第 9 章

2 群の差に関するノンパラメトリックな検定

9.1　ノンパラメトリックな検定とは？

第 4 章で説明した通り，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。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す。

問題を定式化すると，次のようになる。

1. 標本データ $X_1, X_2, \ldots, X_n$ が互いに独立に分布 $F(x)$ に従い，別の標本デー

検定の区別

ただし，厳密に考えると区分はそれほど明確ではない。たとえば，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，そう言ってしまえば正規近似する場合の順位和検定もデータ数が多いことが必要なので，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，すなわち順位の情報だけを使う分析をノンパラメトリックな分析と呼ぶことが多い。

タ $Y_1, Y_2, \ldots, Y_n$ 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（$H : F(x) \equiv G(x)$）を検定する。

つまり，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。

2群の差に関するノンパラメトリックな検定の具体的な方法としては，前章でも軽く触れたように，Wilcoxon の順位和検定（Mann-Whitney の U 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[*1]。

パラメトリックな分析法が使える前提としては，理想的には母集団の分布は正規分布に従っていなくてはならない。しかし，実際には正規分布に従っていない場合もある。この場合の戦略としては，(1) 対数正規分布とかガンマ分布のような別な分布を考える，(2) 正規分布に近づくような変換を施す，といったことが考えられるが，真の分布がわかっていないためにうまく行くとは限らない。そこで，ヒストグラムを描いてみて，どうも正規分布ではなさそうだと思ったら，分布によらない方法を試してみるというのも一案である。2群の分布の位置の差に関する検定の場合，Wilcoxon の順位和検定の検出力は，最良の場合の t 検定の 95% 程度だが，分布が歪んでいる場合には t 検定よりも検出力が良くなる場合もある。

本章のテーマについて詳しく知りたい場合は，竹内，大橋 (1981) の第4章や，伊藤ら (1984) の第2章を参照されたい。

9.2 Wilcoxon の順位和検定

Wilcoxon の順位和検定は，Mann-Whitney の U 検定と（見かけはちょっと違うが）同じ内容の検定である（詳しくは囲みを見よ）。

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，手順を箇条書きする。

1. 変数 X のデータを $x_1, x_2, \ldots, x_m$ とし，変数 Y のデータを $y_1, y_2, \ldots, y_n$ とする。

[*1] 説明は省略するが，R では `ks.test(変数1, 変数2)` で実行可能である。

2. まず，これらをまぜこぜにして小さい方から順に番号をつける[*2]。たとえば，$x_8[1], y_2[2], y_{17}[3], \ldots, x_4[N]$ のようになる（ただし $N = m + n$）。
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，順位の総合計は $(N+1)N/2$ 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 $(i = 1, 2, \ldots, m)$ の順位を R_i と書くと，X の順位の合計は
$$R_X = \sum_{i=1}^{m} R_i$$
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？
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，$1, 2, 3, \ldots, N$ の順位から m 個の数値を取り出すことになる。ありうる組み合わせは，${}_NC_m$ 通りある[*3]。
6. $X > Y$ の場合には，${}_NC_m$ 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. $k/{}_NC_m$ が有意水準 α 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が，N が大きすぎると計算が大変面倒である[*4]。そこで，正規

Wilcoxon の順位和検定

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。有意確率を求めるために参照する表は違うが，数学的には同じ意味をもつ。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，たと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。

[*2] 同順位がある場合の扱いは後述する。

[*3] R では `choose(N,m)`。

[*4] もっとも，今ではコンピュータにやらせればよい。たとえば R であれば，`wilcox.test(X,Y,`

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

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

（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，

$$var(R) = E(R^2) - (E(R)^2)$$

から，

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

となるので[*5]，

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

と

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

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

---

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，総当りして正確な確率を計算してくれる。が，つい 15 年くらい前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標本サイズが大きいと，総当りでは計算時間がかかりすぎて実用的でない。

[*5] 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，

$$E((\sum_{i=1}^{2} R_i)^2) = E((R_1 + R_2)^2) = E(R_1^2 + R_2^2 + 2R_1R_2) = \sum_{i=1}^{2} E(R_i^2) + 2\sum_{i<j} E(R_i R_j)$$

となる。

9. 標準化[*6]して連続性の補正[*7]し，$z_0 = \{|R_X - E(R_X)| - 1/2\}/\sqrt{var(R_X)}$ 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，たとえば $z_0 > 1.96$ 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 `z0` と書けば，`2*(1-pnorm(z0,0,1)` とすればよい。
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。たとえば，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，とりあえず解決できる。

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[*8]。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べきである（詳細はより専門的なテキストを参照されたいが，たとえば Cochran-Armitage 検定などが考えられる)。

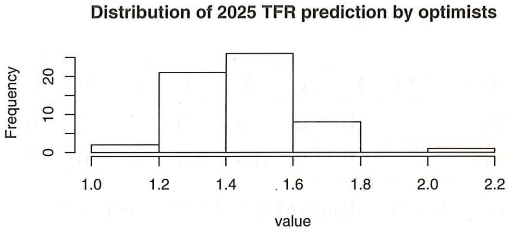
[*6] 何度も出てくるが，平均（期待値）を引いて分散の平方根で割る操作である。

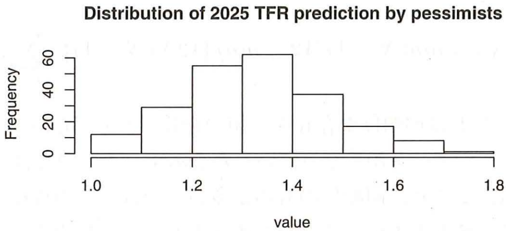
[*7] これも何度も出てくるが，連続分布に近づけるために 1/2 を引く操作である。

[*8] 正確な確率を求めるならば問題ない。

例題 1

前章で紹介した，少子化の見通しに関する専門家調査の結果を再び取り上げることにする。低出生率は今後回復すると見る 58 人と，このまま出生率は低下し続けると見る 221 人の間で，2025 年の合計出生率の予測値には t 検定で差があったわけだが，予測値の分布を調べると，下図のようにかなり右裾を引いた歪んだ形になっていることがわかる。ノンパラメトリックな検定である Wilcoxon の順位和検定で検定をやり直してみよう。





まず，それぞれの順位を計算する。R で 2 群を合わせた順位を計算するには，データフレームに戻すと便利である。今後回復すると見る人の予測値を示す変数を opt，低下し続けると見る人の予測値を示す変数を pes とすると，

```
d<-data.frame(gr=c(rep(1,NROW(opt)),rep(2,NROW(pes)),val=c(opt,pes))
```

とすれば，d というデータフレームができる。次に，順位を得る関数 `rank()` を使って，`rnk <- rank(d$val)` とし，これを元のデータフレームに結合して，`dd<-data.frame(d,rk=rnk)` とすれば，opt の順位の合計が `sum(dd$rk[dd$gr==1])` によって得られ，pes の順位の合計が `sum(dd$rk[dd$gr==2])` で得られる。opt と pes の順位はそれぞれ，

```
(opt) 165.5 165.5 245.5 218.0 165.5 277.0 131.0 ...（後略）
(pes) 141.0  83.0  17.5 103.0 131.0 218.0 241.0 ...（後略）
```

となり，それぞれの合計は，11157，27903 となる[*9]。つまり，$R_{\mathrm{opt}} = 11157$ である。ここでは，opt の方が pes より確率的に大きいと考えられるので，opt$\leq$pes を帰無仮説として片側検定をする。本当は同順位がいくつもあるので分散の計算が面倒なのだが，とりあえず同順位の存在を無視すると，R_{opt} の期待値 $E(R_{\mathrm{opt}})$ は，$E(R_{\mathrm{opt}})$=$58 * (58 + 221 + 1)/2 = 8120$ となり，分散 $V(R_{\mathrm{opt}})$ は，$V(R_{\mathrm{opt}})$=$58 * 221 * (58 + 221 + 1)/12 = 299086.7$ となるので，

$$\frac{R_{\mathrm{opt}} - E(R_{\mathrm{opt}}) - 1/2}{\sqrt{V(R_{\mathrm{opt}})}} = 5.552323$$

より，`1-pnorm(5.552323,0,1)`$=1.4 \times 10^{-8}$ である。この値はほとんどゼロなので帰無仮説は棄却され，opt は pes より大きいといえる。なお，R で同順位を考慮して計算すると `wilcox.test(opt,pes,alternative="greater",exact=F)` で良く，$p = 1.248 \times 10^{-8}$ となり，大差ないことがわかる。これくらいデータ数が大きくて同順位が少なければ，同順位を考慮せずに計算しても，有意確率のオーダーは変わらない。

練習問題

ある大学の学生実習で水道水の水質検査をしたところ，遊離残留塩素濃度 (mg/L) が，集合住宅群 (A) では {0.3346, 0.6230, 0.8580, 1.1031, 0.4580, 0.6210, 0.9071, 0.4760, 0.5020, 0.9670, 0.7100, 0.1350, 1.1390, 0.5741, 0.9090, 1.0400, 0.4190, 0.6296, 1.1080, 0.5793, 1.0420, 1.2826, 1.8280, 0.1630} で，一戸建て群 (B) では {0.8583, 0.9320, 0.4220, 0.3570, 0.0641, 0.5338, 0.8280, 1.1400, 0.7229, 0.0000} であったとしよう（下表のように，この 2 群間には同順位はない)。この 2 群の間で遊離残留塩素濃度に差があるかどうか，Wilcoxon の順位和検定で調べてみよ。

(A)										
値	0.3346	0.6230	0.8580	1.1031	0.4580	0.6210	0.9071	0.4760	0.5020	0.9670
順位	5	16	21	29	9	15	23	10	11	26
値	0.7100	0.1350	1.1390	0.5741	0.9090	1.0400	0.4190	0.6296	1.1080	0.5793
順位	18	3	31	13	24	27	7	17	30	14
値	1.0420	1.2826	1.8280	0.1630						
順位	28	33	34	4						
(B)										
値	0.8583	0.9320	0.4220	0.3570	0.0641	0.5338	0.8280	1.1400	0.7229	0.0000
順位	22	25	8	6	2	12	20	32	19	1

詳しい計算手順は省くが，実際に試されたい。なお，R の `wilcox.test()` 関数を使って検定すると，正規近似では $p = 0.2986$，正確な確率では $p = 0.3040$ となる。すなわち，(A) と (B) の 2 群間に 5% 水準で有意差はないことがわかる。

[*9] もっとも，`d<-cbind(c(rep(1,NROW(opt)),rep(2,NROW(pes))),c(opt,pes))` かつ `x<-rank(d[,2])` としてから，`x[d[,1]==1]` として opt 群の順位を，`x[d[,1]==2]` として pes 群の順位を得ることもできる。

有意

「有意」という考え方は重要なので，念のため復習しておく。この例題では，どちらが高いとか低いとかいった事前情報はないので，「集合住宅群と一戸建て群の間で水道水の遊離残留塩素濃度に差はない」を帰無仮説として両側検定をする。「有意水準を 5% にする」とは，「帰無仮説が偶然に成り立つ確率が 5% 未満であれば，統計的に意味があるほど稀な現象なので帰無仮説は成り立たないとみなす」ということなので，「5% 水準で有意でない」といえば，「帰無仮説が偶然に成り立つ確率が 5 %未満であれば，統計的に意味があるほど稀な現象なので帰無仮説は成り立たないとみなす，としたのに，データから計算するとその確率が 5% より大きくなってしまったので，統計的に意味があるほど稀ではなく，帰無仮説が成り立たないとはみなせない」ということになる。この例でいえば，有意水準を 5% にしたのに，「集合住宅群と一戸建て群の間で水道水の遊離残留塩素濃度に差がない」条件下で，実際に得られているデータが偶然得られる確率は 5% より大きいので，「差がない」という帰無仮説が棄却されなかったということを意味するわけである。

9.3　正規スコア検定

Wilcoxon の順位和検定では R_i として順位そのものを用いたが，これは大小関係が保存されるならば順位の代わりに適当なスコア $s(R_i)$ を使って構わない。スコアとして正規スコアを用いるのが，正規スコア検定である。

正規スコアとしては，$Z_{(1|N)} \le Z_{(2|N)} \le \ldots \le Z_{(N|N)}$ を標準正規分布からの大きさ N の順序統計量としたとき，

$$s(R_i) = E(Z_{(i|N)}) \simeq \Phi^{-1}\left(\frac{i}{N+1}\right)$$

を用いる。正規スコア検定は，ズレのモデルと分布の正規性の仮定の下では，t 検定と漸近的に同等である。

9.4　メディアン検定

$s(R_i)$ として $i \ge [(N+1)/2]$ 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 2×2 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，$Pr(H = h') =_n C_{h'} \cdot_n C_{(m+n)/2-h'}/_{m+n}C_{(m+n)/2}$ 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。

9.5 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，paired-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 θ_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て，$U_i = x_i - y_i = e_i - e\prime_i$ が得られる。このとき帰無仮説は，e_i と $e\prime_1$ 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（ε_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[*10]，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

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

[*10] この正確な確率の計算法は，R. A. Fisher が考案した「並べかえ検定」(permutation test) と呼ばれている。後述する。

分散は

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

となるので，標準化と連続性の補正をして，

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

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=TRUE)` とすればこの検定ができる。

9.6 符号検定

対応がない場合と同様，対応がある場合でも，R_i という順位そのものを用いる代わりに，スコア $s(R_i)$ を使うことが可能である。ただし，すべての $1 \leq i, j \leq n$ について，$R_i \leq R_j$ ならば $s(R_i) \leq s(R_j)$ となっている必要がある。

もっとも単純なスコアとして，すべての i について $s(R_i) = 1$ とすることを考えると，

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

となるので，これは $U_i = x_i - y_i$ が正となるオブザーベーション数 K から負のオブザーベーション数を引いた値となり，$2K - n$ に等しくなるので，K をそのまま検定統計量としてもよい。

帰無仮説の下での K の分布を考える。X と Y に差がなければ U_i が正となるか負となるかは確率 1/2 で起こるので，n 個のオブザーベーション中で正のオブザーベーション数が x 個になる確率 $p(x)$ は 2 項分布で表され，$p(x) =_n C_x \cdot (1/2)^x \cdot (1/2)^{n-x} =_n C_x \cdot (1/2)^n$ となる。

したがって，K よりも稀な値が偶然得られる確率は，$K > n/2$ の場合には，$2 \times \sum_{x=K}^{n} p(x)$ であり，この値が両側検定の有意水準と考えられる。なお，対立仮説が

$X > Y$ である片側検定の場合は，有意確率は $\displaystyle\sum_{x=K}^{n} p(x)$ となる。

もちろん，n が大きければ，2 項分布は正規近似することができるが，n が小さいときに近似を用いずに確率を簡単に計算できるのが符号検定の利点である。

9.7 並べかえ検定

前述の通り R. A. Fisher が考案したが，コンピュータが発達するまでは計算量が多すぎて，データが少ない場合にしか使えなかった。コンピュータ集約型統計学の代表的な手法の 1 つで，分布に依存しない正確な確率が出せる。

たとえば，符号付き順位和検定で正確な確率を計算するには，すべてのありうる R^* を計算して，その絶対値が R^* の実現値の絶対値以上の値になる場合の数 M を数えれば，$M/2^n$ が有意確率となる。R^* が正の場合だけ考えて 2 倍しても，ほぼ同じ値になる（片側検定の場合は 2 倍しなくてよい)。

R では，CRAN から `install.packages("exactRankTests")` として追加パッケージをインストールし，`library(exactRankTests)` とすれば，`perm.test()` という関数で並べかえ検定が可能になる。