Journal of Empirical Research in Accounting & Auditing
ISSN (2384-4787)
J. Emp. Res. Acc. Aud. 2, No. 2 (Oct-2015)

Analysis of Total Differences with Application to Productivity Improvement

Elsayed A. H. Elamir

Department of Statistics and Mathematics, Benha University, Egypt
Currently, Management & Marketing Department, College of Business, University of Bahrain, Kingdom of Bahrain

Received: 28 June 2015, Revised: 15 August 2015, Accepted: 6 Sept. 2015, Published: (Oct) 2015

Abstracts: One way analysis of Gini's mean difference (ANOMD) about mean and median is derived where the total sum of differences is partition into exact between sum of differences and exact within sum of differences. ANOMD has advantages: ensures stability in statistical inferences; has flexibility to test for any location measure and total sum difference does not depend on any fixed location. However, the variance-gamma distribution is used to fit the sampling distributions of between sum differences and within sum differences. Consequently, two tests of equal population medians and means are introduced under the assumption of the normal distribution. Moreover, two measures of effect sizes are re-defined and studied in terms of ANOMD. The ANOMD model is applied to productivity improvement data and it is found that the percentage of explained variation given by ANOMD is more than the percentage given by ANOVA.

Keywords: ANOVA; Effect sizes; L-moments; Variance-gamma distribution.

Introduction

Gini's mean difference (GMD) depends on all pairwise distances rather than square of the data and has been used as an alternative to the standard deviation in many fields. Where the standard deviation is motivated from optimality results in independent random sampling from the normal distribution, an analysis dating back to Fisher; see, Stigler (1973), the GMD may be more appropriate in case of a small departure from normality where it is known that the GMD has asymptotic relative efficiency of 98% at the normal distribution and more efficient than standard deviation if the normal distribution is contaminated by a small fraction; see, David (1986) and Gerstenberger and Vogel (2014). It may also offer certain pedagogical advantages; see, Algina et al. (2005). For extensive discussion and comparisons; see Gerstenkorn and Gerstenkon (2003) and Yitzhaki (2003) and the references therein.

The population GMD is defined as

$$\Delta = E|Y_1 - Y_2|$$

It can be estimated from the sample using many formulas such as

$$\delta = \frac{1}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}|y_i - y_j| = \frac{2}{n(n-1)}\sum_{i=1}^{n-1}\sum_{j=i+1}^{n}|y_i - y_j|$$

See, for example, Yitzhaki (2003).

A random variable has a normal distribution with location parameter $-\infty < \mu < \infty$ and scale $\sigma > 0$ if its probability density function is

$$f(y) = \frac{1}{\sigma\sqrt{2\pi}}\exp\left(-\frac{(y-\mu)^2}{2\sigma^2}\right), \quad -\infty < y < \infty$$

The normal distribution has

$$E(Y) = \mu, \qquad V(Y) = \sigma^2 \text{ and } \Delta(Y) = 2\sigma/\sqrt{\pi}$$

Therefore,

$$\sigma = \Delta\sqrt{\pi}/2$$

This will be used later in simulation studies.

Because of the presence of absolute function in GMD, the exact partition is challenging. Therefore, an exact partition of the total sum of differences (TSD)

E-mail:qm.grorup70@gmail.com

about mean and median into exact between sum of differences (BSD) and exact within sum of differences (WSD) is derived by finding form of GMD that does not depend on the absolute function. It is known that the second L-moment is half of GMD; see, Elamir and Seheult (2003). By using this relationship it has been expressed TSD as a sum of weighted data with total of weights is zero. Therefore, by getting rid of absolute function the exact partitions are obtained. Because the sum of weights is zero, the TSD does not depend on any fixed location while BSD and WSD depend on the location and this is logic where the TSD is the total of all pairwise of the distances.

Moreover, the sampling distributions of the BSD and WSA are studied empirically under the assumption of normal distribution using variance-gamma distribution. Consequently, ANOMD is used to test for equal population means and medians. Moreover, two measures of effect sizes are re-expressed and studied in terms of TSD, BSD and WSD.

Exact GMD partitions

Let $Y_1, Y_2, \dots, Y_n$be a random sample from a continuous distribution with, density function $f(y)$, quantile function $y(F) = F^{-1}(y) = Q(F),\ 0 < F < 1$, cumulative distribution function $F(y) = F$ and $Y_{1:n}, \dots, Y_{n:n}$the order statistics. There is a relationship between the second L-moment and GMD where the second L-moment is a half GMD, therefore

$$\lambda_2 = \frac{1}{2} E(Y_{2:2} - Y_{1:2})$$

Hence,

$$\Delta = E(Y_{2:2} - Y_{1:2})$$

From Elamir and Seheult (2004) this can be estimated as

$$\delta = \frac{2}{n(n-1)} \sum_{i=1}^{n} (2i - n - 1) Y_{i:n}$$
$$= \frac{1}{n} \sum_{i=1}^{n} \frac{2(2i - n - 1)}{(n-1)} Y_{i:n}$$

Assume there are G different groups with individuals in each group y_{ig}, $i = 1,2, \dots, n_g$, $n = n_1 + \cdots + n_G$ and $g = 1, \dots, G$. Let $y_{gi} - \bar{y}$ is the total deviation ($\bar{y} = \sum_g^G \sum_i^{n_g} y_{ig} / n$), $\bar{y}_g - \bar{y}$ is the deviation of grouped mean ($\bar{y}_g = \sum_{i=1}^{n_g} y_{ig} / n_g$) around total mean, and $y_{ig} - \bar{y}_g$ is the deviation of individuals around the grouped mean. The GMD is

$$\delta = \frac{1}{n} \sum_{i=1}^{n} \frac{2(2i - n - 1)}{(n-1)} y_{i:n}$$

This can be rewritten without order and taking the rank of y as

$$\delta = \frac{1}{n} \sum_{i=1}^{n} \frac{2(2\,\mathrm{rank(y)} - n - 1)}{(n-1)} Y_i = \frac{1}{n} \sum_{i=1}^{n} w_i Y_i$$

This is a weighted average form where

$$w_i = \frac{2(2\,\mathrm{rank}(y) - n - 1)}{(n-1)}$$

Note that,

$$\sum_{i=1}^{n} w_i = 0$$

Therefore, the total sum of differences is considered as

$$TSD = \sum_{i=1}^{n} w_i Y_i$$

This is the most important equation to obtain the exact analysis of total differences as follows.

Theorem 1

The total sum of differences partitions about mean $(\bar{Y})$ into exact between sum of differences and exact within sum of differences is

$$TSD = BSD + WSD$$

where

$$TSD = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} (Y_{ig} - \bar{Y}),$$

$$BSD = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} (\bar{Y}_g - \bar{Y}),$$

$$WSD = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} (Y_{ig} - \bar{Y}_g)$$

and

$$w_{ig} = \frac{2(2\,\mathrm{rank}(y_{ig}) - n - 1)}{(n-1)}$$

Proof:

Where $\sum w = 0$, the total sum of differences (TSD) is

$$TSD = \sum_{i=1}^{n} w_i Y_i = \sum_{i=1}^{n} w_i (Y_i - \bar{Y})$$

The TSD does not depend on any fixed location. By adding and subtracting $\bar{Y}_g$ and taking the summation over both g and i then

$$TSD = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(Y_{ig} - \bar{Y}_g + \bar{Y}_g - \bar{Y} \right)$$

Therefore,

$$\sum_{g=1}^{G} \sum_{i=1}^{n_g} w_i \left(Y_{ig} - \bar{Y} \right) = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(\bar{Y}_g - \bar{Y} \right) + \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(Y_i - \bar{Y}_g \right)$$

Theorem 2

The total sum of differences partitions about median $(\tilde{Y})$ into between sum of differences and within sum of differences is

$$TSD_{med} = BSD_{med} + WSD_{med}$$

where

$$TSD_{med} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(Y_{ig} - \tilde{Y} \right),$$

$$BSD_{med} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(\tilde{Y}_g - \tilde{Y} \right)$$

and

$$WSD_{med} = \sum_{g=1}^{G} \sum_{i=1}^{n_g} w_{ig} \left(Y_{ig} - \tilde{Y}_g \right)$$

Proof: same as mean.

Explanation of the ANOMD

In one way analysis of variance (ANOVA) the total sum of squares can be written as

$$\sum_{g} \sum_{i} \left[\left(Y_{ig} - \bar{Y} \right) \right]^2 = \sum_{g} \sum_{i} \left(\bar{Y}_g - \bar{Y} \right)^2 + \sum_{g} \sum_{i} \left(Y_{ig} - \bar{Y}_g \right)^2$$

while in ANOMD

$$\sum_{g} \sum_{i} w_{ig} \left(Y_{ig} - \bar{Y} \right) = \sum_{g} \sum_{i} w_{ig} \left(\bar{Y}_g - \bar{Y} \right) + \sum_{g} \sum_{i} w_{ig} \left(Y_{ig} - \bar{Y}_g \right)$$

In ANOMD there is no square where it is replaced by the weights and that ensures stability in statistical inference. Another important property of ANOMD, TSD does not depend on any fixed location while the partitions (BSD and WSD) depend on the location and this is logic where TSD is the total sum of all pairwise distances.

Illustrative example

To have an idea on how the method work. Table 1 shows TSD partitions about mean and median for a hypothetical data. Note that, $TSD = 95$ and $BSD + WSD = 62 + 33 = 95$ by using mean. Also, for median $BSD + WSD = 42 + 53 = 95$. Both of them give exact partitions.

Applications

One way ANOMD is introduced and used to test for equal population means and medians under the following assumptions.

1. The observations are random and independent samples from the populations.
2. The distributions of the populations from which the samples are selected are normal distribution.
3. The Δ's of the distributions in the populations are equal.

It is difficult to obtain the exact sampling distributions for BSD and WSD, therefore, a general distribution is chosen to fit these sampling distributions. One of the families that connected to chi-square and gamma distributions is the variance-gamma distribution; see, Kotz et al. (2001). More recently, the variance gamma model became popular among some financial researchers, due to its simplicity, flexibility, and an excellent fit to empirical data; see, Madan and Seneta (1990) and Madan et al. (1998). The variance-gamma distribution will be used to fit the sampling distributions of BSD and WSD via method of moments where the moments of the data and the distribution will be equated.

Fitting sampling distributions

The random variable Y is said to have Variance-Gamma (VG) with parameters $c, \theta \in R$, $\nu, \sigma > 0$, if it has probability density function given by

$$f(y; c, \sigma, \theta, \nu)$$

$$= \frac{2e^{\frac{\theta(y-c)}{\sigma^2}}}{\sigma\sqrt{2\pi}\nu^{\frac{1}{\nu}}\Gamma\left(\frac{1}{\nu}\right)}\left[\frac{|y-c|}{\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}\right]^{\frac{1}{\nu}-1} K_{\frac{1}{\nu}-\frac{1}{2}}\left[\frac{|y-c|\sqrt{\frac{2\sigma^2}{\nu}+\theta^2}}{\sigma^2}\right],$$

$$y \in R$$

Where $K_\nu(x)$ is a modified Bessel function of the third kind; see, for example, Seneta, E. (2004), Kotz, et al. (2001) and Gradshteyn and Ryzhik (1980).

Note that there are other versions of this distribution available but this version is chosen because there is a software package in R called *gamma-variance* based on this version that be used to obtain all the simulations and graphs. The moments of this distribution are

$$E(Y) = c + \theta,$$

$$V(Y) = \sigma^2 + \nu\theta^2,$$

$$sk = \frac{2\theta^3\nu^2 + 3\sigma^2\theta\nu}{\sqrt{(\theta^2\nu + \sigma^2)^3}},$$

And

$$ku = 3 + \frac{3\sigma^4\nu + 12\sigma^2\theta^2\nu^2 + 6\theta^4\nu^3}{(\theta^2\nu + \sigma^2)^2}$$

Table 1 TSD partition into BSD and WSD for a hypothetical data using mean and median

				TSD	BSD	WSD	BSD_{med}	WSD_{med}
g	i	y_{gi}	w_{ig}	wy	$w(\bar{y}_g - \bar{y})$	$w(y - \bar{y}_g)$	$w(\tilde{y}_g - \tilde{y})$	$w(y - \tilde{y}_g)$
1	1	10	0.5	5	4.67	-5	3.5	-2.5
$\bar{y}_1 = 20, \tilde{y}_1 = 15$	2	15	1.5	22.5	14	-7.5	10.5	0
	3	35	2	70	18.67	30	14	40
2	1	3	-1.5	-4.5	4	7.5	0	7.5
$\bar{y}_2 = 8, \tilde{y}_2 = 8$	2	8	0	0	0	0	0	0
	3	13	1	13	-2.67	5	0	5
3	1	2	-2	-4	13.33	4	3	4
$\bar{y}_3 = 4, \tilde{y}_3 = 4$	2	4	-1	-4	6.67	0	4	0
	3	6	-0.5	-3	3.33	-1	2	-1
Total			0	95	62	33	42	53
$\bar{y}$ = 10.7								

Table 2 simulated mean, variance, skewness and kurtosis for $U_1$and U_2 with different values of G and n from normal distribution(μ, $\Delta\sqrt{\pi}/2$) and the number of replications is 10000

	Simulated moments for U_1				Simulated moments for U_2			
n	G	$mean$	var	sk	ku	$mean$	var	sk	ku
30	3	2.058	4.013	1.685	6.672	27.980	17.842	0.104	3.081
50	5	4.047	8.112	1.289	5.271	46.001	29.870	0.081	3.073
80	8	7.042	13.84	0.973	4.325	73.011	47.085	0.030	3.051
100	10	9.031	18.11	0.881	4.050	90.974	60.712	0.016	3.031
150	15	14.02	27.83	0.652	3.671	135.98	87.834	0.012	3.019
45	3	2.015	4.032	1.709	6.932	42.940	26.03	0.052	3.051
75	5	4.021	7.987	1.302	5.328	70.897	42.97	0.045	3.039
120	8	7.032	13.98	0.975	4.271	113.01	67.84	0.031	3.024
150	10	9.023	18.05	0.896	3.747	140.81	86.06	0.018	3.015
225	15	14.02	27.95	0.721	3.671	211.07	128.88	0.011	3.010
75	3	2.004	4.10	1.88	7.91	72.97	41.33	0.053	3.036
125	5	3.998	8.03	1.265	5.131	120.96	69.22	0.050	3.024
200	8	7.010	13.95	0.991	4.371	193.07	111.76	0.023	3.022
250	10	9.030	18.18	0.910	3.984	240.88	135.37	0.022	3.011
375	15	14.04	27.98	0.705	3.766	360.89	200.83	0.018	3.007

This distribution is defined over the real line and has many distributions as special cases or limiting distributions such as Gamma distribution in the limit $\sigma \downarrow 0$ and $c = 0$, Laplace distribution as $\theta = 0$ and $\upsilon = 2$ and normal distribution as $\theta = 0$, $\nu = 1/r$ and $r \to \infty$.

Note that if $a > 0$ then

$$aY \sim VG(ac, a\sigma, a\theta, \nu)$$

The Gamma distribution for the random variable $Y$is defined as

$$f(y;k,\omega) = \frac{1}{\Gamma(k)\omega^k} y^{k-1} e^{-\frac{y}{\omega}},\ \ y > 0, k, \omega > 0$$

where k and ω are the shape and scale parameters and the moments are

$$E(Y) = k\omega,\ \ V(Y) = k\omega^2,\ sk = \frac{2}{\sqrt{k}}\quad \text{and}\ \ ku = 3 + \frac{6}{k}$$

GMD about mean

The scaled BSD can be written as

$$U_1 = \frac{BSD}{\Delta} = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} w_{ig}\left(\bar{Y}_g - \bar{Y}\right)}{\Delta}$$

Since BSD depends on one parameter G , the moments of BSD is used to fit the sampling distribution of U_1 based on VG distribution. For this purpose a simulation study is conducted to obtain the first four moments for U_1 based on simulated data from normal distribution with different values of n and G. Table 2 gives the simulated first four moments of U_1.

From Table 2 it is noted that there is a pattern between the mean and the variance for different n and G values where the mean is approximately $G - 1$ and the variance is twice the mean $(2(G-1) + 1/G)$ whatever the values of n. Therefore VG distribution is used to fit the sampling distribution of U_1 as

$$U_1 = \frac{BSG}{\Delta} \approx \text{VG}\left(c = 0, \sigma = \frac{1}{G}, \theta = (G-1), \nu = \frac{2}{(G-1)}\right) \approx \Gamma\left(k = \frac{G-1}{2}, 2\right)$$

Hence,

$$T_1 = \frac{BSG}{\Delta} \approx \text{VG}\left(c = 0, \sigma \downarrow 0, \theta = 1, \nu = \frac{2}{(G-1)}\right) \cong \Gamma\left(\frac{G-1}{2}, \frac{2}{G-1}\right)$$

The first two moments are

$$E(T_1) = 1 \text{ and } V(T_1) = \frac{2}{(G-1)}$$

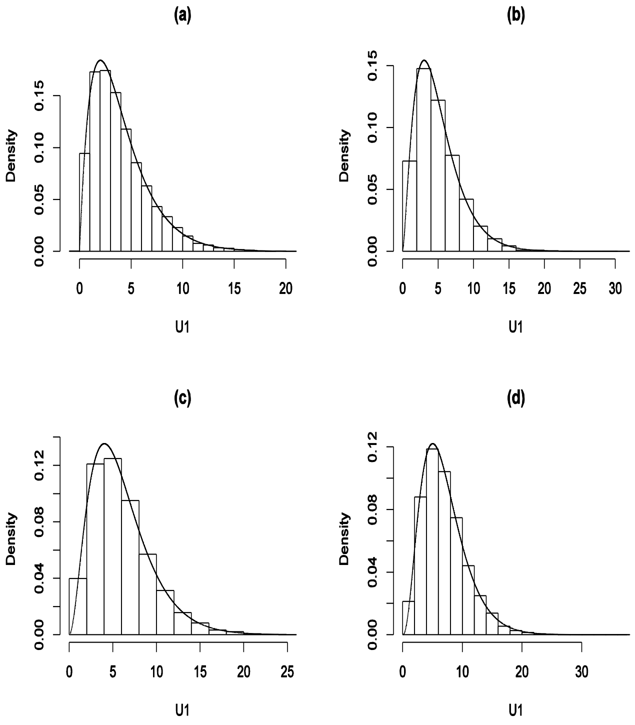

Figure 1 histogram of U_1 based on simulated data from normal distribution with VG distribution superimposed and (a) $G = 5$ and $n = 150$(b) $G = 6$ and $n = 180$ (c) $G = 7$ and $n = 210$ and (d) $G = 8$ and $n = 240$.

Moreover, Figure 1 shows the histogram of U_1 based on simulated data from normal distribution with fitting VG superimposed. The VG clearly gives a very good fit to U_1 for different values of G and n. The scaled WSD can be written as

$$U_2 = \frac{WSD}{\Delta} = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} w_{ig}\left(Y_{ig} - \bar{Y}_g\right)}{\Delta}$$

Since WSG depends on two parameters G and n, the moments of WSG is used to fit the sampling distribution of U_2 based on VG distribution. For this purpose a simulation study is conducted to obtain the first four moments for U_2 based on simulated data from normal distribution (μ,) with different values of n and G. Table 2 gives the simulated first four moments of U_2 and it is noted that as expected there is a pattern between the mean and the variance for different n and G values where the mean is approximately $n - G + 1$ and the variance is half n plus G. Consequently VG distribution is used to fit the sampling distribution of U_2 as

$$U_2 = \frac{WSD}{\Delta} \approx \text{VG}\left(c = 0, \sigma \downarrow 0, \theta = (n - G + 1), \nu = \frac{n + 2G}{2(n - G + 1)^2}\right) \approx \Gamma\left(k = \frac{2(n - G + 1)^2}{(n + 2G)}, \frac{(n + 2G)}{2(n - G + 1)}\right)$$

Therefore,

$$T_2 = \frac{WSD}{(n - G + 1)\Delta} \approx \text{VG}\left(c = 0, \sigma \downarrow 0, \theta = 1, \nu = \frac{n + 2G}{2(n - G + 1)^2}\right)$$

The first two moments are

$$E(T_2) = 1 \text{ and } V(T_2) = \frac{n + 2G}{2(n - G + 1)^2}$$

Moreover, Figure 2 shows the histogram of U_2 based on simulated data from normal distribution with fitting VG superimposed. The VG clearly gives a very good fit to U_2 for different values of G and n.

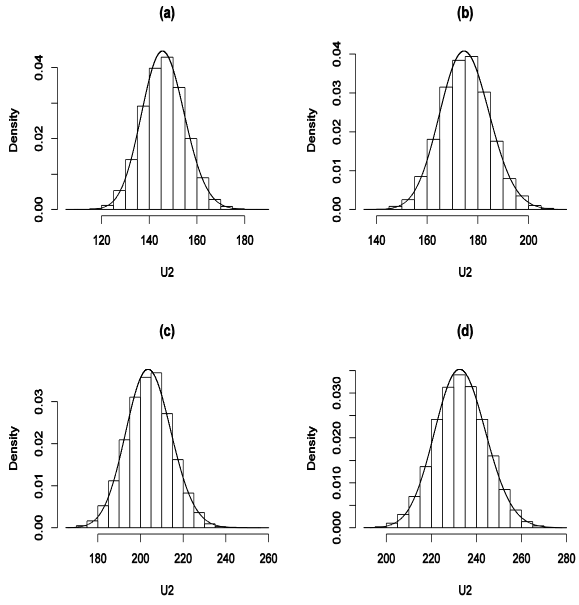

Figure 2 histogram of U_2 based on simulated data from normal distribution with VG distribution superimposed (a) $G = 6, n = 48$, (b) $G = 5, n = 40$, (c) $G = 4, n = 32$, and (d) $G = 3, n = 24$

The ratio of mean BSD to mean WSD can be expressed as

$$R = T_1/T_2 = \frac{(n-G+1)BSD}{(G-1)WSD} = \frac{MBSD}{MWSD}$$

An approximation is obtained based on gamma-variance distribution by using

$$E(R_{mean}) \approx \frac{E(T_1)}{E(T_2)}$$

and

$$V(R_{mean}) \approx \frac{V(T_2)E^2(T_1)}{E^4(T_2)} + \frac{V(T_1)}{E^2(T_2)}$$

This gives

$$R \approx \mathrm{VG}\left(c = 0, \sigma = \frac{\sqrt{0.5n+G}}{(n-G+1)}, \theta = 1, \nu = \frac{2}{(G-1)}\right)$$

Table 3 simulated mean, variance, skewness and kurtosis for $U_3$and U_4 with different values of G and n from normal distribution($\mu, \Delta\sqrt{\pi}/2$) and number of replications is 10000

		Simulated moments for U_3				Simulated moments for U_4			
n	G	$mean$	var	sk	ku	$mean$	var	sk	ku
30	3	2.277	5.121	1.700	7.101	27.755	18.293	0.090	3.081
50	5	4.501	10.10	1.170	5.332	45.620	30.652	0.063	3.056
80	8	7.754	17.80	0.976	4.493	72.384	49.307	0.031	3.050
100	10	9.961	22.13	0.773	3.715	90.168	63.336	0.020	3.034
150	15	15.37	35.95	0.628	3.556	135.02	90.886	0.013	3.016
45	3	2.214	5.1960	1.745	7.078	42.772	26.354	0.048	3.084
75	5	4.440	10.311	1.252	5.317	70.561	44.378	0.035	3.067
120	8	7.747	17.766	0.966	4.456	112.39	71.199	0.030	3.049
150	10	9.943	22.819	0.803	3.955	140.02	89.325	0.028	3.035
225	15	15.42	36.406	0.683	3..658	209.96	132.33	0.020	3.019
75	3	2.254	5.390	1.831	8.017	72.885	42.469	0.043	3.064
125	5	4.450	10.340	1.218	5.056	120.70	70.175	0.031	3.070
200	8	7.724	17.871	0.953	4.231	192.18	112.12	0.027	3.055
250	10	9.908	23.120	0.835	3.850	240.08	137.74	0.021	3.021
375	15	15.47	37.212	0.687	3.683	359.53	211.95	0.019	3.001

GMD about median

The scaled BSD about median can be written as

$$U_3 = \frac{BSG_{med}}{\Delta} = \frac{\sum_{g=1}^{G}\sum_{i=1}^{n_g} w_{ig}\left(\tilde{Y}_g - \tilde{Y}\right)}{\Delta}$$

Since BSD_{med} depends on one parameter G, the moments of BSD_{med} is a good choice to be used to fit the sampling distribution of U_3 based on VG distribution. For this purpose a simulation study is conducted to obtain the first four moments for U_2 using simulated data from normal distribution $(\mu,$ G) with different values of n and G. Table 4 gives the simulated first four moments of U_3.

From Table 3 it is noted that there is a pattern between the mean and the variance for different n and

G values where the mean is approximately $1.1G - 1$ and the variance is approximately $2.6G - 2.5$ whatever the value of n. Therefore VG distribution is used to fit the sampling distribution of U_3 as

$$U_3 = \frac{BSD_{med}}{\Delta} \approx \text{VG}\left(c = 0, \sigma = \sqrt{(4g-5)/10}, \theta = (1.1G - 1), \nu = \frac{2}{(1.1G - 1)}\right)$$

Hence,

$$T_3 = \frac{BSD_{med}}{(1.1G - 1)\Delta} \approx \text{VG}\left(c = 0, \sigma = \frac{\sqrt{(4G-5)/10}}{(1.1G - 1)}, \theta = 1, \nu = \frac{2}{(1.1G - 1)}\right)$$

The first two moments are

$$E(T_3) = 1, \text{and} \quad V(T_3) = \frac{(4G - 5)}{10(1.1G - 1)^2} + \frac{2}{(1.1G - 1)}$$

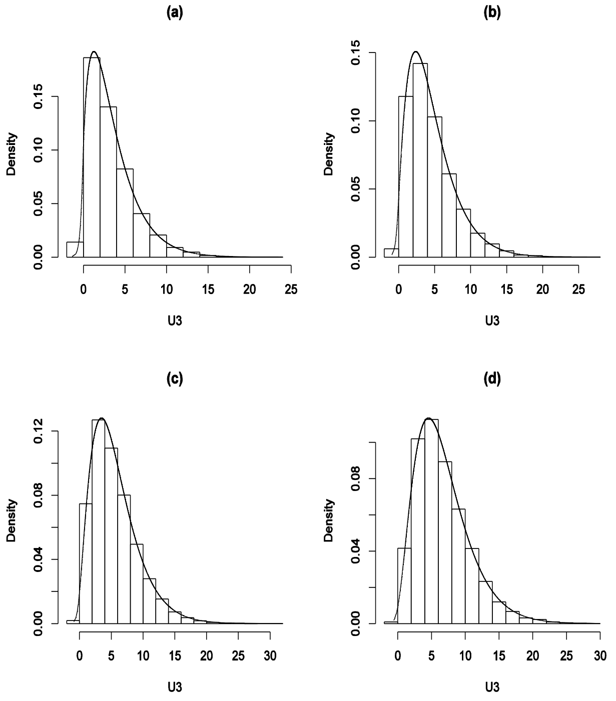

Figure 3 histogram of U_3 based on simulated data from normal distribution with VG distribution superimposed and (a) G = 4 and n = 100 (b) G = 5 and n = 125 (c) G = 6 and n = 150 and (d) G = 7 and n = 175

Moreover, Figure 3 shows the histogram of U_3 based on simulated data from normal distribution with fitting VG superimposed. The VG clearly gives a very good fit to U_3 for different values of G and n. The WSD about median can be written as

$$U_4 = \frac{WSD_{med}}{\Delta} = \frac{\sum_{i=1}^{n}\sum_{g=1}^{G} w_{ig}\left(Y_{ig} - \tilde{Y}_g\right)}{\Delta}$$

Since WSD_{med} depends on two parameters G and n, the moments of WSD_{med} could be used to fit the sampling distribution of U_4 based on VG distribution. For this purpose a simulation study is conducted to obtain the first four moments for U_4 based on simulated data from normal distribution with different values of n and G. Table 3 gives the simulated first four moments of U_4. From Table 3 it is noted that there is a pattern between the mean and the variance for different n and G values where the mean is approximately $n - 1.075G + 1$ and the variance is approximately $\frac{n}{2} + g + 0.025n$.

Therefore VG distribution is used to fit the sampling distribution of U_4 as

$$\begin{aligned} &U_4 \\ &\approx \text{ VG}\left(c = 0, \sigma \downarrow 0, \theta = (n - 1.1G + 1), \nu \right. \\ &\left. = \frac{(0.525n + g)}{(n - 1.1G + 1)^2}\right) \\ &\approx \Gamma\left(\frac{(n - 1.1G + 1)^2}{(0.525n + g)}, \frac{(0.525n + g)}{(n - 1.1G + 1)}\right) \end{aligned}$$

Therefore,

$$\begin{aligned} T_4 = \frac{PWSA_M}{(n - 1.1G + 1)\Delta} \\ &\approx \text{ VG}\left(c = 0, \sigma \downarrow 0, \theta = 1, \nu \right. \\ &\left. = \frac{(0.525n + G)}{(n - 1.1G + 1)^2}\right) \end{aligned}$$

The first two moments are

$$E(T_4) = 1 \quad \text{and} \quad V(T_4) = \frac{(0.525n + G)}{(n - 1.1G + 1)^2}$$

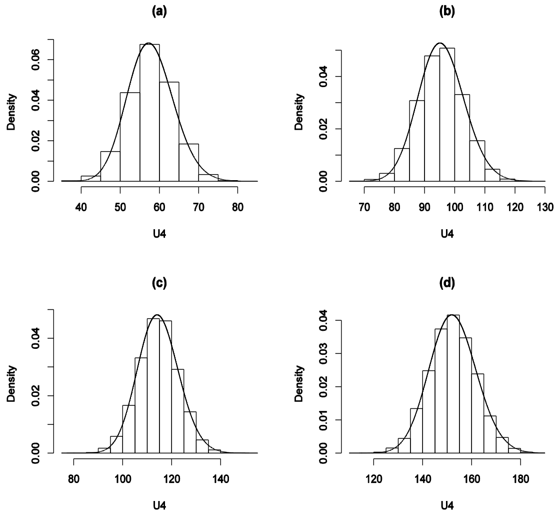

Figure 4 histogram of U_4 based on simulated data from normal distribution with VG distribution superimposed and (a) $G = 3$ and $n = 60$(b) $G = 5$ and $n = 100$ (c) $G = 6$ and $n = 120$ and (d) $G = 8$ and $n = 160$

Moreover, Figure 4 shows the histogram of U_4 based on simulated data from normal distribution with fitting VG superimposed. The VG clearly gives a very good fit to U_4 for different values of G and n. Note that more simulation results for different G and n are available from the author upon request.

The ratio between BSD and WSD can be expressed as

$$R_{med} = T_3/T_4 = \frac{(n - 1.1G + 1)BSD}{(1.1G - 1)WSD} = \frac{MBSD_{med}}{MWSD_{med}}$$

An approximation to R_{med} is obtained based on gamma-variance distribution by using

$$E(R_{med}) \approx \frac{E(T_3)}{E(T_4)} = 1$$

And

$$V(R_{med}) \approx \frac{V(T_4)E^2(T_3)}{E^4(T_4)} + \frac{V(T_3)}{E^2(T_4)}$$

Hence, the approximation gamma-variance is

$$\begin{aligned} &R_{med} \\ &\approx \text{VG}\left(c = 0, \sigma \right. \\ &= \sqrt{\frac{0.525n + G}{(n - 1.1G + 1)^2} + \frac{(4G - 5)}{10(1.1G - 1)^2}}, \theta = 1, \nu \\ &\left. = \frac{2}{(1.1G - 1)}\right) \end{aligned}$$

Tests for equal means and medians

Test for equal medians

The null hypothesis H_0tested in one way ANOMD is that the population medians from which the G samples are selected are equal

$$H_0: \nu_1 = \nu_2 = \nu_3 = \cdots = \nu_G$$

The alternatively hypothesis H_a is that at least two of the group medians are significantly different. Table 4 gives a summary of ANOMD for medians.

Table 4 summary ANOMD for medians

variation	Sum of MD	Divisor	MD estimate (mean difference)	R_{med}
Between	BSD_{med}	$1.1G-1$	$MBSD_{med}=\frac{BSD_{med}}{1.1G-1}$	$\frac{MBSD_{med}}{MWSD_{med}}$
Within	WSD_{med}	$n-1.1G+1$	$MWSD_{med}=\frac{WSD_{med}}{n-1.1G+1}$	
Total	TSA			

Table 5 research and development expenditures for three different Levels and normal goodness of fit using Shapiro-Wilk test

Level			normal test			
Low	**Moderate**	**High**	**Shapiro-Wilk test**			
71.0	72.5	78.0	Group	p-value		
66.5	70.0	84.0	Low	0.563		
77.5	77.0	85.0	Moderate	0.97		
69.0	71.5	90.5	High	0.34		
73.5	73.0	75.5				
68.0	71.0	78.5		Low	Moderate	High
70.0	69.5	81.5	Mean	71.6	73	81
68.5	78.0	79.0	Median	71	72.5	80
73.5	71.5	79.5	GMD	4.75	5.29	4.72
65.5	81.0	80.0				
72.0	64.5	84.5				
77.0	78.5	77.0				
70.5	74.0	84.5				
77.0	67.0	80.5				
75.0	76.5	75.5				

To test for the assumption of normal distribution, the function *Shapiro.test()* in *R-software* is used. One way ANOMD is used to examine whether or not the median and mean productivity improvement differs according to the level of research and development expenditures for a sample of firms producing electronic computing equipment. The firms were classified according to the level of their average expenditure (Low, moderate, high).

The productivity improvement is measured on a scale from 0 to 100. See Table 5 for the sample data with means, medians and GMD reported for each of the three groups. To test the assumption of normal distribution, Shapiro-Wilk test is used from R-software. The results for the three groups are given in Table 5 where p-values more than 0.01, 0.05 and 0.10, therefore, the assumption of normality cannot be rejected. Because the maximum GMD to minimum GMD is 1.1, the assumption of homogeneity of GMD's could not be rejected.

Table 6 ANOMD of testing equal medians for research and development expenditures

variation	Sum of difference	Divisor	GMD estimate (mean difference)	R_{med}	$qVG^{*}_{0.95}$
Between	150.61	2.3	65.48	18.84	3.12
Within	148.38	42.7	3.47		
Total	299				

*This value from Variance-Gamma package in R-software

where $R_{med} = 18.84 > Critical = qVG_{0.95} = 3.12$, H_o is rejected, i.e., this indicates that not all three groups resulted in the same research and development expenditures.

Test for equal means

The null hypothesis H_0 tested in one way ANOMD is that the population means from which the G samples are selected are equal

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \cdots = \mu_G$$

The alternatively hypothesis H_a is that at least two of the group means are significantly different.

Table 7 gives summary of ANOMGD for means.

Table 7 summary ANOMD for means

variation	Sum of Difference	Divisor	GMD estimate (mean difference)	R_M
Between	BSD_M	$G-1$	$MBSD_M = \frac{BSD_M}{G-1}$	$\frac{MBSD_M}{MWSD_M}$
Within	WSD_M	$n-G+1$	$MWSD_M = \frac{WSD_M}{n-G+1}$	
Total	TSD			

Table 8 gives ANOMD to test equal means for standardized test scores.

Table 8 ANOMD of testing equal means for research and development expenditures

Variation	Sum of difference	Divisor	GMD estimate (mean difference)	R_{mean}	$qG^{*}_{0.95}$
Between	155.98	2	77.99	23.4	3.07
Within	143.02	43	3.326		
Total	299				

*This value from Variance-Gamma package in R-software

where $R_{mean} = 23.4 > Critical = qVG_{0.95} = 3.08$, H_o is rejected, i.e. this indicates that not all three groups resulted in the same average expenditures.

Effect size

Effect size (ES) is a measure of practical significance where it is defined as the degree to which a phenomenon exists where any observed difference between, for example, two sample means can be found to be statistically significant when the sample sizes are sufficiently large. In such a case, a small difference with little practical importance can be statistically significant. On the other hand, a large difference with apparent practical importance can be non-significant when the sample sizes are small. Therefore, ES provide another measure of the magnitude of the difference expressed in standard deviation units in the original measurement. Thus, with the test of statistical significance and the interpretation of the effect size (ES), the researcher can address issues of both statistical significance and practical importance. Standardized ES measures typically employed in behavioural and social sciences research; see, Algina et al. (2005) and Cohen (1988).

The first type of standardized ES measure is

$$\eta^2 = \frac{S_B^2}{S_B^2 + S_W^2}$$

Where S_B^2 and S_W^2 are between group variance and within group variance; see, for example, Cohen (1988, 1994). Another measure of the strength of the association between the independent variable and the dependent variable in ANOVA is ω^2 that indicates the proportion of the total variance in the dependent variable that is accounted for by the levels of the independent variable. This is analogous to the coefficients of determination r^2. The formula for ω^2 is

$$\omega^2 = \frac{SS_B - (G-1)MS_W}{SS_T + MS_W}$$

See, for example Cohen (1988).

These two measures are redefined in terms of ANOMD as

$$\eta_{MD} = \frac{BSD}{BSD + WSD}$$

and

$$\omega_{MD} = \frac{BSD - (G-1)MWSD}{TSD + MWSD}$$

Table 9: Effect sizes of ANOVA and ANOMD for research and development expenditures

	Measure	
	η^2	ω^2
ANOVA	0.505	0.475
	η_{MD}	ω_{MD}
ANOMD mean	0.522	0.494
ANOMD median	0.504	0.472

From Table 9 the independent variable in ANOVA accounts for 50.5% of the total variation in the dependent variable while the independent variable in ANOMD accounts for 52.2% of the total variation in the dependent variable.

Randomized block designs

When the available experimental units are not homogeneous, grouping the experimental units into blocks of homogeneous units will reduce the experimental errors and increase the range of validity for inferences about the treatment effects. A randomized block design is a restricted randomization design in which the experimental units are first sorted into homogeneous groups, called blocks, and the treatments are then assigned at random within the blocks; see, Neter et al. (1996).

The model for a randomized complete block design containing the comparison of no interaction effects, when both the block and treatment effects are fixed and there are n blocks (BL) and r treatments (TR), is as

$$Y_{ij} = \mu_{..} + \rho_i + \tau_j + \varepsilon_{ij}$$

The ANOMD for a randomized complete block design can be written as

$$TSD = SDBL + SDTR + SDBL.TR$$

where

$$TSD = \sum_{j=1}^{r}\sum_{i=1}^{n} w_{ji}\left(Y_{ij} - \bar{Y}_{..}\right)$$

$$SDBL = \sum_{j=1}^{r}\sum_{i=1}^{n} w_{ij}(\bar{Y}_{i.} - \bar{Y}_{..}),$$

$$SDTR = \sum_{j=1}^{r}\sum_{i=1}^{n} w_{ij}\left(\bar{Y}_{.j} - \bar{Y}_{..}\right)$$

and

$$SDBL.TR = \sum_{j=1}^{r}\sum_{i=1}^{n} w_{ij}\left(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..}\right)$$

The investigation is under way to find the sampling distributions for $SDBL$ and $SDBL.TR$.

Conclusion

The ANOMD was derived by partition total sum of differences into exact between sum of differences and exact within sum of differences. It had been shown that the TSD had been expressed as a linear combination of the data instead of square in ANOVA. ANOMD is used to test for population means and medians. Moreover, the variance-gamma distribution was used to fit the sampling distribution of BSD and WSD. Also, ANOMD offered a very effective way to find out the shifts in means and medians graphically for all groups and for each group.

However, the effect sizes were re-expressed and studied using ANOMD where it had been shown that the independent variable using ANOMD might account for the total variation in the dependent variable more than using ANOVA Model. Most importantly, it is easy to generalize ANOMD to other designs.

References

Algina, J., Keselman, H.J., & Penfield, R.D. (2005). An alternative to Cohen's standardized mean difference effect size: A robust parameter and confidence interval in the two independent group's case. Psychological Methods, Vol. 10, pp. 317-328.

Cohen, J. (1988). Statistical power analysis for the behavioral sciences(2nd ed.). Hillsdale, NJ: Erlbaum.

Cohen, J. (1994). The earth is round (p<. 05). American Psychologist, Vol.49, pp.997-1003.

David, H. A. (1968). Gini's mean difference rediscovered. Biometrika, Vol.55, pp.573-575.

Elamir, E.A.H. and Seheult, H. (2003). Trimmed L-moments. Computational Statistics & Data Analysis, Vol.4, pp.299-314.

Elamir, E.A.H and Seheult, H. (2004). Exact variance structure of sample L-moments. Journal of Statistical Planning and Inference, 124, pp.337-359.

Gradshteyn, I.S., and Ryzhik, I.M., 1980. Table of Integrals, Series, and Products. Academic Press.

Gerstenberger, C. and Vogel, D. (2014). On the efficiency of Gini's mean difference. Cornel University Library, arxiv.org/abs/1405.5027.

Gerstenkorn, T. and Gerstenkorn, J. (2003). Gini's mean difference in the theory and application. Statistica, IXIII, pp.469-488.

Kotz, S, Kozubowski, T. J., and Podgórski, K. (2001). The Laplace Distribution and Generalizations. *Birkhauser*, Boston.

Madan, D.B., Carr, P. and Chang, E.C. (1998). The variance gamma process and option pricing, European Finance Review, 2, pp. 74-105.

Madan, D.B. and Seneta, E. (1990). The variance gamma (V.G.) model for share markets returns, Journal of Business, Vol.63, pp.511-524.

Neter, J., Kutner, H., Nachtsheim, C. and Wasserman, W. (1996). Applied linear statistical models. 4th ed., McGraw-Hill.

Seneta, E. (2004). Fitting the variance-gamma model to financial data. Journal of Applied Probability. 41A: pp.177-187

Stigler, S. M. (1973). Studies in the history of probability and statistics. XXXII. Laplace, Fisher, and the discovery of the concept of sufficiency. Biometrika, Vol. 60, pp.439–445.

S. Yitzhaki. (2003). Gini's mean difference: A superior measure of variability for non-normal distributions. Metron, Vol.61, pp. 285-316.