The Impact of non-Equidistance on Anova and Alternative Methods

Bjorn Lantz
Chalmers University of Technology, Gothenburg, Sweden
Bjorn.Lantz@chalmers.se

Abstract: The normality assumption behind ANOVA and other parametric methods implies not only mound shape, symmetry, and zero excess kurtosis, but also that data are equidistant. This paper uses a simulation approach to explore the impact of non-equidistance on the performance of statistical methods commonly used to compare locations across several groups. These include the one-way ANOVA and its robust alternatives, the Brown-Forsythe test, and the Welch test. We show that non-equidistance does affect these methods with respect to both significance level and power, but the impact differs between the methods. In general, the ANOVA is less sensitive to non-equidistance than the other two methods are and should therefore be the primary choice when analyzing potentially non-equidistant data.

Keywords: Likert-type scale; equidistance; Monte Carlo simulation; ANOVA

1. Introduction

Likert items (usually referred to as Likert-type scales) are often assumed by business researchers to produce data with interval properties. The reason, in most cases, seems to be that there is a wider range of statistical methods available for data analyses if the interval assumption is valid, that is, if data are equidistant. However, research shows that subjects seldom perceive Likert-type scales as intervals, neither in general nor with respect to specific constructs. For example, it has been shown that subjects' perceptions of such scales are related to the way verbal anchors are used (e.g., Lantz, 2013a; Weiters & Baumgartner, 2012), and correspondence analysis has been used to calculate the perceived non-equidistant gaps between scale points in individual studies (e.g., Bendixen & Sandler, 1995; Lee & Soutar, 2010). Hence, the assumption that Likert-type scales have interval properties seems questionable, at best. Even though the debate on whether parametric statistics can be used to analyze Likert-type data without invalidating the results goes many years back (e.g., Bradley, 1978), this paper does not explicitly aim to take a stand in that debate. Instead, we explore the sensitivity to non-equidistance of common parametric methods that applied researchers often use in analyses of Likert-type data. The results can be used to choose the most suitable method for different situations.

Basic parametric test statistics, such as the omnibus one-way ANOVA, rely on assumptions of normality and homoscedasticity. Alternative parametric test procedures that compare locations, such as the Brown-Forsythe test (Brown & Forsythe, 1974) and the Welch test (Welch, 1951), are considered robust against heteroscedasticity, especially in the case of unequal sample sizes. However, they have slightly less power than the ANOVA when the homoscedasticity assumption is not violated (Tomarken & Serlin, 1986). Alternative non-parametric methods for comparing locations, such as the Mann-Whithey test (Mann & Whithey, 1947) and the Kruskal-Wallis test (Kruskal & Wallis, 1952), are robust against the violation of the normality assumption, as they only assume that data can be ranked, but they also have less power than the ANOVA when normality actually holds.

The ANOVA and its alternatives have, for many years, been subject to Monte Carlo-based analyses of their robustness against violations of normality and homoscedasticity (see e.g., Harwell, Rubinstein, Hayes, & Olds, 1992, for an extensive historical review). In short, the ANOVA seems relatively robust against minor heteroscedasticity when normality holds (e.g., Glass, Peckham, & Sanders, 1972), and the Brown-Forsythe test and Welch test appear to be superior under various types of major heteroscedasticity (e.g., Tomarken & Serlin, 1986). However, violation of the normality assumption seems to be a more complex issue, since normality can be violated in several different ways, which may have different impacts on the ANOVA and its alternatives (Harwell et al., 1992). The parent distributions may, for example, be positively skewed, negatively skewed, platykurtic, leptokurtic, multimodal, and/or discrete. In some of these cases, the violation of normality seems to be of minor importance, while the difference in performance becomes substantial in other cases (e.g., Lantz, 2013b; Schmider, Ziegler, Danay, Beyer, & Buhner, 2010; Tomarken & Serlin, 1986; Bevan, Denton, & Myers, 1973). Hence, the preferred choice of statistical test procedure when normality has been violated depends a great deal on how this has occurred.

Reference this paper as: Lantz B. "The Impact of non-Equidistance on Anova and Alternative Methods" *The Electronic Journal of Business Research Methods Volume 12 Issue 1 2014 (pp 16-26),* available online at www.ejbrm.com

Non-equidistance constitutes yet another way to violate the assumption of normality. A violation of equidistance means the researcher assumes that the distance between adjacent scale points is constant along the scale, even though it is not perceived as such by subjects. Since Likert-type scales have, in many cases, been proven to be non-equidistant, the phenomenon is obviously a potential validity issue within parametric statistics. Based on Greenacre (1984), Bendixen and Sandler (1995) showed how to use correspondence analysis to rescale data from specific Likert-type scales to interval scales and discovered that all the scales they examined were non-equidistant. The same approach has been used to determine the location of ordinal scale steps in other studies (e.g., Abratt, Bendixen, & Drop, 1999; Abratt & Russel, 1999; Bendixen, Sandler, & Cohen, 1993; Bendixen, Sandler, & Seligman, 1994; Bick, Brown, & Abratt, 2004; Heimerl, 1994; Kennedy, Riquier, & Sharp, 1996; Lee and Soutar, 2010; Yates and Firer, 1997). In all cases, the results were characterized by distinct non-equidistance. Table 1 displays the perceived location of the points for some specific five-point Likert-type scales from these studies, and the lack of equidistance is apparent. Similar results can be found in other studies where the Bendixen and Sandler (1995) approach was used.

Table 1: Some non-equidistant Likert-type scales

	Scale 1	Scale 2	Scale 3	Scale 4	Scale 5	Scale 6
Notional scale value	Bendixen and Sandler (1995)	Bendixen et al. (1993)	Heimerl (1994)	Bendixen et al. (1994)	Bendixen et al (1994)	Kennedy et al. (1996)
0	1.00	1.00	1.00	1.00	1.00	1.00
1	2.34	2.45	1.39	2.00	1.56	2.23
2	3.11	3.42	3.04	3.18	2.25	3.14
3	4.15	3.98	4.50	4.51	3.46	4.14
4	5.00	5.00	5.00	5.00	5.00	5.00

The sensitivity of parametric statistics to non-equidistance seems essentially unexplored in the literature. The only exception appears to be Lantz (2013a), who used simulations to show that the preferred choice of statistical method seems to depend on the current type of non-equidistance. However, Lantz (2013a) only tested two arbitrarily chosen types of non-equidistance. The purpose of this paper is to explore the impact of actual and empirically measured non-equidistance on parametric statistical methods commonly used in applied research to compare mean values across several groups, that is, the ANOVA, Brown-Forsythe test, and Welch test. It should be noted that non-parametric alternatives, such as the Kruskal-Wallis test, are not included since they are perfectly insensitive to non-equidistance by definition. The remainder of the paper is organized as follows. In the next section, the methodology of the study is described. The simulation results are then presented and discussed. Finally, the paper concludes with the implications of the results for statistical analysis.

2. Methodology

The binomial distribution is probably the discrete standard distribution that most closely resembles the normal distribution in terms of shape. It is a mound-shaped discrete distribution that exists for any number of steps, and it can have any mean value between 0 and the number of steps minus 1. It also approaches the normal distribution when the number of steps becomes large (Bowerman, O'Connell, & Murphree, 2009). Hence, it is a suitable basis for an experiment in which the mound shape is important, even though the data are on a discrete scale.

An experimental design with three populations and four different combinations of small (defined as five observations) and large (defined as 25 observations) sample sizes was used in the current study. Six different non-equidistant scales were used in each case (see table 1), representing actual empirically measured violations of equidistance. For each combination, the number of significant ANOVA, Brown-Forsythe, and Welch tests was compared to the number of significant tests when data was actually equidistant. Four different effect sizes (see Cohen, 1992) were used for each combination of sample size, test procedure, and scale: no, small, medium, and large effect. Furthermore, each effect size was created in two different ways, firstly, with equally spaced means, and secondly, with one extreme and two equal means (Tomarken & Serlin, 1986). Finally, all combinations were tested with symmetric data (defined as the case in which the mean is equal to the third scale point on the five-point scale), as well as with skewed data (defined as the case in which the mean is equal to the second scale point on the five-point scale). Table 2 displays the mean values for the underlying binomial distributions required to achieve the different effect sizes.

Table 2: Effect sizes and population means

Panel A: Equally spaced means				
Shape	f	Mean 1	Mean 2	Mean 3
Symmetric	0.00	2.00	2.00	2.00
	0.10	1.88	2.00	2.12
	0.25	1.70	2.00	2.30
	0.40	1.53	2.00	2.47
Skewed	0.00	1.00	1.00	1.00
	0.10	0.91	1.00	1.09
	0.25	0.77	1.00	1.23
	0.40	0.64	1.00	1.36
Panel B: One extreme and two equal means				
Shape	f	Mean 1	Mean 2	Mean 3
Symmetric	0.00	2.00	2.00	2.00
	0.10	2.00	2.21	2.00
	0.25	2.00	2.52	2.00
	0.40	2.00	2.81	2.00
Skewed	0.00	1.00	1.00	1.00
	0.10	1.00	1.16	1.00
	0.25	1.00	1.43	1.00
	0.40	1.00	1.71	1.00

Note that the binomial distribution with five possible values goes from 0 to 4, while the regular Likert-type scale goes from 1 to 5. In other words, the mean value, 2.12, in table 2, for example, corresponds with a Likert-type scale mean value of 3.12. Table 3 displays the means and standard deviations for all six non-equidistant scales under these circumstances, as well as for the corresponding equidistant scale. All six non-equidistant scales were measured with correspondence analysis (Greenacre, 1984) based on real likert-type data from different types of research. The basic idea behind the application of correspondence analysis in this context is to conduct a principal component analysis of the contingency table that the AxB (where A is the number of steps on the scale, and B is the number of items in the study) data points constitutes, in order to measure the strength of the participants' responses (see Bendixen and Sandler, 1995, for a detailed technical description of the algorithm).

Table 3: Means and standard deviations for the non-equidistant scales

		Equi	S1	S2	S3	S4	S5	S6
Symmetric	Mean	3.00	3.16	3.27	2.99	3.20	2.47	3.15
	S.D.	1.00	0.96	0.90	1.31	1.14	1.00	0.98
Skewed	Mean	2.00	2.17	2.28	1.77	2.06	1.63	2.13
	S.D.	0.87	0.92	0.98	0.99	0.97	0.64	0.92

For each combination of sample size, test procedure, scale, skewness, and effect size, 50,000 hypothesis tests based on simulated random numbers were conducted, where the null hypothesis that corresponds to no difference between the locations of the populations was challenged at an alpha level of 0.05 in all cases. All simulations and analytic procedures were conducted using Microsoft Excel 2010.

3. Results

In this section, the results from the simulations are presented, subdivided according to the three statistical methods under scrutiny.

3.1 ANOVA

The ANOVA results are presented in tables 4a–4d. The numbers in the tables represent the proportion of significant tests out of the 50,000 tests conducted under the different conditions. Note that a higher value on the mean absolute percentage error (MAPE) indicates a higher sensitivity to non-equidistance. The overall pattern is similar in all four cases. When the effect size is 0.00, the MAPE values are generally low when at least one sample size is large. In these cases, there are very few significant differences to the equidistant case, indicating that the actual significance level is not severely affected by non-equidistance. However, when all sample sizes are small, the MAPE is considerably higher. However, there is no obvious tendency in how the probability of a type I error is affected by non-equidistance in these cases.

Table 4a: ANOVA, symmetric case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.052	0.051	0.050	0.051	0.051	0.051	0.051	1.3%
	5,5,5	0.053	0.053	0.049*	0.057*	0.056	0.051	0.053	4.8%
	5,5,25	0.050	0.051	0.050	0.049	0.049	0.048	0.051	2.2%
	5,25,25	0.051	0.050	0.051	0.049	0.049	0.049	0.050	1.7%
0.10	25,25,25	0.106	0.104	0.102	0.104	0.105	0.102	0.105	1.9%
	5,5,5	0.060	0.060	0.056*	0.065*	0.064*	0.057	0.060	4.8%
	5,5,25	0.071	0.073	0.081*	0.071	0.072	0.050*	0.074	9.1%
	5,25,25	0.073	0.075	0.079*	0.072	0.074	0.059*	0.075	5.8%
0.25	25,25,25	0.453	0.448	0.436*	0.438*	0.448	0.431*	0.452	2.4%
	5,5,5	0.109	0.110	0.103*	0.118*	0.116*	0.103*	0.111	4.6%
	5,5,25	0.190	0.194	0.211*	0.193	0.199*	0.128*	0.198*	9.2%
	5,25,25	0.226	0.228	0.233	0.221	0.227	0.185*	0.231	4.5%
0.40	25,25,25	0,850	0,846	0,833*	0,833*	0,844	0,828*	0,848	1.3%
	5,5,5	0,205	0,207	0,191*	0,213*	0,215*	0,194*	0,207	3.9%
	5,5,25	0,416	0,420	0,436*	0,420	0,431*	0,314*	0,426*	6.1%
	5,25,25	0,496	0,495	0,496	0,482*	0,495	0,438*	0,499	2.6%
Note: '*' indicates a significant difference from the equidistant case									

Table 4b: ANOVA, symmetric case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.050	0.050	0.050	0.051	0.050	0.050	0.050	0.7%
	5,5,5	0.051	0.052	0.048	0.055*	0.056*	0.049	0.052	5.0%
	5,5,25	0.050	0.050	0.051	0.049	0.050	0.048	0.050	1.8%
	5,25,25	0.048	0.049	0.049	0.048	0.049	0.047	0.049	1.3%
0.10	25,25,25	0.109	0.108	0.105	0.108	0.109	0.107	0.108	1.2%
	5,5,5	0.058	0.060	0.054*	0.063*	0.063*	0.057	0.059	4.8%
	5,5,25	0.064	0.062	0.056*	0.060	0.061	0.071*	0.062	6.3%
	5,25,25	0.100	0.100	0.099	0.099	0.100	0.090*	0.100	2.0%
0.25	25,25,25	0.452	0.447	0.428*	0.433*	0.440*	0.447	0.448	2.5%
	5,5,5	0.106	0.109	0.096*	0.110	0.111	0.109	0.107	4.1%
	5,5,25	0.135	0.130	0.110*	0.122*	0.123*	0.161*	0.128*	10.6%
	5,25,25	0.379	0.376	0.366*	0.368*	0.374	0.356*	0.378	2.4%
0.40	25,25,25	0.857	0.853	0.837*	0.835*	0.843*	0.850*	0.855	1.4%
	5,5,5	0.202	0.206	0.180*	0.200	0.202	0.215*	0.203	3.6%
	5,5,25	0.276	0.266*	0.231*	0.241*	0.247*	0.325*	0.263*	10.9%
	5,25,25	0.774	0.771	0.756*	0.756*	0.763*	0.755*	0.773	1.6%
Note: '*' indicates a significant difference from the equidistant case									

Table 4c: ANOVA, skewed case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.049	0.050	0.050	0.050	0.049	0.047	0.050	1.9%
	5,5,5	0.049	0.052*	0.053*	0.039*	0.049	0.044*	0.052	8.9%
	5,5,25	0.048	0.048	0.048	0.045	0.047	0.049	0.048	2.2%
	5,25,25	0.048	0.049	0.048	0.047	0.048	0.049	0.049	0.9%
0.10	25,25,25	0.091	0.090	0.091	0.085*	0.090	0.085*	0.091	2.4%
	5,5,5	0.055	0.059*	0.060*	0.044*	0.056	0.049*	0.059*	9.6%
	5,5,25	0.050	0.058*	0.060*	0.030*	0.046*	0.039*	0.055*	19.5%
	5,25,25	0.060	0.064*	0.066*	0.047*	0.056	0.051*	0.063	10.7%
0.25	25,25,25	0.360	0.354	0.353	0.315*	0.358	0.337*	0.358	3.9%
	5,5,5	0.094	0.099*	0.101*	0.072*	0.094	0.082*	0.099*	9.0%
	5,5,25	0.110	0.133*	0.144*	0.048*	0.098*	0.063*	0.127*	29.5%
	5,25,25	0.151	0.163*	0.167*	0.107*	0.143*	0.120*	0.16*	13.2%
0.40	25,25,25	0.748	0.739*	0.735*	0.678*	0.744	0.718*	0.745	2.8%
	5,5,5	0.163	0.173*	0.176*	0.124*	0.163	0.142*	0.172*	9.4%
	5,5,25	0.245	0.282*	0.303*	0.121*	0.224*	0.143*	0.275*	25.3%
	5,25,25	0.340	0.355*	0.361*	0.251*	0.327*	0.280*	0.354*	10.4%
Note: '*' indicates a significant difference from the equidistant case									

Table 4d: ANOVA, skewed case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.050	0.051	0.052	0.049	0.050	0.048	0.050	2.2%
	5,5,5	0.049	0.052	0.054*	0.041*	0.050	0.045*	0.052	7.6%
	5,5,25	0.051	0.050	0.050	0.047*	0.050	0.052	0.051	2.4%
	5,25,25	0.050	0.050	0.049	0.047	0.048	0.049	0.050	2.2%
0.10	25,25,25	0.092	0.091	0.090	0.089	0.092	0.089	0.091	2.2%
	5,5,5	0.056	0.059	0.060*	0.047*	0.056	0.049*	0.059	8.1%
	5,5,25	0.066	0.062*	0.058*	0.070	0.067	0.073*	0.064	6.6%
	5,25,25	0.079	0.080	0.080	0.068*	0.077	0.071*	0.079	4.9%
0.25	25,25,25	0.379	0.366*	0.357*	0.356*	0.378	0.366*	0.373	3.5%
	5,5,5	0.095	0.097	0.098	0.085*	0.098	0.085*	0.099	5.6%
	5,5,25	0.140	0.125*	0.115*	0.153*	0.144	0.160*	0.130*	10.4%
	5,25,25	0.300	0.297	0.296	0.265*	0.296	0.271*	0.300	4.1%
0.40	25,25,25	0.788	0.775*	0.761*	0.761*	0.788	0.773*	0.782	1.9%
	5,5,5	0.185	0.183	0.181	0.176*	0.192*	0.171*	0.188	3.5%
	5,5,25	0.291	0.262*	0.236*	0.306*	0.297	0.320*	0.272*	8.8%
	5,25,25	0.680	0.673	0.664*	0.638*	0.676	0.643*	0.679	2.7%
Note: '*' indicates a significant difference from the equidistant case									

Non-equidistance seems to affect power more than significance. Many results-based non-equidistant scales differ significantly from their equidistant analogues. When all sample sizes are large, power is often slightly, but significantly decreased by the presence of non-equidistance. However, when at least one sample size is small, power responds erratically to non-equidistance, especially when exactly one sample size is large. It should also be noted that the ANOVA seems more sensitive to non-equidistance when data are skewed and means are equally spaced than in the other three cases (the MAPE values are about twice as high).

3.2 Brown-Forsythe test

The results regarding the Brown-Forsythe test are presented in tables 5a–5d. When the effect size is 0.00, the MAPE values are generally low when all sample sizes are large (albeit higher than when ANOVA is used to analyze data), and generally high when all sample sizes are small. When sample sizes are unequal, the MAPE values are considerably higher under the Brown-Forsythe test than under the ANOVA test. In these cases, many results differ significantly from the equidistant case. This indicates that the actual significance level is rather sensitive to non-equidistance. The MAPE values are also generally higher when data are skewed, for all combinations of sample sizes.

Table 5a: Brown-Forsythe test, symmetric case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.051	0.050	0.050	0.051	0.051	0.050	0.051	2.1%
	5,5,5	0.046	0.042*	0.034*	0.047	0.047	0.036*	0.044	21.0%
	5,5,25	0.050	0.047	0.044*	0.054	0.052	0.051	0.049	5.7%
	5,25,25	0.055	0.052	0.049*	0.063*	0.059*	0.056	0.054	11.1%
0.10	25,25,25	0.106	0.103	0.101	0.104	0.105	0.101	0.104	2.5%
	5,5,5	0.053	0.048*	0.040*	0.054	0.053	0.040*	0.050	20.2%
	5,5,25	0.064	0.060*	0.044*	0.066	0.065	0.083*	0.060*	16.7%
	5,25,25	0.077	0.073	0.061*	0.082*	0.081	0.091*	0.074	10.5%
0.25	25,25,25	0.452	0.447	0.435*	0.437*	0.447	0.429*	0.450	2.7%
	5,5,5	0.098	0.090*	0.079*	0.100	0.099	0.077*	0.094	17.8%
	5,5,25	0.152	0.143*	0.104*	0.145*	0.145*	0.195*	0.141*	24.4%
	5,25,25	0.215	0.206*	0.167*	0.209	0.211	0.251*	0.206*	11.6%
0.40	25,25,25	0.849	0.845	0.832*	0.832*	0.844	0.827*	0.848	1.4%
	5,5,5	0.188	0.175*	0.154*	0.186	0.188	0.150*	0.180*	16.1%
	5,5,25	0.318	0.302*	0.241*	0.296*	0.299*	0.378*	0.302*	27.2%
	5,25,25	0.453	0.437*	0.372*	0.434*	0.435*	0.508*	0.437*	12.6%
Note: '*' indicates a significant difference from the equidistant case									

Table 5b: Brown-Forsythe test, symmetric case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.050	0.049	0.049	0.051	0.050	0.049	0.049	1.3%
	5,5,5	0.045	0.041*	0.034*	0.046	0.046	0.036*	0.043	20.0%
	5,5,25	0.048	0.045	0.043*	0.053*	0.051	0.050	0.047	6.6%
	5,25,25	0.054	0.052	0.048*	0.061*	0.059*	0.054	0.053	12.9%
0.10	25,25,25	0.108	0.107	0.104	0.108	0.109	0.106	0.107	1.6%
	5,5,5	0.052	0.048*	0.040*	0.053	0.053	0.042*	0.050	18.1%
	5,5,25	0.061	0.059	0.062	0.067*	0.065	0.050*	0.062	7.0%
	5,25,25	0.098	0.096	0.087*	0.098	0.100	0.106*	0.096	4.6%
0.25	25,25,25	0.451	0.446	0.426*	0.432*	0.439*	0.445	0.447	2.8%
	5,5,5	0.095	0.090*	0.075*	0.090*	0.091	0.084*	0.091	17.9%
	5,5,25	0.123	0.120	0.123	0.127	0.127	0.095*	0.124	11.3%
	5,25,25	0.335	0.329	0.308*	0.306*	0.313*	0.358*	0.328	14.5%
0.40	25,25,25	0.856	0.852	0.836*	0.835*	0.842*	0.849*	0.854	1.5%
	5,5,5	0.182	0.173*	0.147*	0.165*	0.167*	0.178	0.174*	16.9%
	5,5,25	0.247	0.244	0.240*	0.241	0.243	0.206*	0.248	14.2%
	5,25,25	0.686	0.676*	0.647*	0.639*	0.649*	0.707*	0.674*	14.1%
Note: '*' indicates a significant difference from the equidistant case									

Table 5c: Brown-Forsythe test, skewed case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.049	0.049	0.050	0.049	0.048	0.045	0.049	2.3%
	5,5,5	0.043	0.042	0.044	0.019*	0.037*	0.029*	0.042	27.1%
	5,5,25	0.051	0.048	0.051	0.060*	0.051	0.051	0.049	8.3%
	5,25,25	0.056	0.056	0.061*	0.055	0.056	0.051*	0.057	16.3%
0.10	25,25,25	0.090	0.089	0.090	0.083*	0.089	0.083*	0.090	3.3%
	5,5,5	0.048	0.048	0.050	0.021*	0.041*	0.033*	0.048	27.2%
	5,5,25	0.070	0.057*	0.058*	0.089*	0.075*	0.084*	0.063*	41.5%
	5,25,25	0.080	0.072*	0.073*	0.081	0.082	0.078	0.076	29.0%
0.25	25,25,25	0.358	0.353	0.352	0.312*	0.356	0.333*	0.357	4.5%
	5,5,5	0.083	0.083	0.085	0.038*	0.073*	0.055*	0.083	26.2%
	5,5,25	0.157	0.127*	0.122*	0.178*	0.168*	0.197*	0.139*	41.0%
	5,25,25	0.198	0.172*	0.167*	0.202	0.206*	0.203	0.183*	25.3%
0.40	25,25,25	0.746	0.738*	0.735*	0.675*	0.742	0.714*	0.744	3.1%
	5,5,5	0.144	0.146	0.151*	0.071*	0.127*	0.097*	0.145	24.9%
	5,5,25	0.306	0.261*	0.246*	0.315*	0.320*	0.366*	0.279*	21.6%
	5,25,25	0.408	0.364*	0.351*	0.412	0.420*	0.417*	0.383*	15.0%
Note: '*' indicates a significant difference from the equidistant case									

Table 5d: Brown-Forsythe test, skewed case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.050	0.051	0.052	0.048	0.050	0.047	0.050	2.6%
	5,5,5	0.044	0.041	0.045	0.019*	0.039*	0.031*	0.044	26.5%
	5,5,25	0.052	0.050	0.053	0.061*	0.052	0.051	0.051	5.2%
	5,25,25	0.056	0.057	0.061*	0.054	0.055	0.051*	0.057	13.2%
0.10	25,25,25	0.091	0.090	0.089	0.087	0.092	0.087*	0.091	3.2%
	5,5,5	0.048	0.048	0.050	0.024*	0.043*	0.033*	0.049	26.0%
	5,5,25	0.053	0.057*	0.061*	0.049*	0.050	0.044*	0.056	20.5%
	5,25,25	0.091	0.084*	0.084*	0.094	0.094	0.092	0.087	13.6%
0.25	25,25,25	0.378	0.365*	0.357*	0.354*	0.377	0.362*	0.372	4.0%
	5,5,5	0.084	0.082	0.082	0.054*	0.077*	0.058*	0.084	23.5%
	5,5,25	0.094	0.104*	0.113*	0.068*	0.086*	0.063*	0.102*	36.1%
	5,25,25	0.299	0.274*	0.260*	0.304	0.306	0.318*	0.285*	6.1%
0.40	25,25,25	0.787	0.774*	0.761*	0.760*	0.787	0.770*	0.781	2.1%
	5,5,5	0.165	0.159*	0.154*	0.130*	0.158*	0.120*	0.162	20.4%
	5,5,25	0.191	0.206*	0.213*	0.146*	0.181*	0.125*	0.203*	38.5%
	5,25,25	0.646	0.619*	0.587*	0.635*	0.648	0.665*	0.630*	7.3%
Note: '*' indicates a significant difference from the equidistant case									

Unlike the ANOVA test, non-equidistance does not seem to affect the power of the Brown-Forsythe test more than it does the significance, even though power often decreases. Similar to the ANOVA, the exception is when exactly one sample size is large, in which case, many results differ significantly from the equidistant case. This indicates high sensitivity to non-equidistance. While the ANOVA test displays an increased sensitivity to non-equidistance when data are skewed and the means are equally spaced, the Brown-Forsythe test exhibits a similar increased sensitivity when data are skewed and two means are equal.

3.3 Welch test

The results of the Welch test are presented in tables 6a–6d. When the effect size is 0.00, the MAPE values are rather low when all sample sizes are large (albeit higher than when using the ANOVA or Brown-Forsythe tests). There are only occasional cases in which sample sizes are large and the results differ significantly from the equidistant case. This indicates relatively low sensitivity to non-equidistance. However, when at least one sample size is small, the actual significance is heavily affected by non-equidistance. This is particularly true when data are skewed. In this case, almost all results differ significantly from the equidistant case.

Table 6a: Welch test, symmetric case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.051	0.050	0.049	0.051	0.051	0.051	0.050	2.3%
	5,5,5	0.039	0.034*	0.027*	0.048*	0.049*	0.035*	0.037	27.4%
	5,5,25	0.049	0.045*	0.049	0.064*	0.060*	0.074*	0.047	19.1%
	5,25,25	0.054	0.052	0.060*	0.064*	0.060*	0.081*	0.053	22.1%
0.10	25,25,25	0.104	0.102	0.101	0.104	0.104	0.103	0.103	2.5%
	5,5,5	0.045	0.039*	0.032*	0.055*	0.055*	0.040*	0.043	26.0%
	5,5,25	0.063	0.062	0.040*	0.074*	0.073*	0.126*	0.059*	26.4%
	5,25,25	0.075	0.076	0.057*	0.081*	0.081*	0.135*	0.072	22.3%
0.25	25,25,25	0.447	0.441	0.430*	0.436*	0.444	0.426*	0.445	3.5%
	5,5,5	0.085	0.076*	0.063*	0.105*	0.105*	0.077*	0.083	22.4%
	5,5,25	0.147	0.151	0.079*	0.148	0.157*	0.274*	0.139*	31.6%
	5,25,25	0.196	0.199	0.140*	0.195	0.202	0.291*	0.190	19.9%
0.40	25,25,25	0.844	0.841	0.827*	0.830*	0.840	0.823*	0.843	1.9%
	5,5,5	0.165	0.149*	0.124*	0.193*	0.196*	0.149*	0.160	21.1%
	5,5,25	0.304	0.305	0.181*	0.292*	0.308	0.453*	0.289*	29.8%
	5,25,25	0.423	0.424	0.343*	0.411*	0.422	0.502*	0.415*	15.8%
Note: '*' indicates a significant difference from the equidistant case									

Table 6b: Welch test, symmetric case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.049	0.049	0.049	0.050	0.049	0.051	0.049	1.8%
	5,5,5	0.038	0.033*	0.026*	0.047*	0.047*	0.033*	0.036	27.4%
	5,5,25	0.047	0.044	0.047	0.063*	0.058*	0.073*	0.045	19.5%
	5,25,25	0.054	0.052	0.059*	0.064*	0.061*	0.081*	0.052	27.5%
0.10	25,25,25	0.109	0.108	0.106	0.109	0.110	0.107	0.108	1.1%
	5,5,5	0.045	0.042*	0.030*	0.056*	0.057*	0.042	0.044	22.7%
	5,5,25	0.060	0.056*	0.065*	0.083*	0.077*	0.068*	0.060	12.7%
	5,25,25	0.086	0.090	0.074*	0.091*	0.093*	0.130*	0.085	15.9%
0.25	25,25,25	0.452	0.449	0.438*	0.438*	0.446	0.429*	0.451	2.2%
	5,5,5	0.086	0.082	0.058*	0.102*	0.105*	0.082*	0.085	19.3%
	5,5,25	0.121	0.115*	0.128*	0.164*	0.157*	0.102*	0.124	14.9%
	5,25,25	0.292	0.293	0.259*	0.280*	0.289	0.317*	0.288	24.1%
0.40	25,25,25	0.858	0.856	0.844*	0.842*	0.850*	0.839*	0.857	1.1%
	5,5,5	0.172	0.168	0.123*	0.189*	0.194*	0.163*	0.172	16.6%
	5,5,25	0.244	0.237*	0.246	0.316*	0.306*	0.190*	0.252*	15.1%
	5,25,25	0.667	0.663	0.638*	0.635*	0.646*	0.670	0.663	15.7%
Note: '*' indicates a significant difference from the equidistant case									

Table 6c: Welch test, skewed case with equally spaced means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.050	0.050	0.051	0.055*	0.051	0.050	0.050	3.5%
	5,5,5	0.032	0.032	0.037*	0.014*	0.030	0.025*	0.035	40.3%
	5,5,25	0.047	0.051*	0.054*	0.158*	0.052*	0.053*	0.046	45.6%
	5,25,25	0.054	0.061*	0.063*	0.185*	0.058*	0.061*	0.054	66.9%
0.10	25,25,25	0.089	0.089	0.090	0.088	0.090	0.087	0.090	1.9%
	5,5,5	0.036	0.036	0.042*	0.016*	0.034	0.028*	0.039*	41.2%
	5,5,25	0.067	0.051*	0.054*	0.236*	0.081*	0.088*	0.055*	87.2%
	5,25,25	0.077	0.066*	0.068*	0.247*	0.090*	0.099*	0.066*	77.2%
0.25	25,25,25	0.356	0.349	0.348*	0.323*	0.356	0.341*	0.355	4.1%
	5,5,5	0.064	0.064	0.073*	0.031*	0.059*	0.048*	0.069*	39.1%
	5,5,25	0.153	0.098*	0.099*	0.395*	0.186*	0.205*	0.115*	73.3%
	5,25,25	0.181	0.132*	0.134*	0.381*	0.212*	0.232*	0.147*	45.5%
0.40	25,25,25	0.746	0.734*	0.733*	0.684*	0.744	0.072*	0.742	2.8%
	5,5,5	0.111	0.114	0.127*	0.058*	0.105*	0.083*	0.120*	38.2%
	5,5,25	0.298	0.196*	0.192*	0.533*	0.345*	0.375*	0.232*	43.1%
	5,25,25	0.363	0.283*	0.282*	0.521*	0.401*	0.420*	0.311*	22.8%
Note: '*' indicates a significant difference from the equidistant case									

Table 6d: Welch test, skewed case with one extreme and two equal means

f	n1,n2,n3	Eq	V1	V2	V3	V4	V5	V6	MAPE
0.00	25,25,25	0.051	0.052	0.052	0.055*	0.051	0.050	0.050	3.3%
	5,5,5	0.033	0.034	0.038*	0.014*	0.030	0.025*	0.035	40.1%
	5,5,25	0.048	0.054*	0.056*	0.157*	0.052*	0.053*	0.048	40.2%
	5,25,25	0.054	0.064*	0.064*	0.182*	0.059*	0.062*	0.054	63.1%
0.10	25,25,25	0.090	0.089	0.090	0.090	0.090	0.087	0.090	3.1%
	5,5,5	0.038	0.037	0.043*	0.020*	0.036	0.029*	0.041	38.4%
	5,5,25	0.050	0.062*	0.070*	0.136*	0.050	0.049	0.054*	31.1%
	5,25,25	0.081	0.072*	0.075*	0.210*	0.091*	0.098*	0.071*	38.6%
0.25	25,25,25	0.357	0.351	0.350	0.328*	0.356	0.331*	0.356	9.0%
	5,5,5	0.070	0.064*	0.077*	0.047*	0.067	0.050*	0.074	33.4%
	5,5,25	0.083	0.108*	0.131*	0.144*	0.078*	0.065*	0.097*	26.9%
	5,25,25	0.250	0.216*	0.215*	0.307*	0.268*	0.266*	0.228*	17.5%
0.40	25,25,25	0.765	0.756*	0.751*	0.724*	0.763	0.731*	0.762	5.2%
	5,5,5	0.134	0.125*	0.146*	0.108*	0.132	0.909*	0.138	33.3%
	5,5,25	0.159	0.194*	0.241*	0.191*	0.149*	1.076*	0.183*	38.9%
	5,25,25	0.579	0.550*	0.536*	0.559*	0.585	0.571	0.564*	17.6%
Note: '*' indicates a significant difference from the equidistant case									

Regarding power, the Welch test performs in a similar manner to the Brown-Forsythe test. That is, the MAPE values are not notably higher than those cases in which the effect size is 0.00. However, the apparent pattern is that the Welch test loses power in more cases than the other two tests when data are non-equidistant, even for large sample sizes. The MAPE values are constantly and substantially higher for the Welch test than they are for the ANOVA test, and notably higher than they are for the Brown-Forsythe test.

4. Discussion

The normality assumption behind parametric methods such as the ANOVA, the Brown-Forsythe test, and the Welch test requires data to be equidistant. Through rescaling, different kinds of Likert-type data have been shown to be non-equidistant to different degrees. Hence, the common practice of using parametric methods on such data without rescaling them is doubtful. In Lantz (2013a), the effects of two different arbitrarily chosen types of non-equidistance were explored. This study contributes further to the topic by exploring the impact of several actual and empirically measured types of non-equidistance on the ANOVA, Brown-Forsythe test, and Welch test.

In line with previous research (e.g., Tomarken & Serlin, 1986), the Brown-Forsythe test and the Welch test generally exhibit somewhat lower actual power than the ANOVA test, which can be seen as an indication of the reliability of the results in this study. As regards both significance level and power, all three test procedures are relatively insensitive to non-equidistance when all sample sizes are large, even though the ANOVA test seems somewhat less sensitive than the other two methods. However, there are substantial differences between the methods when at least one sample size is small, which indicates that the overall sensitivity to non-equidistance differs between the methods.

One main result of this study is that the ANOVA test seems substantially less sensitive to non-equidistance than the Brown-Forsythe test and the Welch test with regard to the actual significance level. For the ANOVA test, there were essentially no significant differences between the equidistant case and the non-equidistant cases, unless all sample sizes were small. For the Brown-Forsythe test and the Welch test, many such differences were found when sample sizes were unequal. For the Welch test, there were even a few cases of differences when all sample sizes were large. Hence, as regards the actual significance level, the Welch test seems more sensitive to non-equidistance than the Brown-Forsythe test, but the Brown-Forsythe test still seems more sensitive than the ANOVA test.

Another general observation is that actual power often decreases for all three methods as a result of non-equidistance, even though all sample sizes are large. This is particularly true at the medium and large effect size. When at least one sample size is small, an erratic pattern for actual power appears for all three methods, already at the small effect size, even though it is less apparent for the ANOVA test than for the other two methods.

The sensitivity to non-equidistance seems to increase for all three methods when data are skewed. This is not surprising, since skewness in itself constitutes a violation of the normality assumption that all three methods rely on. However, whether a certain power level depends on equally spaced means or on one extreme mean does not seem to matter much for any of the methods.

5. Conclusion

In this study, we used a simulation-based approach to explore the impact of actual and empirically measured non-equidistance on parametric statistical methods commonly used in applied research to compare mean values across several groups, that is, the ANOVA test, Brown-Forsythe test, and Welch test. The results are rather unambiguous. In general, the ANOVA test is less sensitive to non-equidistance than the Brown-Forsythe test and Welch test. Hence, unless there are other reasons for not selecting it (e.g., heteroscedasticity), the ANOVA test should be the primary choice among these three methods when analyzing Likert-type or other potentially non-equidistant data, especially if not all sample sizes are large. When potentially non-equidistant data are characterized by heteroscedasticity and all sample sizes are large, the Brown-Forsythe test and Welch test can still only be seen as a possible but not natural alternative to the ANOVA test because of their larger sensitivity to non-equidistance. In other cases, non-parametric alternatives to the ANOVA test, such as the Kruskal-Wallis test, should probably be preferred because of their total insensitivity to non-equidistance, even though they have lower power, in general.

Future research in this area should explore the effects of non-equidistance further by combining the violation of equidistance with, for example, heteroscedasticity. In general, the effects of concurrent violations can produce anomalous effects not observed in separate violations (see, e.g. Zimmerman, 1998). Other types of parametric methods should also be tested for their robustness against non-equidistance.

References

Abratt, R., Bendixen, M., & Drop, K. (1999). Ethical perceptions of South African retailers: management and sales personnel. *International Journal of Retail & Distribution Management*, 27, 91-105.

Abratt, R., & Russell, J. (1999). Relationship marketing in private banking in South Africa. *International Journal of Bank Marketing*, 17, 5-19.

Bendixen, M. T., Sandler, M., & Cohen, D. W. (1993). Environmental Issues and the Use of PVC in the Packaging Industry. *Management Dynamics: Contemporary Research*, 2, 69-86.

Bendixen, M. T., Sandler, M., & Seligman, D. (1994). Consumer Perceptions of Environmentally Friendly Products. *South African Journal of Business Management*, 25, 59-64.

Bendixen, M. T., & Sandler, M. (1995). Converting verbal scales to interval scales using correspondence analysis. *Management Dynamics: Contemporary Research*, 4, 31-49.

Bevan, M. F., Denton, J. Q., & Myers, J. L. (1974). The robustness of the F test to violations of continuity and form of treatment population. *British Journal of Mathematical and Statistical Psychology*, 27, 199-204.

Bick, G., Brown, A, B., & Abratt, R. (2004). Customer perceptions of the value delivered by
retail banks in South Africa. *International Journal of Bank Marketing*, 22, 300-318.

Bowerman, B. L., O'Connell, R. T., & Murphree, E. S. (2009). *Business Statistics in Practice*. McGraw-Hill Irwin, New York, NY.

Bradley, J. V. (1978). Robustness?. *British Journal of Mathematical and Statistical Psychology*, 31, 144-152.

Brown, M. B. & Forsythe, A. B. (1974). The ANOVA and multiple comparisons for data with heterogeneous variances. *Biometrics*, 30, 719-724.

Cohen, J. (1992). A Power Primer. *Psychological Bulletin*, 112, 155-159.

Dawes, J. (2008). Do data characteristics change according to the number of scale points used?. International Journal of Market Research, 50, 61-77.

Glass, G. V., Peckham, P. D., & Sanders, J. R. (1972). Consequences of Failure to Meet Assumptions Underlying the Fixed Effects Analyses of Variance and Covariance. *Review of Educational Research*, 42, 237-288.

Greenacre, M. J. (1984). *Theory and Application of Correspondence Analysis*, Academic Press, London.

Harwell, M. R., Rubinstein, E. N., Hayes, W. S., & Olds, C. C. (1992). Summarizing Monte Carlo Results in Methodological Research: The One- and Two-Factor Fixed Effects ANOVA Cases. *Journal of Educational Statistics*, 17, 315-339.

Heimerl, J. (1994). *A Comparative Study of Brand Loyalty between Urban Blacks and Urban Whites*, Unpublished MBA Research Report, University of the Witwatersrand, Johannesburg.

Kennedy, R., Riquier, C., & Sharp, B. (1996). Practical Applications of Correspondence Analysis to Categorical Data in Market Research. *Journal of Targeting, Measurement and Analysis for Marketing*, 5, 56-70.

Kruskal, W. H., & Wallis, W. A. (1952). Use of Ranks in One-Criterion Variance Analysis. *Journal of the American Statistical Association*, 47, 583–621.

Lantz, B. (2013a). Equidistance of Likert-type scales and validation of inferential methods using experiments and simulations, *Electronic Journal of Business Research Methods*, 10, 16–28.

Lantz, B. (2013b). The impact of sample non-normality on ANOVA and alternative methods, *British Journal of Mathematical and Statistical Psychology*, 66, 224-244.

Lee, J. A., & Soutar, G. N. (2010). Is Schwartz's value survey an interval scale, and does it really matter? *Journal of Cross-Cultural Psychology*, 41, 76-86.

Mann, H. B. & Whitney, D. R. (1947). On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. *Annals of Mathematical Statistics*, 18, 50-60.

Schmider, E., Ziegler, M., Danay, E., Beyer, L., & Buhner, M. (2010). Is it really robust? Reinvestigating the robustness of ANOVA against violations of the normal distribution assumption. *Methodology: European Journal of ResearchMethods for the Behavioral and Social Sciences*, 6, 147–151.

Tomarken, A. J., & Serlin, R. C. (1986). Comparison of ANOVA Alternatives Under Variance Heterogeneity and Specific Noncentrality Structures. *Psychological Bulletin*, 99, 90-99.

Weijters, B. & Baumgartner, H. (2012). Misresponse to Reversed and Negated Items in Surveys: A Review. *Journal of Marketing Research*, 49, 737-747.

Welch, B. L. (1951). On the Comparison of Several Mean Values: An Alternative Approach. *Biometrika*, 38, 330-336.

Yates, A., & Firer, C. (1997). The Determinants of the Risk Perceptions of Investors, *Investment Analysts Journal*, 44, 61-69.