JOURNAL OF APPLIED MATHEMATICS AND DECISION SCIENCES, 7(4), 187–206

Robustness to Non-Normality of Common Tests for the Many-Sample Location Problem

AZMERI KHAN† azmeri@deakin.edu.au
School of Computing and Mathematics, Deakin University, Waurn Ponds, VIC3217, Australia

GLEN D. RAYNER*
National Australia Bank, Australia

Abstract. This paper studies the effect of deviating from the normal distribution assumption when considering the power of two many-sample location test procedures: ANOVA (parametric) and Kruskal-Wallis (non-parametric). Power functions for these tests under various conditions are produced using simulation, where the simulated data are produced using MacGillivray and Cannon's [10] recently suggested *g-and-k* distribution. This distribution can provide data with selected amounts of skewness and kurtosis by varying two nearly independent parameters.

Keywords: ANOVA, *g-and-k* distribution, Kruskal-Wallis test, Quantile function.

1. Introduction

Analysis of variance (ANOVA) is a popular and widely used technique. Besides being the appropriate procedure for testing location equality within several populations (under certain assumptions about the data), the ANOVA model is basic to a wide variety of statistical applications.

The data used in an ANOVA model are traditionally assumed to be independent identically distributed (IID) normal random variables (with constant variance). These assumptions can be violated in many more ways than they can be satisfied! However, the normality assumption of the data is a pre-requisite for appropriate estimation and hypothesis testing with every variation of the model (e.g. homo-scedastic, hetero-scedastic).

In situations where the normality assumption is unjustified, the ANOVA procedure is of no use and can be dangerously misleading. Fortunately, nonparametric methods such as the Kruskal-Wallis test (cf Montgomery

† Requests for reprints should be sent to Azmeri Khan, School of Computing and Mathematics, Deakin University, Waurn Ponds, VIC3217, Australia.
* Fellow of the University of Wollongong Institute of Mathematical Modelling and Computational Systems University of Wollongong, Wollongong NSW2522, Australia

[12]) are available to deal with such situations. The Kruskal-Wallis test also tests whether several populations have equal locations.

The Kruskal-Wallis test, while less sensitive to distributional assumptions, does not utilize all the information collected as it is based on a rank transformation of the data. The ANOVA procedure, when its assumptions are satisfied, utilizes all available information. Naturally the extent of the departure from IID normality is the key factor that regulates the strength (or weakness) of the ANOVA procedure. Various measures of skewness and kurtosis have traditionally been used to measure the extent of non-normality, where skewness measures the departure from symmetry and kurtosis the thickness of the tails of the distribution.

The problem of robustness of the ANOVA test to non-normality has been extensively studied. Tiku [17] obtains an approximate expression for the ANOVA test power function. Tan [16] provides an excellent supplement to the reviews of Tiku [18] and Ito [6]. These approaches seem to revolve around examining the properties of various approximations to the exact distribution of the test statistic. This allows the authors to make useful conclusions about the general behaviour of the ANOVA test. However, our approach is based around the fact that a practitioner will usually need to analyze the data (in this case, test for location differences) regardless of its distribution, and will therefore be interested in the relative performance of the two major tools applicable here: ANOVA and Kruskal-Wallis.

We present a size and power simulation study of the ANOVA and Kruskal-Wallis tests for data with varying degrees of non-normality. Data were simulated using MacGillivray and Cannon's [10] recent *g-and-k* family of distributions, chosen for its ability to smoothly parameterize departures from normality in terms of skewness and kurtosis measures. The skewness and kurtosis characteristics of the *g-and-k* distribution are controlled in terms of the (reasonably) independent parameters g (for skewness) and k (for kurtosis). For $g = 0$ (no skewness) the *g-and-k* distribution is symmetric, and for $g = k = 0$ it coincides with the normal distribution (see Figure 1). Using the *g-and-k* distribution also allows the investigation to consider departures from normality that are not necessarily from the exponential family of distributions, unlike previous investigations.

Section 2 defines the ANOVA and Kruskal-Wallis tests, section 3 introduces quantile distributions, section 4 introduces the *g-and-k* distribution used in this paper, and section 5 describes the simulation study undertaken for this paper and discusses its results. Finally, section 6 concludes the paper.

2. Multiple Sample Location Tests

One-way or one-factor analysis of variance (ANOVA) is the most commonly used method of comparing the location of several populations. Here, each level of the single factor corresponds to a different population (or treatment). ANOVA tests whether or not the treatment *means* are identical, and requires that the data be independent and identically distributed (IID), with this identical distribution being the normal distribution.

Alternatively, for situations with no prior assumption about the particular distribution of the data, the Kruskal-Wallis test is often used. This test examines the hypothesis that treatment *medians* are identical, against the alternative that some of the treatment medians are different. The Kruskal-Wallis test also assumes the data are IID, but does not require that this particular distribution be normal, just that it be continuous.

Although ANOVA and Kruskal-Wallis do not test *precisely* the same hypothesis, operationally they are both used to test the same problem: whether or not the *location* of the treatments are the same. The fact that each test measures location in different ways is often not really relevant to the way they are used in practice.

2.1. ANOVA

The simplest one-way ANOVA model for t populations (or treatments) can be presented ([12]) as

$$y_{ij} = \mu_i + \epsilon_{ij}, \quad i = 1, 2, \ldots, t; \quad j = 1, 2, \ldots, n_i, \tag{1}$$

where y_{ij} is the j-th of the n_i observations in the i-th treatment (class or category); μ_i is the *mean* or expected response of data in the i-th treatment (often called the treatment mean); and ϵ_{ij} are independent, identically distributed *normal* random errors. The treatment means $\mu_i = \mu + \tau_i$ are sometimes also expressed as a combination of an overall expected response μ and a treatment effect τ_i (where $\sum_{i=1}^{t} n_i \tau_i = 0$).

Under the traditional assumption that the model errors ϵ_{ij} are independently distributed normal random variables with zero mean and constant variance σ^2, we can test the hypothesis

$$H_0 : \mu_1 = \mu_2 = \ldots = \mu_t; \quad H_a : \text{all } \mu_i\text{'s are not equal.} \tag{2}$$

The classical ANOVA statistic (cf [12]) can be used for testing this hypothesis.

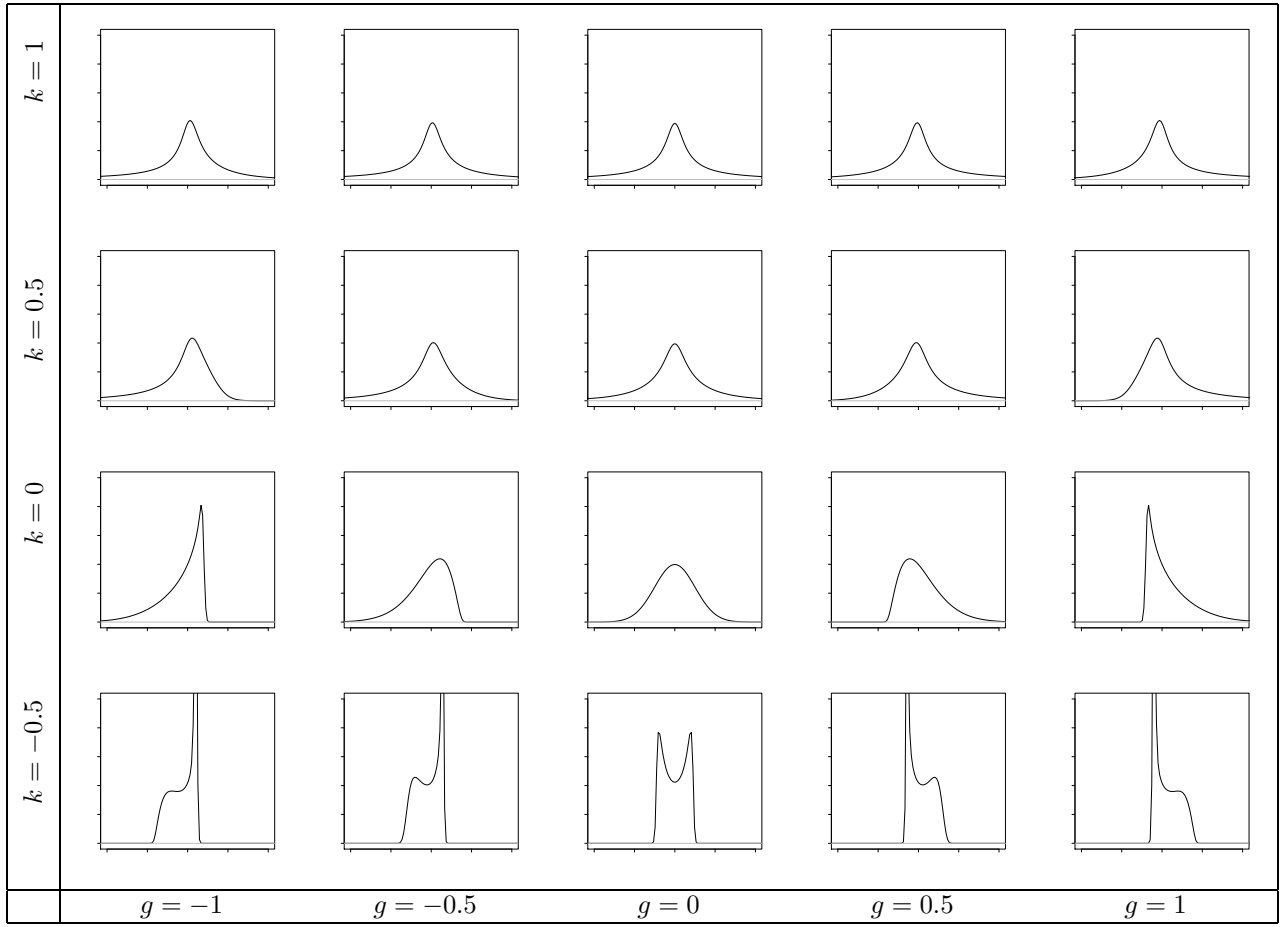

Figure 1. Shows *g-and-k* distribution densities corresponding to various values of the skewness parameter g and the kurtosis parameter k. Here the location and scale parameters are $A = 0$ and $B = 1$ respectively.

2.2. The Kruskal-Wallis test

The Kruskal-Wallis test models the data somewhat differently. As with ANOVA, the data are assumed to be IID, but now this may be any continuous distribution, rather than only the normal. The data are modeled as

$$y_{ij} = \eta_i + \phi_{ij}, \quad i = 1, 2, \ldots, t; \quad j = 1, 2, \ldots, n_i, \tag{3}$$

where as before, y_{ij} is the j-th of the n_i observations in the i-th treatment (class or category); η_i is the *median* response of data in the i-th treatment (often called the treatment mean); and ϕ_{ij} are independent, identically distributed *continuous* random errors.

This model allows us to test the hypothesis

$$H_0 : \eta_1 = \eta_2 = \ldots = \eta_t; \quad H_a : \text{all } \eta_i\text{'s are not equal.} \tag{4}$$

The Kruskal-Wallis test statistic (cf [12]) can be used for testing this hypothesis.

3. Quantile Distributions

Distributional families tend to fall into two broad classes:

1. those corresponding to general forms of a density or distribution function, and
2. those defined by a family of transformations of a base distribution, and hence by their quantile function.

Some of those defined by their quantile function are the Johnson system, with a base of normal or logistic, the Tukey lambda and its various asymmetric generalizations, with a uniform base, and Hoaglin's *g-and-h*, with a normal base.

Distributions defined by transformations can represent a wide variety of density shapes, particularly if that is the purpose of the transformation. For example, Hoaglin's [5] original *g-and-h* distribution was introduced as a transformation of the normal distribution to control the amount of skewness (through the parameter g) and kurtosis (with the parameter h) added.

A desirable property of a distributional family is that the nature and extent of changes in shape properties are apparent with changes in parameter values, for example when an indication of the departure from normality is required ([3]). It is not generally recognized how difficult this is to achieve

in asymmetric distributions, whether they have one or more than one shape parameters. For the asymmetric Tukey lambda distributions, measures of asymmetry (skewness) and heaviness in the tails (kurtosis) of the distributions are functions of both shape parameters, so that even if approximations to well-known distributions are good, it is not always obvious what shape changes occur as the lambda distributions' shape parameters move between, or away from, the values providing such approximations. An additional difficulty with the use of this distribution when fitting through moments, is that of non-uniqueness, where more than one member of the family may be realized when matching the first four moments to obtain parameters for the distribution ([13]).

In the *g-and-h* distribution of Hoaglin [5] and Martinez and Iglewicz [11] the shape parameters g and h are more interpretable, although the family does not provide quite as wide a variety of distributional types and shapes as the asymmetric lambda. The MacGillivray [10] adaptation of the *g-and-h* distributions called the *generalized g-and-h* distribution and a new family called the *g-and-k* distributions, make greater use of quantile-based measures of skewness and kurtosis, to increase interpretability in terms of distributional shape changes ([7], [8], [1] and [9]). A further advantage of the MacGillivray *generalized g-and-h* and *g-and-k* distributions is the greater independence between the parameters controlling skewness and kurtosis (with respect to some quantile-based measures of distribution shape). The *g-and-k* distributions also allow distributions with more negative kurtosis (ie"shorter tails") than the normal distribution, and even some bimodal distributions.

4. The *generalized g-and-h* and *g-and-k* Distributions

The *generalized g-and-h* and *g-and-k* distribution families have shown considerable ability to fit to data, approximate standard distributions, ([10]), and facility for use in simulation studies ([3], [4], [15]). In an assessment of the robustness to non-normality in ranking and selection procedures, Haynes *et al.* [3] utilize the *g-and-k* distribution for the above reasons as well as for flexibility in the coverage of distributions with both long and short tails. Rayner and MacGillivray [15] examine the effect of non-normality on the distribution of (numerical) maximum likelihood estimators.

The *generalized g-and-h* and *g-and-k* distributions ([10]) are defined in terms of their quantile functions as

$$Q_X(u|A,B,g,h) = A + Bz_u\left(1 + c\frac{1-e^{-gz_u}}{1+e^{-gz_u}}\right)e^{hz_u^2/2}$$

and

$$Q_X(u|A,B,g,k) = A + Bz_u\left(1 + c\frac{1-e^{-gz_u}}{1+e^{-gz_u}}\right)(1+z_u^2)^k$$

respectively. Here A and $B > 0$ are location and scale parameters, g measures skewness in the distribution, $h > 0$ and $k > -\frac{1}{2}$ are measures of kurtosis (in the general sense of peakedness/tailedness) of the distributions, $z_u = \Phi^{-1}(u)$ is the uth standard normal quantile, and c is a constant chosen to help produce proper distributions.

For both the *generalized g-and-h* and *g-and-k* distributions the sign of the skewness parameter indicates the direction of skewness: $g < 0$ indicates the distribution is skewed to the left, and $g > 0$ indicates skewness to the right. Increasing/decreasing the unsigned value of g increases/decreases the skewness in the indicated direction. When $g = 0$ the distribution is symmetric.

Increasing the *generalized g-and-h* kurtosis parameter, h, adds more kurtosis to the distribution. The value $h = 0$ corresponds to no extra kurtosis added to the standard normal base distribution, which is the minimum kurtosis the *generalized g-and-h* distribution can represent.

The kurtosis parameter k, for the *g-and-k* distribution, behaves similarly. Increasing k increases the level of kurtosis and vice versa. As before, $k = 0$ corresponds to no kurtosis added to the standard normal base distribution. However this distribution can represent less kurtosis than the normal distribution, as $k > -\frac{1}{2}$ can assume negative values. If curves with even more negative kurtosis are required then a base distribution with less kurtosis than the standardized normal can be used. See Figure 1.

For these distributions c is the value of the "overall asymmetry" ([7]). For an arbitrary distribution, theoretically the overall asymmetry can be as great as one, so it would appear that for $c < 1$, data or distributions could occur with skewness that cannot be matched by these distributions. However for $g \neq 0$, the larger the value chosen for c, the more restrictions on k or h are required to produce a completely proper distribution. Real data seldom produce overall asymmetry values greater then 0.8 ([10]). We have used $c = 0.8$ throughout this paper. For this value of c the *generalized g-and-h* distribution is proper for any g and $h \geq 0$. The restrictions placed on the *g-and-k* distribution to be proper are more complicated, but for our

purposes (using these distributions to generate data) the distribution need not be proper.

For more information about the properties of these distributions, see [10], [14], and [15].

5. Simulation Study

When data are not IID normally distributed then the error term ϵ_{ij} in equation (1) is also not normally distributed. In such situations, the assumptions required for the ANOVA model (see section 2.1) are not satisfied, and it is not appropriate or optimal to use the ANOVA test statistic to test for equality of treatment locations. However, we can still *calculate* the ANOVA test statistic and, therefore, the resulting power function of this test. Similarly, the Kruskal-Wallis test statistic and its power function can also be found. These power functions will allow us to compare the usefulness of the ANOVA and Kruskal-Wallis tests under various kinds and degrees of non-normality (combinations of the g and k parameter values for data from the *g-and-k* distribution).

To examine how these tests react to the degree of non-normality, we use data distributed according to MacGillivray and Cannon's [10] *g-and-k* distribution. Using the *g-and-k* distribution allows us to quantify how much the data depart from normality in terms of the values chosen for the g (skewness) and k (kurtosis) parameters. For $g = k = 0$, the quantile function for the *g-and-k* distribution is just the quantile function of a standard normal variate, and hence the assumptions required for the ANOVA model (section 2.1) are satisfied.

In order to compare the power of the ANOVA and Kruskal-Wallis tests, expressions for these power functions are required. However, in practice it is very difficult to obtain analytic expressions for these power functions. Instead, we have conducted a Monte-Carlo simulation to estimate these power functions for various combinations of the g and k parameter values for data from the *g-and-k* distribution.

To simplify matters, only $t = 3$ treatments or populations are considered here. We simulate these populations as $y_{ij} = \mu_i + e_{ij}$ for $i = 1, \ldots, t$ where the errors e_{ij} are taken from the *g-and-k* distribution with parameters (A, B, g, k). Here the location and scale parameters of the error variates are $A = 0$ and $B = 1$ respectively.

We examine the ANOVA and Kruskal-Wallis tests for $\mu_3 = 0$ and $\mu_1, \mu_2 = -11 : 1 : 11$ with $R = 5,000$ simulations and the following (g, k) combinations: $(-1, 0), (-0.5, 0), (0, 0), (0.5, 0), (1, 0)$ and $(0, -0.5), (0, 0.5), (0, 1)$. We consider only the case of equal sample sizes $n = 3, 5, 15, 30$ for each

treatment group. Tests are carried out using a nominal size of $\alpha = 1\%$ for $n = 5, 15, 30$. For $n = 3$, the Kruskal-Wallis test is conservative and consideration of nominal $\alpha = 1.1\%$ would produce a test of actual size approximately $\alpha = 0.5\%$ (see for example [2]). This fact is also supported by the results of this study (see Figure 8). That is why the ANOVA test with size $\alpha = 0.5\%$ is carried out to compare with the Kruskal-Wallis test with nominal size $\alpha = 1.1\%$ in simulations.

Figures 2 through 4 show empirical power surfaces for the ANOVA test using sample sizes $n = 3, 5, 15$ respectively. Figures 5 through 7 show empirical power surfaces for the Kruskal-Wallis test, also using these sample sizes. These Figures show power surfaces as a function of the population location parameters μ_2 and μ_3 for different amounts of skewness and kurtosis in the error terms. Power surfaces for $n = 30$ are omitted since the power surfaces are almost everywhere 1, although showing some decrease in power as k increases, when they begin to look more like the $n = 15$ power surfaces.

It is also interesting to consider how power varies with the skewness and kurtosis parameters g and k, for different combinations of values of the true locations. We examine the ANOVA and Kruskal-Wallis tests for $g = -5 : \frac{10}{14} : 5$ and $k = -0.5 : \frac{1.5}{14} : 1$ with the combinations $(2, 2, 0), (-2, 2, 0), (2, 0, 0)$ and $(0, 0, 0)$. Tests are carried out using equal sample sizes of $n = 3, 5, 15, 30$ for each treatment group, $R = 5,000$ simulations and sizes as mentioned before. Figures 8 and 9 show power surfaces as a function of the skewness and kurtosis parameters g and k for different combinations of the population locations.

Examining Figures 2 to 7 it can be seen that power functions for both ANOVA and Kruskal-Wallis tests seem:

1. insensitive to variations in skewness;
2. worse (less sharply discriminating) for larger k (heavier tailed error distributions); and
3. better as the sample size n increases.

This confirms the evidence supplied in earlier studies of ANOVA robustness (eg [16]). Interestingly, the Kruskal-Wallis test does seem to be affected by the shape of the error distribution.

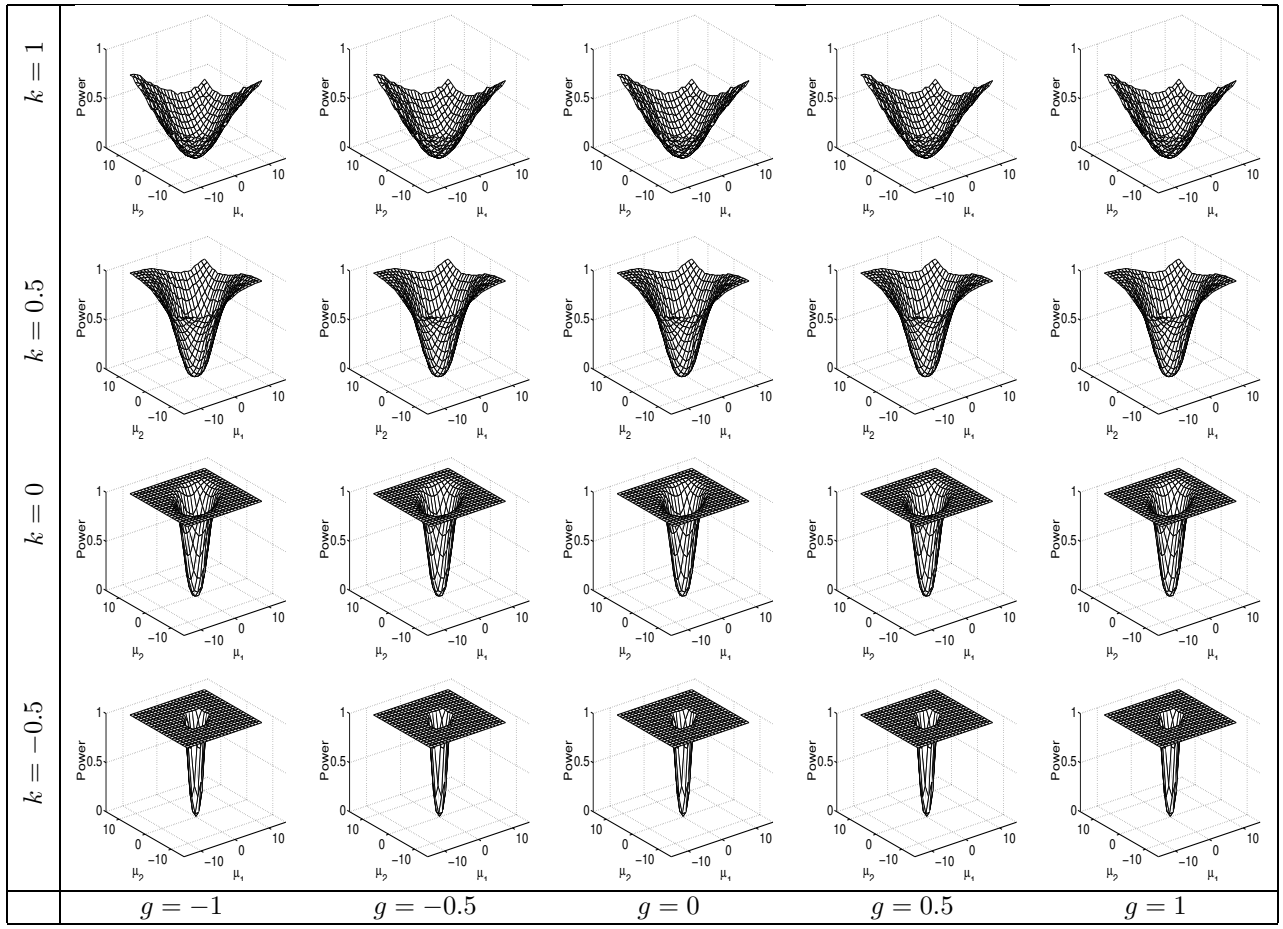

Figure 2. Each of these figures shows a power surface for the ANOVA F-test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 3$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 0.5\%$).

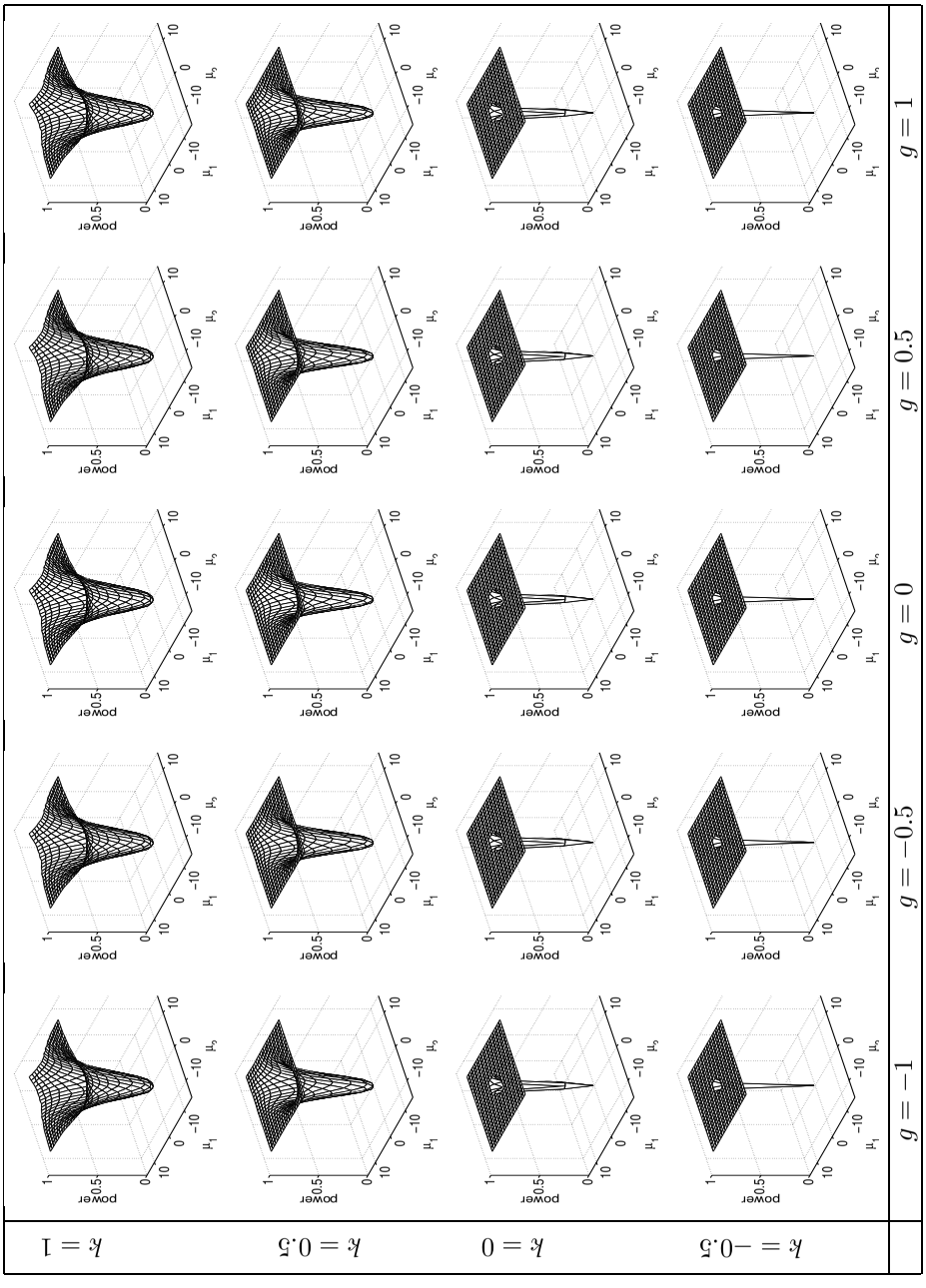

Figure 3. Each of these figures shows a power surface for the ANOVA F-test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 5$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 1\%$).

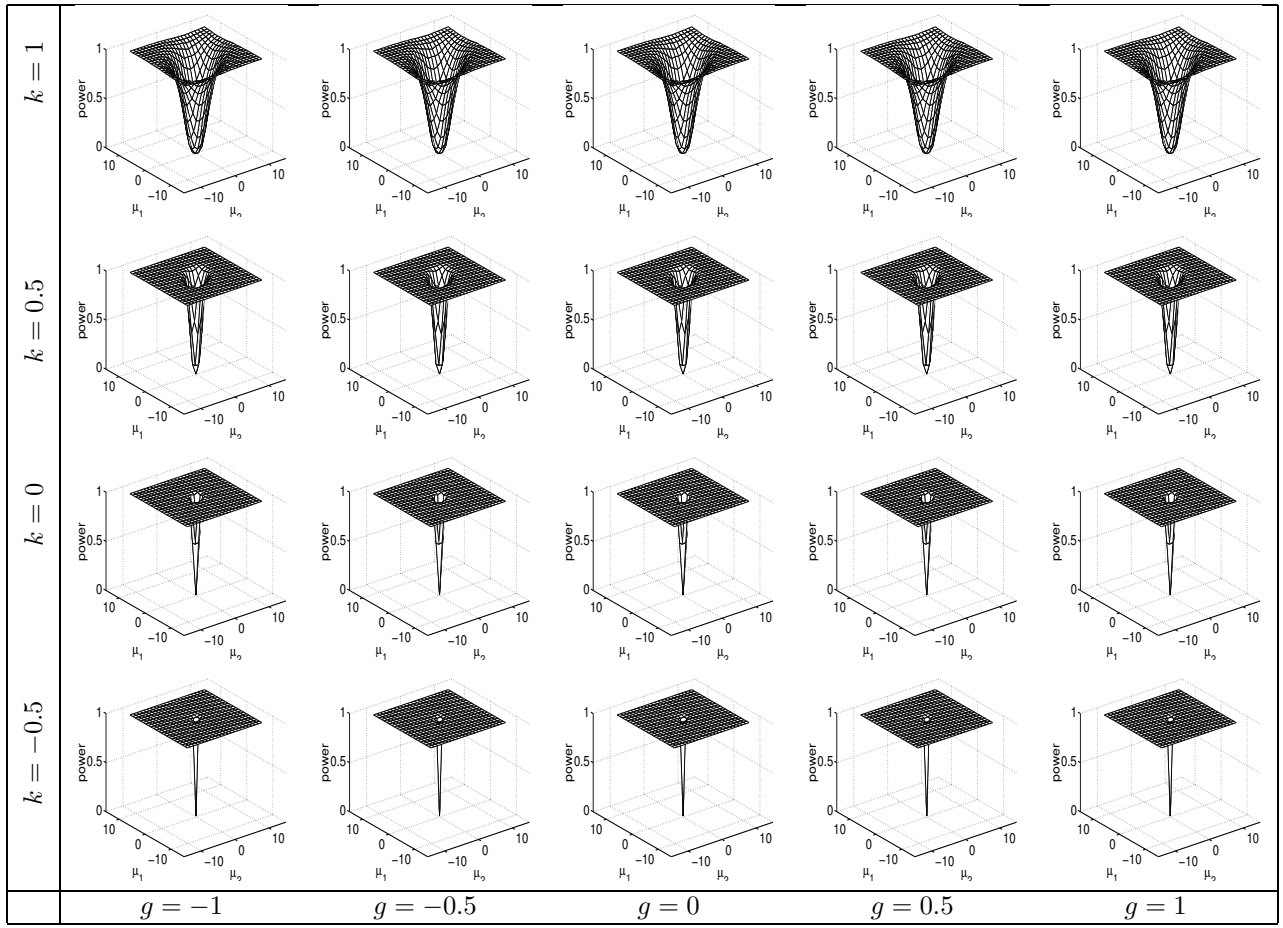

Figure 4. Each of these figures shows a power surface for the ANOVA F-test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 15$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 1\%$).

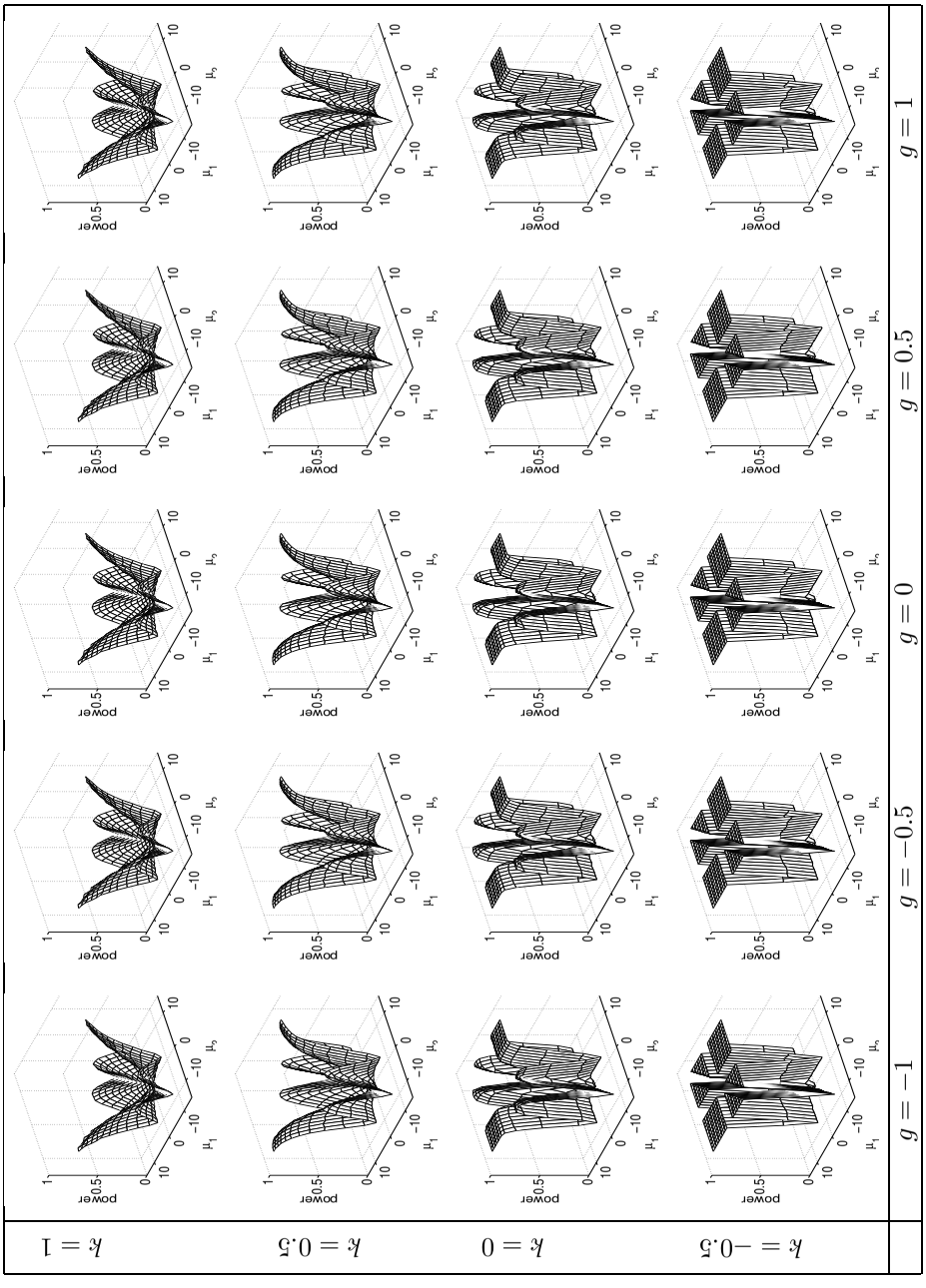

Figure 5. Each of these figures shows a power surface for the Kruskal-Wallis test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 3$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 1.1\%$).

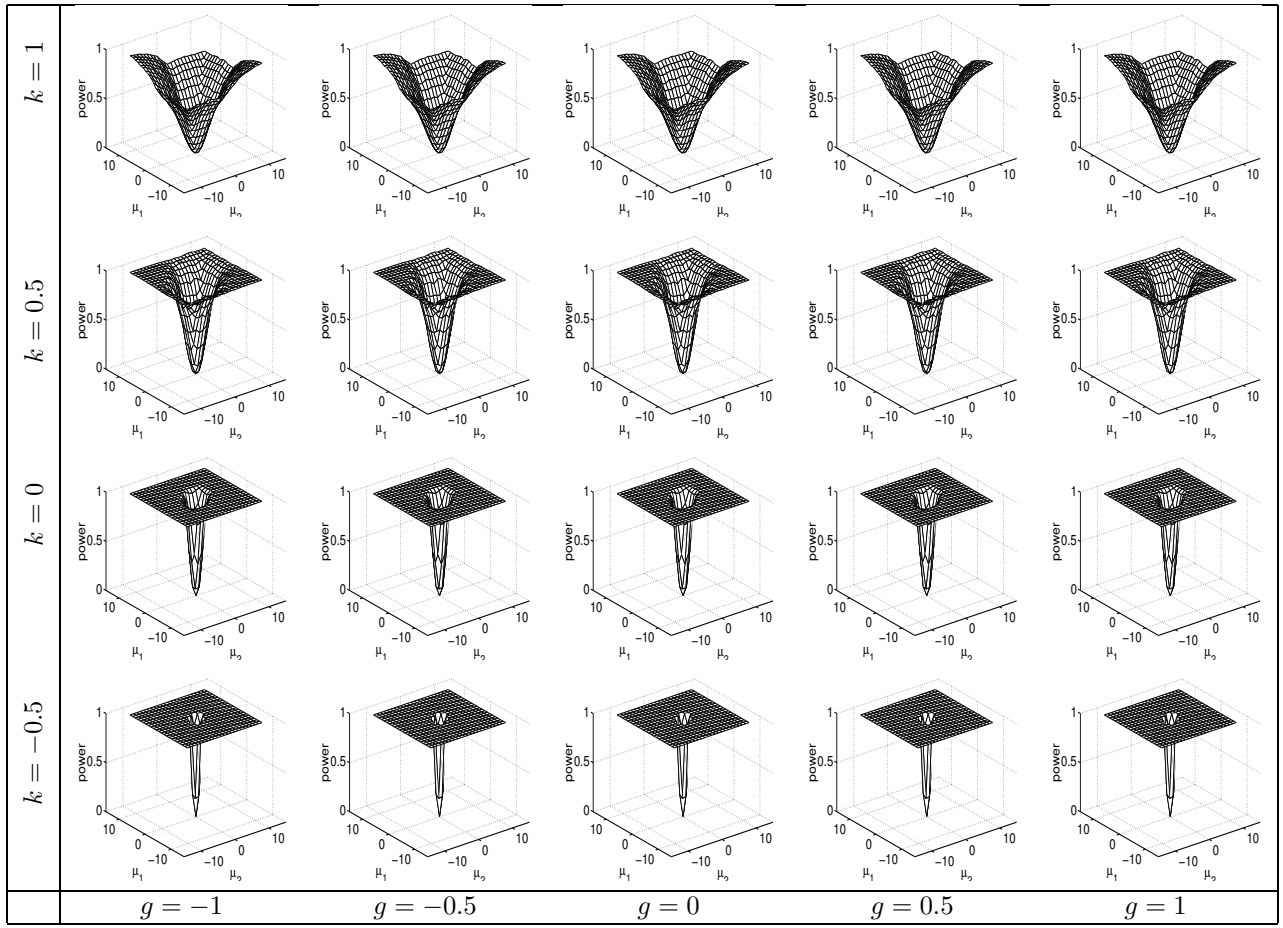

Figure 6. Each of these figures shows a power surface for the Kruskal-Wallis test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 5$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 1\%$).

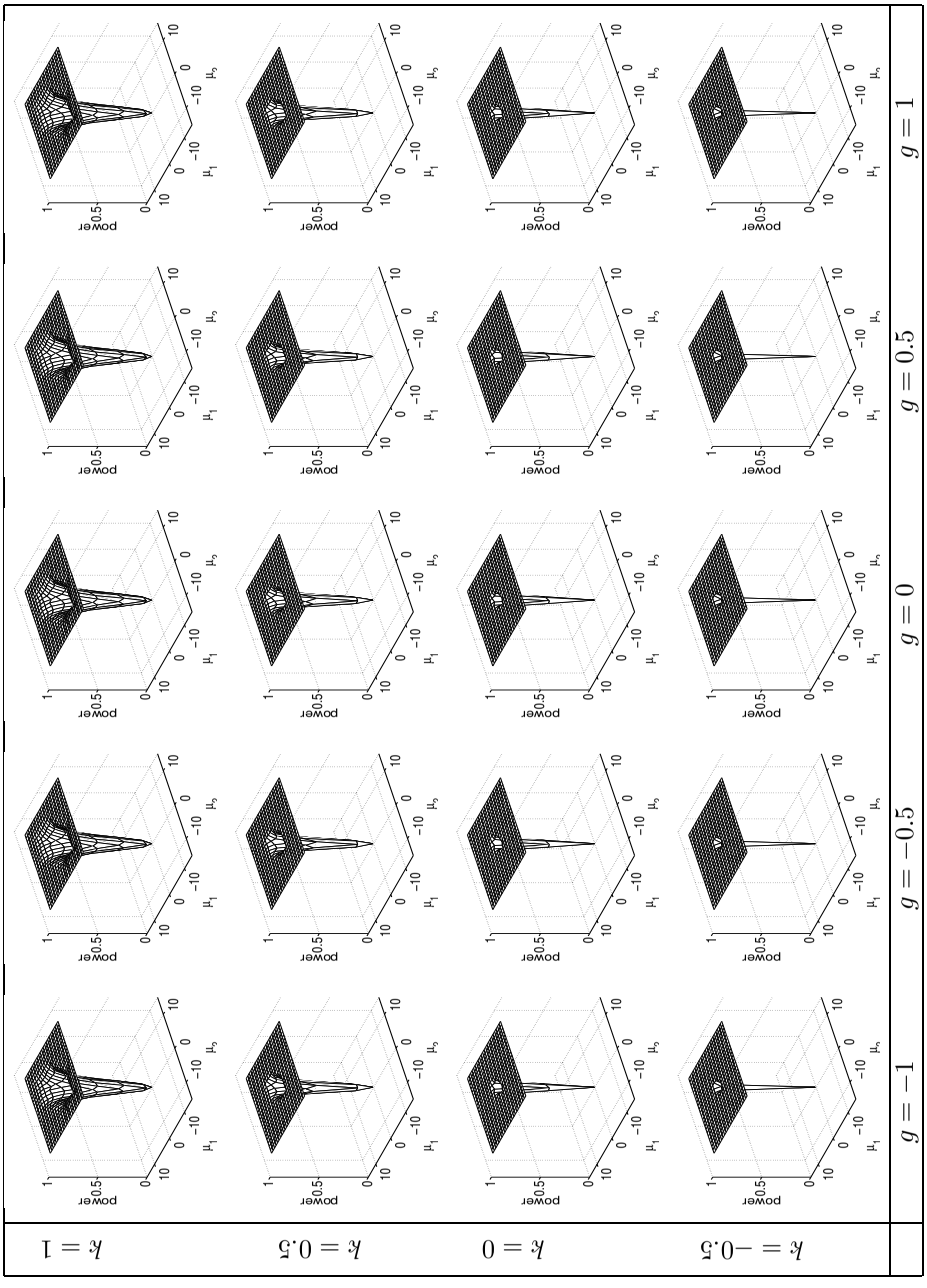

Figure 7. Each of these figures shows a power surface for the Kruskal-Wallis test of the null hypothesis that $\mu_1 = \mu_2 = \mu_3$ given $\mu_3 = 0$ and various different true values of μ_1, μ_2 and $n = 15$ observations per sample. Comparing these power surfaces shows the effect of data with non-normal skewness or kurtosis ($\alpha = 1\%$).

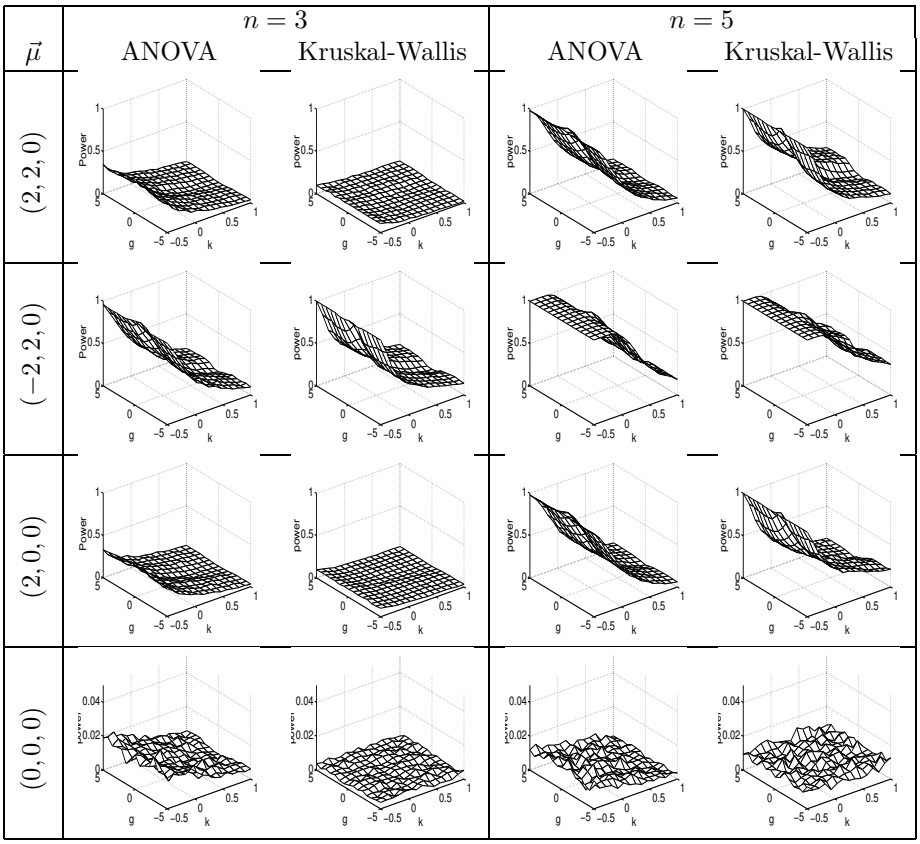

Figure 8. Power surfaces for both the ANOVA F-test and the Kruskal-Wallis test for sample sizes $n = 3, 5$. These plots show how power varies with skewness and kurtosis parameters $g = -5 : 5$ and $k = -0.5 : 1$ for different values of the true treatment means $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$. Note that the power surface for $\vec{\mu} = (\mu_1, \mu_2, \mu_3) = (0, 0, 0)$ shows how the true size of the test (for $n = 3$, $\alpha = 0.5\%$ for ANOVA and $\alpha = 1.1\%$ for Kruskal-Waliis and for $n = 5$, $\alpha = 1\%$ for both tests) varies with skewness and kurtosis. The jaggedness of these true size surfaces is due to sampling variation.

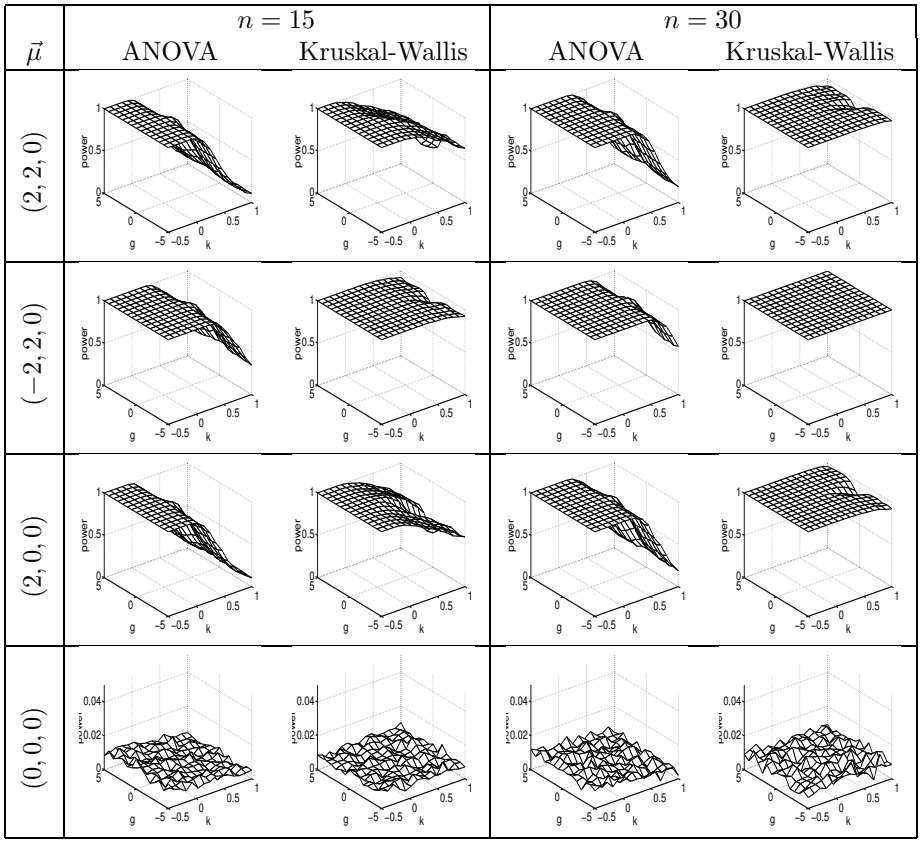

Figure 9. Power surfaces for both the ANOVA F-test and the Kruskal-Wallis test for sample sizes $n = 15, 30$. These plots show how power varies with skewness and kurtosis parameters $g = -5 : 5$ and $k = -0.5 : 1$ for different values of the true treatment means $\vec{\mu} = (\mu_1, \mu_2, \mu_3)$. Note that the power surface for $\vec{\mu} = (\mu_1, \mu_2, \mu_3) = (0, 0, 0)$ shows how the true size of the test ($\alpha = 1\%$) varies with skewness and kurtosis. The jaggedness of these true size surfaces is due to sampling variation.

Figure 5 shows the power surface for the Kruskal-Wallis test with samples of size $n = 3$. Strangely, when $\mu_1 = \pm\mu_2$, $\mu_1 = 0$, or $\mu_2 = 0$ the power function is very low. This means that for $n = 3$ the Kruskal-Wallis test concludes that all treatment locations are equal, when in fact only the absolute value of any two treatment means are equal.

A different way in which the ANOVA and Kruskal-Wallis tests depend on skewness and kurtosis is provided by Figures 8 and 9. Here we examine the power surfaces as the skewness and kurtosis parameters g and k vary, for a variety of different combinations of (μ_1, μ_2, μ_3). Once again, Figures 8 and 9 confirm the dotpoints made above, although some small dependence on skewness (g) is revealed here for both the ANOVA and Kruskal-Wallis tests. As one would expect, whenever the power surface changes with skewness it is symmetric through $g = 0$, showing that only the degree of skewness is important rather than its direction.

The true size of the ANOVA and Kruskal-Wallis tests is revealed by the $\vec{\mu} = (\mu_1, \mu_2, \mu_3) = (0, 0, 0)$ power surfaces in Figures 8 and 9. For both tests, size seems indifferent to variations in the error distribution shape. However, for the Kruskal-Wallis test with $n = 3$ the true size is much smaller than the nominal size.

Examining Figure 8 more closely reveals that for $n = 3$ and $n = 5$ both the ANOVA and Kruskal-Wallis tests have similar trouble appropriately rejecting the null hypothesis when only one location parameter is different. Figure 9 shows that for $n = 15$ and $n = 30$ the ANOVA power surface is still heavily dependent on kurtosis (k), whereas the Kruskal-Wallis test is far less affected and by $n = 30$ the Kruskal-Wallis test is all but indifferent to the error distribution shape.

6. Conclusion

We distil our study into four observations/recommendations.

1. Both the ANOVA and Kruskal-Wallis tests are vastly more affected by the kurtosis of the error distribution rather than by its skewness, and the effect of skewness is unrelated to its direction.

2. Both the ANOVA and Kruskal-Wallis test sizes do not seem to be particularly affected by the shape of the error distribution.

3. The Kruskal-Wallis test does not seem to be an appropriate test for small samples (say $n < 5$). Even for non-normal data, the ANOVA test is a better option than the Kruskal-Wallis test for small sample sizes (say $n = 3$). This comment is made on the basis of the comparison

between a Kruskal-Wallis test of nominal size 1.1% and an ANOVA test of size 0.5%. It is clearly understandable that a comparison between the Kruskal-Wallis and ANOVA tests of same size is a more reasonable procedure.

4. The Kruskal-Wallis tests clearly performs better than the ANOVA test if the sample sizes are large and kurtosis is high. Increasing sample size drastically improves the performance of the Kruskal-Wallis test, whereas the ANOVA test does not seem to improve as much or as quickly.

The first result above reflects commonly held wisdom as well as the results presented in [16].

While the simulation results included here are for only three treatment groups, it is not unreasonable to use these as a guide in the case of larger or more complex ANOVA-based models where normality would probably be a more rather than less critical assumption.

7. Acknowledgments

The authors would like to thank several helpful referees for their comments and suggestions.

References

1. K. P. Balanda and H. L. MacGillivray. Kurtosis and spread. *Canadian Journal of Statistics*, 18:17–30, 1990.
2. K. R. Gabriel, and P. A. Lachenbruch. Non-parametric ANOVA in Small samples: A Monte Carlo study of the adequacy of the asymptotic approximation. *Biometrics*, Vol 25, 593-596, 1969.
3. M. A. Haynes, M. L. Gatton, and K. L. Mengersen. Generalized control charts for non-normal data. *Technical Report Number 97/4*, School of Mathematical Sciences, Queensland University of Technology, Brisbane, Australia, 1997a.
4. M. A. Haynes, H. L. MacGillivray, and K. L. Mengersen. Robustness of ranking and selection rules using generalized g-and-k distributions. *Journal of Statistical Planning and Inference*, 65:45–66, 1997b.
5. D. C. Hoaglin. Summarizing shape numerically: the g-and-h distributions. In D. C. Hoaglin, F. Mosteller, and J. W. Tukey (eds) *Understanding Robust and Exploratory Data Analysis*, Wiley, New York, 1985.
6. K. Ito. Robustness of ANOVA and MANOVA test procedures. In *H*andbook of Statistics, Vol. 1, (P. R. Krishnainh, ed.), Amsterdam, Holand, 1980.
7. H. L. MacGillivray. Skewness and asymmetry: measures and orderings. *The Annals of Statistics*, 14:994–1011, 1986.
8. H. L. MacGillivray and K. P. Balanda. The relationships between skewness and kurtosis. *Australian Journal of Statistics*, 30:319–337, 1988.

9. H. L. MacGillivray. Shape properties of the g-and-h and Johnson families. *Communications in Statistics - Theory and Methods*, 21:1233-1250, 1992.
10. H. L. MacGillivray and W. H. Cannon. Generalizations of the g-and-h distributions and their uses. *Preprint*, 2002.
11. J. Martinez and B. Iglewicz. Some properties of the Tukey g-and-h family of distributions. *Communications in Statistics - Theory and Methods*, 13:353–369, 1984.
12. D. C. Montgomery. *Design and Analysis of Experiments, 2nd ed.* John Wiley, 1984.
13. J. S. Ramberg, P. R. Tadikamalla, E. J. Dudewicz, and E. F. Mykytka. A probability distribution and its uses in fitting data. *Technometrics*, 21:201–214, 1979.
14. G. D. Rayner. *Statistical Methodologies for Quantile-based Distributional Families.* PhD thesis, Queensland University of Technology, Australia, 2000.
15. G. D. Rayner and H. L. MacGillivray. Numerical maximum likelihood estimation for the g-and-k and generalized g-and-h distributions. *Statistics & Computing*, 12:57–75, 2002.
16. W. Y. Tan. Sampling distributions and robustness of t, f and variance-ratio in two samples and ANOVA Models with respect to departure from normality. *Communications in Statistics - Theory and Methods*, 11:2485–2511, 1982.
17. M. L. Tiku. Power function of the F-test under non-normal situations. *Journal of the American Statistical Association*, 66:913-915, 1971.
18. M. L. Tiku. Laguerre series forms of the distributions of classical test statistics and their robustness in non-normal situations. In *A*pplied Statistics (R. P. Gupta, eds), New York, American Elsevier, 1975.