Computational Statistics & Data Analysis 40 (2002) 435–445
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

A test of normality with high uniform power

Douglas G. Bonett[a,*], Edith Seier[b]

[a]Department of Statistics, Iowa State University, Snedecor Hall, 50011-1210 Ames, IA, USA
[b]Department of Mathematics, East Tennessee State University, Johnson City, TN 37614, USA

Abstract

Kurtosis can be measured in more than one way. A modification of Geary's measure of kurtosis is shown to be more sensitive to kurtosis in the center of the distribution while Pearson's measure of kurtosis is more sensitive to kurtosis in the tails of the distribution. The modified Geary measure and the Pearson measure are used to define a joint test of kurtosis that has high uniform power across a very wide range of symmetric nonnormal distributions. © 2002 Elsevier Science B.V. All rights reserved.

Keywords: Geary; Kurtosis; Leptokurtosis; Normality; Shapiro–Wilk

1. Introduction

Karl Pearson introduced the idea of kurtosis to describe distributions that differed from normal distributions in terms of peakedness. Pearson (1905) referred to *leptokurtic* distributions as being more peaked and *platykurtic* distributions as being less peaked than normal distributions, but modern definitions of kurtosis acknowledge the fact that kurtosis will be influenced by both the peakedness and the tail weight of a distribution (Ruppert, 1987) and can be "formalized in many ways" (Balanda and MacGillivray, 1988). According to van Zwet (1964), only symmetric distributions should be compared in terms of kurtosis and kurtosis should not be measured by a single number.

The renewed interest in assessing kurtosis may be due to the increasing use of normal-theory covariance structure methods which are known to perform poorly in leptokurtic distributions (Bollen, 1989, pp. 415–418; Hu et al., 1990) and Micceri's (1989) findings that the majority of real data sets exhibit leptokurtosis. Normal-theory inferential procedures for variances and Pearson correlations also perform poorly in the

* Corresponding author.
E-mail address: dgbonett@iastate.edu (D.G. Bonett).

0167-9473/02/$ - see front matter © 2002 Elsevier Science B.V. All rights reserved.
PII: S0167-9473(02)00074-9

presence of leptokurtosis (Scheffé, 1959, p. 336; Kowalski, 1972). Kurtosis also affects the power of tests regarding means. Although inferential methods for means are very robust to kurtosis, the median will be a more efficient measure of location than the mean in many leptokurtic distributions such as the Laplace and Student's t distribution with df < 5. Nonparametric tests of location such as the Mann–Whitney test can be far more powerful than the t-test in certain leptokurtic distributions (Hodges and Lehmann, 1956).

Geary (1936) introduced a measure of kurtosis which he defined as τ/σ where $\tau = E(|Y-\mu|)$. Under normality, it can be shown that $\tau/\sigma = (2/\pi)^{1/2} = 0.7979$. Geary's measure of kurtosis never gained widespread usage, possibly because $0 < \tau/\sigma \leqslant (2/\pi)^{1/2}$ in leptokurtic distributions so that large increases in leptokurtosis have small numerical effects on τ/σ. We propose a simple transformation of Geary's measure that, like Pearson's measure, equals 3.0 under normality and increases without bound with increasing leptokurtosis. The transformation also improves the small-sample sampling distribution of the test statistic under the null hypothesis of normality.

2. Two measures of kurtosis

The population value of Pearson's measure of kurtosis may be defined as

$$\beta_2 = E(X-\mu)^4/\{E(X-\mu)^2\}^2 \tag{1}$$

which is equal to 3 in normal distributions and is greater than 3 in leptokurtic distributions. The usual estimator of β_2 is

$$\hat{\beta}_2 = n\Sigma(X_i-\hat{\mu})^4/\{\Sigma(X_i-\hat{\mu})^2\}^2, \tag{2}$$

where $\hat{\mu} = \Sigma X_i/n$ and n is the sample size.

Extensive tables of lower and upper critical values of β_2 that can be used for one-sided and two-sided tests of $\beta_2 = 3$ are available (D'Agostino and Stephens, 1986; Ramsey and Ramsey, 1993, 1995). Most statistical packages report $\beta_2 - 3$ or Fisher's estimator of $\beta_2 - 3$ (see DeCarlo, 1997) which is unbiased in normal distributions. Anscombe and Glynn (1983) proposed the following transformation of $\hat{\beta}_2$:

$$z_\beta = [1 - 2/9c_1 - \{(1-2/c_1)/(1+c_3\{2/(c_1-4)\}^{1/2}\}^{1/3}]/(2/9c_1)^{1/2}, \tag{3}$$

where

$$c_1 = 6 + (8/c_2)\{2/c_2 + (1+4/c_2)^{1/2}\},$$

$$c_2 = \{6(n^2-5n+2)/(n+7)(n+9)\}\{6(n+3)(n+5)/n(n-2)(n-3)\}^{1/2},$$

$$c_3 = \{\hat{\beta}_2 - 3(n-1)/(n+1)\}/\{24n(n-2)(n-3)/(n+1)^2(n+3)(n+5)\}^{1/2}.$$

Reject normality if $|z_\beta| > z_{\alpha/2}$ where $z_{\alpha/2}$ is the point on the standard normal distribution exceeded with probability $\alpha/2$. A one-sided test of leptokurtosis rejects normality if $z_\beta > z_\alpha$.

Under normality $\sigma/\tau = (\pi/2)^{1/2}$ (Stuart and Ord, 1994, p. 361) and it follows that

$$\omega = 13.29(\ln \sigma - \ln \tau) \tag{4}$$

is approximately equal to 3 in normal distributions. The following estimator of ω is proposed:

$$\hat{\omega} = 13.29(\ln \hat{\sigma} - \ln \hat{\tau}), \tag{5}$$

where $\hat{\tau} = \Sigma|X_i - \hat{\mu}|/n$ and $\hat{\sigma}^2 = \Sigma(X_i - \hat{\mu})^2/n$.

Geary (1936) shows that var $(\hat{\tau}/\hat{\sigma}) \doteq 0.2123^2/n$. Applying the δ-method gives $\text{var}(\ln \hat{\tau} - \ln \hat{\sigma}) \doteq 0.2123^2(\sigma^2/n\tau^2)$. Under normality $\sigma/\tau = (\pi/2)^{1/2}$ and thus $\text{var}(\ln \hat{\sigma} - \ln \hat{\tau}) \doteq 0.2123^2\pi/2n$ and $\text{SE}(\hat{\omega}) \doteq 13.29\{0.2123^2\pi/2n\}^{1/2} \doteq 3.54/n^{1/2}$. In a preliminary empirical investigation, we found that the asymptotic standard error of $\hat{\omega}$ can be estimated more accurately by $3.54/(n+2)^{1/2}$.

Applying standard asymptotic results, it can be shown that the test statistic

$$z_\omega = (n+2)^{1/2}(\hat{\omega} - 3)/3.54 \tag{6}$$

has an approximate standard normal distribution under the null hypothesis of normality. Reject normality if $|z_\omega| > z_{\alpha/2}$. A one-sided test of leptokurtosis rejects normality if $z_\omega > z_\alpha$.

3. Values of β_2 and ω for some symmetric distributions

Values of β_2 and ω are given in Table 1 for a variety of symmetric distributions. Most of the values of β_2 were obtained from standard sources such as Evans et al. (1993), Zwillinger and Kokoska (2000) and Johnson et al. (1994). The values of ω were obtained by computing τ for each distribution and then computing (4).

According to Oja (1981), a valid measure of kurtosis must be location and scale invariant and also must obey van Zwet ordering which rank orders the distributions in Set A of Table 1 from smallest to largest. Both β_2 and ω are location and scale invariant and both rank the distributions in Set A from smallest to largest. Note that β_2 exists in distributions where $E(z^4)$ exists while ω exists in a larger class of distributions where $E(z^2)$ exists.

Because β_2 is more sensitive to tail weight than ω, it is more likely than ω to incorrectly classify distributions that differ primarily in terms of peakedness. DeCarlo (1997) and others have pointed that the Laplace distribution is clearly more peaked than the t_5 distribution but $\beta_2 = 6$ for the Laplace and $\beta_2 = 9$ for the t_5. In contrast, $\omega = 4.61$ for the Laplace and $\omega = 4.09$ for the t_5 and thus ω correctly classifies these distributions according to peakedness.

4. Small-sample Type I error rates of kurtosis-based normality tests

The z_β test statistic has been studied extensively and is known to have excellent Type I error control in small samples (D'Agostino et al., 1990). A z-test for Geary's

Table 1
Values of β_2 and ω for some symmetric distributions

Set	Member	Distribution	β_2	ω
A	1	Beta (1/2, 1/2)	1.50	1.37
	2	Uniform	1.80	1.91
	3	Normal	3.00	3.00
	4	Logistic	4.20	3.55
	5	Laplace	6.00	4.61
B	1	$t(6)$	6.00	3.82
	2	$t(5)$	9.00	4.09
	3	$t(4)$	∞	4.61
	4	$t(3)$	∞	6.00
	5	$t(2)$	∞	∞
C	1	Lambda (0.50, 0.60)	1.99	2.16
	2	Lambda (−0.33, −0.14)	9.36	4.52
	3	Lambda (−0.38, −0.16)	11.61	4.71
	4	Lambda (−0.43, −0.17)	13.21	4.81
	5	Lambda (−0.55, −0.20)	22.21	5.14
D	1	ScConN (0.2, 2.0)	4.68	3.70
	2	ScConN (0.2, 2.5)	6.15	4.28
	3	ScConN (0.2, 3.0)	7.54	4.89
	4	ScConN (0.2, 4.0)	9.75	5.97
	5	ScConN (0.2, 5.0)	11.14	6.87
E	1	ScConN (0.4, 0.30)	4.47	4.36
	2	ScConN (0.4, 0.25)	4.62	4.62
	3	ScConN (0.4, 0.20)	4.75	4.91
	4	ScConN (0.4, 0.15)	4.86	5.23
	5	ScConN (0.4, 0.10)	4.93	5.58
F	1	$S_U(0, 1.30)$	10.10	4.64
	2	$S_U(0, 1.20)$	13.55	4.99
	3	$S_U(0, 1.10)$	20.36	5.49
	4	$S_U(0, 1.00)$	36.19	6.22
	5	$S_U(0, 0.95)$	57.73	6.59

Note: Descriptions of the distributions in Sets A and Bcan be found in Evans et al. (1993). Set C contains a familyof four-parameter $(\lambda_1, \lambda_2, \lambda_3, \lambda_4)$Lambda distributions described by Ramberg et al. (1979). Symmetric Lambda distributions with $\lambda_1 = 0$ and$\lambda_3 = \lambda_4$ are specified here in terms ofλ_2 and λ_3. Sets D and E containScConN(p, σ) distributions in which anobservation is sampled with probability $1 - p$ from N(0, 1) andprobability p from N$(0, \sigma^2)$(Gleason, 1993). Set F contains a set of symmetricdistributions from Johnson's S_U family of distributions(Stuart and Ord, 1994, p. 247).

index was developed by D'Agostino (1970a, b) and also was found to perform very well in small samples. The Type I error rate of the new z_ω test will now be examined.

Table 2 gives the results of a computer simulation of the empirical one-sided and two-sided Type I error rates for the z_β test and the z_ω test at $\alpha = [0.10\ 0.05\ 0.01]$

Table 2
Empirical Type I error rates for kurtosis tests

	z_β test			z_ω test			Joint test		
n	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$	$\alpha = 0.10$	$\alpha = 0.05$	$\alpha = 0.01$
Two-sided test									
10	0.089	0.041	0.005	0.099	0.042	0.008	0.081	0.035	0.006
20	0.095	0.047	0.009	0.099	0.048	0.010	0.087	0.042	0.010
30	0.099	0.050	0.011	0.100	0.049	0.011	0.092	0.047	0.011
40	0.101	0.053	0.012	0.100	0.049	0.011	0.096	0.049	0.012
50	0.102	0.053	0.013	0.100	0.050	0.011	0.097	0.050	0.012
100	0.103	0.053	0.013	0.100	0.050	0.010	0.099	0.052	0.013
One-sided test									
10	0.106	0.056	0.011	0.063	0.036	0.012	0.084	0.046	0.012
20	0.100	0.051	0.011	0.077	0.044	0.014	0.089	0.049	0.013
30	0.100	0.049	0.010	0.084	0.047	0.014	0.092	0.049	0.013
40	0.099	0.049	0.009	0.087	0.048	0.014	0.093	0.049	0.013
50	0.099	0.049	0.009	0.088	0.048	0.014	0.094	0.050	0.013
100	0.098	0.048	0.009	0.090	0.050	0.013	0.097	0.051	0.012

and $n = [10\ 20\ 30\ 40\ 50\ 100]$ based on 300,000 Monte Carlo samples from a normal population. The simulation program was written in Gauss and executed in a Pentium III class computer.

Table 2 shows that the two-sided Type I error rate for the z_β test is slightly conservative for $n < 30$ and slightly liberal for $n > 30$. The two-sided Type I error rates for the z_ω test are very close to α for all $n \geqslant 20$. The z_ω test has slightly better two-sided Type I error control than the z_β test or the D'Agostino test based on $\hat{\tau}/\hat{\sigma}$ (see D'Agostino, 1970a).

An upper one-sided test provides an important test of leptokurtosis. Table 2 shows that the upper one-sided Type I errors rates for the z_β test are very close to α for all $n \geqslant 20$. The upper one-sided Type I error rates for the z_ω test are slightly conservative at $\alpha = 0.10$ and $\alpha = 0.05$ but slightly liberal at $\alpha = 0.01$. The z_β test has slightly better upper one-sided Type I error control than the z_ω test.

5. Residual kurtosis in multiple-group designs

Bonett and Woodward (1990) found that the z_β test can be applied to the residual scores $(y_{ij} - \hat{\mu}_j)$ in multiple-group designs if $n_j \geqslant 10$. We found that the z_ω test also retains its excellent Type I error control with ANOVA residual scores if $n_j \geqslant 10$. This result is important because the sample size may be too small to test normality with adequate power within each group. In multiple-group designs, a more powerful test of normality can be obtained by computing z_ω or z_β from the total sample of residual scores.

Table 3
Power of two-sided kurtosis tests and Shapiro–Wilk test at $\alpha = 0.05$

Set	Member	n	z_β	z_ω	Joint	Shapiro–Wilk
A	1	23	0.761	0.601	0.714	0.817
	2	43	0.809	0.538	0.754	0.636
	4	390	0.807	0.792	0.827	0.744
	5	75	0.650	0.812	0.792	0.681
B	1	240	0.806	0.780	0.818	0.758
	2	175	0.805	0.793	0.822	0.766
	3	120	0.805	0.810	0.828	0.780
	4	72	0.784	0.805	0.813	0.774
	5	40	0.785	0.818	0.819	0.792
C	1	66	0.807	0.499	0.747	0.522
	2	105	0.774	0.802	0.809	0.756
	3	95	0.782	0.814	0.820	0.766
	4	87	0.767	0.800	0.806	0.753
	5	75	0.769	0.805	0.809	0.759
D	1	240	0.818	0.732	0.809	0.753
	2	100	0.801	0.760	0.803	0.767
	3	63	0.808	0.790	0.819	0.791
	4	36	0.799	0.804	0.822	0.810
	5	25	0.762	0.783	0.795	0.801
E	1	80	0.467	0.812	0.773	0.619
	2	60	0.427	0.806	0.765	0.744
	3	50	0.418	0.826	0.787	0.651
	4	40	0.392	0.808	0.772	0.662
	5	35	0.381	0.802	0.770	0.704
F	1	95	0.772	0.804	0.812	0.752
	2	77	0.766	0.805	0.809	0.756
	3	61	0.753	0.800	0.800	0.749
	4	50	0.758	0.811	0.809	0.762
	5	45	0.756	0.812	0.809	0.767
Mean (SD)			0.717 (0.143)	0.773 (0.082)	0.798 (0.027)	0.737 (0.065)
Mean (SD) at		$n = 25$	0.368 (0.171)	0.397 (0.175)	0.408 (0.175)	0.375 (0.185)
		$n = 50$	0.585 (0.213)	0.644 (0.219)	0.662 (0.207)	0.586 (0.217)
		$n = 100$	0.795 (0.178)	0.844 (0.184)	0.861 (0.171)	0.806 (0.183)

6. Power of kurtosis tests

The z_β test is known to be a powerful test of normality (D'Agostino et al., 1990). Table 3 compares the empirical powers of the two-sided z_β and z_ω tests at $\alpha = 0.05$ for all of the distributions in Table 1 except the normal. Table 4 compares the empirical

Table 4
Power of upper one-sided kurtosis tests and Shapiro–Wilk test at $\alpha = 0.05$

Set	Member	n	z_β	z_ω	Joint	Shapiro–Wilk
A	4	320	0.816	0.795	0.829	0.666
	5	63	0.691	0.806	0.788	0.611
B	1	195	0.805	0.770	0.810	0.682
	2	150	0.805	0.793	0.822	0.710
	3	99	0.804	0.796	0.816	0.709
	4	63	0.803	0.807	0.819	0.723
	5	35	0.798	0.812	0.818	0.746
C	2	90	0.795	0.804	0.817	0.695
	3	80	0.790	0.804	0.814	0.699
	4	77	0.796	0.809	0.820	0.706
	5	65	0.788	0.803	0.812	0.709
D	1	190	0.811	0.724	0.798	0.667
	2	83	0.809	0.755	0.804	0.696
	3	50	0.799	0.760	0.796	0.715
	4	30	0.799	0.781	0.805	0.743
	5	23	0.801	0.789	0.810	0.765
E	1	67	0.558	0.813	0.775	0.573
	2	50	0.523	0.800	0.761	0.753
	3	40	0.496	0.797	0.758	0.548
	4	34	0.486	0.797	0.760	0.589
	5	30	0.480	0.789	0.761	0.638
F	1	80	0.782	0.796	0.806	0.689
	2	67	0.790	0.808	0.816	0.702
	3	55	0.787	0.812	0.814	0.711
	4	45	0.789	0.817	0.818	0.722
	5	40	0.786	0.818	0.818	0.722
Mean (SD)			0.737 (0.117)	0.794 (0.022)	0.803 (0.022)	0.688 (0.055)
Mean (SD) at		$n = 25$	0.440 (0.154)	0.462 (0.173)	0.474 (0.165)	0.385 (0.161)
		$n = 50$	0.645 (0.177)	0.700 (0.202)	0.708 (0.188)	0.596 (0.203)
		$n = 100$	0.839 (0.142)	0.873 (0.165)	0.883 (0.148)	0.805 (0.187)

powers of the one-sided z_β and z_ω tests at $\alpha = 0.05$ for all leptokurtic distributions in Table 1. The Shapiro–Wilk test (Shapiro and Wilk, 1965) is also a very powerful test of normality (Gan and Koehler, 1990) and this popular test, based on Royston's (1992) computational method, is included for comparison. The results in Tables 3 and 4 are based on 100,000 Monte Carlo samples from each distribution. To make the comparisons among the three tests more meaningful and less complicated, the empirical powers are reported for the sample size at which the most powerful test achieves power of about 0.8. The sample sizes in Tables 3 and 4 also provide sample size requirements to reject the null hypothesis of normality with power of 0.8 for a wide variety of symmetric nonnormal distributions. Tables 3 and 4 also report the mean power at $n = 25, 50$, and 100.

Table 3 shows that the mean power of the two-sided z_ω test is higher than the mean power of the two-sided z_β test and the Shapiro–Wilk test across the 29 nonnormal distributions in Table 1. However, the z_β and z_ω tests are sensitive to different types of kurtosis. Specifically, the z_β test is far more powerful than the z_ω test in platykurtic distributions while the z_ω test tends to be more powerful than the z_β test in highly peaked leptokurtic distributions such as the Laplace and the scale-contaminated normals in Set E.

Table 4 shows that the mean power of the one-sided z_ω test is higher than the mean power of the one-sided z_β test and the Shapiro–Wilk test. The z_ω test is more powerful than the Shapiro–Wilk test in all 26 distributions and more powerful than the z_β test in 17 of the 26 distributions. These findings extend the results of D'Agostino and Roseman (1974) who found that D'Agostino's test based on $\hat{\tau}/\hat{\sigma}$ was more powerful than the Shapiro–Wilk test in detecting certain leptokurtic distributions.

The z_ω test is a powerful test of leptokurtosis in symmetric distributions. If the distribution is skewed, the z_ω test should be accompanied by the standard test of skewness (DeCarlo, 1997) which together provide a powerful and informative test of normality in a manner described by D'Agostino et al. (1990).

7. Joint test of kurtosis

The z_β test is more powerful than the z_ω test in all platykurtic distributions we examined and the z_ω test is more powerful than the z_β test in most leptokurtic distributions. A more uniformly powerful test of kurtosis can be obtained by using both Pearson's measure of kurtosis and Geary's measure of kurtosis. Although a joint Bonferroni test of H_0: $\beta_2 = 3$ and $\omega = 3$ could be used in an obvious way, a more powerful joint test can be obtained by exploiting the correlation between z_β and z_ω under the null hypothesis of normality.

Under the null hypothesis of normality, Geary (1936) showed that $\hat{\beta}_2$ and $\hat{\tau}/\hat{\sigma}$ have an asymptotic bivariate normal distribution with $\rho = -\{12(\pi - 3)\}^{-1} = -0.588$. Given the complexity of z_β, we were not able to derive the correlation between z_β and z_ω but we were able to estimate it very accurately using Monte Carlo methods. The large-sample correlation between z_β and z_ω was found to be approximately 0.77. One-sided and two-sided critical z-values were obtained by numerically integrating over a standardized bivariate normal distribution with $\rho = 0.77$ are given in Table 5.

Let $z^*_{\alpha/2}$ denote a two-sided critical bivariate z-value and let z^*_α denote a one-sided critical bivariate z-value from Table 5. Reject normality if $|z_\beta| > z^*_{\alpha/2}$ or $|z_\omega| > z^*_{\alpha/2}$. A one-sided test of leptokurtosis rejects normality if $z_\beta > z^*_\alpha$ or $z_\omega > z^*_\alpha$.

Table 2 shows that the joint test exhibits Type I error control similar to the z_β test. Table 3 shows that the joint test is more powerful, on average, than any of the other tests. Furthermore, the joint test is more powerful than the Shapiro–Wilk test for 28 of the 29 distributions and more powerful than the z_β test for 25 of the 29 distributions

Table 5
Critical z-values for joint test

α	One-sided	Two-sided
0.10	1.505	1.857
0.05	1.857	2.162
0.01	2.517	2.759

in Table 3. Table 4 shows that the joint test is a more powerful test of leptokurtosis than any of the other tests for all 27 leptokurtic distributions in Table 4. Although no currently available test of normality is known to be uniformly most powerful across all alternative distributions, the joint test of kurtosis has very high uniform power across the set of leptokurtic distributions considered here. Tables 3 and 4 also report the mean power across all distributions at $n = 25, 50$ and 100. Note that the joint test has the highest mean power at all three sample sizes. The variability in power across distributions for the joint test at $n = 25, 50$ and 100 is low and this suggests that its high mean power is not simply the result of a few cases where it has unusually high power.

8. Summary

Kurtosis can be measured in more than one way, just as centrality or dispersion can be measured in more than one way. Compared to β_2, ω is influenced more by scores close to the center of the distribution, and compared to ω, β_2 is influenced more by scores in the tails of the distribution. Consequently, the z_β test is more powerful than the z_ω test in platykurtic distributions, but the z_ω test is more powerful than the z_β test in all highly peaked symmetric leptokurtic distributions we examined. The joint kurtosis test introduced here exploits the unique strengths of two different measures of kurtosis and has high uniform power across a wide range of symmetric nonnormal distributions.

Although the Shapiro–Wilk test is among the most powerful and widely used tests of normality, when this test rejects the null hypothesis of normality it provides no information regarding the shape of the distribution. D'Agostino et al. (1990) argue convincingly that tests based on skewness and kurtosis measures are especially useful in assessing the shape of a nonnormal distribution. Since ω and β_2 are sensitive to different types of kurtosis, some researchers will now want to use the new joint test of kurtosis along with the standard test of skewness.

The z_ω test is easy to compute, has excellent power and Type I error control, and does not require special tables of critical values. The simplicity of ω and z_ω make them ideal for textbook and classroom applications. Although it is now common to avoid the discussion of kurtosis in introductory treatments of statistics, the fact that ω is a function of two common measures of dispersion and that the z_ω test does not require special tables of critical values might encourage some authors to include this important topic in future versions of their texts.

In conclusion, we suggest that ω be referred to as *G-kurtosis* in honor of Geary and to distinguish this type of kurtosis from Pearson's measure. We also suggest that statistical packages report G-kurtosis, z_ω and its p-value in addition to Pearson's kurtosis, z_β and its p-value. The joint test described here can then be easily obtained from z_ω and z_β. If the joint test leads to a rejection of the null hypothesis, many researchers will want to examine a normal probability plot to obtain additional descriptive information regarding the shape of the distribution.

Acknowledgements

We would like to thank the referees for their many helpful comments and suggestions.

References

Anscombe, F.J., Glynn, W.J., 1983. Distribution of kurtosis statistic b_2 for normal statistics. Biometrika 70, 227–234.

Balanda, K., MacGillivray, H.L., 1988. Kurtosis: a critical review. Amer. Statist. 42, 111–119.

Bollen, K.A., 1989. Structural Equation Models with Latent Variables. Wiley, New York.

Bonett, D.G., Woodward, J.A., 1990. Testing residual normality in the ANOVA model. J. Appl. Statist. 17, 383–387.

D'Agostino, R.B., 1970a. Geary's test revisited. Psychol. Bull. 74, 138–140.

D'Agostino, R.B., 1970b. Transformation to normality of the null distribution of g_1. Biometrika 57, 679–681.

D'Agostino, R.B., Roseman, B., 1974. The power of Geary's test of normality. Biometrika 61, 181–184.

D'Agostino, R.B., Stephens, M.A., 1986. Goodness-of-fit Techniques. Marcel Dekker, New York.

D'Agostino, R.B., Belanger, A., D'Agostino Jr., R.B., 1990. A suggestion for using powerful and informative tests of normality. Amer. Statist. 44, 316–322.

DeCarlo, L.T., 1997. On the meaning and use of kurtosis. Psychol. Meth. 3, 292–307.

Evans, M., Hastings, N., Peacock, B., 1993. Statistical Distributions, 2nd Edition. Wiley, New York.

Gan, F.F., Koehler, K.L., 1990. Goodness-of-fit tests based on P–P probability plots. Technometrics 32, 289–303.

Geary, R.C., 1936. Moments of the ratio of the mean deviation to the standard deviation for normal samples. Biometrika 28, 295–307.

Gleason, J.R., 1993. Understanding elongation: the scale contaminated normal family. J. Amer. Statist. Assoc. 88, 327–337.

Hodges, J.L., Lehmann, E.L., 1956. The efficiency of some nonparametric competitors of the t-test. Ann. Math. Statist. 27, 224–235.

Hu, L.-T., Bentler, P.M., Kano, Y., 1990. Can test statistics in covariance structure models be trusted? Psychol. Bull. 112, 351–362.

Johnson, N.L., Kotz, S., Balakrishnan, N., 1994. Continuous Univariate Distributions, Vol. 1. Wiley, New York.

Kowalski, C.J., 1972. On the effects of non-normality on the distribution of the sample product moment correlation coefficient. Appl. Statist. 21, 1–12.

Micceri, T., 1989. The unicorn, the normal curve, and other improbable creatures. Psychol. Bull. 105, 156–166.

Oja, H., 1981. On location, scale, skewness and kurtosis of univariate distributions. Scand. J. Statist. 8, 154–168.

Pearson, K., 1905. Skew variation, a rejoinder. Biometrika 4, 169–212.

Ramberg, J.S., Tadikamalla, P.R., Dudewicz, E.J., Mykytka, C.F., 1979. A probability distribution and its uses in fitting data. Technometrics 21, 201–214.

Ramsey, P.H., Ramsey, P.P., 1993. Updated version of the critical values of the standardized fourth moment. J. Statist. Comput. Simulation 44, 231–241.

Ramsey, P.H., Ramsey, P.P., 1995. Correction to Updated version of the critical values of the standardized fourth moment (V44, 231–241) J. Statist. Comput. Simulation 53, 127.

Royston, P., 1992. Approximating the Shapiro–Wilk W-test for non-normality. Statist. Comput. 2, 117–119.

Ruppert, D., 1987. What is kurtosis? An influence function approach. Amer. Statist. 41, 1–5.

Scheffé, H., 1959. The Analysis of Variance. Wiley, New York.

Shapiro, S.S., Wilk, M.B., 1965. An analysis of variance test for normality (complete samples) Biometrika 52, 591–611.

Stuart, A., Ord, J.K., 1994. Kendall's Advanced Theory of Statistics, Vol. 1: Distribution Theory. Edward Arnold, London.

van Zwet, W.R., 1964. Convex Transformations of Random Variables. Mathematical Centre Tract 7. Mathematisch Centrum Amsterdam.

Zwillinger, D., Kokoska, S., 2000. CRC Standard Probability and Statistics Tables and Formulae. Chapman & Hall, New York.