中央値の（差の）信頼区間

- 機能
  1. 対応のある場合および対応のない場合の，中央値の差の信頼区間を推定する。
  2. 1 標本の場合の，中央値の信頼区間を求める。

     以下の 3 通りの解析法のうちいずれの方法をとるかは解析条件と指定により決まる。
     - 対応のある場合に帰着させる方法
     - 二項分布に基づく方法
     - 正規近似に基づく方法
- 関連プログラム

  間隔尺度以上の場合は t 検定，対応のある場合の t 検定により，平均値の信頼区間を求めることができる。
- 解析理論
  1. 対応のないデータの中央値の差の信頼区間

     2 つの母集団の中央値の差の信頼区間を推定するときには，分布の形が同じで位置の母数だけが異なることを仮定する。このように仮定すると，以下に述べるノンパラメトリックな信頼区間は 2 つの中央値の差であると共に，平均値や特定のパーセンタイル値の差とみなせる。この仮定は分布に差がないという帰無仮説には必ずしも必要ではないが，検定結果を解釈する上では困難が生じる。

     2 つの必ずしも正規分布に従わない母集団から，標本の大きさ n_1，n_2 の データを抽出し，それぞれ X_1，X_2，$\ldots$，X_{n_1}，および Y_1，Y_2，$\ldots$，Y_{n_2} とする。母集団の中央値（平均値）および差の信頼区間は，データのあらゆる組合せ $n_1 \times n_2$ の差 $X_i - Y_j$（$i = 1$，2，$\ldots$，n_1；$j = 1$，2，$\ldots$，n_2）の中央値から推定される。

     (a) $60 \geq n_1 \geq n_2$ の場合

     ほぼ $(1-\alpha)\,100\,\%$ の信頼区間を求めるには，Mann–Whitney 統計量の $100\alpha/2\,\%$ 点を K とすると，$n_1 \times n_2$ 個の観測値間の差の，小さい方から数えて K 番目と，大きい方から数えて K 番目の値が信頼限界値である。

     (b) $n_1 > 60$ または $n_2 > 60$ の場合

     正規分布の上側確率が $\alpha/2$ となるパーセント点を $Z_{\alpha/2}$ として，前項で示した K を，(1) 式で近似する。

     $$K = \frac{n_1\, n_2}{2} - Z_{\alpha/2}\sqrt{\frac{n_1\, n_2\,(n_1 + n_2 + 1)}{12}} \tag{1}$$
  2. 対応のあるデータの中央値の差の信頼区間

     同一個体の繰り返し測定，またはマッチングされたケース–コントロールの比較などにより得られる対応のあるデータの差の信頼区間を求める場合を考える。この場合，2 つの分布は位置の母数以外は同じであり対応付けられたデータの差の分布は対称であると仮定する。

     標本の大きさを n，対応づけられた測定値の差を d_1，d_2，$\ldots$，d_n とする。2 つの母集団の差の中央値およびその信頼区間は，あらゆる 2 つの測定値の差の組合せについての $n\,(n+1)\,/\,2$ 個の平均値 $(d_i + d_j)\,/\,2$ に基づいて求められる。

     (a) $n \leq 250$ までの場合

     ほぼ $(1-\alpha)\,100\,\%$ の信頼区間を求めるには，まず Wilcoxon 統計量の $100\,\alpha/2\,\%$ 点を $K^\star$ としたとき，$n\,(n+1)\,/\,2$ 個の測定値の差の平均値の小さい方から数えて $K^\star$ 番目と，大きい方から数えて $K^\star$ 番目の値が信頼限界値である。

     (b) $n > 250$ の場合

     正規分布の上側確率が $\alpha/2$ となるパーセント点を $Z_{\alpha/2}$ として，前項で示した $K^\star$ を，(2) 式

で近似する。

$$K^{\star} = \frac{n\,(n+1)}{4} - Z_{\alpha/2}\,\sqrt{\frac{n\,(n+1)\,(2\,n+1)}{24}} \tag{2}$$

3. 1 標本の場合の中央値の信頼区間

1 標本データの場合には 3 つのアプローチがある。

(a) 対応のある場合に帰着させる方法

前項 2 の「対応のあるデータの中央値の差の信頼区間」を参照。

(b) 二項分布に基づく方法

ある測定値が中央値より大きいか小さいかが，母比率 0.5 の 二項分布に従うと仮定する。信頼率を $(1-\alpha)\,100\,\%$ とする。$\Pr\{X < r\} < \alpha/2$ となる最大の r と，$\Pr\{X > s\} < \alpha/2$ となる 最小の s を求め，測定値の小さい方から r 番目と s 番目を信頼限界値とする。

(c) 正規近似に基づく方法

正規分布の上側確率が $\alpha/2$ となるパーセント点を $Z_{\alpha/2}$ として，(3) 式による r，s を求め (最も近い整数に丸める)，測定値の小さい方から r 番目と s 番目を信頼限界値とする。

$$\begin{cases} r = \dfrac{n}{2} - Z_{\alpha/2}\,\dfrac{\sqrt{n}}{2} \\ s = 1 + \dfrac{n}{2} + Z_{\alpha/2}\,\dfrac{\sqrt{n}}{2} \end{cases} \tag{3}$$

- 参考文献

1) Gardner, M. J.: Statistics with Confidence — Confidence Intervals and Statistical Guidelines, British Medical Journal, London, 1989.