Colloquium Biometricum 43
2013, 73–80

POWER OF PARAMETRIC AND NONPARAMETRIC TESTS OF SEVERAL MEAN POPULATIONS

Magdalena Ćwiklińska, Zofia Hanusz

Department of Applied Mathematics and Computer Science
University of Life Sciences
Akademicka 13, 20-950 Lublin, Poland
e-mails: magdalena.cwiklinska@up.lublin.pl, zofia.hanusz@up.lublin.pl

Summary

One of the main statistical problem is testing the hypothesis about equality of means of several populations. In practice, to test this hypothesis the F test is used. However, the F test needs normality of populations and homogeneity of variances. In practice, generally these assumptions are not fulfilled. In such cases nonparametric tests should be used. Practitioners apply the F-test without verifying assumptions, with confidence that this test is the best one. In the paper we show in simulation study that nonparametric Kruskal-Wallis test is not worse than the F test by mean of their sample significance levels, and power under several alternative conditions.

Keywords and phrases: normality, Kruskal-Wallis test, the F test, power

Classification AMS 2010: 62F03, 62G10

1. Introduction

Planning experiments we should know which hypothesis will be verified and which test will be used. In the literature of this subject, to the same testing problem several tests are frequently proposed. However, we should apply the most proper test. Such a test should keep a significance level α, and should have a high power. The power of the test says how frequent wrong alternative hypothesis is rejected. The power depends on significance level, sample sizes and distances of population means (Cohen, 1988).

In the paper we focus on testing the hypothesis about equality of several mean populations. In this problem, we can use the F test where populations are normal and homogeneous, or the Kruskal-Wallis test. The second test can be used even when the assumptions for the F test are not met. In Section 2 we describe the Kruskal-Wallis test. Simulation results on sample significance levels are presented in Section 3, and power of both test in Section 4. Some concluding remarks are given in Section 5.

2. Parametric and nonparametric tests for equality of several mean populations

Let us assume that we are interested in testing a hypothesis about equality of means of *k* independent populations $\pi_1, \pi_2, \ldots, \pi_k$ ($k > 2$), namely,

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_k , \tag{2.1}$$

where $\mu_1, \mu_2, \ldots, \mu_k$ denote unknown means of populations. The hypothesis (2.1) will be verified against the alternative that not all means are the same i.e. $H_1 : \underset{i \neq j = 1..k}{\exists} \mu_i \neq \mu_j$,.

Let $X_{11}, \ldots, X_{1n_1}; X_{21}, \ldots, X_{2n_2}; \ldots; X_{k1}, \ldots, X_{kn_k}$ denote *k* random samples of the sizes $n_1, n_2, \ldots, n_k$ from $\pi_1, \pi_2, \ldots, \pi_k$. When *X*'s are normally distributed with the same variances then to test (2.1) the F test can be used. However, when the normality assumption and homogeneity are not fulfill then to test (2.1), the nonparametric test has to be used. In the paper, we consider the Kruskal-Wallis statistic of the form (Hollander and Wolfe, 1999)

$$H = \frac{12}{N(N+1)} \sum_{i=1}^{k} \frac{R_i^2}{n_i} - 3(N+1), \tag{2.2}$$

where $N = \sum_{i=1}^{k} n_i$, n_i is the size of a random sample from *i*th population, $R_i = \sum_{j=1}^{n_k} R_{ij}$, R_{ij} denotes a rank in the join ranking of X_{ij} , and X_{ij} is the *j*th observation from *i*th population ($i = 1,\dots,k$, $j = 1,\dots,n_i$).

The hypothesis (2.1) is rejected on significance level α if $H \ge h_\alpha$, where h_α is the critical point given e.g. in Hollander and Wolfe (1999) or Zieliński and Zieliński (1990). When minimum of sample sizes n_i ($i = 1,\dots,k$) tends to infinity then (2.1) is rejected if $H \ge \chi^2_{\alpha,k-1}$, where $\chi^2_{\alpha,k-1}$ is the upper α percentile point of a chi-square distribution with (*k*-1) degrees of freedom (Hollander and Wolfe, 1999).

If there are ties among the *X*'s, assign each observations in a tied group the average of the integer ranks that are associated with the tied group, the following modification is needed to apply

$$H^* = \frac{H}{1 - \frac{1}{N^3 - N} \sum_{i=1}^{g} \left(t_i^3 - t_i\right)}, \qquad (2.3)$$

where *H* is defined in (2.2), *g* denotes the number of tied *X* groups, t_i is the size of tied group $(i = 1,\dots,g)$.

3. Sample significance level of the F and Kruskal-Wallis tests

In this section, we compare in simulation study sample significance levels of the F test and the Kruskal-Wallis test. In simulations we determine the significance level $\alpha = 0.05$ and consider sample sizes from 4 to 100 with the step 2, generated from $k = 3$ and $k = 4$ populations of different distributions fulfilled or not the assumption for the F test using. For each of the case 10,000 runs were done. Sample significance level was calculated as the ratio of rejected true hypothesis (2.1) to 10,000 runs. All simulation were carried out in R program (R Development Core Team, 2008). The results of sample significance levels for samples from 3 and 4 normal populations are illustrated in Figures 1 and 2. The dotted line denotes fixed significance level $\alpha = 0.05$.

The simulated results presented in Figure 1 and 2 are very similar for 3 and 4 populations. Namely, when samples are generated from standard normal distribution, i.e. the assumption for the F test are fulfilled, both tests preserve the significance level $\alpha = 0.05$ for sample sizes greater than 10 (left panels at the top of Fig. 1, 2). In the case where variances differ little (right panels at the top), then the sample significance levels for the F tests is greater than $\alpha = 0.05$. When variances are different (both panels at the bottom), then both tests do not preserve the significance level. The sample significance level of the F tests is smaller than $\alpha = 0.05$ but for the Kruskal-Wallis test is bigger than 0.05.

The results for 3 and 4 populations presented in Figure 1 and 2 are very similar therefore in further simulations we consider only samples generated from 3 different populations. The results for samples generated from distributions with "heavy tails" are presented in Figure 3. We note that the sample significance level of the Kruskal-Wallis test preserve $\alpha = 0.05$ but the F test does not.

The results obtained for samples generated from Uniform, Student, Beta and Gamma distributions are presented in Figure 4. It can be noticed that both tests behave very similar and for sample sizes greater than 10 both preserve $\alpha = 0.05$.

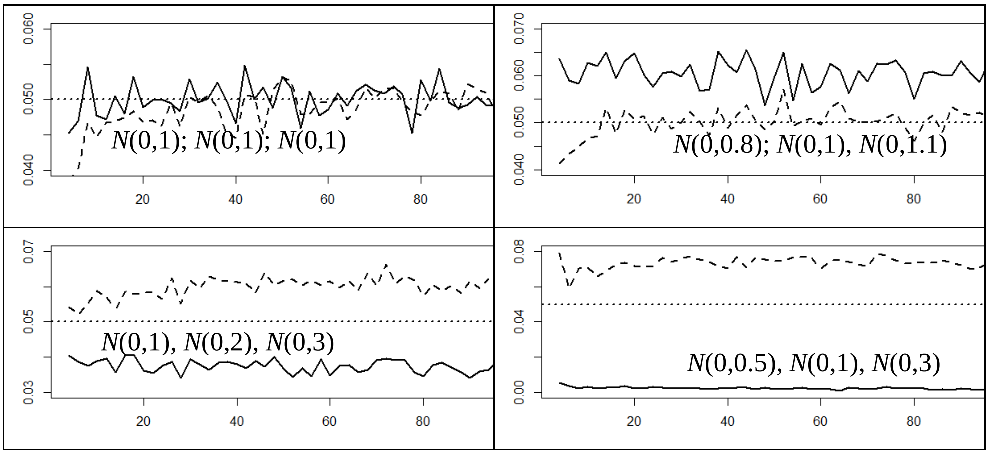

Fig. 1. Sample significance level of the F test (solid line) and the Kruskal-Wallis test (dashed line) from k=3 different normal populations

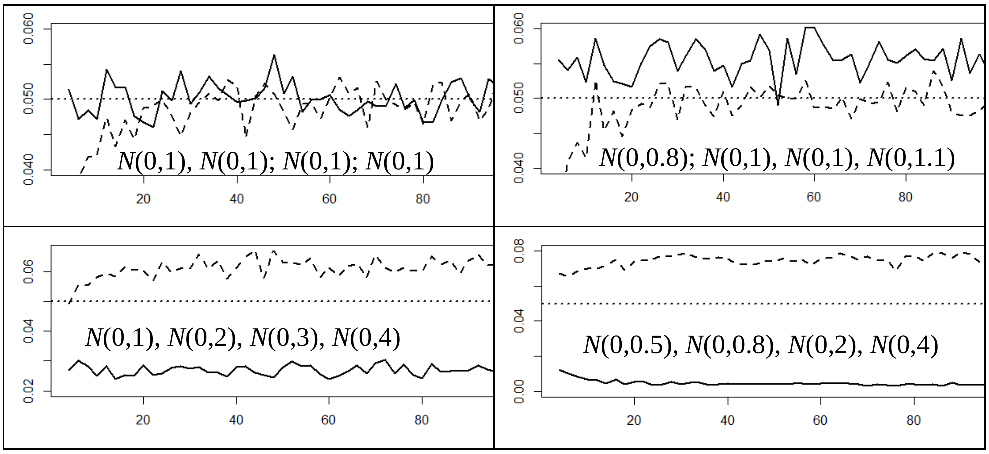

Fig. 2. Sample significance level of the F test (solid line) and the Kruskal-Wallis test (dashed line) from k=4 different normal populations

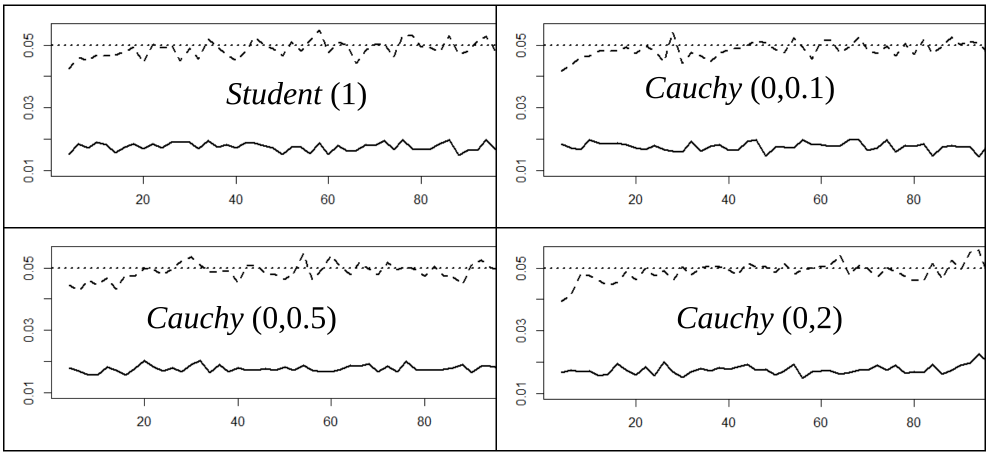

Fig. 3. Sample significance level of the F test (solid line) and the Kruskal-Wallis test (dashed line) for samples generated from k=3 populations with heavy tails

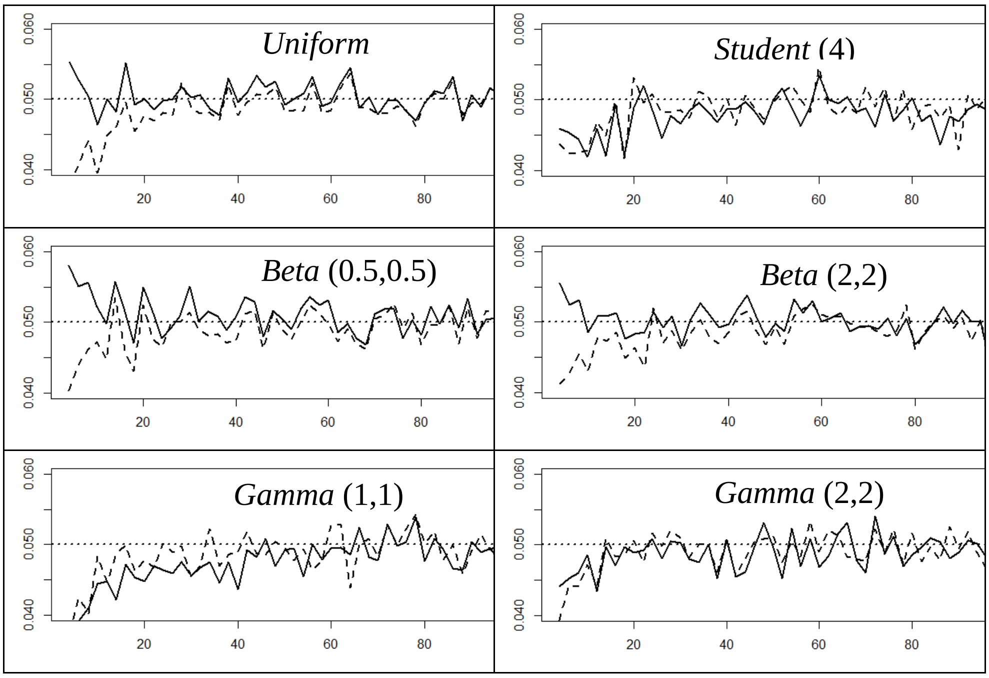

Fig. 4. Sample significance level of the F test (solid line) and the Kruskal-Wallis test (dashed line) for samples generated from k=3 nonnormal populations of Uniform, Student, Beta and Gamma distributions

4. Power of the F and Kruskal-Wallis tests

In simulation study on power of the F and the Kruskal-Wallis tests we consider only a case of $k = 3$ populations of different distributions and sample sizes 10, 30 and 100 generated from them. Moreover, we regard the case where two populations have null scale parameter, but the third one is shifted by x=0.0, 0.5, 1.0, 1.5 and 2.0. For each case 10,000 testing of the hypothesis (1) were done. Power of both tests was calculated as the proportion of the rejected hypotheses on the significance level $\alpha = 0.05$. The results are presented in Figure 5. It is easy to see that if the samples were generated from selected distributions, the Kruskal-Wallis test was more powerful than F test. Only for normal distribution for sample sizes 10 and 30, the F test turned out to be more powerful.

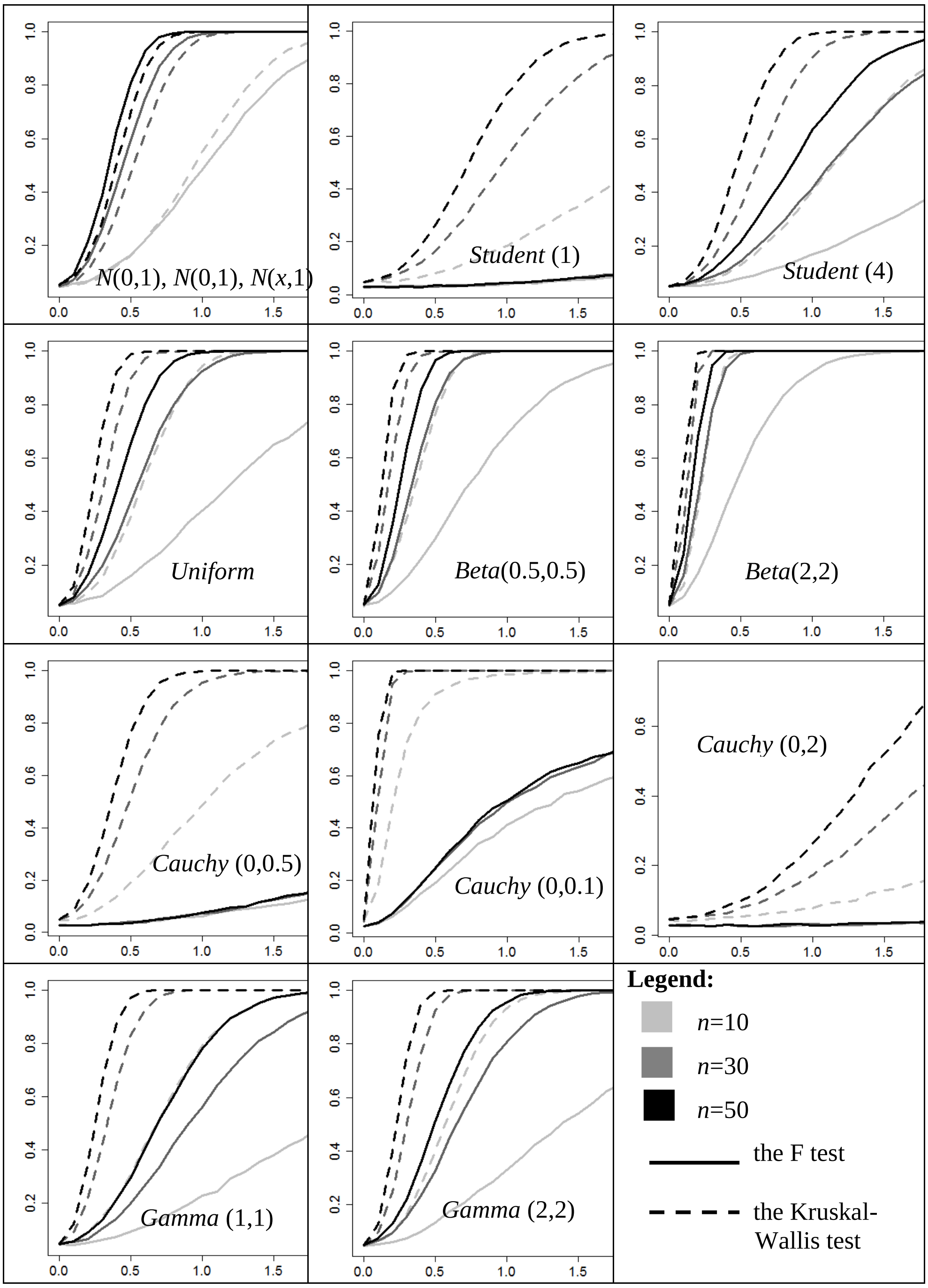

Fig. 5. Power of the F test (solid line) and the Kruskal-Wallis test (dashed line) for samples generated from k=3 different populations

Power comparison of Kruskal-Wallis test and F-test for populations of normal, exponential or Poisson distributions can be found in Adams et al. (2009). More simulation results on sample significance levels and power can be also found in Ćwiklińska (2013).

5. Conclusions

In the paper we showed that nonparametric Kruskal-Wallis test applying to test the hypothesis about equality of several mean populations is not worse than the parametric F test. Both tests behave similarly regards to sample significance level an power. Even in the case where assumptions for use of the F test are met, the simulation results showed (Figure 1) that the Kruskal-Wallis test is as good as the F test.

In the case when the homogeneity assumption is not fulfilled, the Kruskal-Wallis test preserve the significance level $\alpha = 0.05$ pretty much (Figure 2) but the F test not. Similar results we get when the assumption about normality is not fulfilled, namely, for heavy tailed distribution (Figure 3). For the Uniform, Student with 4 d.f., Beta and Gamma distributions, both compared tests preserve the significance level.

Sample significance levels of the F test and the Kruskal-Wallis test for samples generated from three nonnormal populations as Uniform, Student, Beta and Gamma distributions are similar for n>10 (Figure 4). Power of both tests, for three samples generated from non normal population showed that almost everywhere the Kruskal-Wallis test is more powerful than the F test (Figure 5).

All simulations carried out in the paper also showed that the Kruskal-Wallis test should be applied both in the case where the assumptions for the F test are not fulfilled and when the assumptions are achieved.

References

Adams S. O., Gayawan E., Garba M. K. (2009). Empirical Comparison of the Kruskal Wallis Statistic and its Parametric Counterpart. *Journal of Modern Mathematics and Statistics,* 3(2), 38–42.

Cohen J. (1988). *Statistical Power Analysis for the Behavioral Sciences*. Psychology Press Taylor & Francis Group.

Ćwiklińska M. (2012). *Application of statistical nonparametric methods in agriculture engineering*. Ph.D. Dissertation, in Polish.

Hollander M., Wolfe D. A. (1999). *Nonparametric statistical methods*. John Wiley & Sons Inc.

Zieliński R., Zieliński W. (1990). *Statistical Tables*, PWN.

R Development Core Team (2008). R: *A language and environment for statistical computing. R Foundation for Statistical Computing*. Vienna, Austria, URL http://www.R-project.org.