

Kruskal-Wallis 検定: 近似検定の標本サイズ

井口豊*

*生物科学研究所, 長野県岡谷市

The validity of the chi-square approximation of the Kruskal-Wallis test statistic

Yutaka Iguchi*

*Laboratory of Biology, Okaya, Japan

Email: bio.iguchi@gmail.com

OSF Preprint: https://doi.org/10.31219/osf.io/2gprx_v1

Kruskal-Wallis 検定: 近似検定の標本サイズ

井口豊*

*生物科学研究所, 長野県岡谷市

The validity of the chi-square approximation of the Kruskal-Wallis test statistic

Yutaka Iguchi*

*Laboratory of Biology, Okaya, Japan

Email: bio.iguchi@gmail.com

Abstract

This study demonstrates by QQ-plots that the validity of the chi-square approximation of the Kruskal-Wallis H statistic depends on sample size. This approximation requires a sample size of 30 or more.

1. はじめに

ノンパラメトリック検定として有名な Kruskal-Wallis 検定では, しばしば近似計算が使われる (Kruskal and Wallis, 1952)。例えば 3 群の場合, 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。ここで 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。

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

この統計量 H が, 標本サイズが大きくなるにつれて, 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 値が算出される。

ところが, これに気づかないで利用していると思われる研究例もある。例えば, 藤本ほか (1997) 表 3 では, 4 群で合計 $n = 22$ であるが, 近似検定しているようである (井口, 2025)。

では, Kruskal-Wallis 検定の近似計算を使う場合, どのくらいの標本サイズが必要だろうか? 名取 (2014) は, Meyer and Seaman (2008) により 3 群の各サンプルサイズ 35 までの正確検定表が作成されたことを受けて, これより大きな標本の場合を近似計算可能と考えたようである。しかし実際にシミュレーションで調べた近似計算の妥当性ではなかった。そこで本稿では, 3 群で標本サイ

ズの違いによる Kruskal-Wallis 検定の検定統計量 H による近似の妥当性を調べた。なお名取 (2014) は冒頭で Kruskal and Wallis (1952) を引用し、「クラスカル・ウォリス検定は、3 群以上を対象」と書いているが、これは 2 群以上の検定である (井口, 2025)。Kruskal and Wallis (1952) の論文でも 2 群も適用できる議論になっている。

2. Kruskal-Wallis 検定統計量 H 近似と標本サイズ

ここでは、いわゆる 5 件法、5 段階リッカート尺度 (Likert scale) から成る 3 群を想定した。1 から 5 の整数値が、母比率 2:4:2:1:1 である集団を考え、大きさ 5, 10, 20, 30, 50, 100 の標本を 10 万回繰り返し取り出し、 H を計算し、自由度 2 のカイ二乗分布との適合度を QQ plot で調べた。検定と並び替え検定を行い、 p 値の出現状況を調べるシミュレーションを試みる。

以下が、R スクリプトである。パッケージ EnvStats (Millard, 2013) の qqPlot 関数が使われている。

```
#####  
n<- c(5, 10, 20, 30, 50, 100)  
h<- hn<- as.numeric(NULL)  
  
set.seed(123)  
  
for (i in 1: length(n)) {  
  h<- replicate(1e+5, {  
    grp<- factor(rep(1:3, each = n[i]))  
    dat<- sample(  
      1:5, 3*n[i], prob = c(2, 4, 2, 1, 1),  
      replace = TRUE)  
  
    kruskal.test(dat ~ grp)$statistic  
  })  
  
  hn<- c(hn, h)  
}  
  
hmat<- matrix(hn, ncol = length(n))  
head(hmat)  
  
library(EnvStats)
```

```

qq<- function(x) {
  qpt<- qqPlot(
    x, dist = "chisq",
    param.list=list(df = 2),
    add.line = TRUE,
    line.col = "red",
    line.lwd = 2,
    equal.axes = TRUE,
    ylab ="Observed quantiles",
    main = "Chi-square QQ Plot
           for KW test with 3 groups"
  )
}

par(mfrow = c(3, 2), mar = c(5, 5, 5, 3))

qq(hmat[, 1])
text(8, 23, "n = 5 per group")

qq(hmat[, 2])
text(8, 23, "n = 10 per group")

qq(hmat[, 3])
text(8, 23, "n = 20 per group")

qq(hmat[, 4])
text(8, 23, "n = 30 per group")

qq(hmat[, 5])
text(8, 23, "n = 50 per group")

qq(hmat[, 6])
text(8, 23, "n = 100 per group")
#####

```

シミュレーション結果を次の図 1 に示した。Kruskal-Wallis 検定統計量 H のカイ二乗近似計算を適用するためには、1 群あたり標本サイズ n が 30 またはそれ以上必要だと思われる。もちろん、これ以外の条件になれば結果も変わってくる。

逆に言えば、1 群あたり 10 個前後のデータで、Kruskal-Wallis 検定の近似計算に気づかないと問題になるかもしれない。実際、前述の藤本ほか (1997) の場合、4 群 22 症例の Kruskal-Wallis 検定が近似検定だったために、井口 (2025) が正確検定で再計算した結果と異なった可能性がある。

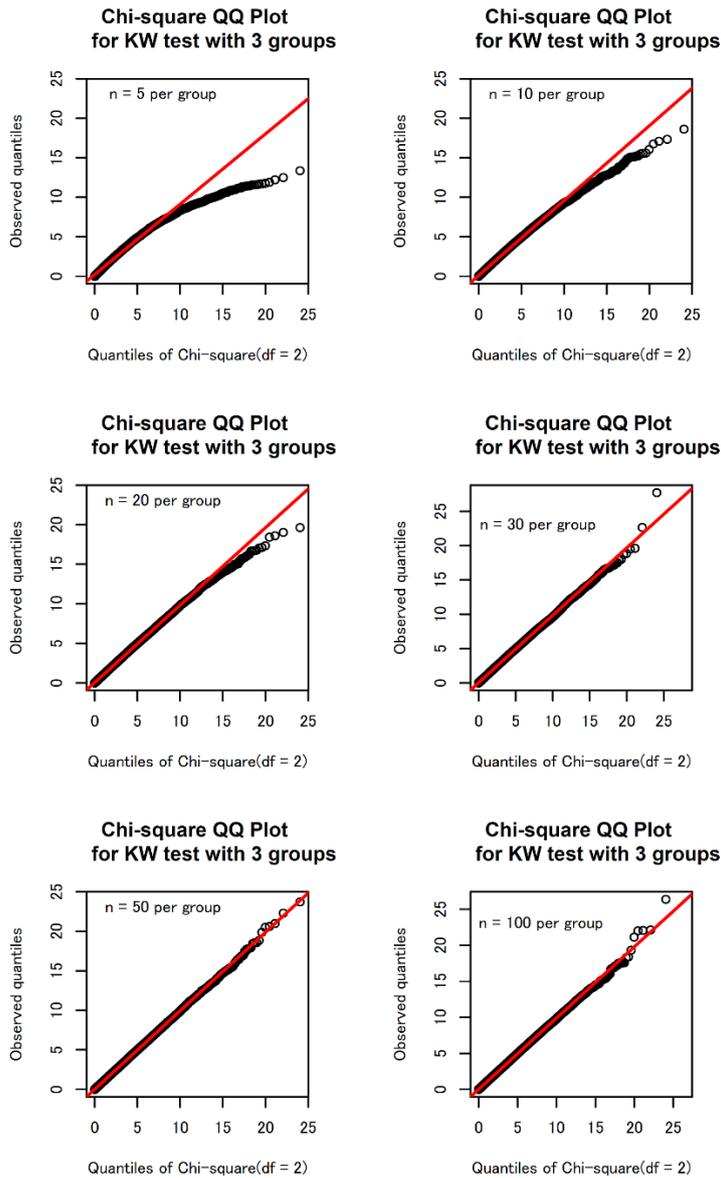


Figure 1. The chi-square approximation of the Kruskal-Wallis H statistic depending on sample size.

図 1. Kruskal-Wallis 検定統計量 H カイ二乗近似と標本サイズ

参考文献

藤本保志・松浦秀博・田山二郎・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 検定を使えば U 検定は不要: 漸近と正確検定. 生物科学研究所 研究報告 2025 年 3 月 8 日. <https://doi.org/10.5281/zenodo.14993412>

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>

Kruskal, W. H. and Wallis, W. A. (1952) Use of ranks in one-criterion variance analysis. *Journal of the American statistical Association* 47: 583–621.
<https://doi.org/10.1080/01621459.1952.10483441>

Millard S. P. (2013) *EnvStats: An R Package for Environmental Statistics*. Springer, New York.