推測統計 [数理統計学]VI

石綿 元

第１１回講義

目次

1　母数によらない方法

1.1　パラメータによらない統計（ノンパラメトリック法）

前回講義の「適合度検定」を除き、これまでの検定・推定では、母集団の分布をあらかじめ想定しており、その前提のもとで母数（パラメータ）を検定・推定するものであった。これをパラメトリックモデルと呼んだ（「第６回講義　**推測統計概論**」の講義回　「**パラメトリックモデルと推定量**」で解説した）。おおむね正規分布に従う母集団を対象としてきた訳だが、現実世界のデータが、いつもそんなにきれいな訳がない。自然科学におけるデータにおいても測定条件が統制できない場合もある。正規分布からはほど遠い分布の場合や、分布の形状がまったく未知の場合だってある。いつも正規分布に近似できるとは限らないのである。

母集団の分布を仮定できず、標本の大きさも小さいなど、データが従っている分布の特定が困難であるなら、パラメトリックモデルを用いることができない。

また、そもそも統計では得られるデータが常に数値とは限らない。質的データを扱う場合もある。データが順序尺度である場合も、数値で表せてもパラメトリックモデルは使えない。（「第６回講義　**推測統計概論**」の講義回「**統計学の考え方　尺度水準**」　参照）

そのような場合でも使用できる統計的手法についていくつか紹介しておく。母集団の分布を仮定しない場合の統計学の方法を総称して「ノンパラメトリック法」、「母数によらない方法」、「分布によらない統計」などと呼ぶ。

これらの方法は、パラメトリックモデルと異なり、利用する際の前提条件が無いため、いつでも利用できるが、パラメトリックモデルを利用できる場面で用いると、検出力 $1-\beta$ が低下する場合があることが知られている。データの性質に応じて適切な検定方法を選択するようにしなければならない。

1.2　順序統計量

ノンパラメトリック法では、母集団についていっさいの前提を仮定しない代わりに “順序” を尺度として導入する。

標本データ $a_1, a_2, \cdots, a_n$ があったとき、各 a_i について小さい順に並べるとき、小さい順に $a_{(1)}, a_{(2)}, \cdots, a_{(n)}$ と表記するとすると、

$$a_{(1)} \leq a_{(2)} \leq \cdots \leq a_{(n)}$$

このように表記するとき、添字の (j) は、小さい順に並べ替えたときの順序である。この順序に応じた標本そのものが順序統計量であって、単に順番に並べ替えただけである。

順序統計量は、ノンパラメトリック法における最も基本的な統計量である。

1.3　符号検定（中央値の検定）

この検定の目的は、メディアンを検定することにある。すなわち、母集団から取得した n 個の標本のメディアンが ξ であるかとうかを検定する。

n 個の各データ a_i について、ξ との差がプラスであるかマイナスであるかを考え、もし、ξ がメディアンであれば、各データとの差がプラスである確率もマイナスである確率も $\frac{1}{2}$ であるはずで、このことを利用し、比率の検定を行う。

■手順１　仮説をたてる。

$$\begin{cases} H_0 & \text{：帰無仮説}\quad \xi \text{は} \xi_0 \text{とみなせる。} \\ H_1 & \text{：対立仮説}\quad \xi \text{は} \xi_0 \text{とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

■手順２　すべてのデータについて、$a_i - \xi_0$ を計算し、正負の符号をつける。ただし、$a_i - \xi_0 = 0$ となったならば、その値は除外しておく。

■手順３　プラスの符号の出現確率 $\hat{p}$ を計算する。すなわち、プラスの符号の個数 x を総数 n で割ったもので、

$$\hat{p} = \frac{x}{n}$$

ただし、ここで、総数 n は 手順２で除外した個数を除く数とする。

以降の手順は比率の検定である。

真の仮説は次である。

$$\begin{cases} H_0 & \text{：帰無仮説}\quad \xi \text{は} \xi_0 \text{とみなせる。} \\ H_1 & \text{：対立仮説}\quad \xi \text{は} \xi_0 \text{とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

これを、符号を導入したことによって、$p = \frac{1}{2}$ であるかを検定する比率の検定に持ち込むことができたから、次の仮説に取り替えて考えることになる。

$$\begin{cases} H_0 & \text{：帰無仮説}\quad p \text{ は} \frac{1}{2} \text{とみなせる。} \\ H_1 & \text{：対立仮説}\quad p \text{ は} \frac{1}{2} \text{とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

■手順４　統計量を導入する。

$$Z = \frac{\hat{p} - p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} = \frac{\frac{x}{n} - \frac{1}{2}}{\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}}} = \frac{n\left(\frac{x}{n} - \frac{1}{2}\right)}{n\left(\sqrt{\frac{\frac{1}{2}(1-\frac{1}{2})}{n}}\right)} = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

この Z は、$\mathrm{N}(0,1)$ に従う。

■手順５　有意水準 α として導入した統計量について棄却域を求める。

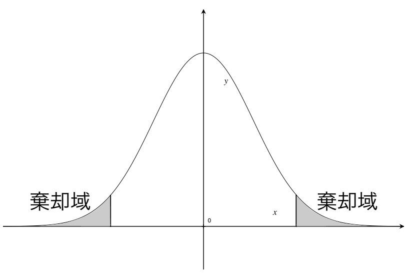

図１　両側検定

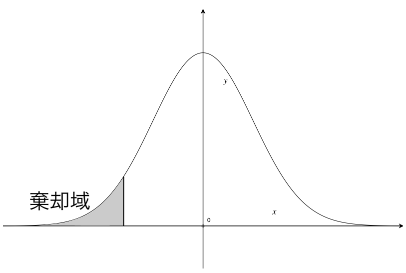

図２　下側片側検定

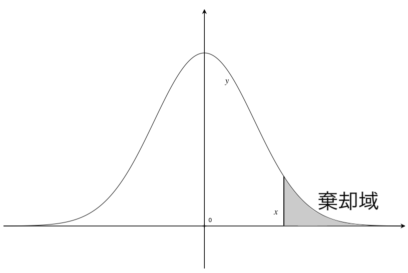

図３　上側片側検定

■手順６　手順４に従って標本から統計量を求め、それが棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

1.4 符号検定（二つの集団の差の検定）

対応のある二つの集団について、集団間に差があるかの検定について符号検定が応用できる。

■手順１　仮説をたてる。

$$\begin{cases} H_0 & \text{：帰無仮説　二つの集団には差がない。} \\ H_1 & \text{：対立仮説　二つの集団には差がないとはみなせない。} \end{cases}$$

真の仮説は、二つの集団の対応するデータの差のメディアンを ξ としたとき、次である。

$$\begin{cases} H_0 & \text{：帰無仮説　} \xi \text{は } 0 \text{ とみなせる。} \\ H_1 & \text{：対立仮説　} \xi \text{は } 0 \text{ とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

■手順２　対となるすべてのデータについて、$a_i - b_i$ を計算し、正負の符号をつける。ただし、$a_i - b_i = 0$ となったならば、その値は除外しておく。

■手順３　プラスの符号の出現確率 $\hat{p}$ を計算する。すなわち、プラスの符号の個数 x を総数 n で割ったもので、

$$\hat{p} = \frac{x}{n}$$

ただし、ここで、総数 n は 手順２で除外した個数を除く数とする。
以降の手順は比率の検定である。

真の仮説は次である。

$$\begin{cases} H_0 & \text{：帰無仮説　} \xi \text{は } 0 \text{ とみなせる。} \\ H_1 & \text{：対立仮説　} \xi \text{は } 0 \text{ とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

これを、符号を導入したことによって、$p = \frac{1}{2}$ であるかを検定する比率の検定に持ち込むことができたから、次の仮説に取り替えて考えることになる。

$$\begin{cases} H_0 & \text{：帰無仮説　} p \text{ は} \frac{1}{2} \text{とみなせる。} \\ H_1 & \text{：対立仮説　} p \text{ は} \frac{1}{2} \text{とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

■手順４　統計量を導入する。

$$Z = \frac{x - \frac{n}{2}}{\sqrt{\frac{n}{4}}}$$

この Z は、$\mathrm{N}(0,1)$ に従う。

■手順５　有意水準 α として導入した統計量について棄却域を求める。

■手順６　手順４に従って標本から統計量を求め、それが棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

1.5 ウィルコクソン符号付き順位検定

符号検定では、メディアンとの差の符号のみに着目し、差そのものの大きさは関係ない。そこで、差の絶対値について大きさについても考慮するように改良したものが順位検定である。目的は、対応する二つの集団の代表値（中央値）に差がないことを検定することである。これは、パラメトリックモデルの「二つの集団の平均の差の検定 (分散既知なら正規分布、分散未知なら t-分布)」で「分散が等しい場合」に相当するものである。

■手順１　仮説をたてる。目的どおり、本来の帰無仮説は、「二つの集団には差がない。」である。

二つの集団の対応するデータの差のメディアンを ξ としたとき、

$$\begin{cases} H_0 & \text{：帰無仮説}\quad \xi \text{は } 0 \text{ とみなせる。} \\ H_1 & \text{：対立仮説}\quad \xi \text{は } 0 \text{ とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

■手順２　対となるすべてのデータについて、$|a_i - b_i|$ を計算し、小さい順に並べる（差の絶対値の順序統計量を求める）。

$$|a_i - b_i| = d_i$$

この d_i を並べ替えて、

$$d_{(1)} \leq d_{(2)} \leq \cdots \leq d_{(n)}$$

ここで、添字 (j) は、d_i の順位である。このとき、

- $|a_i - b_i| = |a_{i'} - b_{i'}| = d_i$ この場合、$|a_i - b_i|$ も $|a_{i'} - b_{i'}|$ も順位 (j) は同じ順位になる。このときは、この 2 つの順位を

$$j' = \frac{(j) + (j+1)}{2}$$

  として両方とも順位は、j' を割り当てる。

  例）3 位のデータ $d_{(3)}$ が二つあったなら、その二つのデータの順位は両方とも 3.5 位を割り当てる。
- $d_i = 0$、つまり $a_i = b_i$ ならば、そのデータの組を除外する。（データ対の総数 n がその分減少する）

■手順３　統計量を計算する。

手順２の順序統計量の順位を符号別に分け、符号ごとに順位を合計し、小さい方を統計量とする。

つまり、$a_i - b_i < 0$ からの $d_{(j)}$ の (j) と $a_i - b_i > 0$ からの $d_{(j)}$ の (j) に分けて、

$$w = \sum_{a_i - b_i < 0 \text{ 条件}} (j), \qquad W = \sum_{a_i - b_i > 0 \text{ 条件}} (j)$$

を計算した上で、統計量は、

$$T = \min(W, w)$$

■手順４　$n \leq 25$ ならば、

有意水準 α として手順３で導入した統計量について「ウィルコクソン符号順位検定：数表」を用いて棄却域を求める。

■手順５　手順３に従って標本から統計量を求め、それが棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

■Notice　手順４について$n > 25$ ならば、手順３で計算した統計量 T は、$\mathrm{N}(\frac{n(n+1)}{4}, \frac{n(n+1)(2n+1)}{24})$ に従う。
よって、標準化変換は、

$$Z = \frac{T - \frac{n(n+1)}{4}}{\sqrt{\frac{n(n+1)(2n+1)}{24}}}$$

この Z は、N(0,1) に従う。したがって、標準正規分布を用いて有意水準 α とした棄却域から判断する。

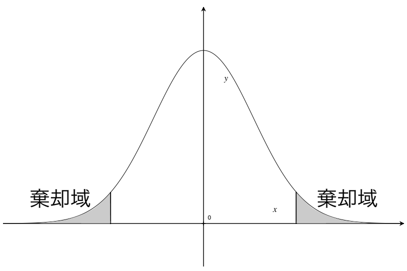

図４　両側検定

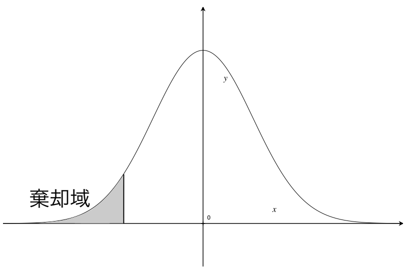

図５　下側片側検定

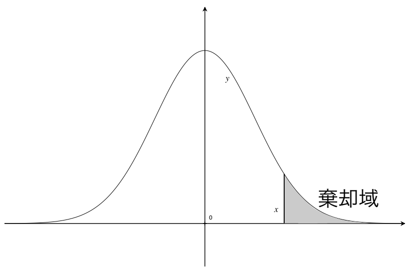

図６　上側片側検定

1.6　順位和検定

目的は、二つの集団の代表値（中央値）に差がないことを検定することである。これは、正規分布もしくは t-分布をもちいたときの「二つの集団の平均の差の検定 (分散既知なら正規分布、分散未知なら t-分布)」で「分散が異なっている場合」に相当するものである。

1.6.1　ウィルコクソン順位和検定

前節の「ウィルコクソン符号付き順位検定」とは、２集団のデータ間に対応があるかないかの違いである。

■手順１　仮説をたてる。

$$\begin{cases} H_0 & \text{：帰無仮説　二つの集団には差がない。} \\ H_1 & \text{：対立仮説　二つの集団には差がないとはみなせない。} \end{cases}$$

■手順２　二つの集団のデータを混ぜ合わせて、小さい順に並べる（混合データの順序統計量を求める)。
データの個数が $m \leq n$ である２つのデータ a_i と b_j

$$\begin{cases} \text{データ 1} & : a_1, a_2, \cdots, a_m \\ \text{データ 2} & : b_1, b_2, \cdots, b_n \end{cases}$$

について、混ぜ合わせる。

$$\text{混合データ} : a_1, a_2, \cdots, a_m, b_1, b_2, \cdots, b_n$$

これを小さい順に並べる。a_i と b_j の混合データを c_k と表すと、c_k の順序統計量 $c_{(l)}$ は、

$$\text{混合データ：} c_{(1)}, c_{(2)}, \cdots, c_{(m)}, \cdots, c_{(n)}, \cdots, c_{(m+n)}$$

■手順３　手順２の混合データ c_k の順序統計量 $c_{(l)}$ での順位 (l) を元のデータ列ごとに分ける。

たとえば､「$c_{(1)}, c_{(2)}, \cdots, c_{(m)}, \cdots, c_{(n)}, \cdots, c_{(m+n)}$」を元のデータで表して､「$b_1, a_5, \cdots, b_8, \cdots, b_2, \cdots, a_1$」であったとすると、順序統計量としては、「$b_1$ は $c_{(1)}$」、「a_5 は $c_{(2)}$」、「a_1 は $c_{(m+n)}$」であるという具合である。

ここで、a_i のときの $c_{(l)}$ の添字 (l) を小さい順に $(1_a), (2_a), \cdots, (m_a)$、$b_j$ のときの $c_{(l)}$ の添字を小さい順に $(1_b), (2_b), \cdots, (n_b)$ と表すとすると、

$$\begin{cases} \text{データ 1} & : (1_a), (2_a), \cdots, (m_a) \\ \text{データ 2} & : (1_b), (2_b), \cdots, (n_b) \end{cases}$$

■手順４　統計量を導入する。データの個数が少ない方のデータに割り当てた順位の和を計算する。いま、データの個数が $m \leq n$ だったから、データの個数は「データ１」の方が少ない。だから、順位和の対象は、手順３で作ったデータ１の順位の列

$$(1_a), (2_a), \cdots, (m_a)$$

これの和を計算する。したがって統計量は、

$$T = \sum_{h=1}^{m} (h_a)$$

■手順５　有意水準 α として「ウィルコクソン順位和検定：数表」より棄却域を次の条件で求める。

$$T \leq w \text{ または } W \leq T$$

ここで、w:下側、W:上側。

■手順６　手順４に従って標本から統計量を求め、それが手順５の棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

■Notice　手順５についてm も n も十分大きいならば、（最低でも $m > 8,\ n > 8$）手順４で計算した統計量 T は、$\mathrm{N}(\frac{m(m+n+1)}{2}, \frac{mn(m+n+1)}{12})$ に従う。よって、標準化変換は、

$$Z = \frac{T - \frac{m(m+n+1)}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}}$$

この Z は、N(0,1) に従う。したがって、標準正規分布を用いて有意水準 α とした棄却域から判断する。

1.6.2　マン-ホイットニー U 検定

ウィルコクソンの順位和検定では、並び替えの操作が面倒なこともある。そこで、マン-ホイットニー U 検定では、二つの集団の標本のうち、データ２の各データについてデータ１の各データと比較してより大きい回数を統計量として導入するが、本質的にウィルコクソン順位和検定と違いはない。

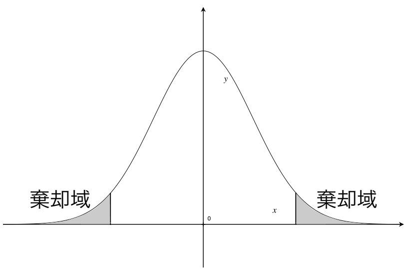

図 7　両側検定

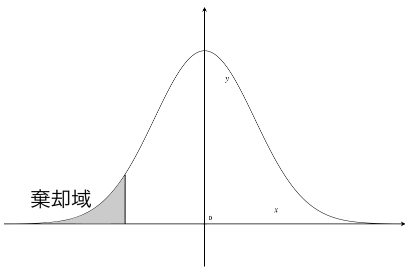

図 8　下側片側検定

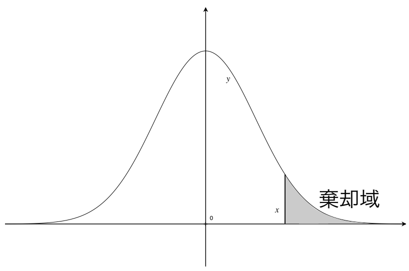

図 9　上側片側検定

■手順 1　仮説をたてる。

$$\begin{cases} H_0 & : \text{帰無仮説　二つの集団には差がない。} \\ H_1 & : \text{対立仮説　二つの集団には差がないとはみなせない。} \end{cases}$$

■手順 2　統計量を導入する。データセットは、ウィルコクソン順位和検定の場合と同じ a_i と b_j とする。

標本数が少ない場合

2 変数関数 $I(a_i, b_j),\ J(a_i, b_j)$ について

$$I(a,b) = \begin{cases} 1 & (a < b) \\ 0.5 & (a = b) \\ 0 & (a > b) \end{cases},\quad J(b,a) = \begin{cases} 1 & (b < a) \\ 0.5 & (b = a) \\ 0 & (b > a) \end{cases}$$

と定義して、さらに U_I と U_J を

$$U_I = \sum_{i=1}^{m}\sum_{j=1}^{n} I(a_i, b_j),\quad U_J = \sum_{j=1}^{n}\sum_{i=1}^{m} J(b_j, a_i)$$

と定義すれば、マン-ホイットニー U 検定統計量 U は、

$$U = \min(U_I, U_J)$$

♣ マン-ホイットニー U 検定統計量 U とウィルコクソン順位和統計量 T

マン-ホイットニー U 検定統計量 U は、ウィルコクソン順位和統計量 T との間に

$$U = m \cdot n + m\frac{m+1}{2} - T$$

の関係が成立するため T と U の統計量は単に並行なだけで本質的な違いはない。

■手順 3　有意水準 α として「マン-ホイットニー順位和検定：数表」より棄却域を求める。
該当する表の値以下であれば棄却域である。

■手順 4　手順 2 で求めた統計量が、手順 3 の棄却域に含まれれば、帰無仮説を棄却する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

■Notice　標本数が多い場合（$m > 20,\ n > 20$）

　ウィルコクソン順位和検定の場合と同じく混合データの順序統計量を求め、データに割り当てた順位の和を計算する。データの順位の列

$$\begin{cases} \text{データ 1} & :(1_a),(2_a),\cdots,(m_a) \\ \text{データ 2} & :(1_b),(2_b),\cdots,(n_b) \end{cases}$$

のとき、

$$T_1 = \sum_{h=1}^{m}(h_a), \quad T_2 = \sum_{k=1}^{n}(k_b)$$

これを用いて、

$$U_1 = m \cdot n + m\frac{m+1}{2} - T_1, \quad U_2 = m \cdot n + n\frac{n+1}{2} - T_2$$

と定義すれば、マン-ホイットニー U 検定統計量 U は、

$$U = \min\left(U_1, U_2\right)$$

この U は、N$(\frac{mn}{2}, \frac{mn(m+n+1)}{12})$ に従う。よって、標準化変換は、

$$Z = \frac{U - \frac{mn}{2}}{\sqrt{\frac{mn(m+n+1)}{12}}}$$

この Z は、N(0,1) に従う。したがって、標準正規分布を用いて有意水準 α とした棄却域から判断する。

1.7　順位相関係数

「第９回講義　**相関分析**」の講義回で定義した相関係数は、二つの変数の間の相関関係を尺度としたものであった。

■相関係数の復習　「第９回講義　**相関分析**」の講義回で定義した相関係数とは、

$$r \equiv r_{XY} \overset{\text{def}}{=} \frac{s_{XY}}{s_X s_Y} = \frac{s_{XY}}{\sqrt{s_X^2 s_Y^2}}$$

ここで、

$$s_X^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$
$$s_Y^2 = \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \bar{Y})^2$$
$$s_{XY} = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})$$

であった。つまり、表現を変えると、

$$r = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}\sqrt{\sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

この r を相関係数と定義して導入したのであった。これを特にピアソンの積率相関係数とも呼ぶ。

現実問題においては、量的データが得られず質的データの順序なら得られることがある。(「第６回講義　**尺度水準**」で解説した質的データの順序尺度）たとえば、「１０個の果物を好きな順に並べてください。」のようにアンケートを取るときなど対象の” 質的データ” について得られるデータが単なる順序だけである。このようなときは順序を用いた相関関係の尺度を導入したい。

1.7.1　スピアマンの順位相関係数

２つの対応するデータの順序統計量について、その順位について普通に相関係数を計算する。

いま、２つの標本について順位を測定し、それぞれの順位が $r_1, r_2, \cdots, r_n$ と $s_1, s_2, \cdots, s_n$ であったとする。

このとき

$$\rho = \frac{\sum_{i=1}^{n}(r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(r_i - \bar{r})^2}\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2}}$$

ただし、

$$\bar{r} = \bar{s} = \frac{n+1}{2}$$
$$\sum_{i=1}^{n}(r_i - \bar{r})^2 = \sum_{i=1}^{n}(s_i - \bar{s})^2 = \frac{1}{12}(n^3 - n)$$
$$\sum_{i=1}^{n}(r_i - \bar{r})(s_i - \bar{s}) = \sum_{i=1}^{n}(r_i - \bar{r})^2 - \frac{1}{2}\sum_{i=1}^{n}(r_i - s_i)^2 = \frac{1}{12}(n^3 - n) - \frac{1}{2}\sum_{i=1}^{n}(r_i - s_i)^2$$

を用いて書き換えると、

$$\rho = \frac{\sum_{i=1}^{n}(r_i - \bar{r})(s_i - \bar{s})}{\sqrt{\sum_{i=1}^{n}(r_i - \bar{r})^2}\sqrt{\sum_{i=1}^{n}(s_i - \bar{s})^2}} = \frac{\frac{1}{12}(n^3 - n) - \frac{1}{2}\sum_{i=1}^{n}(r_i - s_i)^2}{\frac{1}{12}(n^3 - n)}$$
$$= 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n}$$

ここで、$r_i - s_i = d_i$ とおくと、

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

この ρ をスピアマンの順位相関係数（スピアマンの ρ）と呼ぶ。

■スピアマンの順位相関係数の無相関検定　標本 X と標本 Y の間に相関がないことを検定する。

手順１）仮説をたてる。

$$\begin{cases} H_0 & ：帰無仮説\ \ 母相関係数は\ 0\ とみなせる。\\ H_1 & ：対立仮説\ \ 母相関係数は\ 0\ とみなせない。（両側検定もしくは片側検定）\end{cases}$$

手順２）統計量を導入する。

$$\rho = 1 - \frac{6\sum_{i=1}^{n}(r_i - s_i)^2}{n^3 - n} = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}$$

手順３）有意水準 α として導入した統計量について「スピアマンの順位相関係数の無相関検定：数表」を用いて棄却域を求める。標本数 n が十分大きいならば、「20　相関分析」の講義回の「無相関性の検定」をそのまま適用でき、統計量 T には r を上記 ρ と読み替えて代入し、通常通り t-分布を用いる。

手順４）手順２および手順３に従って標本から統計量を求め、それが棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、正または負の相関があると結論する。
- 棄却域に含まれないならば H_0 を採択する。

1.7.2　ケンドールの順位相関係数

まず、n 個の対象について 2 つの標本ごとの順位をつける。つまり、n 個の x における順序統計量 $x_{(1)_x}, x_{(2)_x}, \cdots, x_{(n)_x}$ の順位 $(1)_x, (2)_x, \cdots, (n)_x$ と、n 個の y における順序統計量 $y_{(1)_y}, y_{(2)_y}, \cdots, y_{(n)_y}$ の順位 $(1)_y, (2)_y, \cdots, (n)_y$ ができる。ここでは、順位 (i) に x か y を区別するための添字 $(i)_x, (i)_y$ をつけた。

n 個の対象から 2 個を選ぶ選び方は、${}_nC_2$ である。このときの 2 個とは、(n_i, n_j) と表せるとする。この ${}_nC_2$ 個の組み合わせすべてについて x の順位の組と y の順位の組にする。

$$(n_i, n_j) \Rightarrow \begin{cases} \{(i)_x, (j)_x\} \\ \{(i)_y, (j)_y\} \end{cases}$$

このとき、それぞれの組で x の大小関係と y の大小関係が一致するなら $+1$、しないなら -1 とする。

$$\left.\begin{matrix} (i)_x & > (j)_x \\ (i)_y & < (j)_y \end{matrix}\right\} \Rightarrow -1, \quad \left.\begin{matrix} (i)_x & < (j)_x \\ (i)_y & > (j)_y \end{matrix}\right\} \Rightarrow -1, \quad \left.\begin{matrix} (i)_x & < (j)_x \\ (i)_y & < (j)_y \end{matrix}\right\} \Rightarrow +1, \quad \left.\begin{matrix} (i)_x & > (j)_x \\ (i)_y & > (j)_y \end{matrix}\right\} \Rightarrow +1$$

このようにしてすべての組に -1 か $+1$ の点数を与える。

つぎに、-1 を与えた数と $+1$ を与えた数を数え上げる。ここで、-1 を与えた数を L、$+1$ を与えた数を K とする。全組数は ${}_nC_2$ だったので、具体的に書くと ${}_nC_2 = \frac{n(n-1)}{2}$ であるから、$K + L = \frac{n(n-1)}{2}$。

これらを用いて、

$$\tau = \frac{K - L}{\frac{n(n-1)}{2}} = \frac{2K}{\frac{n(n-1)}{2}} - 1 = 1 - \frac{2L}{\frac{n(n-1)}{2}}$$

この τ をケンドールの順位相関係数（ケンドールの τ）と呼ぶ。

■ケンドールの順位相関係数の無相関検定　標本 X と標本 Y の間に相関がないことを検定する。

手順１）仮説をたてる。

$$\begin{cases} H_0 & \text{：帰無仮説　母相関係数は 0 とみなせる。} \\ H_1 & \text{：対立仮説　母相関係数は 0 とみなせない。（両側検定もしくは片側検定）} \end{cases}$$

手順２）統計量を導入する。

$$\tau = \frac{K - L}{\frac{n(n-1)}{2}} = \frac{2K}{\frac{n(n-1)}{2}} - 1 = 1 - \frac{2L}{\frac{n(n-1)}{2}}$$

手順３）有意水準 α として導入した統計量について「ケンドールの順位相関係数の無相関検定：数表」を用いて棄却域を求める。標本数 n が十分大きいならば、τ は、$\mathrm{N}(0, \frac{(4n+10)}{9n(n-1)})$ に従う。よって、標準化変換は、

$$Z = \frac{\tau}{\sqrt{\frac{(4n+10)}{9n(n-1)}}}$$

この Z は、$\mathrm{N}(0,1)$ に従う。したがって、標準正規分布を用いて有意水準 α とした棄却域から判断する。

手順４）手順２および手順３に従って標本から統計量を求め、それが棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、正または負の相関があると結論する。
- 棄却域に含まれないならば H_0 を採択する。

1.8　ランダム性の検定

たとえば、２種類の要素 A, B が５つずつ合計１０個あったとする。このとき、A と B の要素の出現がランダムであるのかを検定したい。

$$\underline{\mathrm{AA}}\ \underline{\mathrm{B}}\ \underline{\mathrm{A}}\ \underline{\mathrm{BBB}}\ \underline{\mathrm{AA}}\ \underline{\mathrm{B}}$$

得られたデータの同一要素の連続を**ラン**といい、その連続の長さを**ランの長さ**と呼ぶ。

２種類の要素 A, B がそれぞれ n_A, n_B 個、合計 n 個あったとする。

■手順１　仮説をたてる。

$$\begin{cases} H_0 & \text{：帰無仮説　A, B の起こり方はランダムとみなせる。} \\ H_1 & \text{：対立仮説　A, B の起こり方はランダムとみなせない。} \end{cases}$$

■手順２　統計量としてランの個数を数える。（ランの個数を R と表すことにする。）

■手順３　有意水準 α として「ラン検定：数表」を用いて棄却域を求める。

■Notice　n_A も n_B も十分大きいならば、$(n_A > 20,\ n_B > 20)$
手順 2 の統計量 R は、$\mathrm{N}(\frac{2n_An_B}{n_A+n_B}+1, \frac{2n_An_B(2n_An_B-n_A-n_B)}{(n_A+n_B)^2(n_A+n_B-1)})$ に従う。
よって、標準化変換は、

$$Z = \frac{R - \left(\frac{2n_An_B}{n_A+n_B}+1\right)}{\sqrt{\frac{2n_An_B(2n_An_B-n_A-n_B)}{(n_A+n_B)^2(n_A+n_B-1)}}}$$

この Z は、N(0,1) に従う。
したがって、標準正規分布を用いて有意水準 α として導入した統計量について棄却域を求める。

■手順４　手順 2 に従って標本から統計量を求め、それが手順 3 の棄却域に含まれるか判断する。

- 棄却域に含まれるならば H_0 を棄却し、H_1 を採択する。
- 棄却域に含まれないならば H_0 を採択する。

■Notice　「AAAAA BBBBB」も「A B A B A B A B A B」でもランダムではないと考えられるから、ランの個数は多すぎても少なすぎてもランダムではない。

1.9 ノンパラメトリック法-数値表

ウィルコクソン符号付き順位検定：数表

検定統計量が数表の値よりも小さい場合、帰無仮説を棄却する。同じ値でも棄却。有意水準は片側で表記。

n	有意水準（片側）			
	0.05	0.025	0.01	0.005
6	2	0	–	–
7	3	2	0	–
8	5	3	1	0
9	8	5	3	1
10	10	8	5	3
11	13	10	7	5
12	17	13	9	7
13	21	17	12	9
14	25	21	15	12
15	30	25	19	15
16	35	29	23	19
17	41	34	27	23
18	47	40	32	27
19	53	46	37	32
20	60	52	43	37
21	67	58	49	42
22	75	65	55	48
23	83	73	62	54
24	91	81	69	61
25	100	89	76	68

ウィルコクソン順位和検定：数表

　各有意水準で、検定統計量が下側の表の値よりも下及び上側の表の値よりも上なら帰無仮説を棄却。有意水準は両側で表記。

α=0.05		m									
w	下側	1	2	3	4	5	6	7	8	9	10
	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	6	-	-	-	-	-	-	-
	4	-	-	6	11	-	-	-	-	-	-
	5	-	3	7	12	19	-	-	-	-	-
n	6	-	3	8	13	20	28	-	-	-	-
	7	-	3	8	14	21	29	39	-	-	-
	8	-	4	9	15	23	31	41	51	-	-
	9	-	4	10	16	24	33	43	54	66	-
	10	-	4	10	17	26	35	45	56	69	82

α=0.05		m									
W	上側	1	2	3	4	5	6	7	8	9	10
	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	15	-	-	-	-	-	-	-
	4	-	-	18	25	-	-	-	-	-	-
	5	-	13	20	28	36	-	-	-	-	-
n	6	-	15	22	31	40	50	-	-	-	-
	7	-	17	25	34	44	55	66	-	-	-
	8	-	18	27	37	47	59	71	85	-	-
	9	-	20	29	40	51	63	76	90	105	-
	10	-	22	32	43	54	67	81	96	111	128

α=0.025		m									
w	下側	1	2	3	4	5	6	7	8	9	10
	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	10	-	-	-	-	-	-
	5	-	-	6	11	17	-	-	-	-	-
	6	-	-	7	12	18	26	-	-	-	-
	7	-	-	7	13	20	27	36	-	-	-
n	8	-	3	8	14	21	29	38	49	-	-
	9	-	3	8	14	22	31	40	51	62	-
	10	-	3	9	15	23	32	42	53	65	78
	11	-	3	9	16	24	34	44	55	68	81
	12	-	4	10	17	26	35	46	58	71	84
	13	-	4	10	18	27	37	48	60	73	88
	14	-	4	11	19	28	38	50	62	76	91
	15	-	4	11	20	29	40	52	65	79	94

α=0.025		m									
W	上側	1	2	3	4	5	6	7	8	9	10
n	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	26	-	-	-	-	-	-
	5	-	-	21	29	38	-	-	-	-	-
	6	-	-	23	32	42	52	-	-	-	-
	7	-	-	26	35	45	57	69	-	-	-
	8	-	19	28	38	49	61	74	87	-	-
	9	-	21	31	42	53	65	79	93	109	-
	10	-	23	33	45	57	70	84	99	115	132
	11	-	25	36	48	61	74	89	105	121	139
	12	-	26	38	51	64	79	94	110	127	146
	13	-	28	41	54	68	83	99	116	134	152
	14	-	30	43	57	72	88	104	122	140	159
	15	-	32	46	60	76	92	109	127	146	166

α=0.01		m									
w	下側	1	2	3	4	5	6	7	8	9	10
n	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	15	-	-	-	-	-
	6	-	-	-	10	16	23	-	-	-	-
	7	-	-	-	10	16	24	32	-	-	-
	8	-	-	-	11	17	25	34	43	-	-
	9	-	-	6	11	18	26	35	45	56	-
	10	-	-	6	12	19	27	37	47	58	71
	11	-	-	6	12	20	28	38	49	61	73
	12	-	-	7	13	21	30	40	51	63	76
	13	-	-	7	13	22	31	41	53	65	79
	14	-	-	7	14	22	32	43	54	67	81
	15	-	-	8	15	23	33	44	56	69	84

α=0.01		m									
W	上側	1	2	3	4	5	6	7	8	9	10
n	1	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-
	3	-	-	-	-	-	-	-	-	-	-
	4	-	-	-	-	-	-	-	-	-	-
	5	-	-	-	-	40	-	-	-	-	-
	6	-	-	-	34	44	55	-	-	-	-
	7	-	-	-	38	49	60	73	-	-	-
	8	-	-	-	41	53	65	78	93	-	-
	9	-	-	33	45	57	70	84	99	115	-
	10	-	-	36	48	61	75	89	105	122	139
	11	-	-	39	52	65	80	95	111	128	147
	12	-	-	41	55	69	84	100	117	135	154
	13	-	-	44	59	73	89	106	123	142	161
	14	-	-	47	62	78	94	111	130	149	169
	15	-	-	49	65	82	99	117	136	156	176

マン-ホイットニー U 検定：数表

検定統計量が数表の値よりも小さい場合、帰無仮説を棄却する。同じ値でも棄却。有意水準は片側で表記。

α=0.025		n																			
片 側		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	0	0	0	0	1	1	1	1	1	2	2	2	2
	3	-	-	-	-	0	1	1	2	2	3	3	4	4	5	5	6	6	7	7	8
	4	-	-	-	0	1	2	3	4	4	5	6	7	8	9	10	11	11	12	13	13
	5	-	-	0	1	2	3	5	6	7	8	9	11	12	13	14	15	17	18	19	20
	6	-	-	1	2	3	5	6	8	10	11	13	14	16	17	19	21	22	24	25	27
	7	-	-	1	3	5	6	8	10	12	14	16	18	20	22	24	26	28	30	32	34
	8	-	0	2	4	6	8	10	13	15	17	19	22	24	26	29	31	34	36	38	41
	9	-	0	2	4	7	10	12	15	17	20	23	26	28	31	34	37	39	42	45	48
m	10	-	0	3	5	8	11	14	17	20	23	26	29	33	36	39	42	45	48	52	55
	11	-	0	3	6	9	13	16	19	23	26	30	33	37	40	44	47	51	55	58	62
	12	-	1	4	7	11	14	18	22	26	29	33	37	41	45	49	53	57	61	65	69
	13	-	1	4	8	12	16	20	24	28	33	37	41	45	50	54	59	63	67	72	76
	14	-	1	5	9	13	17	22	26	31	36	40	45	50	55	59	64	67	74	78	83
	15	-	1	5	10	14	19	24	29	34	39	44	49	54	59	64	70	75	80	85	90
	16	-	1	6	11	15	21	26	31	37	42	47	53	59	64	70	75	81	86	92	98
	17	-	2	6	11	17	22	28	34	39	45	51	57	63	67	75	81	87	93	99	105
	18	-	2	7	12	18	24	30	36	42	48	55	61	67	74	80	86	93	99	106	112
	19	-	2	7	13	19	25	32	38	45	52	58	65	72	78	85	92	99	106	113	119
	20	-	2	8	13	20	27	34	41	48	55	62	69	76	83	90	98	105	112	119	127

α=0.05		n																			
片 側		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	2	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
	3	-	-	0	0	1	2	2	3	4	4	5	5	6	7	7	8	9	9	10	11
	4	-	-	0	1	2	3	4	5	6	7	8	9	10	11	12	14	15	16	17	18
	5	-	-	1	2	4	5	6	8	9	11	12	13	15	16	18	19	20	22	23	25
	6	-	-	2	3	5	7	8	10	12	14	16	17	19	21	23	25	26	28	30	32
	7	-	-	2	4	6	8	11	13	15	17	19	21	24	26	28	30	33	35	37	39
	8	-	-	3	5	8	10	13	15	18	20	23	26	28	31	33	36	39	41	44	47
	9	-	-	4	6	9	12	15	18	21	24	27	30	33	36	39	42	45	48	51	54
m	10	-	-	4	7	11	14	17	20	24	27	31	34	37	41	44	48	51	55	58	62
	11	-	-	5	8	12	16	19	23	27	31	34	38	42	46	50	54	57	61	65	69
	12	-	-	5	9	13	17	21	26	30	34	38	42	47	51	55	60	64	68	72	77
	13	-	-	6	10	15	19	24	28	33	37	42	47	51	56	61	65	70	75	80	84
	14	-	-	7	11	16	21	26	31	36	41	46	51	56	61	66	71	77	82	87	92
	15	-	-	7	12	18	23	28	33	39	44	50	55	61	66	72	77	83	88	94	100
	16	-	-	8	14	19	25	30	36	42	48	54	60	65	71	77	83	89	95	101	107
	17	-	-	9	15	20	26	33	39	45	51	57	64	70	77	83	89	96	102	109	115
	18	-	-	9	16	22	28	35	41	48	55	61	68	75	82	88	95	102	109	116	123
	19	-	-	10	17	23	30	37	44	51	58	65	72	80	87	94	101	109	116	123	130
	20	-	-	11	18	25	32	39	47	54	62	69	77	84	92	100	107	115	123	130	138

順位相関係数：数表

検定統計量が数表の値よりも大きい場合、帰無仮説を棄却する。有意水準は片側で表記。

■スピアマンの順位相関係数の無相関検定

n	有意水準（片側）			
	0.05	0.025	0.01	0.005
4	1.000	–	–	–
5	0.900	1.000	1.000	–
6	0.829	0.886	0.943	1.000
7	0.714	0.786	0.893	0.929
8	0.643	0.738	0.833	0.881
9	0.600	0.700	0.783	0.833
10	0.564	0.648	0.746	0.794
11	0.536	0.618	0.709	0.755
12	0.503	0.587	0.678	0.727
13	0.484	0.560	0.648	0.703
14	0.464	0.538	0.626	0.679
15	0.446	0.521	0.604	0.654
16	0.429	0.503	0.582	0.635
17	0.414	0.488	0.566	0.618
18	0.401	0.472	0.550	0.600
19	0.391	0.460	0.535	0.584
20	0.380	0.447	0.522	0.570
21	0.370	0.436	0.509	0.556
22	0.361	0.425	0.497	0.544
23	0.353	0.416	0.486	0.532
24	0.344	0.406	0.476	0.521
25	0.337	0.398	0.466	0.511
26	0.331	0.390	0.457	0.501
27	0.324	0.383	0.449	0.492
28	0.318	0.375	0.441	0.483
29	0.312	0.368	0.433	0.475
30	0.306	0.362	0.425	0.467

■ケンドールの順位相関係数の無相関検定

n	有意水準（片側）			
	0.05	0.025	0.01	0.005
4	1.000	–	–	–
5	0.800	1.000	1.000	–
6	0.733	0.867	0.867	1.000
7	0.619	0.714	0.810	0.905
8	0.571	0.643	0.714	0.786
9	0.500	0.556	0.667	0.722
10	0.467	0.511	0.600	0.644
11	0.418	0.491	0.564	0.600
12	0.394	0.455	0.545	0.576
13	0.359	0.436	0.513	0.564
14	0.363	0.407	0.473	0.516
15	0.333	0.390	0.467	0.505
16	0.317	0.383	0.433	0.483
17	0.309	0.368	0.426	0.471
18	0.294	0.346	0.412	0.451
19	0.287	0.333	0.392	0.439
20	0.274	0.326	0.379	0.421
21	0.267	0.314	0.371	0.410
22	0.264	0.307	0.359	0.394
23	0.257	0.296	0.352	0.391
24	0.246	0.290	0.341	0.377
25	0.240	0.287	0.333	0.367
26	0.237	0.280	0.329	0.360
27	0.231	0.271	0.322	0.356
28	0.228	0.265	0.312	0.344
29	0.222	0.261	0.310	0.340
30	0.218	0.255	0.301	0.333

ラン検定：数表

有意水準：両側 $\alpha = 0.05$、片側 $\alpha = 0.025$

表 I は、確率 $P(R \leq R_0) \leq 0.025$ となる R_0 の最大値。R が表 I の値よりも小さくなれば H_0 を棄却。

表 II は、確率 $P(R \geq R_0') \leq 0.025$ となる R_0' の最小値。R が表 II の値よりも大きくなれば H_0 を棄却。

I		nA																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	2	-	-	-	-	-	-	-	-	-	-	2	2	2	2	2	2	2	2	2
	3		-	-	-	2	2	2	2	2	2	2	2	2	3	3	3	3	3	3
	4			-	2	2	2	3	3	3	3	3	3	3	3	4	4	4	4	4
	5				2	3	3	3	3	3	4	4	4	4	4	4	4	5	5	5
	6					3	3	3	4	4	4	4	5	5	5	5	5	5	6	6
	7						3	4	4	5	5	5	5	5	6	6	6	6	6	6
	8							4	5	5	5	6	6	6	6	6	7	7	7	7
	9								5	5	6	6	6	7	7	7	7	8	8	8
	10									6	6	7	7	7	7	8	8	8	8	9
nB	11										7	7	7	8	8	8	9	9	9	9
	12											7	8	8	8	9	9	9	10	10
	13												8	9	9	9	10	10	10	10
	14													9	9	10	10	10	11	11
	15														10	10	11	11	11	12
	16				(	上	半	と	同	じ	)					11	11	11	12	12
	17																11	12	12	13
	18																	12	13	13
	19																		13	13
	20																			14

II		nA																		
		2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
	2	4																		
	3	5	6																	
	4	5	7	8																
	5	5	7	8	9															
	6	5	7	8	9	10					(	下	半	と	同	じ	)			
	7	5	7	9	10	11	12													
	8	5	7	9	10	11	12	13												
	9	5	7	9	11	12	13	13	14											
	10	5	7	9	11	12	13	14	15	15										
nB	11	5	7	9	11	12	13	14	15	16	16									
	12	5	7	9	11	12	13	15	15	16	17	18								
	13	5	7	9	11	13	14	15	16	17	18	18	19							
	14	5	7	9	11	13	14	15	16	17	18	19	19	20						
	15	5	7	9	11	13	14	15	17	17	18	19	20	21	21					
	16	5	7	9	11	13	15	16	17	18	19	20	20	21	22	22				
	17	5	7	9	11	13	15	16	17	18	19	20	21	22	22	23	24			
	18	5	7	9	11	13	15	16	17	18	19	20	21	22	23	24	24	25		
	19	5	7	9	11	13	15	16	17	19	20	21	22	22	23	24	25	25	26	
	20	5	7	9	11	13	15	16	17	19	20	21	22	23	24	24	25	26	26	27

索引