに睡眠時間が何時間長くなったかという値をもつ。変数 `extra` が睡眠時間の増分を示し、`group` が催眠剤の異なる 2 群を示す要因型の変数である。2 群間で睡眠時間の増加量に差はあるか Wilcoxon の順位和検定をせよ。

手順は以下の通り。実際に入力して結果をみてみよう。

<sup>\*11</sup> この場合は 2 群のサンプルサイズが等しいので、期待値や分散は RX と RY のどちらを検定統計量にしても変わらないが、連続修正の効く方向が異なるので合計順位の大きい方を検定統計量とする。

<sup>\*12</sup> `wilcox.test(X,Y,exact=F)` の結果の p 値は 0.01388 と表示され、この計算結果と若干異なるが同順位の扱い方の違いによるものだろう。

```
data(sleep)
wilcox.test(sleep$extra~sleep$group,exact=F)
```

exact=F オプションをつけないと、同順位があるので正確な  $p$  値が計算できないという警告メッセージが表示される。サンプルサイズが 50 未満なのでデフォルトで正確な確率を計算しようとするのだけれども同順位の値があるので正確な確率が計算できず、正規近似で計算されているためである。

## メディアン検定

検定統計量を計算するために、順位そのものでなくても、大小関係を反映するスコアならば使うことができる。順位が  $i$  番目のオブザーベーションのスコア  $s(R_i)$  として、順位そのものを使う代わりに、 $i \geq [(N+1)/2]$  のとき 1、 $i < [(N+1)/2]$  のとき 0 を用いるのがメディアン検定である。次のように言い換えることもできる。

$m$  個のデータからなる  $X$  と  $n$  個のデータからなる  $Y$  を合わせた  $N = m + n$  個のデータを、全体のメディアン以上かメディアンより小さいかによって分類すると、以下の  $2 \times 2$  クロス集計表が得られる。

	$X$	$Y$	合計
メディアン以上	$H$	$(m+n/2) - H$	$(m+n)/2$
メディアンより小さい	$m - H$	$H + (n-m)/2$	$(m+n)/2$
合計	$m$	$n$	$m+n$

帰無仮説の下では  $H$  は  $m/2$  の周りに分布する（超幾何分布）ので、 $Pr(H = h') = {}_n C_{h'} \cdot {}_m C_{(m+n)/2-h'}/{}_{m+n} C_{(m+n)/2}$  より、 $Pr(H \geq h')$  をすべて合計して 2 倍すれば、両側検定での有意確率が得られる。下枠内のように関数定義しておけば、例えば、`median.test(rnorm(100),rnorm(100))` のように 2 つのベクトルを与えることによって、簡単にメディアン検定をすることができる。ただし、通常はデータが歪んでいても Wilcoxon の順位和検定ができれば十分なので、メディアン検定はあまり使われない。

```
median.test <- function(X,Y) {
  M <- median(c(X,Y))
  fisher.test(cbind(table(X>=M),table(Y>=M)))
}
```

### 例題 3

例題 2 で取り上げたデータ `sleep` の 2 群間でメディアン検定をせよ。

上で定義した関数では  $X \sim Y$  型の引数は与えられないので、上のように定義した後で `median.test(sleep$extra[sleep$group==1],sleep$extra[sleep$group==2])` とするか、または、下枠内のようにする。

```
X <- subset(sleep$extra,sleep$group==1)
Y <- subset(sleep$extra,sleep$group==2)
M <- median(sleep$extra)
fisher.test(cbind(table(X>=M),table(Y>=M)))
```

## 符号付き順位和検定

2群間の各サンプルに対応がある場合には、単純な順位和検定よりも切れ味がよい方法がある。符号化順位検定とも呼ばれるこの方法は、対応のある  $t$  検定の場合と同じような考え方に基づく。

変数  $X$  の任意の  $i$  番目 ( $i$  は 1 から  $n$  までの整数値) のデータが  $x_i = e_i + \theta_i$  のように、誤差変動  $e_i$  と真の効果  $\theta_i$  の和であると捉えれば、もし  $X$  と  $Y$  が同じ母集団からのサンプルであるならば  $X - Y$  により  $X$  と  $Y$  に共通する真の効果を打ち消すことができ、 $U_i = x_i - y_i = e_i - e_i$  が得られる。このとき帰無仮説は、 $e_i$  と  $e_{i_1}$  の分布が同じということなので、 $U_i$  は原点に対して対称になるはずである。そこで、 $U_i$  の絶対値が小さい方から順に順位  $R_i$  をつける。さらに、 $\varepsilon_i = 1(U_i > 0), \varepsilon_i = -1(U_i < 0)$  とすれば、帰無仮説の下で  $Pr(\varepsilon_i = 1) = Pr(\varepsilon_i = -1) = 1/2$  となる。いま、

$$R^* = \sum_{i=1}^n \varepsilon_i R_i$$

とおけば、 $R^*$  の大きさによって検定ができる。

すべての場合 ( $\varepsilon_i$  の値が各  $i$  について 2 通りあるので、 $2^n$  通り) を計算してやれば正確な確率が計算できるが<sup>\*13</sup>、 $n$  が大きくなると計算が大変なので、 $n \geq 15$  ならば近似を行ってよいことになっている。 $R^*$  の期待値は

$$E(R^*) = \sum_{i=1}^n R_i E(\varepsilon_i) = \sum_{i=1}^n R_i (1 \times 1/2 + (-1) \times 1/2) = 0$$

分散は

$$\begin{aligned} \text{var}(R^*) &= \sum_{i=1}^n R_i^2 \text{var}(\varepsilon_i) \\ &= \sum_{i=1}^n R_i^2 (1^2 \times 1/2 + (-1)^2 \times 1/2) \\ &= \sum_{i=1}^n R_i^2 = n(n+1)(2n+1)/6 \end{aligned}$$

となるので、標準化と連続性の補正をして、

$$\frac{|R^*| - 1/2}{\sqrt{\text{var}(R^*)}}$$

が標準正規分布に従うことを利用して検定する。なお、R では、対応のある 2 群の生のデータを  $X$  と  $Y$  に付値しておき、`wilcox.test(X,Y,paired=T)` とすればこの検定ができる。

### 例題 4

以下は、躁鬱の患者 9 人についてのハミルトンの抑うつ尺度の測定値である。治療前と精神安定剤で治療を開始した後の値を示す<sup>a</sup>。治療によって抑うつが改善されたといえるか、Wilcoxon の符号付き順位和検定で検定せよ。

治療前	1.83	0.50	1.62	2.48	1.68	1.88	1.55	3.06	1.30
治療後	0.878	0.647	0.598	2.05	1.06	1.29	1.06	3.14	1.29

<sup>a</sup> 出典: Hollander M, Wolfe DA (1973) *Nonparametric statistical inference*, New York: John Wiley & Sons, pp.27-33.

<sup>\*13</sup> この正確な確率の計算法は、R. A. Fisher が考案した「並べ換え検定」(permutation test) と呼ばれている。R で実行するには `exactRankTests` ライブラリの `perm.test()` 関数を用いる。

下枠内のように入力すればよい。注意すべきことは、治療後の方が抑うつ尺度は改善していると期待されるので、対立仮説を「XがYより大きい」とする片側検定にすべきだということである。なお、対応のあるt検定と同じく、差を計算して、その期待値がゼロより大きいという対立仮説を検定する形にしても同等である。

```
X <- c(1.83, 0.50, 1.62, 2.48, 1.68, 1.88, 1.55, 3.06, 1.30)
Y <- c(0.878, 0.647, 0.598, 2.05, 1.06, 1.29, 1.06, 3.14, 1.29)
wilcox.test(X, Y, paired=T, alt="greater")
# wilcox.test(X-Y, alt="greater")と同じ
```

## 多群間の分布の位置の差の検定

### クラスカル=ウォリス (Kruskal-Wallis) の検定

ノンパラメトリックな分析でも、多群間の分布の位置の差を調べるためのアプローチとしては、パラメトリックな場合と同じく大別して2つある。1つは、一元配置分散分析と同じく、群分け変数が量的な変数に与える効果のみをみるという形にする方法で、もう1つは、2群ずつすべての組み合わせについて検定の多重性を調整してWilcoxonの順位和検定を繰り返す方法である。

前者を行う代表的な方法がクラスカル=ウォリス (Kruskal-Wallis) の検定である。量的な変数X, 群分け変数Cにデータが付値されているとして、`kruskal.test(X~C)` とすれば実行できる。なお、一元配置分散分析を行うときに前もってパートレットの検定によって分散の同等性をしておく必要があったのと同様、厳密に言えば、Fligner-Killeenの検定 (`fligner.test(X~C)`) によって、前もってばらつきの同等性を検定しておかねばならない。以下、Kruskal-Wallisの検定の仕組みを箇条書きで説明する。

- 「少なくともどれか1組の群間で大小の差がある」という対立仮説に対する「すべての群の間で大小の差がない」という帰無仮説を検定する。
- まず2群の比較の場合の順位和検定と同じく、すべてのデータを込みにして小さい方から順に順位をつける (同順位がある場合は平均順位を与える)。
- 次に、各群ごとに順位を足し合わせて、順位和  $R_i (i = 1, 2, \dots, k; k \text{ は群の数})$  を求める。
- 各群のオブザーベーションの数をそれぞれ  $n_i$  とし、全オブザーベーション数を  $N$  としたとき、各群について統計量  $B_i$  を  $B_i = n_i \{R_i/n_i - (N+1)/2\}^2$  として計算し、

$$B = \sum_{i=1}^k B_i$$

として  $B$  を求め、 $H = 12 \cdot B / \{N(N+1)\}$  として  $H$  を求める。同順位を含むときは、すべての同順位の値について、その個数に個数の2乗から1を引いた値を掛けたものを計算し、その総和を  $A$  として、

$$H' = \frac{H}{1 - \frac{A}{N(N^2-1)}}$$

により  $H$  を補正した値  $H'$  を求める。

- $H$  または  $H'$  から表を使って (データ数が少なければ並べかえ検定によって) 有意確率を求めるのが普通だが、 $k \geq 4$  で各群のオブザーベーション数が最低でも4以上か、または  $k = 3$  で各群のオブザーベーション数が最低でも5以上なら、 $H$  や  $H'$  が自由度  $k-1$  のカイ二乗分布に従うものとして検定できる。



#### 例題 5

R の組み込みデータ `chickwts` (詳細は第 7 回の資料参照) を使って、餌の種類 (`feed`) によって、6 週間飼育後の鶏の体重 (`weight`) に差がどうかを、ノンパラメトリックな方法で分析せよ。

上で説明したとおり、下枠内のように入力すればよい。グラフ表示としては層別箱ヒゲ図を使うのが普通である (2 群の場合も)。

```
data(chickwts)
attach(chickwts)
boxplot(weight~feed)
fligner.test(weight~feed)
kruskal.test(weight~feed)
detach(chickwts)
```

### フリードマン (Friedman) の検定

対応のある多群間の差をノンパラメトリックな方法で調べるには、フリードマン (Friedman) の検定と呼ばれる手法を用いた方が切れ味がよい。簡単に説明すると、まず同じ個体について群間で順位をつける (群といっても、対応がある場合だから、例えば 2005 年の予測値と 2010 年の予測値と 2025 年の予測値というように、個々の個体について順位をつけることが可能である)。次に、群ごとに順位の合計 (順位和) を計算する。順位和の二乗和から順位和の平均の二乗を引いた値を統計量  $S$  として、サンプル数が少ない場合は表によって (コンピュータシミュレーションによってもよい) 有意確率を計算し、サンプル数が多い場合は自由度が群数より 1 少ないカイ二乗分布に従う統計量  $Q$  を  $S$  の 12 倍を個体数と群数と「群数 + 1」の積で割った値として計算して有意確率を計算する。ただし同順位がある場合は調整が必要であり、煩雑なので、通常はコンピュータソフトウェアに計算させる。R では `friedman.test()` 関数を用いる。使い方は `example(friedman.test)` を参照されたい。

### 多重比較

多群があるときに 2 群ずつのすべての組み合わせについて分布の位置の差を検定するには、パラメトリックな分析でやったのと同様、多重比較を行って第一種の過誤を調整する必要がある。第一種の過誤の調整法としては、Holm の方法と Bonferroni の方法は問題なく使えるが、Tukey の HSD のように母集団の分布を仮定した方法は使えない。R では量的変数を  $X$ 、群分け変数を  $C$  として、`pairwise.wilcox.test(X,C)` 関数を用いればよい ( $X \sim C$  ではなく  $X, C$  であることに注意)。なお、サンプルサイズが小さいときは正確な確率を求めるのがデフォルト動作だが、同順位の値があると正確な確率を求めることができず、警告メッセージが表示される。警告を見たくないときは `exact=F` オプションをつければよい。

#### 例題 6

例題 5 の `chickwts` データを使って、どの餌とどの餌の間で体重が異なるかを、検定の多重性を調整して検定せよ。

上で説明したとおり、下枠内のように入力すればよい (同順位の値が含まれているので `exact=F` オプションをつけないと警告メッセージがでる)。

```
data(chickwts)
attach(chickwts)
pairwise.wilcox.test(weight,feed,exact=F)
detach(chickwts)
```

## 課題

(データは第6回の例題の再掲) 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? ノンパラメトリックな方法で有意水準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

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。