International Journal of Mathematics and Statistics Invention (IJMSI)
E-ISSN: 2321 – 4767 P-ISSN: 2321 - 4759
www.ijmsi.org Volume 6 Issue 3 || March, 2018 || PP-32-44

A New Transformed Test forAnalysis of Variance forSkewed DistributionswithaUnivariate Goodness of Fit

Khairul Islam and TanweerShapla
Department of Mathematics and Statistics, Eastern Michigan University, Ypsilanti, MI 48197, USA
Corresponding Author: Khairul Islam

ABSTRACT: *Analysis of variance (ANOVA) is one of the most popular statistical techniques for comparing different groups or treatments with respect to their means. One of the important assumptions for the validity of ANOVA F test is the assumption of normality of the groups being compared. However, many real-life data do not follow normal distributions. In the violation of normality, the non-parametric Kruskal-Wallis test is often preferable. In this paper, we propose a new transformed test for one way ANOVA for skewed distributions. The performance of the new test is compared with the standard F and the non-parametric analogue of ANOVA by examples and simulations. Our results suggest that the new transformed test is appropriate for estimating the level of significance and is more powerful than standard F test and the non-parametric Kruskal-Wallis test for skewed distributions.*
KEYWORDS**:** *ANOVA F test, Transformed test, Kruskal-Wallis test, Level of significance, Power of the test, Simulation*.

DATE OF SUBMISSION: 23-04-2018 DATE OF ACCEPTANCE: 08-05-2018

I INTRODUCTION

Let us consider the test of equality of k population means given samples $\{X_{ij}: i = 1, 2, \ldots, n_j; j = 1, 2, \ldots, k\}$from jth population with meanμ_j, variance σ^2 and distribution $F\{\sigma^{-1}(x - \mu_j)\}$. We want to test

$$H_0: \mu_1 = \mu_2 = \cdots = \mu_k$$

versus

$$H_1: \mu_j \neq \mu_{j'} \text{ for some } j \neq j'$$

Under the assumption that the samples come from normal distributions with common variance, the test statistic to test $H_0$is given by

$$F = \frac{SS_{Treat}\ /(k-1)}{SS_{Error}\ /(n-k)} \sim F(k-1, n-k) \tag{1}$$

where,$n = \sum_{j=1}^{k} n_j$, $SS_{Treat} = \sum_{j=1}^{k} n_j (\bar{x}_j - \bar{x})^2$ and $SS_{Error} = \sum_{j=1}^{k} \sum_{i=1}^{n_j} (x_{ij} - \bar{x}_j)^2$.

For a level α test, reject H_0 if $F > F_{k-1,n-k;1-\alpha}$or $P(F_{k-1,n-k} > F) < \alpha$, where $F_{k-1,n-k;1-\alpha}$ is the $(1-\alpha)th$ percentile of F distribution with numerator degrees of freedom $(k-1)$ and the denominator degrees of freedom $(n-k)$.

In real-life data, however, the assumption of normality is often invalid. As such, the usual ANOVA F test fails and we proceed with the widely used non-parametric Kruskal–Wallis test[1]. This test compares$k$ sample means by using ranks of the combined dataset from $k$samples. When there are no ties, the test statistic is given by

$$K = K^* = \frac{12}{n(n+1)} \times \sum_{j=1}^{k} \frac{R_j^2}{n_j} - 3(n+1) \sim \chi_{k-1}^2 \tag{2}$$

whereR_j is the sum of ranks of$j$th sample in the combined set of n observations from k samples.

If there are ties, the test statistic is given by

$$K = \frac{K^*}{1 - \frac{\sum_{i=1}^{g}(t_i^3 - t_i)}{n^3 - n}} \sim \chi_{k-1}^2 \tag{3}$$

where$g$ is the number of groups with tied values and t_i is the number of observations with tie in ith group, $i = 1,2, \ldots, g$.For a level α test, reject H_0 if $K > \chi_{k-1,1-\alpha}^2$or $P(\chi_{k-1}^2 > K) < \alpha$.

II THE NEW TRANSFORMED ANOVA F-TEST

If k groups deviate from normality, an alternative to Kruskal-Wallis test, we propose a new transformed test with the Box-Cox type power transformation [2]. The idea is to estimate the transformation parameter by a univariate normal goodness-of-fit (GOF). We provide an algorithm to perform the new test, and compare the performance of the new test with the traditional ANOVA F test and the non-parametric Kruskal-Wallis test with examples and a simulation study.

Given k samples $\{X_{ij} : i = 1, 2, \dots, n_j; j = 1, 2, \dots, k\}$ and a scalarλ, the Box-Cox power transformation [2]to jth sample$\boldsymbol{X}_j = \{X_{1j}, X_{2j}, \dots, X_{n_j j}\}$is defined by

$$X_{ij}(\lambda) = \begin{cases} \left(X_{ij}^{\lambda} - 1\right)/\lambda, & \text{if } \lambda \neq 0 \\ \log\left(X_{ij}\right), & \text{if } \lambda = 0 \end{cases} \quad ; i = 1, 2, \dots, n_j; j = 1, 2, \dots, k \tag{4}$$

As expected with Box-Cox transformation, the transformed$X_{ij}(\lambda), j = 1, 2, \dots, k$,follows a normal distribution, say,$X_{ij}(\lambda) \sim N(\mu_j(\lambda), \sigma_j^2(\lambda))$.Then,$Z_{ij}(\lambda) = \frac{X_{ij}(\lambda) - \mu_j(\lambda)}{\sigma_j(\lambda)}$is independent N(0,1). Therefore, $\boldsymbol{Z}(\lambda) = \left(\boldsymbol{Z}_1(\lambda), \boldsymbol{Z}_2(\lambda), \dots, \boldsymbol{Z}_k(\lambda)\right) = \left(Z_1(\lambda), Z_2(\lambda), \dots, Z_n(\lambda)\right)$ represents a sample of size $n = \sum_{j=1}^{k} n_j$ from N(0,1) distribution.

In order to estimate λ, we use the fact that the $\boldsymbol{Z}(\lambda)$ is as close as possible to a N(0,1) distribution by a goodness of fit criteria.Viewing this problem as a goodness-of-fit to a normal distribution, we test the hypothesis:

H_0: $Z_1(\lambda), Z_2(\lambda), \dots, Z_n(\lambda)$is coming from a N(0,1) distribution, against
H_1: $Z_1(\lambda), Z_2(\lambda), \dots, Z_n(\lambda)$is not a N(0,1) distribution.

Following Shapiro and Wilk [3], we use the test statistic $W_{\boldsymbol{Z}}(\lambda)$ to test H_0, which is given by

$$W_{\boldsymbol{Z}}(\lambda) = \frac{\left[\sum_{i=1}^{n} a_i Z_{(i)}(\lambda)\right]^2}{\sum_{i=1}^{n} (Z_i(\lambda) - \bar{Z}(\lambda))^2} \tag{5}$$

where

$Z_{(i)}(\lambda), i = 1, \dots, n$represents the ith order statistic of the sample $\boldsymbol{Z}(\lambda)$,
$\bar{Z}(\lambda) = (\sum_{i=1}^{n} Z_i(\lambda))/n$,
$(a_1, \dots, a_n) = \frac{m^T V^{-1}}{\left(m^T V^{-1} V^{-1} m\right)^{1/2}}$,
$m = (m_1, \dots, m_n)^T$,
$m_i = E\left(Z_{(i)}(\lambda)\right), i = 1, \dots, n$, is the expected value of the ith order statistic $Z_{(i)}(\lambda)$,
$V = \left(v_{i,i'}\right)$is the variance-covariance matrix of order $n \times n$, and
$v_{i,i'} = Cov\left(Z_{(i)}(\lambda), Z_{(i')}(\lambda)\right), i, i' = 1, \dots n$,is the covariance between ith and i'th order statistics.

While the value of $W_{\boldsymbol{Z}}(\lambda)$ lies between zero and one, the small value of $W_{\boldsymbol{Z}}(\lambda)$ leads to the rejection of normality, whereas a value close to one indicates normality. In other words, given a level of significance α one may reject the null hypothesis if p-value $p(\lambda) = P(W \leq w_{\boldsymbol{Z}}(\lambda)) \leq \alpha$ and accept otherwise. We propose to estimate λ by observing the maximum p-value associated with $W_Z(\lambda)$ over all possible values ofλ to achieve the desired normality of the transformed data. In other words, the new estimate $\hat{\lambda}_n$using the goodness-of-fit to N(0,1) distribution satisfies the equation

$$p(\hat{\lambda}_n) = \max_{\lambda \in [a,b]} P(W \leq w_{\boldsymbol{Z}}(\lambda)) \tag{6}$$

Once $\hat{\lambda}_n$is obtained, we re-express the original samples and apply an ANOVAF-test to the transformed data to compare the group means. This test is termed as Transformed F test (Trans F).

In this article, we employed the software R [4] in all examples and simulation to obtain the optimum$\hat{\lambda}_n$. The search for $\hat{\lambda}_n$is made over the interval [-1,1] with an increment of 0.1 written hereafter as $\lambda \in \{-1: 0.1: 1\}$.

Below is an algorithm for the estimate $\hat{\lambda}_n$ and the transformed test using$\hat{\lambda}_n$.

Given X_j and a fixed λ:

1) Obtain the transformation $X_{ij}(\lambda)$ using equation (1).
2) Find$Z_{ij}(\lambda) = \frac{X_{ij}(\lambda) - \overline{X_j}(\lambda)}{S_{X_j}(\lambda)}$;$i = 1, 2, \dots, n_j; j = 1, 2, \dots, k$, where$S_{\boldsymbol{X}_j}(\lambda) = \sqrt{\sum_{i=1}^{n_j} \left(X_{ij}(\lambda) - \overline{X}_j(\lambda)\right)^2 / n_j}$
3) Combine k samples together to form $\boldsymbol{Z}(\lambda) = \left(Z_1(\lambda), Z_2(\lambda), \dots, Z_n(\lambda)\right)$, where$n = \sum_{j=1}^{k} n_j$.

4) Compare $\boldsymbol{Z}(\lambda)$ with the N(0,1) distribution using the Shapiro-Wilk goodness-of-fit $W_{\boldsymbol{Z}}(\lambda)$and find the p-value.

5) Repeat steps (1) through (4) for all $\lambda \in \{-1: 0.1: 1\}$.
6) Select the maximum p-value among all p-values from steps (1) through (5).
7) Identify the $\hat{\lambda}_n$corresponding to the maximum p-value in step (6).
8) Obtain $\boldsymbol{X}_j\left(\hat{\lambda}_n\right), j = 1, 2, \ldots, k$.
9) Perform usual F-test on the basis of transformed data in step (8) and decide on the acceptance and rejection of the null hypothesis comparing observed value of F with critical value of$F_{k-1,n-k;1-\alpha}$distribution for a given level of significance α.

The theoretical aspects of the Box-Cox transformed data analysis described above have been reported in literature. For examples, [5] and [6] investigated the asymptotic properties of the parameter estimates; Bickel and Doksum[7] critically examined the behavior of the asymptotic variances of the parameter estimates for regression and analysis of variance situations; Chen and Loh[8] and Chen [9] proved that the Box-Cox transformed t-test is typically more efficient asymptotically than the t-test without transformation. The use of transformed t-test is also justified by Chen and Islam [10] by fitting a t distribution to transformed data. In this paper, we empirically assess the performance of the transformed ANOVA (**Trans F**) as compared to traditional ANOVA F and non-parametric Kruskal-Wallis (**Kruskal**) test using a Monte Carlo simulation in terms of the estimated size of the test and power of the test, and a real-life example.

III SIMULATION AND RESULT DISCUSSION

In this section, we carry out a simulation study to compare the finite sample performance of the three ANOVA tests described in this article. All simulations are performed by using the statistical software R. For the transformed F-test (Trans F), we estimate $\hat{\lambda}_n$ from values of λ between -1 and 1 with an increment of 0.1, denoted symbolically as: $\lambda \in \{-1: 0.1: 1\}$. Thesamples$\{X_{ij}: i = 1, 2, \ldots, n_j; j = 1, 2, \ldots, k\}$, with $k = 3$chosen arbitrarily,are simulated from $G(\theta_1, \theta_2)$ population where θ_1 is the shape parameter and θ_2 is the scale parameter.

Note that the skewness of $G(\theta_1, \theta_2)$ distribution is $\gamma_1 = 2/\sqrt{\theta_1}$. In simulations, we choose different values of the parameters$\theta_1$and $\theta_2$to allow varying levels of skewness at 0.5, 1, 2 and 4 chosen arbitrarily keeping means of all simulated distributions fixed at 1 under the null model. The mean difference (Δ) of 0.15, 0.30, 0.45 and 0.60 are considered under the alternative models to ensure a testing power away from 0 and 1 for the purpose of the comparisons.

In all simulations, the Monte Carlo size is 2,000. The power of a test is estimated from the proportion of rejection of null hypothesis under alternative models over a Monte Carlo simulation of size 2,000 at 5% level of significance. In a similar manner, the level of significance is estimated from the proportion of the rejection of the null hypothesis over a Monte Carlo simulation of size 2,000 at 5% level of significance when the null hypothesis is true. Table 1 provides the values of the parameter $\theta_1$and $\theta_2$used in the simulation of samples to allow varying values of the skewness with fixed mean.

TABLE 1: Values of θ_1, θ_2 and γ_1 used in simulations of samples under four null models (M1-M4)

Models	θ_1	θ_2	γ_1	mean
M1	16	0.0625	0.5	1
M2	4	0.25	1	1
M3	1	1	2	1
M4	0.25	4	4	1

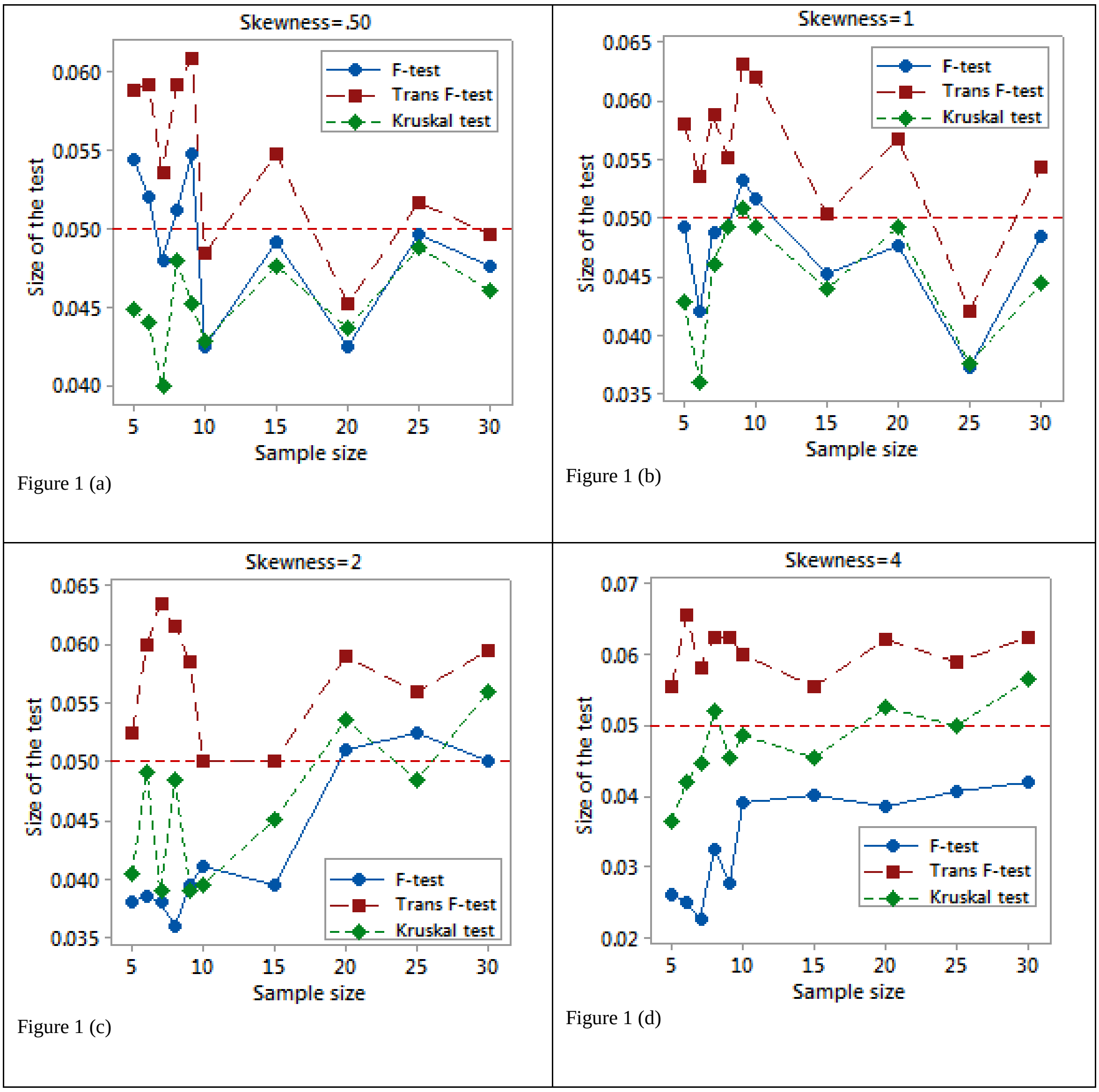
Tables 2 provides estimated size of the test (i.e., the estimated level of significance) from the simulated samples for varying values of skewness (γ_1) and sample size (n) under the four null models, along with the estimated mean and the standard deviation of the estimate$\hat{\lambda}_n$.

Fig. 1 shows the estimated size reported in Table 2 graphically so as to better understand the performance of the three underlying tests in controlling the size of the test at $\alpha = 0.05$.

TABLE 2: Estimated α at 5% significance level for F, transformed F and Kruskal-Wallis tests.

γ_1	n	F	Trans F	Kruskal	mean ($\hat{\lambda}_n$)	sd($\hat{\lambda}_n$)
	5	0.054	0.059	0.045	0.118	0.838
	6	0.052	0.059	0.044	0.135	0.807
	7	0.048	0.054	0.040	0.163	0.756
	8	0.051	0.059	0.048	0.207	0.726
0.5	9	0.055	0.061	0.045	0.226	0.690
	10	0.042	0.048	0.043	0.220	0.667
	15	0.049	0.055	0.048	0.264	0.560
	20	0.042	0.045	0.044	0.301	0.486
	25	0.050	0.052	0.049	0.305	0.419
	30	0.048	0.050	0.046	0.313	0.381
	5	0.049	0.058	0.043	0.196	0.692
	6	0.042	0.054	0.036	0.251	0.610
	7	0.049	0.059	0.046	0.257	0.556
	8	0.049	0.055	0.049	0.271	0.495
1	9	0.053	0.063	0.051	0.275	0.458
	10	0.052	0.062	0.049	0.293	0.424
	15	0.045	0.050	0.044	0.313	0.311
	20	0.048	0.057	0.049	0.317	0.257
	25	0.037	0.042	0.038	0.309	0.216
	30	0.048	0.054	0.044	0.307	0.194
	5	0.038	0.053	0.041	0.237	0.438
	6	0.039	0.060	0.049	0.254	0.369
	7	0.038	0.064	0.039	0.271	0.302
	8	0.036	0.062	0.049	0.263	0.267
2	9	0.040	0.059	0.039	0.270	0.243
	10	0.041	0.050	0.040	0.268	0.206
	15	0.040	0.050	0.045	0.273	0.142
	20	0.051	0.059	0.054	0.274	0.114
	25	0.053	0.056	0.049	0.271	0.098
	30	0.050	0.060	0.056	0.271	0.088
	5	0.026	0.056	0.037	0.175	0.206
	6	0.025	0.066	0.042	0.172	0.152
	7	0.023	0.058	0.045	0.176	0.129
	8	0.033	0.063	0.052	0.169	0.102
4	9	0.028	0.063	0.046	0.167	0.094
	10	0.039	0.060	0.049	0.167	0.083
	15	0.040	0.056	0.046	0.165	0.063
	20	0.039	0.062	0.053	0.166	0.055
	25	0.041	0.059	0.050	0.164	0.051
	30	0.042	0.063	0.057	0.166	0.049

FIGURE 1: Estimated size at 5% significance level for F, transformed F and Kruskal-Wallis tests



From the results of Table 2 and the performance of three tests in Fig. 1(a)-1(d), it is very clear that for a skewness of 0.5 with small sample size ($n \leq 10$),all tests have estimated size of the test have within 0.015 from the desired level of 0.05. For sample size $n > 10$, the estimated size of the test scattered around within 0.005 from the desired level 0.05, and the performance of all tests are comparable. Indeed, all three tests have acceptable and comparable control over the estimated size of the test at 5% level of significance for skewness$<$ 4. When skewness$\geq$ 4, the Kruskal-Wallis test seems to have the best control, and the transformed test has the second best control over the estimated size, with F test showing the underestimation of estimated test size when $n \leq 10$.

Overall, all tests are within 0.015% in either side of the desired nominal level of 5% with a few exceptions for F test when$n \leq 10$.

Now, let us have a look at the power of the three underlying tests reported in Tables 3-6 for varying values of skewness at 0.5, 1, 2 and 4, and mean difference at 0.15, 3, 4.5 and 6, chosen arbitrarily so that a meaningful comparison of performance can be made. For the better understanding of the power of the three tests, the estimated power reported in Table 3 (skewness=0.5) and Table 6 (skewness=4) are shown in Fig. 2 and 3, respectively.

TABLE 3: Estimated power of F, transformed F and Kruskal-Wallis tests when skewness γ_1=0.5

Δ	n	F	Trans F	Kruskal	mean ($\hat{\lambda}_n$)	sd($\hat{\lambda}_n$)
	5	0.138	0.148	0.120	0.101	0.854
	6	0.154	0.160	0.132	0.151	0.808
	7	0.184	0.188	0.160	0.150	0.778
	8	0.198	0.210	0.185	0.172	0.738
0.15	9	0.226	0.243	0.211	0.166	0.721
	10	0.240	0.256	0.223	0.215	0.677
	15	0.366	0.388	0.347	0.250	0.574
	20	0.471	0.496	0.465	0.254	0.507
	25	0.572	0.601	0.567	0.278	0.441
	30	0.633	0.667	0.634	0.277	0.393
	5	0.404	0.398	0.362	0.063	0.860
	6	0.481	0.484	0.430	0.126	0.814
	7	0.568	0.580	0.550	0.138	0.782
	8	0.645	0.653	0.614	0.127	0.764
0.3	9	0.683	0.702	0.664	0.153	0.735
	10	0.761	0.776	0.750	0.158	0.701
	15	0.924	0.935	0.918	0.243	0.575
	20	0.973	0.979	0.975	0.232	0.504
	25	0.994	0.996	0.995	0.249	0.454
	30	0.998	0.999	0.999	0.243	0.411
	5	0.730	0.701	0.684	0.069	0.858
	6	0.843	0.838	0.806	0.078	0.833
	7	0.909	0.911	0.896	0.121	0.796
	8	0.945	0.947	0.938	0.163	0.762
0.45	9	0.968	0.971	0.964	0.152	0.733
	10	0.983	0.985	0.981	0.148	0.697
	15	0.999	1.000	1.000	0.191	0.611
	20	1.000	1.000	1.000	0.210	0.520
	25	1.000	1.000	1.000	0.222	0.478
	30	1.000	1.000	1.000	0.245	0.425
	5	0.938	0.920	0.922	0.069	0.865
	6	0.973	0.967	0.965	0.081	0.833
	7	0.994	0.992	0.992	0.099	0.798
	8	0.997	0.998	0.996	0.108	0.768
0.6	9	1.000	1.000	0.999	0.128	0.750
	10	0.999	1.000	1.000	0.138	0.715
	15	1.000	1.000	1.000	0.175	0.607
	20	1.000	1.000	1.000	0.200	0.535
	25	1.000	1.000	1.000	0.204	0.495

30	1.000	1.000	1.000	0.207	0.438

TABLE 4: Estimated power of F, transformed F and Kruskal-Wallis tests when skewness γ_1=1

Δ	n	F	Trans F	Kruskal	mean ($\hat{\lambda}_n$)	sd($\hat{\lambda}_n$)
	5	0.060	0.070	0.059	0.202	0.693
	6	0.075	0.087	0.066	0.208	0.636
	7	0.076	0.092	0.072	0.210	0.580
	8	0.082	0.098	0.085	0.245	0.521
0.15	9	0.078	0.102	0.083	0.239	0.478
	10	0.092	0.106	0.090	0.256	0.436
	15	0.115	0.142	0.120	0.265	0.326
	20	0.157	0.194	0.168	0.275	0.268
	25	0.189	0.226	0.200	0.278	0.225
	30	0.188	0.238	0.215	0.273	0.198
	5	0.138	0.151	0.136	0.151	0.718
	6	0.148	0.168	0.136	0.170	0.647
	7	0.185	0.208	0.180	0.209	0.589
	8	0.202	0.236	0.206	0.199	0.552
0.3	9	0.222	0.249	0.218	0.230	0.499
	10	0.244	0.283	0.253	0.235	0.469
	15	0.353	0.428	0.386	0.251	0.343
	20	0.468	0.559	0.514	0.243	0.273
	25	0.575	0.683	0.634	0.247	0.237
	30	0.646	0.750	0.715	0.249	0.210
	5	0.257	0.258	0.228	0.131	0.729
	6	0.305	0.320	0.282	0.157	0.674
	7	0.340	0.370	0.339	0.155	0.611
	8	0.413	0.446	0.411	0.192	0.565
0.45	9	0.457	0.506	0.471	0.192	0.514
	10	0.496	0.556	0.526	0.214	0.481
	15	0.704	0.784	0.753	0.220	0.355
	20	0.821	0.896	0.878	0.226	0.283
	25	0.909	0.960	0.940	0.209	0.252
	30	0.955	0.986	0.979	0.224	0.215
	5	0.402	0.377	0.374	0.111	0.736
	6	0.496	0.492	0.462	0.141	0.667
	7	0.567	0.588	0.570	0.127	0.620
	8	0.653	0.689	0.667	0.154	0.567
0.6	9	0.723	0.772	0.750	0.155	0.523
	10	0.763	0.813	0.794	0.185	0.489
	15	0.917	0.959	0.948	0.192	0.360
	20	0.976	0.990	0.987	0.199	0.299
	25	0.992	0.999	0.998	0.204	0.250

30	0.998	1.000	1.000	0.204	0.222

TABLE 5: Estimated power of F, transformed F and Kruskal-Wallis tests when skewness γ_1=2

Δ	n	F	Trans F	Kruskal	mean ($\hat{\lambda}_n$)	sd($\hat{\lambda}_n$)
	5	0.046	0.064	0.049	0.194	0.481
	6	0.049	0.075	0.050	0.191	0.400
	7	0.054	0.086	0.066	0.198	0.324
	8	0.052	0.080	0.059	0.214	0.292
0.15	9	0.050	0.081	0.055	0.209	0.249
	10	0.053	0.084	0.063	0.210	0.230
	15	0.067	0.114	0.084	0.206	0.157
	20	0.058	0.119	0.090	0.209	0.126
	25	0.075	0.160	0.121	0.212	0.110
	30	0.081	0.186	0.140	0.213	0.093
	5	0.067	0.105	0.089	0.156	0.511
	6	0.075	0.119	0.093	0.156	0.429
	7	0.090	0.139	0.115	0.185	0.363
	8	0.097	0.162	0.129	0.169	0.316
0.3	9	0.089	0.170	0.135	0.168	0.270
	10	0.088	0.180	0.134	0.168	0.242
	15	0.120	0.257	0.203	0.182	0.170
	20	0.159	0.356	0.303	0.180	0.132
	25	0.185	0.427	0.335	0.184	0.109
	30	0.224	0.515	0.432	0.186	0.099
	5	0.094	0.136	0.109	0.124	0.506
	6	0.122	0.184	0.142	0.120	0.438
	7	0.136	0.214	0.176	0.143	0.380
	8	0.142	0.238	0.208	0.131	0.323
0.45	9	0.154	0.272	0.231	0.136	0.282
	10	0.168	0.304	0.269	0.144	0.255
	15	0.241	0.494	0.415	0.148	0.175
	20	0.296	0.626	0.548	0.163	0.137
	25	0.351	0.749	0.646	0.162	0.117
	30	0.415	0.830	0.761	0.166	0.104
	5	0.151	0.179	0.174	0.085	0.535
	6	0.175	0.243	0.215	0.105	0.462
	7	0.204	0.287	0.272	0.092	0.386
	8	0.216	0.333	0.310	0.122	0.337
0.6	9	0.264	0.403	0.389	0.127	0.293
	10	0.274	0.441	0.410	0.128	0.260
	15	0.374	0.675	0.615	0.136	0.176
	20	0.486	0.831	0.769	0.151	0.140
	25	0.582	0.917	0.864	0.155	0.119

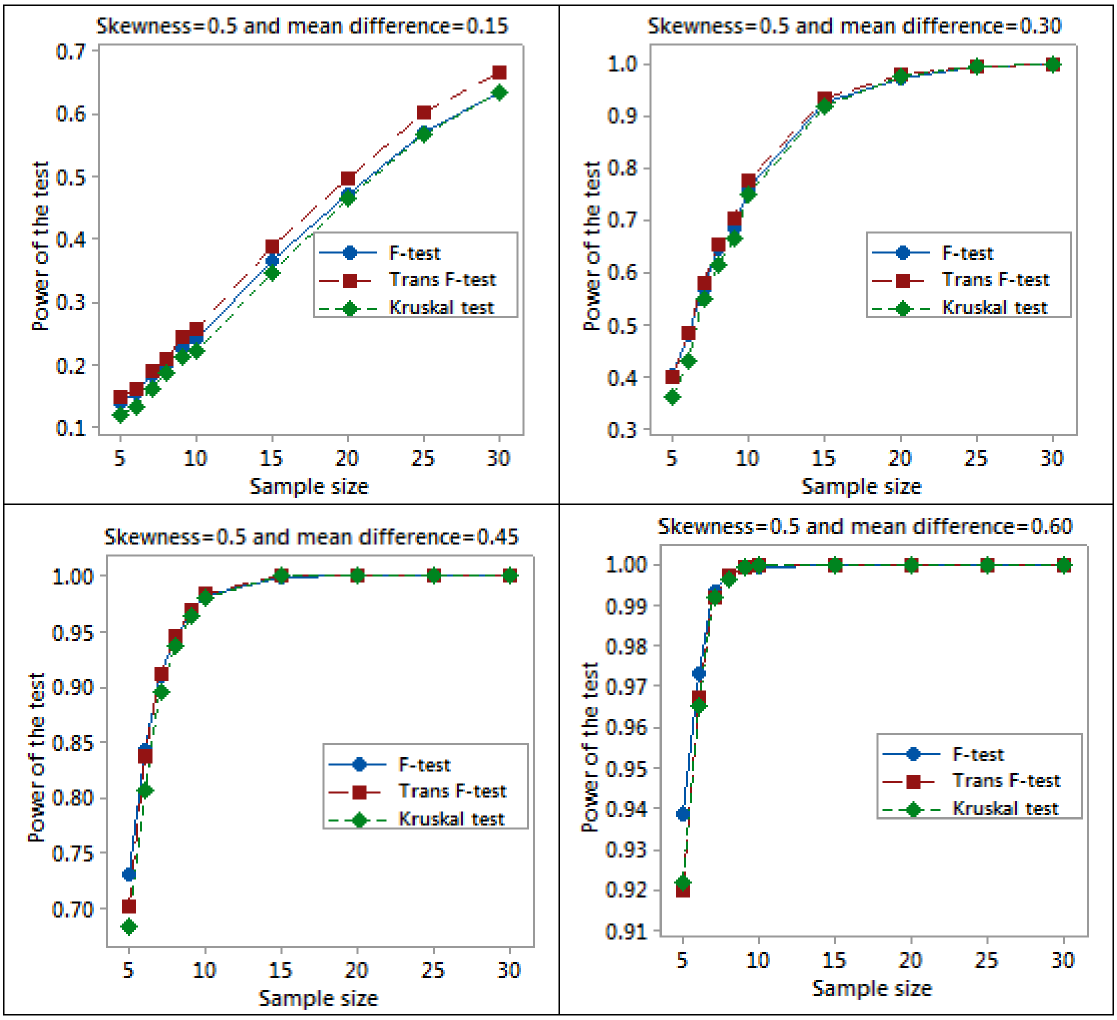
30	0.701	0.971	0.935	0.154	0.103

TABLE 6: Estimated power of F, transformed F and Kruskal-Wallis tests when skewness γ_1=4

Δ	n	F	Trans F	Kruskal	mean ($\hat{\lambda}_n$)	sd($\hat{\lambda}_n$)
	5	0.031	0.099	0.094	0.034	0.275
	6	0.031	0.120	0.111	0.046	0.214
	7	0.034	0.144	0.117	0.049	0.185
	8	0.034	0.165	0.138	0.059	0.139
0.15	9	0.035	0.208	0.150	0.063	0.125
	10	0.037	0.220	0.172	0.069	0.108
	15	0.035	0.356	0.251	0.081	0.075
	20	0.051	0.496	0.338	0.089	0.056
	25	0.037	0.628	0.422	0.089	0.047
	30	0.055	0.726	0.489	0.093	0.039
	5	0.042	0.123	0.131	0.009	0.293
	6	0.043	0.158	0.160	0.023	0.227
	7	0.045	0.222	0.213	0.028	0.184
	8	0.049	0.260	0.235	0.046	0.145
0.3	9	0.040	0.298	0.264	0.051	0.122
	10	0.061	0.330	0.282	0.062	0.106
	15	0.066	0.626	0.483	0.072	0.074
	20	0.070	0.775	0.601	0.078	0.058
	25	0.065	0.867	0.690	0.085	0.049
	30	0.083	0.934	0.804	0.087	0.042
	5	0.067	0.147	0.191	-0.007	0.294
	6	0.075	0.215	0.241	0.007	0.237
	7	0.065	0.248	0.262	0.020	0.177
	8	0.072	0.343	0.332	0.032	0.148
0.45	9	0.080	0.392	0.378	0.047	0.122
	10	0.090	0.467	0.408	0.052	0.106
	15	0.108	0.762	0.637	0.070	0.073
	20	0.118	0.896	0.781	0.077	0.058
	25	0.144	0.960	0.860	0.083	0.050
	30	0.157	0.985	0.923	0.087	0.042
	5	0.102	0.196	0.249	-0.017	0.302
	6	0.098	0.231	0.277	-0.011	0.240
	7	0.097	0.316	0.346	0.007	0.189
	8	0.123	0.395	0.430	0.030	0.145
0.6	9	0.117	0.481	0.472	0.037	0.126
	10	0.117	0.558	0.513	0.047	0.110
	15	0.142	0.848	0.749	0.068	0.073
	20	0.190	0.960	0.883	0.078	0.057
	25	0.214	0.990	0.941	0.084	0.047

30	0.221	0.997	0.974	0.087	0.040

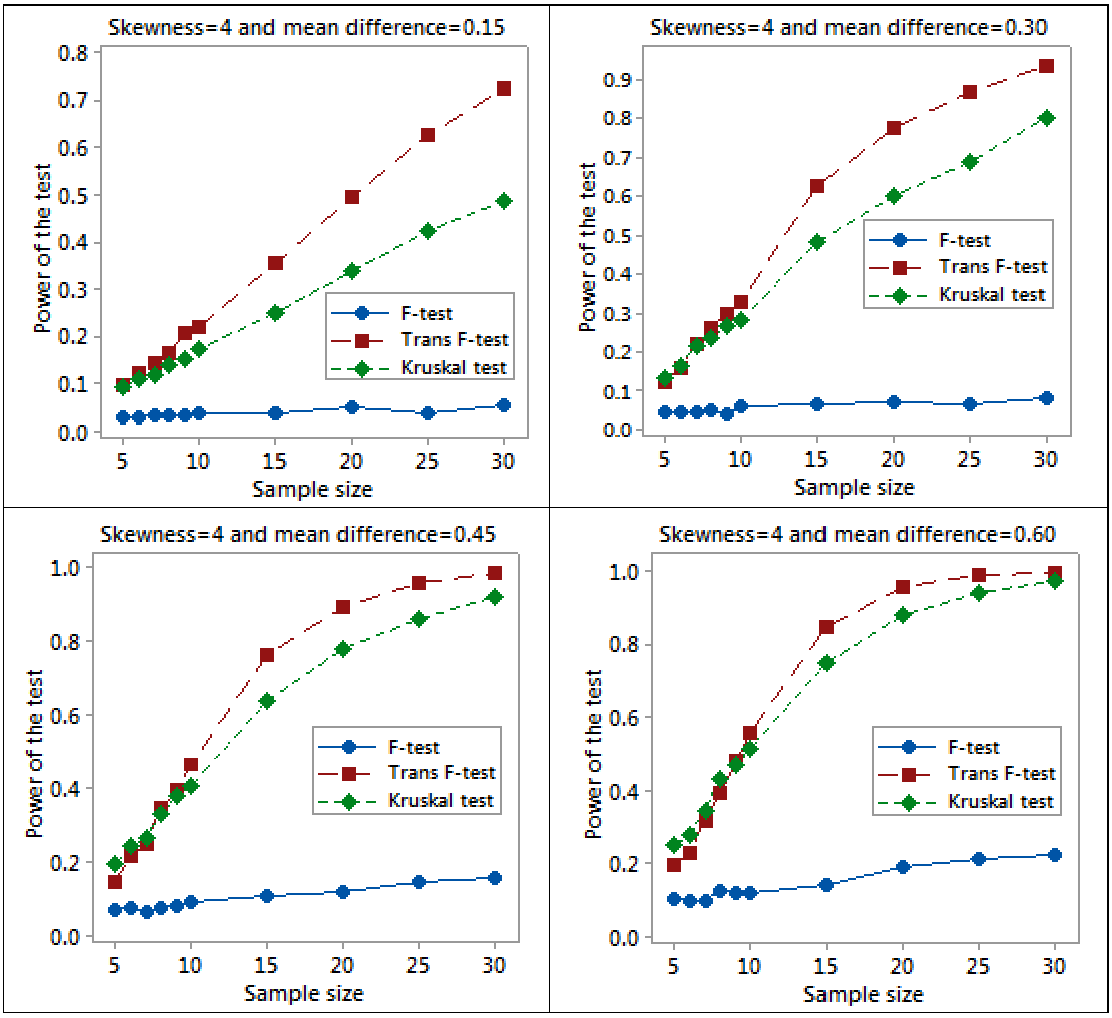
FIGURE 2:Comparing power of three tests when γ_1=0.5 and Δ=0.15, 0.30, 0.45 and 0.60



As we look at the power of the three underlying tests reported in Table 3 and Fig. 2, it is evident that when the mean difference is 0.15, the transformed F test (Trans F) has highest power compared to F and Kruskal-Wallis tests. As difference of the means goes higher, the power of all three tests seem comparable. It is also evident that the power of all tests are sensitive to the sample size in that the power increases, as the sample size increases.

Looking at Fig. 3, the Trans F test performs the best with respect to the power as compared to the F or Kruskal-Wallis test. Note that, the traditional ANOVA F test performs poorly as skewness reaches 4, and apparently, there is no significant improvement in the power when the mean differencesare 0.15 and 0.30, even with the increase in the sample size. However, the transformed F and Kruskal-Wallis tests show significant improvement in power as the sample size gets larger, and the mean difference gets higher, with the trans F-test showing the best performance.Given these facts, it can be concluded that the ANOVA F tests breaks down in terms of the estimated power when the skewness gets higher and the transformed F-test performs the best when the skewness gets higher.

FIGURE 3: Comparing power of three tests when γ_1=4 and Δ=0.15, 0.30, 0.45 and 0.60



IV EXAMPLE

In this section, an example using a real-life data is used to evaluate the performance of the three underlying tests. The data for this example refer to the prices (in dollars) of 30-count packages of 3randomly selected store-band vitamin/mineral supplements from three different sources appeared in [11]:

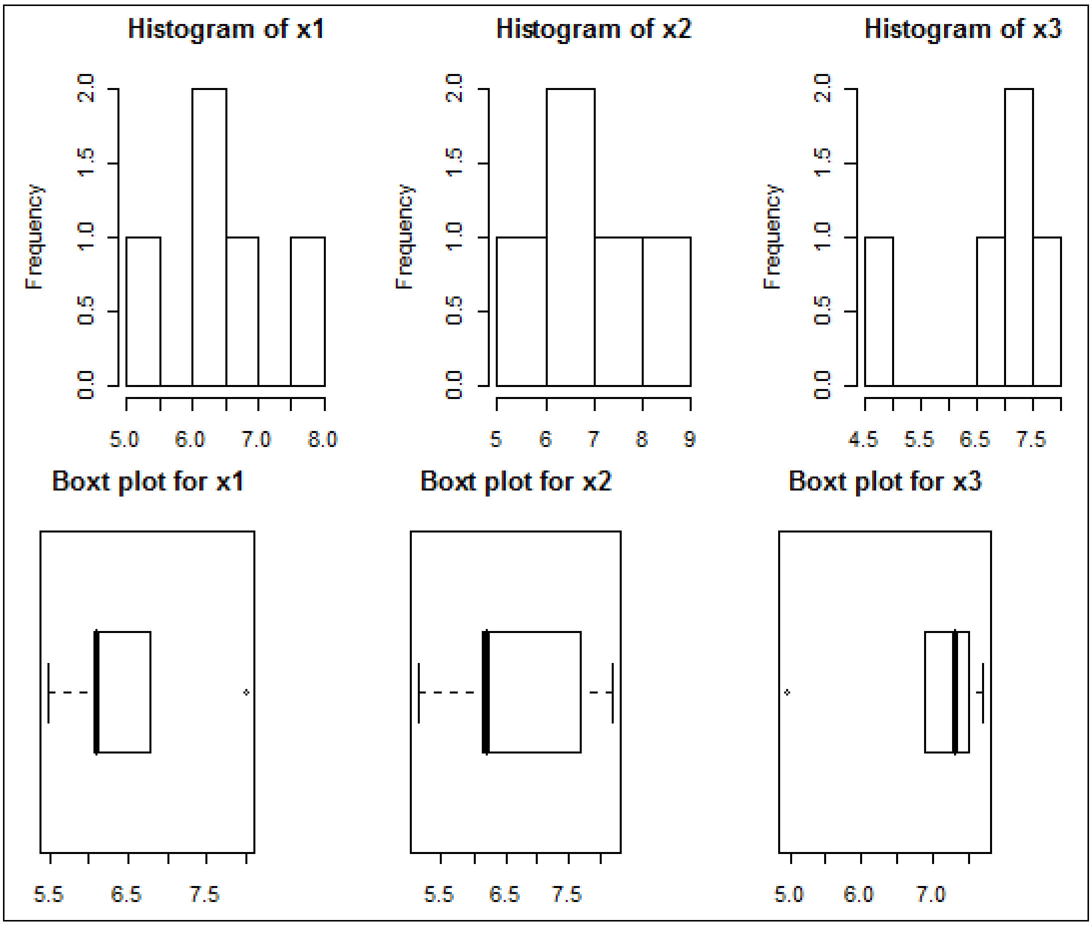
Grocery store (X_1)	6.79	6.09	5.49	7.99	6.10
Drugstore (X_2)	7.69	8.19	6.19	5.15	6.14
Discount store (X_3)	7.49	6.89	7.69	7.29	4.95

The means and skewness for the three samples are as follows:
$\bar{x}_1 = 6.49$, skewness = 1.08
$\bar{x}_2 = 6.67$, skewness = 0.17
$\bar{x}_3 = 6.86$, skewness = -1.86

From the results of the skewness, it appears that the sample might have come from skewed populations.

An attempt is also made to graphically assess the normality of the data using the histograms and boxplots for the three sample data. The histograms and boxplots have been presented in Fig. 4, which also provide evidence that three samples X_1, X_2 and X_3 might have been drawn from skewed populations. Thus, the validity of ANOVA F test is invalid due to the violation of the normality assumption of the data.

FIGURE 4: Histograms and box plots of three store-band prices for 3 samples of 30-count packages.



We wish to test
$H_0: \mu_1 = \mu_2 = \mu_3$ versus $H_1: \mu_j \neq \mu_{j'}$ for some $j \neq j'$

For transformed test, the search for $\hat{\lambda}_n$ is made over the interval [-2, 2] with an increment of 0.1. In other words, $\lambda \in \{-2: 0.1: 2\}$, following the notations of Section III. Using the algorithm of Section II, the estimated value of λ is $\hat{\lambda}_n = 1.93$. The results of the three tests for this example are as follows:

Tests	Value of test statistic and P-value
F test	F = 0.1393; $df(n; d)^* = (2,12)$; p-value = 0.8713
Kruskal test	Chi-squared = 0.546, df = 2; p-value = 0.7611
Trans F test	F = 0.1565; $df(n, d) = (2,12)$; p-value = 0.8568

$*\ df(n; d)$ refers to the numerator and denominator degrees of freedom for F distribution.

From the summary of the analysis results presented in the table above, it follows that all three tests failed to reject the null hypothesis of equality of average prices for the three store-band vitamin/mineral supplements.

V CONCLUSION

Under the assumption of normalityof $k$populations with a common variance,the test of comparing $k$means using independent samplesby means of the traditional ANOVA F test is the most powerful test.However, due to the violation of the normality, if k groups are far from the normality with higher skewness, the Kruskal-Wallis test is more robust than the F test in controlling the estimated size of the test.This article considers evaluation of the performance of a new transformed test as compared to the ANOVA F test and Kruskal-Wallis test for comparing k means from skewed populations, using simulation and real-life example. Noting the fact that the gamma distribution is one of the popular choices for simulating data from the skewed distributions, we consider a gamma distribution $G(\theta_1, \theta_2)$ having the skewness $\gamma_1 = 2/\sqrt{\theta_1}$ for simulating data. To control for the skewness efficiently, different values of the parameters θ_1 and $\theta_2$were considered to allow varying levels of skewness at 0.5, 1, 2 and 4, chosen arbitrarily, keeping means of all simulated distributions fixed at 1 under the null model. The mean difference (Δ) of 0.15, 0.30, 0.45 and 0.60 are considered under the alternative models to ensure a testing power away from 0 and 1 for the purpose of the comparisons.

From the simulation results, it follows that Transformed F test (Trans F) has comparable control over the estimated size of the test similar to Kruskal-Wallis and F-test. In terms of the estimated power from a Monte Carlo simulation, it appears that the new transformed test outperforms traditional F test and the non-parametric Kruskal-Wallis test for highly skewed distributions. The new proposed test is equally appropriate for level of significance for the test.The new test also performed equally well in real-life example tested as an application. Given the facts of the study, the new test could be implemented successfully for controlling estimated testing size and retaining higher power efficiency compared to other tests for skewed distributions.

REFERENCES

[1]. Kruskal, W.H. and Wallis, W.A. (1952). Use of ranks in one-criterion variance analysis. *Journal of the American Statistical Association*;47 (260): 583–621. Doi:10.1080/01621459.1952.10483441

[2]. Box, G.E.P. and Cox, D.R. (1964). Analysis of transformation.*Journal of Royal Statistical Society*. Ser. B 26 211-246.

[3]. Shapiro, S.S. and Wilk, M.B (1965). An analysis of variance test for normality (complete samples). *Biometrika*. 52, 591-611.

[4]. R version 2.10.1 (2009). The R Foundation for Statistical Computing.

[5]. Hinkley, D.Y. (1975). On power transformations to symmetry. Biometrika. 62, 101-111.

[6]. Hernandez, F. and Johnson, R.A. (1980). The large-sample behavior of transformations to normality. J. Amer. Statist. Assoc. 75, 855-861.

[7]. Bickel, P.J andDoksum, K.A (1981): An analysis of transformations revisited. J. Amer. Statist. Assoc. 76, 296-311.

[8]. Chen, H and Loh, W.Y (1992). Bounds on ARE's of tests following Box-Cox transformations. Ann Statist. 20, 1485-1500.

[9]. Chen, H. (1995). Tests following transformations. Ann. Statist. 23, 1587-1593.

[10]. Islam, M. K. and Chen, H. (2007). Transformed Test for Homogeneity of Variances. *Journal of Probability and Statistical Science*, 5(2): 151-170.

[11]. Bluman, A. G. (2014). Elementary Statistics, 9th Edition, McGraw Hill Education.

Khairul Islam "A New Transformed Test for Analysis of Variance for Skewed