

## Empirical Comparison of the Kruskal Wallis Statistics and its Parametric Counterpart

<sup>1</sup>Samuel O. Adams, <sup>2</sup>Ezra Gayawan and <sup>3</sup>Mohammed K. Garba

<sup>1</sup>Department of Mathematics, Statistics and Computer Science, Abuja University, Abuja

<sup>2</sup>Department of Mathematical Sciences, Redeemer's University,  
Redemption, Ogun State, Nigeria

<sup>3</sup>School of Arts and Sciences, American University of Nigeria, Yola, Nigeria

---

**Abstract:** The nonparametric Kruskal-Wallis statistics (H-test) and its parametric counterpart, the one way ANOVA (F-test) are two powerful statistics commonly used to compare  $k$  ( $>2$ ) sample. The Monte Carlo approached was used in this study to compare the performances of the two statistics especially when the assumptions of normality and homogeneity of variance are violated. Data were generated from normal, exponential and Poisson distributions. It was discovered that when the samples are from normal and Poisson distributions, the F-test is more powerful than the H-test but the reverse is the case when they are from exponential distribution. However, when the sample size increases to say  $\geq 15$ , the two statistics perform equally well in terms of their power irrespective of the distribution from which the samples are drawn.

**Key words:** Classical F-test, H-test, normal population, exponential distribution, poisson distribution, level of significance

---

### INTRODUCTION

A common problem in many areas of the application of practical statistics is that of deciding whether several samples are to be regarded as coming from the same population. Often, the sample differ and the question is whether such difference is as a result of differences among the populations or that due to chance expected among random variables. Two statistical methods of dealing with this problem of comparing means of  $k$  ( $>2$ ) populations are the parametric one-way analysis of variance (F-test) and the non-parametric H-test proposed by Kruskal and Wallis (1952) and is therefore, commonly referred to as the Kruskal and Wallis (1952) statistics.

The classical F-test is more stringent in the use of assumptions than the H-test. Its most stringent assumption is that it requires all samples to be drawn from a normal population. Also, the measurement scale should be at least interval. Other assumptions are that all samples must be mutually independent and the homoscedastic of variance (Daniel, 1990). The H-test does not require the samples to come from normal population and the measurement scale can be at least ordinal. In many experimental situations, the normality assumption is not very realistic and sometimes, the samples may not be continuous though this is commonly assumed and the

classical approach of comparing the samples is often used (Laan and Verdooren, 1987). This undoubtedly, may yield a misleading result.

Many researchers have tried to look at the performances of these two statistics in terms of their strength and limitations. Laan and Verdooren (1987) are of the view that where the experimental is not sure whether the normality assumption is realistic for the classical F-test, a non-parametric test should be used as according to them, it is possible that the non-parametric test has a larger power than its classical counterpart in this situation. Rust and Fligner (1984) proposed a modification to the H-test in comparing several samples and this they contended requires fewer assumptions about the shape of the population. Other researchers, who have commented on the classical F-test and its non-parametric counterpart, include Conover (1999), Hettmansperger and McKean (1978), Puri and Sen (1971), Randles and Wolfe (1979), Buringer *et al.* (1980) and Hollander and Wolfe (1999).

In this study, we investigate the robustness of the classical F-test over the non-parametric H-test when the assumptions of normality and homogeneity of variance are violated under three different scenarios using the Monte Carlo approach. The power function for each scenario was computed at the significance level ( $\alpha$ ) of 0.01, 0.05 and 0.1.

### TEST STATISTICS FOR COMPARING K (>2) POPULATIONS

**The one way ANOVA (F-test):** Let a random sample of size  $n_i$  be drawn from population  $i$  ( $i = 1, 2, \dots, k$ ). The resulting random variables  $X_{ij}$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$ ) are assumed to be normally distributed  $N(\mu_i, \sigma^2)$ . The linear model of the form  $X_{ij} = \mu_i + E_{ij}$  is usually used with independent  $E_{ij}$ , which are  $N(0, \sigma^2)$  (Laan and Verdooren, 1987; Oyejola, 2003). The observations can thus be classified according to a classification A with classes  $A_1, A_2, \dots, A_k$ , where the  $n_i$  observations  $X_{ij}$  ( $i = 1, 2, \dots, k$ ) belong to class  $A_i$ . For testing the null hypothesis  $H_0: \mu^1 = \mu^2 = \dots = \mu_k$  against its alternate  $H_1$ : at least a pair of  $\mu$ 's is unequal, the test statistics  $F = MSA/MSE$  is used, where  $MSA = SSA/(k-1)$  and  $SSA$ , the sum of squares of A equals:

$$\sum_{i=1}^k \left( \sum_{j=1}^{n_i} X_{ij} \right)^2 / n_i - \left( \sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij} \right)^2 / \sum_{i=1}^k n_i$$

and

$$MSE = SSE / (\sum n_i - k)$$

and SSE, the sum of square error, equals

$$\sum_{i=1}^k \sum_{j=1}^{n_i} X_{ij}^2 - \sum_{i=1}^k \left( \sum_{j=1}^{n_i} X_{ij} \right)^2 / n_i$$

Under  $H_0$ ,  $F \sim F(a, e)$  with  $a = k-1$  and

$$e = \sum_{i=1}^k n_i - k$$

degrees of freedom.

**The Kruskal-Wallis test (H-test):** Let a random sample of size  $n_i$  be drawn from population  $i$  ( $i = 1, 2, \dots, k$ ). The resulting random variables  $X_{ij}$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$ ) are assumed to come from populations with continuous distribution function  $F_i$ . The hypothesis to test is of the form  $H_0: F_1(x) = F_2(x) = \dots = F_k(x)$ . In applying the H-test, all the

$$N = \sum_{i=1}^k n_i$$

observations will be ranked in increasing order of magnitude with ranks  $1, 2, \dots, N$ . The H statistic is

$$H = \frac{12}{N(N+1)} \sum_{i=1}^k n_i (\bar{R}_i - \bar{R}_{..})^2$$

where,  $R_{ij}$  is the rank of  $X_{ij}$  ( $i = 1, 2, \dots, k; j = 1, 2, \dots, n_i$ ) and with the usual notation

$$R_i = \sum_{j=1}^{n_i} R_{ij}, \quad \bar{R}_i = R_i / n_i$$

etc. Details of the critical values of H for few samples with each having  $<5$  observations can be found by Daniel (1990) and Iman *et al.* (1975). When  $n_i \rightarrow \infty$  H has asymptotically a Chi-square ( $\chi^2$ ) distribution with  $k-1$  degrees of freedom ( $\chi^2(k-1)$ ).

**Generation of random sample:** The random samples used for the aim of this study were generated from the normal, exponential and the Poisson distributions using the SPSS package. Four independent samples were generated from the three distributions to make up three scenarios for the study. Sample size of 2, 5, 10, 15, 20, 25 and 30 were used for the three scenarios expect that of normal where we included sample size 8. Each sample size was replicated 100 to allow us compute the power function.

In the first Scenario, the normality and homogeneity of variance assumptions were withheld. The means used in generating the four independent samples were 10.0, 13.0, 15.0 and 18.0, respectively with a common standard deviation of 5. For the second scenario, where the exponential distribution was used, the scale parameters were taken to be 0.2, 0.5, 2.0 and 5.0, respectively. By this, both the normality and homogeneity of variance assumptions were violated. In the third scenario, the means for the Poisson distribution were also taken to be respectively 10.0, 13.0, 15.0 and 18.0. With this, not only that the two assumptions above were violated, the variables so generated were not continuous.

### RESULTS AND DISCUSSION

The classical F statistic and its non-parametric H counterpart were applied to all the data generated in order to determine if the samples can be said to have come from the same populations. The number of rejections out of 100 replications for the three scenarios at the level of significance ( $\alpha$ ) of 0.01, 0.05 and 0.1 are shown in Table 1-3 while, the power functions computed for each of the scenarios are presented in Fig. 1-9.

The results in Table 1 indicate that under the normal distribution, the F-test rejects more often than the H-test and as the level of significance of the test increases, the number of rejections increases consistently for the two tests. For  $n = 15$ , the two tests perform equally in rejecting the null hypothesis except for when  $\alpha = 0.01$ . For the results of the exponential distribution in Table 2, the F-test does better than the H in its number of rejection when  $n = 2$ . As  $n = 5$ , the H-test becomes more powerful though the two test soon become equally powerful as  $n = 10$ . This is consistent for all the  $\alpha$  levels used.

Table 1: Number of rejections out of 100 replications under the normal distribution

Sample size	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
	F-test	H-test	F-test	H-test	F-test	H-test
02	2	0	11	0	22	1
05	17	11	36	35	54	49
08	43	32	73	64	80	76
10	57	45	78	76	90	90
15	92	87	98	98	99	99
20	100	100	100	100	100	100
25	100	100	100	100	100	100
30	100	100	100	100	100	100

Table 2: Number of rejections out of 100 replications under the exponential distribution

Sample size	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
	F-test	H-test	F-test	H-test	F-test	H-test
2	10	0	23	0	34	9
5	33	53	70	89	86	97
10	98	99	100	100	100	100
15	100	100	100	100	100	100
20	100	100	100	100	100	100
25	100	100	100	100	100	100
30	100	100	100	100	100	100

Table 3: Number of rejections out of 100 replications under the Poisson distribution

Sample size	$\alpha = 0.01$		$\alpha = 0.05$		$\alpha = 0.1$	
	F-test	H-test	F-test	H-test	F-test	H-test
2	3	0	17	0	37	7
5	44	21	73	66	81	81
10	88	87	99	96	99	99
15	100	100	100	100	100	100
20	100	100	100	100	100	100
25	100	100	100	100	100	100
30	100	100	100	100	100	100

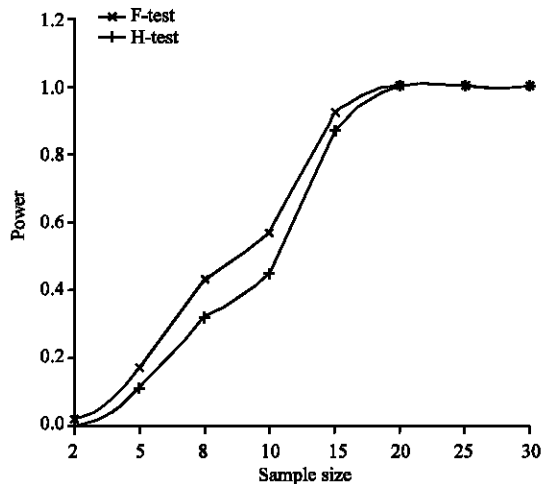


Fig. 1: Power curve at  $\alpha = 0.01$  using normal distribution

In the case of the Poisson distribution, all through the levels of significance, the F-test is more powerful than the H-test. But as  $n = 10$ , the two test statistics can be said to have equal power.

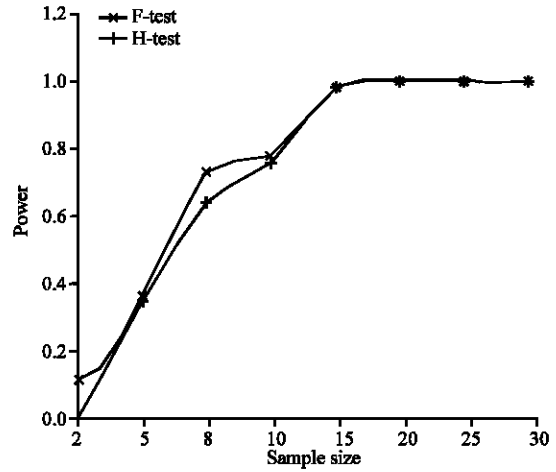


Fig. 2: Power curve at  $\alpha = 0.05$  using normal distribution

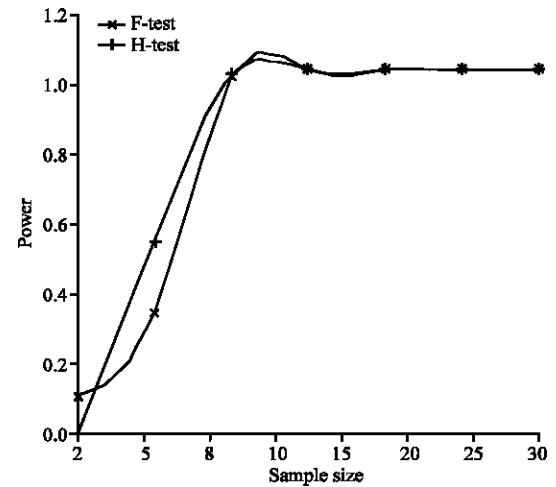


Fig. 3: Power curve at  $\alpha = 0.1$  using normal distribution

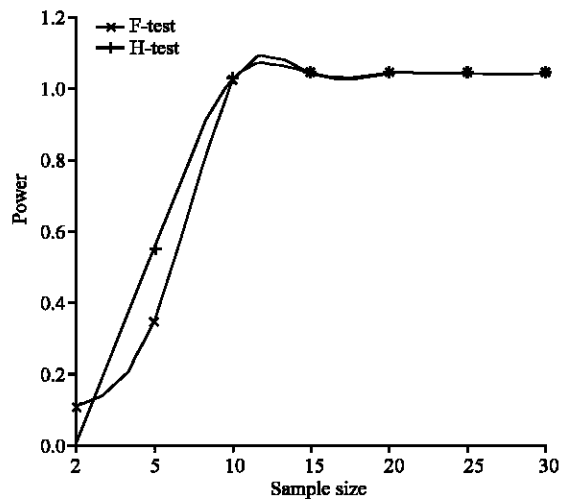


Fig. 4: Power curve at  $\alpha = 0.01$  using exponential distribution

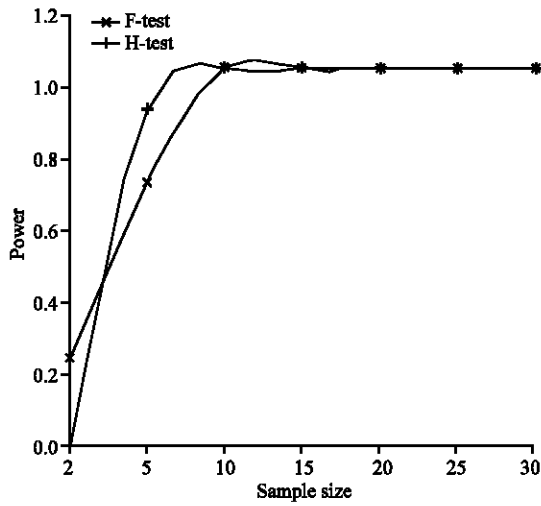


Fig. 5: Power curve at  $\alpha = 0.05$  using exponential distribution

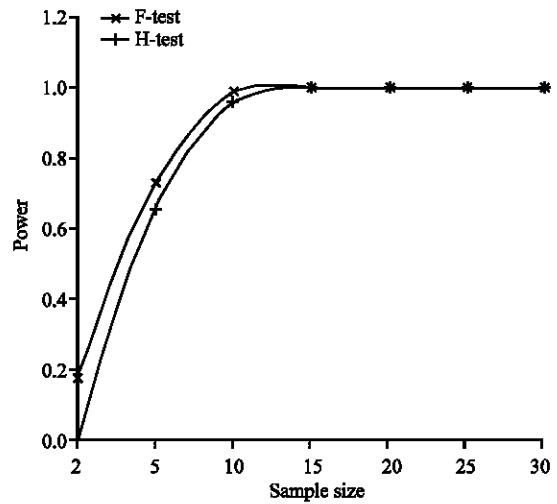


Fig. 8: Power curve at  $\alpha = 0.05$  using poisson distribution

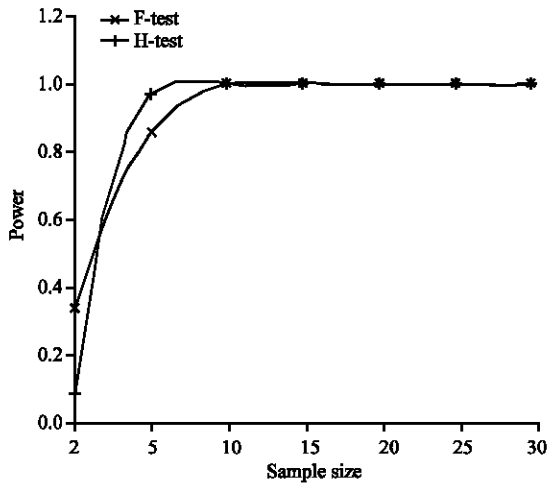


Fig. 6: Power curve at  $\alpha = 0.1$  using exponential distribution

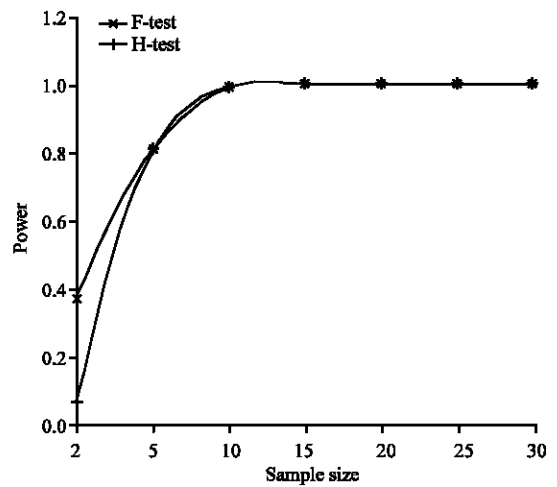


Fig. 9: Power curve at  $\alpha = 0.1$  using POISSON distribution

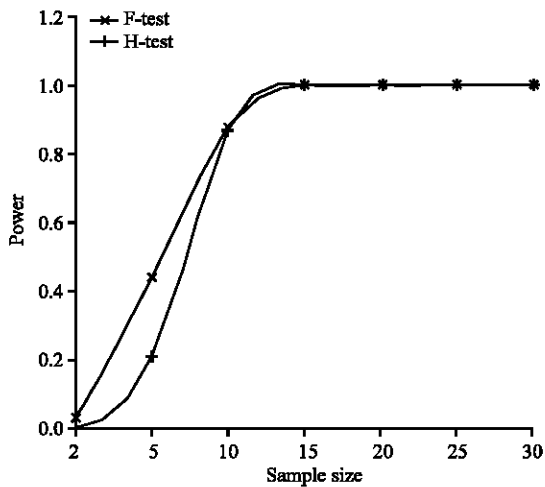


Fig. 7: Power curve at  $\alpha = 0.01$  using poisson distribution

The power curves in the Fig. 1-9 indicate that the parametric F-test is more powerful than the non-parametric H-test when  $n = 10$ . However, in the case of the exponential distribution in Fig. 4-6, the non parametric H-test is more powerful at all the levels of significance considered.

## CONCLUSION

The Monte Carlo study has revealed that when there are a number of samples to be compared whether, they all come from the same population, the parametric F-test should be used when it is known that the samples are from normal distribution. When they are known to have come from the Poisson distribution, which is of the discrete type and the experimental wants to use either of

the tests for comparison, the F-test should be preferred. However, when the samples are continuous but the experimental is not sure whether the assumption of normality and homogeneity of variance are valid, the non-parametric H-test is to be considered for use. But where the sample size is not small say  $n = 15$ , any of the two test will perform well for the purpose of such comparison irrespective of the distribution.

## REFERENCES

- Buringer, H., H. Martin and K.H. Schriever, 1980. Nonparametric Sequential Selection Procedures, Birkhauser. 1st Edn. Boston, pp: 216-297. ISBN: 978-3764330217.
- Conover, W.J., 1999. Practical Nonparametric Statistics. 3rd Edn. John Wiley and Sons Inc., pp: 269-427. ISBN: 978-0-471-16068-7.
- Daniel, W.W., 1990. Applied Nonparametric Statistics. 2nd Edn. Wadsworth. Inc. Boston, pp: 221-255. ISBN: 9780 534919764.
- Hettmansperger, T.P. and J.W. McKean, 1978. Statistical Inference Based on Ranks. Psychometrical, 43 (1): 69-79. DOI: 10.1007/BF022 94090.
- Hollander, M. and D.A. Wolfe, 1999. Nonparametric Statistical Methods. 2nd Edn. Hoboken. NJ: John Wiley and Sons. Inc., pp: 189-269. ISBN: 978-0-471-32986-2.
- Iman, R.L., D. Quade and D.A. Alexander, 1975. Exact probability level for the Kruskal Wallis test in selected tables in mathematical statistics. Bull. Am. Mathe. Soc., 3: 329-384.
- Kruskal, W.H. and W.A. Wallis, 1952. Use of ranks in one-criterion variance analysis. J. Am. Statist. Assoc., 47: 583-621. <http://www.jstor.org/stable/2280779>.
- Laan, P. Van. Der. and L.R. Verdooren, 1987. Classical analysis of variance methods and nonparametric counterparts. Biometrical. J., 29: 635-665. <http://www3.interscience.wiley.com/journal/114076640/issue>.
- Oyejola, B.A., 2003. Design and Analysis of Experiments for Biology and Agriculture Students. 1st Edn. Olad Publishers, Ilorin, pp: 29-72. ISBN: 978-8044-24-7.
- Puri, M.L. and P.K. Sen, 1971. Nonparametric Methods in Multivariate Analysis. 1st Edn. John Wiley and Sons. Inc., pp: 263-387. ISBN: 978-0471702405.
- Randles, R.H. and D.A. Wolfe, 1979. Introduction to the Theory of Nonparametric Statistics. 1st Edn. John Wiley and Sons Inc. New York, pp: 196-245. ISBN: 978-0894 645433.
- Rust, S.W. and M.A. Fligner, 1984. A modification of the Kruskal Wallis statistic for the generalized Behrens\_fisher problem. Commun. Statist. Theory Methods, 13: 2013-2027.