

Empirical Power of Four Statistical Tests in One Way Layout

Lorenzo J. Martínez H.

Universidad de Caldas, Colombia

Copyright © 2014 Lorenzo J. Martínez H. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Abstract

By simulating random samples from continuous distributions the empirical power of four statistical tests is compared when parametric assumptions are satisfied and not satisfied.

Keywords: ANOVA, Parametric and Nonparametric Tests, empirical Power

1 Introduction

This paper is focused on the empirical power of four statistical tests (Anova-F, Kruskal-Wallis, Van der Waerden, and Wilson) which compare several treatments classified by only an experimental factor. Random samples from continuous distributions were taken when parametric assumptions were satisfied and not satisfied.

In one way layout we consider a single factor T with k levels, ($k \geq 3$), and the hypothesis testing

$$\begin{aligned} H_0 : \theta_1 = \cdots = \theta_k \\ \text{versus} \\ H_1 : \theta_i \neq \theta_j, i \neq j \end{aligned} \tag{1}$$

where θ_i is a location parameter. The hypothesis H_0 is rejected on the basis of observations x_{i1}, \cdots, x_{in_i} for $i = 1, 2, \cdots, k$, when the value of the statistics

$$T_n = f(x_{i1}, x_{i2}, \cdots, x_{in_i}), \quad i = 1, 2, \cdots, k \tag{2}$$

falls in the critical region C given by

$$C = \{T_n : T_n \geq c\} \quad (3)$$

where c is a real number, or equivalently, when

$$P[T_n \geq c] = p < \alpha \quad (4)$$

where α is the level of significance of the test and p is the p -value; otherwise we accept H_0 . When we test H_0 , a type I error occurs If a true H_0 is rejected, and a type II error if a false H_0 is accepted, and the power refers to the probability that the test rejects H_0 when in fact H_0 is false [9, 4].

	$H_0 : True$	$H_0 : False$
<i>rejected</i>	Type I error $P[\text{Type I error}] = \alpha$ Level of Significance	Right Decision $\pi = 1 - \beta$ Power of the test
<i>accepted</i>	Right decision Level of confidence= $1-\alpha$	Type II error $P[\text{Type II error}] = \beta$

The power function is defined by,

$$\pi_\gamma(\theta) = P[T_n \geq c | H_0 \text{ false}] \quad (5)$$

In the following a brief description related to the involved tests.

1.1 ANOVA-F Test One Way Layout (Fixed Effects)

This test can be traced back to the beginning of XX century, when R. Fisher (1890-1962) used analysis of variance in agricultural experiments. The mathematical model is given by

$$y_{ij} = \mu_i + \varepsilon_{ij}, \quad i = 1, 2, \dots, k; j = 1, 2, \dots, n_i \quad (6)$$

which can be written in the form

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \quad (7)$$

where y_{ij} is the value of the j -th observation at the i -th level, μ is the general mean, τ_i is the effect of i -th level of the factor, ε_{ij} is the random error associated to y_{ij} and $\sum_{i=1}^k n_i = N$ is the total number of observations[9]. Assumptions of the model are given by

1. $E[\varepsilon_{ij}] = 0$, the errors ε_{ij} s are assumed to be randomly distributed with mean zero,

2. $\sigma_{\varepsilon_{ij}}^2 = \sigma^2$, the errors have a common variance, and
3. $\text{cov}[\varepsilon_{ij}, \varepsilon_{hk}] = 0$, the covariance between every pair of different ε_{ij} is zero, unless $h = i$ and $k = j$. for $i = 1, \dots, k$,

The hypothesis testing for this test has the form

$$\begin{aligned} H_0 : \mu_1 = \dots = \mu_k \\ \text{versus} \\ H_1 : \mu_i \neq \mu_j \text{ } i \neq j \end{aligned} \quad (8)$$

The hypothesis H_0 is tested based on the ratio

$$F_{ratio} = \frac{\sum_{i=1}^k \sum_{j=1}^{n_i} [\bar{y}_{i.} - \bar{y}_{..}]^2}{k-1} \quad (9)$$

$$\frac{\sum_{i=1}^k \sum_{j=1}^{n_i} [y_{ij} - \bar{y}_{i.}]^2}{N-k}$$

where $y_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} y_{ij}$, and $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$. According to [3], under H_0

$$F_{ratio} \sim F_{k-1, N-k} \quad (10)$$

and under H_1

$$F_{ratio} \sim F_{k-1, N-k, \lambda} \quad (11)$$

where the non centrality parameter $\lambda = \frac{\sum_{i=1}^k n_i \tau_i^2}{\sigma^2}$, and $\tau_i = \bar{y}_{i.} - \bar{y}_{..}$ is the effect of the i th-treatment.

1.2 The Kruskal-Wallis test

This test, proposed by W. Kruskal and W. Wallis in 1952, is an alternative to F-ANOVA when parametric assumptions are not satisfied [5]. It is focused on the medians of three or more populations, T_1, \dots, T_k ($k \geq 3$). According to [7], the following are the theoretical assumptions:

1. The k random samples y_{i1}, \dots, y_{in_i} , $i = 1, \dots, k$ are mutually independent.
2. For each fixed $i \in \{1, \dots, k\}$ the random sample is coming from a continuous distribution function F_i .
3. The distribution functions F_1, \dots, F_k are connected through the relationship $F_i = F(t - \tau_i)$, $-\infty < t < \infty$ for $i = 1, \dots, k$, where F is a distribution function for a continuous distribution with unknown median θ and τ_i is the unknown treatment effect for the i th-population.

The mathematical model is given by

$$y_{ij} = \theta + \tau_i + \varepsilon_{ij}, \quad i = 1, \dots, k; \quad j = 1, \dots, n_i \quad (12)$$

where k is the number of samples, n_i is the number of observations in the i th treatment, $N = \sum_{i=1}^k n_i$ is the total number of observations and α is the level of significance. In this test the original observations, y_{i1}, \dots, y_{in_i} are transformed in r_{i1}, \dots, r_{in_i} where r_{ij} is the j -th rank in the i -th treatment. The Kruskal-Wallis statistic H is defined without ties by

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left[\bar{R}_{i.} - \frac{(N+1)}{2} \right]^2 \quad (13)$$

and with ties by

$$H = \frac{\frac{12}{N(N+1)} \sum_{i=1}^k \frac{R_{i.}^2}{n_i} - 3(N+1)}{1 - \sum_t (t^3 - t) / (N^3 - N)}, \quad (14)$$

where t is the number of ties, $R_i = \sum_{j=1}^{n_i} r_{y_{ij}}$, $R_{i.} = \sum_{j=1}^{n_i} r_{ij}$, $\bar{R}_{i.} = \sum_{j=1}^{n_i} \frac{1}{n_i} r_{ij}$ and $R_{..} = \sum_{i=1}^k \sum_{j=1}^{n_i} r_{ij}$ and $\bar{R}_{..} = \frac{1}{N} \sum_{i=1}^k \sum_{j=1}^{n_i} r_{ij}$. The contrast of hypothesis for this test has the form

$$\begin{aligned} H_0 : \quad & \theta_1 = \dots = \theta_k \\ & \text{versus} \\ H_1 : \quad & \theta_i \neq \theta_j, \\ & i \neq j, \quad j = 1, 2, \dots, k \end{aligned} \quad (15)$$

According to [8], under H_0 , $E[r_{ij}] = \frac{N+1}{2}$; $E[\bar{R}_{i.}] = \frac{N+1}{2}$; $\text{var}[r_{ij}] = \frac{N^2-1}{12}$; $\text{var}[\bar{R}_{i.}] = \frac{(N+1)(N-n_i)}{12n_i}$. By using results due to Wald and Walfowitz 1944 [2], Kruskal-Wallis 1952 [5], proofed that

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i \left[\bar{R}_{i.} - \frac{N+1}{2} \right]^2 \quad (16)$$

under H_0 has, as $\min\{n_1, \dots, n_k\}$ tends to infinity, an asymptotic chi-square distribution with $k-1$ degrees of freedom [5]. The distribution of the statistic H under the alternative hypothesis H_1 is unknown, so the power of this test can not be computed directly, instead, by using Monte Carlo simulations the empirical power can be calculated.

1.2.1 Van der Waerden Test

This test is an extension of the test for two populations proposed by B. L. Van der Waerden. The assumptions of the model are the same of the Kruskal-Wallis test [4]. The data consist of k random samples y_{i1}, \dots, y_{in_i} , $i = 1, \dots, k$. All N values form rank 1 to rank N as explained in the Kruskal-Wallis test, namely, r_{ij} denote the rank of the observation y_{ij} and each r_{ij} is transformed into the $R_{ij}/(N+1)$ quantile of a standar random normal variable and it is called $A_{ij} = Z_{(r_{ij}/(N+1))}$. In this case the variance has the form

$$S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{n_i} A_{ij}^2. \quad (17)$$

The Van der Waerden statistical test is defined by

$$V = \frac{1}{S^2} \sum_{i=1}^k n_i \bar{A}_{ij}^2, \quad S^2 = \frac{1}{N-1} \sum_{i=1}^k \sum_{j=1}^{n_i} A_{ij}^2. \quad (18)$$

And the contrast of hypothesis has the form

$$\begin{aligned} H_0 : & \text{All of the } k \text{ population distribution functions are identical} \\ & \text{versus} \\ H_1 : & \text{All of the } k \text{ population distribution functions are identical} \end{aligned} \quad (19)$$

According to [4], under H_0 the statistic V is asymptotically χ_{k-1}^2 , however, under H_1 the distribution is unknown. so the power of this test can not be computed directly, instead, by using Monte Carlo simulations the empirical power is calculated.

1.2.2 Wilson test

Theoretical assumptions related to these tests are the following

1. random samples are independent of each other,
2. the measurement scale is at least ordinal, and
3. If all the population have the same median, all population have the same probability p of an observation exceeding the grand median [6].

In this test, M denote the grand sample median, O_{i1} denote the number of observation in the i th sample greater or equal than M , O_{i2} the number of observation in the i th sample that are less than M , a is the total number of observation above or equal to the the grand median, and b is the total number of observation less than the grand median.

<i>Sample</i>	T_1	T_2	\cdots	T_k	<i>Total</i>
$\geq M$	O_{11}	O_{21}	\cdots	O_{k1}	a
$< M$	O_{12}	O_{22}	\cdots	O_{k2}	b
<i>Total</i>	t_1	t_2	\cdots	t_k	N

The test statistic is defined by

$$W = \frac{4k}{N} \sum_{i=1}^k \left[O_{i2} - \frac{N}{2k} \right]^2 \quad (20)$$

and the hypothesis test is given by

$$\begin{aligned} H_0 : & \text{All } k \text{ populations have the same median} \\ & \text{versus} \\ H_1 : & \text{At least two populations have different medians} \end{aligned} \quad (21)$$

according to [6] under H_0 , the distribution of W is approximately chi-squared with $k-1$ degrees of freedom. The power of this test can be computed by using Monte Carlo simulations.

2 Empirical Power Comparison

2.1 Algorithm

1. Input: level of significance (α); number of treatments (k); initial number of observations per treatment in the first iteration (r_o); number of observations per treatment in the last iteration (r_f); difference between the number of observations per treatment of two consecutive iterations (Δr); number of times that random samples X_1, X_2, \dots, X_k are selected in each iteration (T).
2. The empirical power in the i -th iteration is computed as the ratio between the number of times that H_0 is rejected (number of rejections) and the number of times (T) that random samples are selected

$$pot_i = \frac{\#rejections}{T}. \quad (22)$$

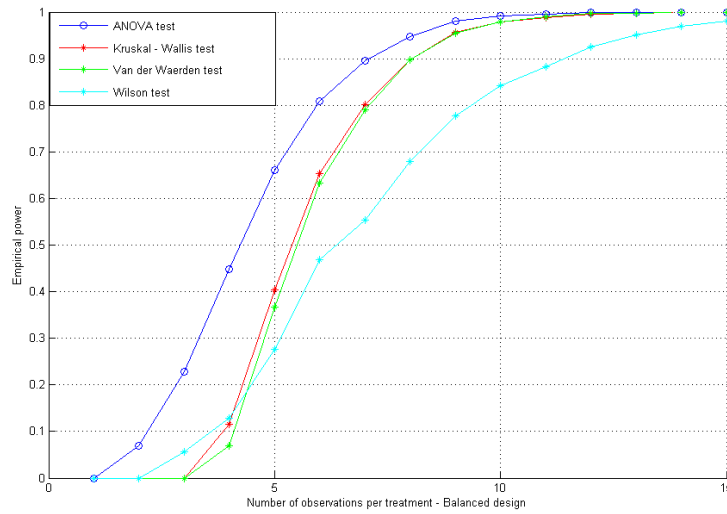
3. The number of rejections increases as the sample size increases from one iteration to another one, In this way the empirical power achieved is obtained.

We consider 4 treatments, and a level of significance $\alpha = 0.01$ in every numerical experimentation.

2.2 Parametric assumptions

$\sigma_1 = 0.6$	$\sigma_2 = 0.6$	$\sigma_3 = 0.6$	$\sigma_4 = 0.6$
$\mu_1 = 1.5$	$\mu_2 = 2.0$	$\mu_3 = 2.5$	$\mu_4 = 3.0$
$CV_1 = 40\%$	$CV_2 = 30\%$	$CV_3 = 24\%$	$CV_4 = 20\%$

Figure 1: Empirical power under parametric assumptions



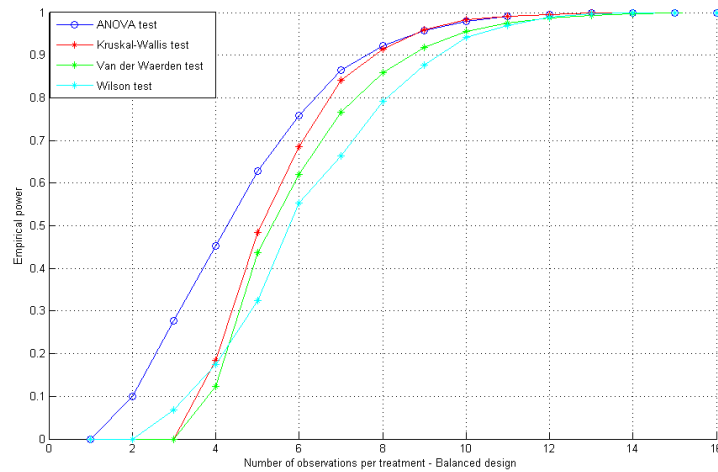
In this context we take samples from normal populations, the number of observations by treatments varies from 1 to 15. Herein the empirical power of the F-test is the best. The performance of Van der Waerden test and Kruskal-Wallis test is very similar, and the empirical power of Wilson test is very poor.

2.3 Normality with moderate departures from homoscedasticity

$\sigma_1 = 0.23$	$\sigma_2 = 0.4$	$\sigma_3 = 0.7$	$\sigma_4 = 0.9$
$\mu_1 = 1.5$	$\mu_2 = 2.0$	$\mu_3 = 2.5$	$\mu_4 = 3.0$
$CV_1 = 18.40\%$	$CV_2 = 20\%$	$CV_3 = 28\%$	$CV_4 = 30\%$

Under the above assumptions the empirical power of the F-test is the best when the number of observations per treatment is less or equal than eight, otherwise, the empirical power of F-test and Kruskal-Wallis test is very similar. The empirical power of Van der Waerden test is less than Kruskal-Wallis but better than the power of Wilson test.

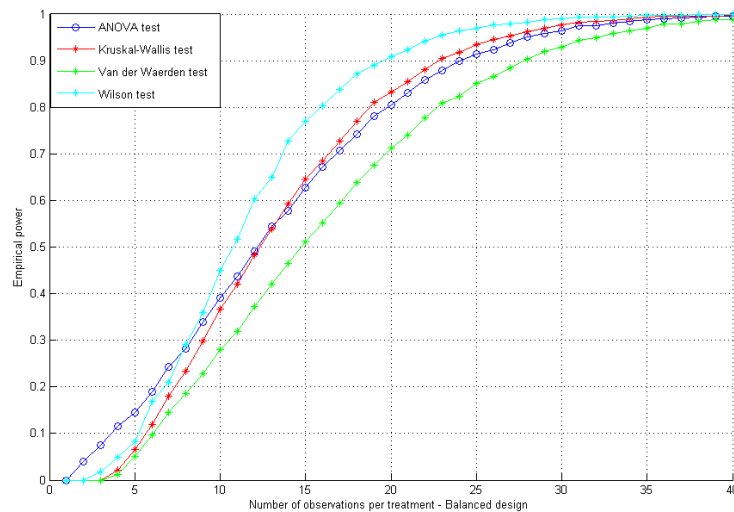
Figure 2: Normality with moderate departures from homoscedasticity



2.4 Normality with severe departures from homoscedasticity

$\sigma_1 = 0.46$	$\sigma_2 = 0.80$	$\sigma_3 = 1.25$	$\sigma_4 = 1.80$
$\mu_1 = 1.5$	$\mu_2 = 2.0$	$\mu_3 = 2.5$	$\mu_4 = 3.0$
$CV_1 = 30.67\%$	$CV_2 = 40\%$	$CV_3 = 50\%$	$CV_4 = 60\%$

Figure 3: Normality with severe departures from homoscedasticity

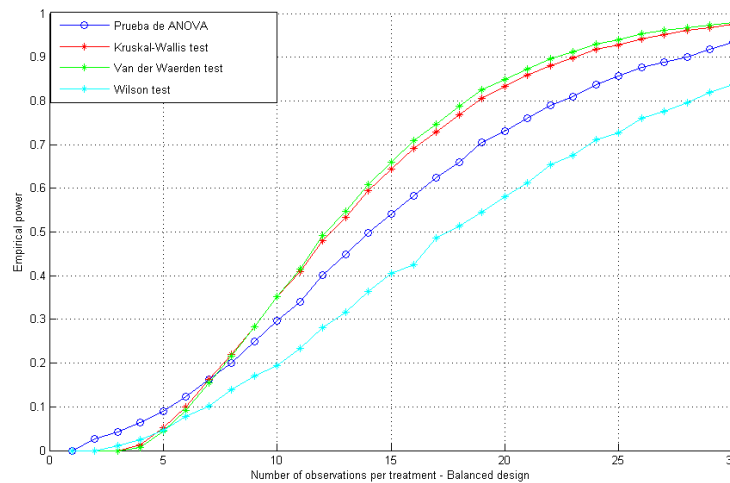


Herein the number of observation by treatments goes from 1 to 40. Under the above assumptions an acceptable value of the empirical power is obtained

by the Wilson test when the number of observations is greater or equal than twelve.

2.5 Random samples taken from chi-squared populations

Figure 4: Empirical power comparison



Herein we take samples from chi-squared populations with 2, 3, 4, 5 degrees of freedom respectively. The Van der Waerden test and Kruskal-Wallis reported better results, with a slight advantage of Vander Waerden test.

2.6 Conclusions and recommendations

1. A wider spectrum of distributions and ranges of values for location and scale parameters must be considered in order to obtain more conclusive results.
2. Classical results related with the power of ANOVA - F test when parametric assumptions are satisfied were corroborated, e.g., the ANOVA - F test detects a false null hypothesis when in fact it is false more frequently than the other ones.
3. When the assumption of normality is satisfied with moderate departures of equal variances, the empirical power of ANOVA-F test and Kruskal - Wallis test are very similar.
4. When the assumption of normality is satisfied with severe departures from equal variances, the empirical power of the Wilson test is the best

and it is acceptable when the number of observations per treatment is greater than eleven.

5. When random samples are taken from chi-squared populations the empirical power of Vander Waerden and Kruskal-Wallis test are the best, with slight advantage of Vander Waerden test.

References

- [1] S.O. Adams, E. Gayawan and M. K. Garba, Empirical Comparison of the Kruskal Wallis Statistics and its Parametric Counterpart. *Journal of Modern Mathematics and Statistics* 3 (2009) 38-42.
- [2] A. Wald and F. Waldfowitz, Statistical tests based on permutations of the observations. *The Annals of Mathematical Statistics* 15 (1944), no. 4, 358-372.
- [3] Sh. Searle. *Linear Models for Unbalanced Data*. John-Wiley & Sons, Inc. 1982.
- [4] W. Conover, *Practical Nonparametric Statistics*. John-Wiley. 1999.
- [5] W.H. Kruskal, W. A. Wallis, Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, Vol. 47, No. 260. (Dec., 1952), pp. 583-621.
- [6] Diseños experimentales en ciencias de la conducta: (Prueba no paramétrica de Wilson). Jimenez, C., Perez, R. *Anuario de Psicología*. Número 42-1989.
- [7] *Nonparametric Methods*. Hollander M., Wolf D. Wiley-Interscience. Second Edition.
- [8] J. Gibbons and S. Chakraborti, *Statistics: Nonparametric Statistical Inference*. Series of TEXBOOKS and MONOGRAPHS.
- [9] S. Hardeo and M. I. Ageeel. *The Analysis of Variance*. Birkhauser Editor, 2000.

Received: July 11, 2014