統計学第９回
「２群の差に関するノンパラメトリックな検定」

中澤　港

http://phi.ypu.jp/stat.html

<minato@ypu.jp>

ノンパラメトリックな検定とは？

- パラメータとは母集団の分布を示す値（母数）である。これまで説明した検定の多く（ｔ検定，Ｆ検定等）は，母数に関して何らかの仮定を置いていた。
- フィッシャーの正確な確率など，母数を仮定しない検定をノンパラメトリックな検定という。
- ２群の差に関するノンパラメトリックな検定の場合，母数を仮定しないといっても，母集団の分布が連続であるとは仮定する。理想的には分布の形が同じで位置だけがずれている「ズレのモデル」が成り立つときに，その差を検出するための方法である。
- ２群の差に関するノンパラメトリックな検定としては，Ｗilcoxon の順位和検定（または Mann-Ｗhitney の U 検定。両者は検定に使う統計量が若干違うが本質的に同じもの）と符号付き順位和検定が代表的。前者は2群間に対応がない場合，後者は対応がある場合に用いる。

R

Wilcoxon の順位和検定 (Rank Sum Test)

- 群 X のデータ数 m，群 Y のデータ数 n， m+n=N とする。
- 2群を混ぜて小さい方から順に順位をつけ（同順位の場合は平均順位をつける），片方の群について，順位を合計する。この値を R とすると，{|R-E(R)|-1/2}/√var(R) が標準正規分布に近似的に従うことを使って検定ができる。
  - 正確な確率は，ありうるすべての順位の組み合わせのうち，R よりも E(R) から遠い組み合わせになる割合（両側検定なら両方の群について計算して確率を合計する）として得られる
- 但し，
  E(R)=m(N+1)/2
  var(R)=mn(N+1)/12-mn/{12N(N-1)}Σ(dt^3-dt)
  dt 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数。同順位がなければ
  var(R)=mn(N+1)/12 となるので簡単。
- R では wilcox.test(X,Y) または，wilcox.test(量的変数 ~ 群分け変数) 。正確な確率を計算したいときは wilcox.test(X,Y,exact=T) または wilcox.test(量的変数 ~ 群分け変数 ,exact=T)

練習問題の解答例

- B 群の方が数が少ないので計算が簡単。そこでB群について順位和を計算する。
- $R=22+25+8+6+2+12+20+32+19+1=147$
- $E(R)=10\times(34+1)/2=175$
- $\mathrm{var}(R)=10*24*(34+1)/12=700$
- $z0=(|147-175|-1/2)/\sqrt{700}=2.75/\sqrt{7}=1.04$
- $1.04<1.96$ なので，両側検定で 5% 水準で有意ではない（ちなみに 2*(1-pnorm(1.04))=0.298 ）。
- 正確な確率は 0.304

順位の代わりにスコアを使う場合

- 正規スコア検定：順位の代わりに標準正規分布の分位点関数を使って検定する。順位そのものを使う場合に比べて，もとの分布が正規分布に近い場合の検出力が良くなるが，計算は面倒になるので，あまり使われていない。
- メディアン検定：順位をざくっと単純化して，メディアンより大きいか小さいかという情報だけを使う。2群のどちらにメディアンより大きい値が相対的に多いかを調べることになる。計算が簡単なので時折使われるが，検出力はよくない。

R

対応のある場合

- データに対応がある場合は，パラメトリックな検定の「対応のあるt 検定」と似た考え方で，2群の差の順位を考えると，より良い検出力をもった分析ができる。Wilcoxon の符号付き順位和検定 (Signed Rank Sum Test) と呼ばれる。
- 変数 X と変数 Y をデータ数 n の対応がある変数とする。
  - 合成変数 U=X-Y を計算する。
  - U の絶対値の小さい方から順位 R をつける。
  - U が負なら -1，正なら 1 となる変数 ε を使って，$R^*=\Sigma\varepsilon R$ を計算する。
  - $E(R^*)=0$，$var(R^*)=n(n+1)(2n+1)/6$ となるので，n が概ね 15 以上なら $(|R^*|-1/2)/\sqrt{var(R^*)}$ が標準正規分布に従うことで検定できる。
  - R では wilcox.test(X,Y,paired=T) でよい。

順位以外のスコアを使う「符号検定」

- 対応のない場合と違って，差の順位については正規スコアを割り当てることは行われない。
- メディアン検定に対応するやり方はあって，ＸとＹの大小関係，つまり差が正か負かという符号だけを使う。これは符号だけを使うので符号検定 (Sign Test) と呼ばれる。
- 符号付き順位和検定で差の絶対値に与える順位Ｒをすべて１とすると，R* は X>Y のデータ数から X<Y のデータ数を引いた値になる。総数は決まっているので，X>Y のデータ数そのものを検定統計量にしても同じである。
- 実際の X>Y のデータ数 K が n/2 より大きい場合の有意確率は，$({}_{n}C_{K} + {}_{n}C_{K+1} + \ldots + {}_{n}C_{n})/2^{n}$ となる。

Fisher の「並べかえ検定」

- 正確な確率を求めることができる。すべてのありうる組み合わせについて順位和を計算し，それが実測値と同じかより珍しい場合の数を全組み合わせ数で割ると有意確率が得られる。
- 例で考えると，X={4,11,3}, Y={2,12,22,54} であるとき，ありうる組み合わせは X={2,3,4}, Y={11,12,22,54} から，X={54,22,12},Y={11,4,3,2} までの $_7C_3$=7*6*5/(3*2)=35 通りある。このうち X={4,11,3} の順位 {3,4,2} の和 9 と同じかより珍しい順位和をもつ組み合わせは，{1,2,3}{1,2,4}{1,3,4}{1,2,5},{1,3,5}{1,2,6} を合わせた小さい側の7通りと{7,6,5}{7,6,4}{7,5,4}{7,6,3}{6,5,4}{7,5,3}{7,6,2} を合わせた大きい側の7通りなので，p=(7+7)/35=0.4 となる。
- 対応のある場合も同様の考え方で計算できる。いずれにせよ，コンピュータに計算させるのが普通。R では，exactRankTests ライブラリを使えば，perm.test(X,Y) などで可能。