生物科学研究所 研究報告 2025 年 3 月 8 日

Kruskal-Wallis 検定を使えば *U* 検定は不要：漸近と正確検定

井口豊*
*生物科学研究所，長野県岡谷市

DOI: https://doi.org/10.5281/zenodo.14993412

1. はじめに

Kruskal-Wallis （クラスカル-ウォリス）検定は，2 群以上の平均順位（Mean rank）の差の検定として用いられ，中央値の検定ではないことに注意する必要がある（井口, 2025）。この検定が 3 群以上と誤って解説されることがあるが（例えば，名取, 2014）， 2 群でも適用可能である。また，この検定には漸近検定（近似検定）と正確検定があることも意外に知られていない。実際，医療系の学術論文でこの点を見落としているのではないか，と思われるものがあり，そのデータを再分析してみた結果を本稿の後半で示した。

まず便宜上，タイ（結び，同点）が無いデータを考え，統計解析ソフト R を用いて，2 群（2 標本，サンプル数 2）の Kruskal-Wallis 検定を行なってみる。多群の検定の場合，その特徴を知るには，2 群比較を行なうのが最適な方法である（注 1，注釈は末尾に一括）。

以下に統計ソフト R を利用して Kruskal-Wallis 検定の漸近検定（近似検定）と正確検定の違いを調べてみよう。

2. Kruskal-Wallis 漸近検定（近似検定）：2 群の場合

まず一般的に使われる kruskal.test 関数で，以下のような，ともに大きさ 4 の 2 標本 a, b の検定をしてみる。以下が R の検定スクリプトである。

```
##############
# Kruskal-Wallis 検定
# データ
x1<- c(1, 3, 5, 6)
x2<- c(2, 4, 7, 9)

dat<- c(x1, x2)
grp<- rep(1:2, c(4, 4))

kruskal.test(dat ~ grp)
```

```
################
```

結果
p = 0.3865

これが，漸近 U 検定における p 値と同値であることは，以下のようにして確認できる。

```
#############
# Mann-Whitney U (Wilcoxon 順位和) 漸近検定，連続補正なし
wilcox.test(x1, x2, exact=F, correct=F)
################
```

したがって，これは 2 群の Steel-Dwass 多重検定の結果とも同値であることは，パッケージ NSM3 の中の pSDCFlig 関数を使うと確認できる。

```
#############
#  2 群の Steel-Dwass 多重検定
library(NSM3)
pSDCFlig(dat, grp, method="Asymptotic")
################
```

つまり，通常使われる Kruskal-Wallis 検定は，2 群の場合，漸近 U 検定であり，漸近 Steel-Dwass 多重検定なのである。2 群であっても，多重検定できるのである。

この場合の漸近検定とは，漸近的にカイ二乗検定分布に近似させた検定である。ただし，標準正規分布 $N(0,\ 1)$ に従う確率変数 Z の 2 乗は，自由度 1 のカイ二乗分布になるので，上記のような 2 群の場合は，正規分布近似を行なっていると言っても良い。

3. Kruskal-Wallis 正確検定: 2 群の場合

さきほどと同じデータで正確検定の場合を考える。まず， U 検定を使う場合は，次のようになる。

```
#############
# Mann-Whitney U (Wilcoxon 順位和) 正確検定
wilcox.test(x1, x2, exact=T)
################
```

結果

p = 0.4857

当然ながら，さきほどの漸近検定の場合とは，結果が異なる。 p 値を詳しく算出するには以下のようにする。

```
#############
# p 値取り出し
wilcox.test(x1, x2, exact=T)$p.value
################
```

結果
p = 0.4857143

これを 2 群の Kruskal-Wallis 検定の正確検定として実行するには，パッケージ kSamples の中の qn.test 関数（Rank Score k-Sample Tests）を使えば良い。この関数は，実に便利な優れものであって，漸近 p 値も，正確 p 値も算出される。

```
#############
# 2 群の Kruskal-Wallis 正確検定
library(kSamples)
qn.test(x1, x2, test="KW", method="exact")$qn
################
```

結果
```
asympt. P-value    exact P-Value
0.3864762          0.4857143
```

したがって，この関数で Kruskal-Wallis 検定を行なえば， Mann-Whitney U 検定は不要であるとも言える。

4. Kruskal-Wallis 漸近検定（近似検定）: 3 群の場合

次に， 3 群（3 標本） x_1, x_2, x_3 の場合を考えてみよう。すなわちサンプル数は 3 であり，それぞれサンプルサイズを 4, 4, 3 とする。

まず，一般的な Kruskal-Wallis 検定，すなわち，漸近検定を行なう。

```
#############
# Kruskal-Wallis 漸近検定， 3 群の場合
x1<- c(1, 3, 5, 6)
x2<- c(2, 4, 7, 9)
x3<- c(8, 10, 12)
```

```
dat<- c(x1, x2, x3)
grp<- rep(1:3, c(4, 4, 3))

kruskal.test(dat ~ grp)
################
```

結果
p = 0.06086

これを理論的に算出するには，まずデータ全体の順位を計算し，その中から，各群内の順位和 R_1, R_2, R_3 を取り出し，以下の式に基づいて，検定統計量 H を求める。なお， N は全体の標本サイズ， n_1, n_2, n_3 は各群の標本サイズである。

$$H = \frac{12}{N(N+1)}\left(\frac{R_1^2}{n_1} + \frac{R_2^2}{n_2} + \frac{R_3^2}{n_3}\right) - 3(N+1)$$

この H が，群数 − 1 の自由度（ここでは， $3 - 1 = 2$） のカイ二乗分布に近似的に従うことを利用して p 値を求めることができる。前述のデータを利用した R のスクリプトは以下のようになる。

```
#############
# 計算式に基づく Kruskal-Wallis 漸近検定
# 全体の順位
rk<- rank(dat)

# 各群内の順位和
r1<- sum(rk[1: 4])
r2<- sum(rk[5: 8])
r3<- sum(rk[9: 11])

# Kruskal-Wallis 検定統計量
h<- 12/(11*12)*(r1^2/4+r2^2/4+r3^2/3)-3*(11+1)

# カイ二乗分布を用いた p 値の算出
pchisq(h, 2, lower=F)
################
```

当然ながら，前述の kruskal.test 関数を用いた場合と同じ近似結果になる。

5. Kruskal-Wallis 正確検定: 3 群の場合

3 群で Kruskal-Wallis 正確検定をするためは，2 群の場合と同じく， qn.test 関数を使えば良い。ただし，正確と言っても，この場合は，シミュレーションで反復回数を多くして計算することになる。

```
#############
# Kruskal-Wallis 正確検定
library(kSamples)
qn.test(x1, x2, x3, test="KW", Nsim = 1000000, method="exact")
################
```

結果を抜粋すると，以下のとおりである。

```
asympt. P-value    exact P-Value
     0.06086          0.04866
```

前述のとおり，漸近 p 値である asympt. P-value は，通常の Kruskal-Wallis 検定の p 値である。この結果は， 5% 有意水準とすると，カイ二乗近似による漸近検定では有意差なしだが，正確検定では有意差ありとなり，異なる結論となる。

Steel-Dwass 検定の場合と同じく，問題は，自分が使うソフトやプログラムが，どちらの方法を採っているか知らずに，あるいは，区別があることすら知らずに，それらを使う場合である。例えば， R ベースの有名なフリーソフト EZR の Kruskal-Wallis 検定も漸近法である。

6. 学術論文で気になるケース

漸近検定か正確検定か，それを考慮せずに Kruskal-Wallis 検定を利用したように思われるの医学論文が藤本保志ほか（1997）である。 Kruskal-Wallis 検定が行なわれた表 3 のデータで，私が気になったのは，項目 D 「逆流」の結果である。

患者を舌切除範囲によって，なし・部切・半切・亜全摘または全摘の 4 群に分けている。 部切は 10 例となっているが，表 1 を見ると，部切は 9 例しかなく，たぶん，表 1 が間違えていて，症例 8 は，なし，ではなく，部切，であろう。

実際に，この逆流項目 D を，Kruskal-Wallis 漸近検定と正確検定で調べてみよう。

```
#######################################
## 藤本ら（1997）データ
```

```
# 症例番号（表 1）
ex<- 1:22

# 舌切除範囲（表 1）
# 0:なし, 1:部切, 2:半切, 3:亜全摘・全摘
r<- c(
1, 1, 1, 0, 0, 0, 1, 1, 1, 1,
1, 1, 1, 2, 2, 2, 2, 2, 2, 3,
3, 3
)

# 逆流スコア D（表 2）
D<- c(
8, 5, 8, 8, 9, 10, 8, 8, 8, 8,
4, 10, 10, 10, 3, 6, 8, 6, 8, 6,
4, 2
)

# データフレーム
dat<- data.frame(ex, r, D)

# 平均
tapply(dat$D, dat$r, mean)

# 標準偏差
tapply(dat$D, dat$r, sd)
###############################
```

結果を抜粋すると，以下のとおりである。

```
平均
       0        1        2        3
9.000000 7.700000 6.833333 4.000000

標準偏差
       0        1        2        3
1.000000 1.888562 2.401388 2.000000
```

それぞれ，論文の表 3 の D の統計量に一致することが確認できる。

次に， Kruskal-Wallis 検定を実行する。先に述べたように， qn.test 関数を使えば，漸近検定でも正確検定でも，その結果が出力される。ただし，データをリスト化して渡すことに注意する。

```
####################################
# Kruskal-Wallis 検定，漸近と正確
library(kSamples)

dat.lt<- tapply(dat$D, dat$r, list) # D データのリスト化
qn.test(dat.lt, test="KW", Nsim = 1000000, method="exact")

################################
```

結果を抜粋すると，以下のとおりである。

```
asympt. P-value     sim. P-Value
        0.05549          0.04130
```

5% 水準で，漸近検定では有意差なし，正確検定では有意差あり，となる。論文の表 3 を見ると， D は， $p < 0.05$ のアスタリスク＊が付いていないので，おそらく漸近検定で分析したものだろうと推測される。本文でも，総合点のみ有意差を認めた，というような表現になっている。

この正確検定の結果が，藤本ら（1997）の結論を変えるものではない。しかしながら，通常の検定が漸近検定であることに気づいていなかったとしたら，データ解析上の問題点として挙げられる部分である。

注

1. 多重比較 Steel-Dwass 正規近似と正確検定： *U* 検定が基準
   https://biolab.sakura.ne.jp/steel-dwass.html 2025 年 3 月 8 日確認.
   ウェルチ多重検定と 2 群の分散分析 *t* 検定， R の pairwise.t.test
   https://biolab.sakura.ne.jp/welch-comparison.html 2025 年 3 月 8 日確認.

参考文献

藤本保志・松浦秀博・田山二朗・中山敏・長谷川泰久 (1997) 口腔・中咽頭癌治療後えん下機能評価基準の提案とその評価成績．日本気管食道科学会会報 48(3): 234–241. https://doi.org/10.2468/jbes.48.234

井口豊 (2025) Kruskal-Wallis は平均順位検定であり，中央値検定ではない．生物科学研究所 研究報告 2025 年 3 月 2 日. https://doi.org/10.5281/zenodo.14956326

名取真人 (2014) カイ二乗近似によるクラスカル・ウォリス検定と小標本．霊長類研究 30(2): 209–215. https://doi.org/10.2354/psj.30.019