X^2 and Its Components as Tests of Normality for Grouped Data

BY D. J. BEST, J. C. W. RAYNER*
School of Mathematical and Physical Sciences,
University of Newcastle, NSW 2308, Australia
John.Rayner@newcastle.edu.au

AND O. THAS
Department of Applied Mathematics, Biometrics and Process Control,
Ghent University, B-9000 Gent, Belgium
olivier.thas@Ugent.be

ABSTRACT *We consider testing for an unobservable normal distribution with unspecified mean and variance. It is only possible to observe the counts in groups with boundaries specified before sighting the data. On the basis of a small power study we recommend the usual X^2 test be used as an omnibus test, augmented by informal examination of the first two non-zero component of X^2. We also recommend use of maximum likelihood and method of moments estimation.*

Key Words: Critical values; improved grouped normal models; maximum likelihood estimation; method of moments estimation; power study

1. Introduction

In the Table 1 below we give counts of 1053 mothers grouped in two inch classes for height. These data are derived from data given in Pearson and Lee (1903). The question of interest is are the underlying data normally distributed?

In the Table 1 scenario we are given a random sample of n observations of a random variable X but all that is known about these observations is into which of K mutually exclusive predetermined groups they fall. This situation occurs in practice because

the measuring instrument only gives readings to a certain accuracy and

only the histogram counts may be available now even if individual measurements were once available.

We assume the K groups have boundaries $k_1, k_2, \dots, k_{K-1}$ specified before sighting the data, and we also take $k_0 = -\infty$ and $k_K = \infty$. The null hypothesis to be tested is that X has a normal distribution with probability density function

$$f(x;\, \mu,\, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left\{-\left(x-\mu\right)^2 / \left(2\sigma^2\right)\right\} \text{ for } -\infty < x < \infty,$$
$$\text{in which } -\infty < \mu < \infty, \text{ and } 0 < \sigma < \infty.$$

For $j = 1, 2, \dots, K$ the probability of an observation in group j is

$$p_j = \Phi((k_j - \mu)/\sigma) - \Phi((k_{j-1} - \mu)/\sigma), \text{ in which } \Phi(z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} \exp\left(-x^2/2\right) dx.$$

Again for $j = 1, 2, \dots, K$ let N_j be the number of the n observations that fall into the jth group and for the Table 1 and Table 5 data and $j = 2, \dots, K-1$, let $x_j = (k_j + k_{j-1})/2$ with $x_1 = k_1 - (k_2$

$- k_1)/2 = (3k_1 - k_2)/2$ and $x_K = k_{K-1} + (k_{K-1} - k_{K-2})/2 = (3k_{K-1} - k_{K-2})/2$. The definitions of x_1 and x_K ensure all the x_js are equi-spaced if the k_j are equi-spaced. This is a common approach; see, for example, Hoel (1984, p.258) and Freund (2004, exercise 14-41, p.355). The sample mean and variance of the grouped data are given by

$$\overline{X} = \sum_{j=1}^{K} N_j x_j / n \text{ and } S^2 = \sum_{j=1}^{K} N_j \left(x_j - \overline{X}\right)^2 / n .$$

The maximum likelihood estimators (MLEs) $\hat{\mu}$ of μ and $\hat{\sigma}$ of σ are obtained by iteratively solving the non-linear equations derived by differentiating the logarithm of the likelihood of the sample. Initial estimates of $\hat{\mu}$ and $\hat{\sigma}$ are taken to be $\hat{\mu}_0 = \overline{X}$ and $\hat{\sigma}_0 = S$ respectively, and then new estimates are obtained by bivariate Newton-Raphson. This process is repeated until convergence is reached. Further details are given in Appendix A.

Once $\hat{\mu}$ and $\hat{\sigma}$ have been obtained it is straightforward to apply the test of normality described subsequently. For the Table 1 data $\hat{\mu}$ = 62.49 and $\hat{\sigma}$ = 2.37. Table 1 also gives the cell expectations, $E_j = n\hat{p}_j$, where the estimated $\hat{\mu}$ and $\hat{\sigma}$ have been used to obtain the class probabilities $\hat{p}_j$. If no pooling is done $X^2 = \sum_{j=1}^{K}\left(N_j - E_j\right)^2 / E_j$ = 13.45 which is asymptotically distributed as chi-squared with six degrees of freedom: χ_6^2. The corresponding p-value is 0.04 using this χ_6^2 approximation. Cochran (1952) suggested that when testing for normality one class expectation of 0.5 still allows a valid χ^2 approximation. It appears these data are not consistent with the grouped normal distribution.

Without access to relevant computer routines, finding the MLEs $\hat{\mu}$ and $\hat{\sigma}$ can be difficult, and so a traditional approach which is still given in textbooks estimates μ and σ by $\tilde{\mu} = \overline{X}$ and $\tilde{\sigma} = S$ respectively. See, for example, Selvanathan et al. (2000, section 17.4). However the X^2 statistic no longer has an asymptotic χ_{k-3}^2 distribution. See Fisher (1924) and Rayner and Best (1989, Chapter 7). The question is, in general what are the consequences of using $(\tilde{\mu}, \tilde{\sigma})$ instead of $(\hat{\mu}, \hat{\sigma})$?

For the Table 1 data $\tilde{\mu}$ = 62.49 and $\tilde{\sigma}$ = 2.44 with X^2 = 12.56. Using χ_6^2 to approximate the distribution of X^2 we find a p-value of 0.051. Thus if we had taken $\alpha = 0.05$, use of $(\hat{\mu}, \hat{\sigma})$ gives a significant X^2 while use of $(\tilde{\mu}, \tilde{\sigma})$ gives a (just) non-significant value. The structure of the paper is as follows. Section 2 considers components of X^2, section 3 gives a study of critical values and powers, while section 4 uses the components to derive a model better than that based on the $\{n\hat{p}_j\}$.

Table 1. Heights of mothers (in inches)

Class interval	$(-\infty, 55)$	(55, 57)	(57, 59)	(59, 61)	(61, 63)
Frequency	3	8	53	215	346
Cell expectations (E_j)	0.8	10.1	63.5	204.3	336.6
Class interval	(63, 65)	(65, 67)	(67, 69)	$(69, \infty)$	
Frequency	277	120	24	7	

Cell expectations (E_j)	284.3	123.0	27.2	3.2	

2. Components of the Chi-Squared Statistic

A more thorough examination of the null grouped normal hypothesis can be made by looking at the components of the X^2 statistic. These components may be calculated via orthonormal polynomials as in Lancaster (1953) or Rayner and Best (1989). The rth component of X^2 is defined as

$$V_r = \sum_{j=1}^{K} N_j g_r(x_j)/\sqrt{n} \text{ for } r = 1, 2, \ldots, K-1,$$

in which $\{g_r(x_j)\}$ are polynomials orthonormal on $\{p_j\}$, defined as follows. For an arbitrary distribution for which the following quantities exist, suppose μ is the familiar population mean while μ_r, $r = 2, 3, 4, \ldots$ are the population central moments:

$$\mu = \sum_{j=1}^{K} x_j p_j \text{ and } \mu_r = \sum_{j=1}^{K} (x_j - \mu)^r p_j \text{ for } r = 2, 3, 4, \ldots .$$

The first four orthonormal polynomials are

$$g_0(x_j) = 1 \text{ for all } x_j,$$
$$g_1(x_j) = (x_j - \mu)/\sqrt{\mu_2}$$
$$g_2(x_j) = \{ (x_j - \mu)^2 - \mu_3(x_j - \mu)/\mu_2 - \mu_2 \}/\sqrt{\mu_4 - \mu_3^2/\mu_2 - \mu_2^2} \text{ and}$$
$$g_3(x_j) = \frac{(x_j - \mu)^3 - a(x_j - \mu)^2 - b(x_j - \mu) - c}{\sqrt{\mu_6 - 2a\mu_5 + (a^2 - 2b)\mu_4 + 2(ab - c)\mu_3 + (b^2 + 2ac)\mu_2 + c^2}}$$

in which

$$a = (\mu_5 - \mu_3\mu_4/\mu_2 - \mu_2\mu_3)/d,\ b = (\mu_4^2/\mu_2 - \mu_2\mu_4 - \mu_3\mu_5/\mu_2 + \mu_3^2)/d,$$
$$c = (2\mu_3\mu_4 - \mu_3^3/\mu_2 - \mu_2\mu_5)/d \text{ and } d = \mu_4 - \mu_3^2/\mu_2 - \mu_2^2.$$

Appendix B gives an explicit formula for $g_4(x_j)$. Further polynomials may be given using the recurrence relations of Emerson (1968). Subsequently if (μ, σ) is estimated by $(\hat{\mu}, \hat{\sigma})$ then we refer to p_j as $\hat{p}_j$, and if (μ, σ) is estimated by $(\tilde{\mu}, \tilde{\sigma})$ then we refer to p_j as $\tilde{p}_j$.
The statistic X^2 may be expressed in terms of the components V_r via

$$X^2 = V_1^2 + \ldots + V_{K-1}^2.$$

See Lancaster (1953).

For the Table 1 data we find $\tilde{V}_1 = -0.05$, $\tilde{V}_2 = -1.20$, $\tilde{V}_3 = 0.52$ and $\tilde{V}_4 = 2.47$, where $\tilde{V}_r$ is V_r using $(\tilde{\mu}, \tilde{\sigma})$. We also find $\hat{V}_1 = -0.005$, $\hat{V}_2 = -0.05$, $\hat{V}_3 = 0.84$ and $\hat{V}_4 = 2.62$, where $\hat{V}_r$ is V_r using $(\hat{\mu}, \hat{\sigma})$. Later we show that, in agreement with the large $\tilde{V}_4$ and $\hat{V}_4$ that possibly reflect kurtosis values, the values of $(N_j - E_j)/\sqrt{E_j}$ are large in the tails. Also we suggest that $\hat{V}_1$ and $\hat{V}_2$ will be close to zero and that the other $\hat{V}_r$ are distributed approximately as $N(0, 1)$; this suggestion is supported by the simulations of the next section. The distributions of the $\tilde{V}_r$ seem to approximate those of the $\hat{V}_r$, but the exact details are unknown to us. In the next section we will look at the critical values of the tests based on both $\hat{V}_r$ and $\tilde{V}_r$ as n increases. We expect $\tilde{V}_3$ and $\tilde{V}_4$ to approximate the standardized sample skewness and kurtosis coefficients for grouped data. Calculation details of these grouped coefficients are given, for example, in Snedecor and Cochran (1989, sections 5.13 and 5.14).

3. Critical Values and Power Comparisons

Table 2 gives some critical values for X^2, V_3^2 and V_4^2 using the grouped frequency estimators ($\tilde{\mu}$, $\tilde{\sigma}$), called GRO in Table 2. Also given are the critical values using the maximum likelihood (ML) estimators ($\hat{\mu}$, $\hat{\sigma}$) and the method of moments (MOM) estimators obtained by solving $V_1 = V_2 = 0$. Both the ML and MOM methods of estimation require the use of the iterative bivariate Newton-Raphson method. See Appendix A for details. The critical values are for a standard normal distribution.

For $K = 10$ the categories were defined as $(-\infty, -3]$, $(-3, -2]$, $(-2, -1.5]$, $(-1.5, -0.5]$, $(-0.5, 0]$, $(0, 0.5]$, $(0.5, 1.5]$, $(1.5, 2]$, $(2, 3]$, $(3, \infty)$, and for $K = 20$ the categories were $(-\infty, -2]$, $(-2, -1.778]$, $(-1.778, -1.556]$, $(-1.556, -1.334]$, $(-1.334, -1.112]$, $(-1.112, -0.890]$, $(-0.890, -0.668]$, $(-0.668, -0.446]$, $(-0.446, -0.224]$, $(-0.224, 0]$, and the reflections of these categories.

In Table 2(a) only sample sizes $n = 500$ and $n = 1000$ meet the Cochran (1952, p. 329) criterion that the smallest class expectation should be 0.5 or greater. All the Table 2(b) sample sizes meet this criterion and we see that all the critical values agree well with the asymptotic critical values when this criterion is satisfied. Further, the ML critical values are generally better than the corresponding GRO values in the sense that they are closer to the χ^2 values. Moreover the corresponding ML and MOM values are generally very similar. The critical values shown in Table 2 are based on N(0, 1) random values, but were little changed for other (μ, σ).

Thus we suggest that in introductory statistics courses when the GRO approach is used it should be emphasised that the approach is approximate and a better method exists. For example, Selvanathan et al. (2000, section 17.4) could mention that what is presented is an approximation to a more efficient method.

Table 2(a). Critical values for X^2, V_3^2 and V_4^2 based on 20,000 simulations and categories as specified in the text for $K = 10$ and $\alpha = 0.5$, 0.25, 0.10, 0.05 and 0.01

Statistic	n	Estimator	0.50	0.25	0.10	0.05	0.01
X^2		ML	5.14	7.69	11.07	14.06	25.66
	50	GRO	4.91	7.24	10.09	12.15	17.62
		MOM	5.17	7.67	10.83	13.27	20.49
		ML	5.39	8.16	11.75	14.88	25.75
	75	GRO	5.17	7.63	10.57	12.78	17.62
		MOM	5.42	8.10	11.44	14.05	20.83
		ML	5.58	8.44	11.96	14.85	23.92
	100	GRO	5.38	7.91	10.75	12.70	17.80
		MOM	5.59	8.35	11.64	14.09	21.00
		ML	5.98	8.71	12.16	14.78	21.61
	200	GRO	5.89	8.34	11.17	13.35	17.72
		MOM	5.96	8.64	11.89	14.27	19.83

	500	ML	6.20	8.90	11.95	14.26	19.44
		GRO	6.91	9.41	12.17	14.12	18.40
		MOM	6.21	8.88	11.85	14.07	18.93
	1000	ML	6.23	8.91	11.97	14.12	19.41
		GRO	8.39	10.86	13.68	15.69	20.09
		MOM	6.24	8.91	11.92	14.07	19.10
	∞	ML	6.35	9.04	12.02	14.07	18.48
V_3^2	50	ML	0.30	0.95	2.22	3.57	7.50
		GRO	0.23	0.71	1.62	2.52	4.80
		MOM	0.30	0.96	2.14	3.24	6.24
	75	ML	0.32	1.03	2.45	3.82	7.50
		GRO	0.25	0.79	1.80	2.72	5.01
		MOM	0.33	1.05	2.36	3.54	6.53
	100	ML	0.34	1.10	2.52	3.90	7.25
		GRO	0.26	0.84	1.85	2.80	4.97
		MOM	0.35	1.11	2.42	3.64	6.41
	200	ML	0.41	1.22	2.65	3.85	6.77
		GRO	0.31	0.93	1.99	2.90	4.90
		MOM	0.41	1.21	2.58	3.75	6.39
	500	ML	0.44	1.28	2.69	3.87	6.96
		GRO	0.34	0.99	2.03	2.93	5.23
		MOM	0.44	1.28	2.67	3.83	6.76
	1000	ML	0.44	1.29	2.70	3.86	6.69
		GRO	0.36	0.99	2.08	2.95	5.10
		MOM	0.44	1.29	2.69	3.85	6.62
	∞	ML	0.45	1.32	2.71	3.84	6.64
V_4^2	50	ML	0.30	0.70	1.36	2.65	9.43
		GRO	0.18	0.48	1.04	2.08	6.00
		MOM	0.32	0.75	1.41	2.37	7.20
	75	ML	0.34	0.82	1.70	3.34	9.31
		GRO	0.21	0.57	1.35	2.56	6.05
		MOM	0.36	0.86	1.71	2.97	7.51
	100	ML	0.41	0.91	1.82	3.37	8.85
		GRO	0.25	0.65	1.48	2.67	6.12
		MOM	0.43	0.96	1.85	3.06	7.47
	200	ML	0.52	1.12	2.04	3.20	7.56
		GRO	0.33	0.80	1.67	2.69	5.79
		MOM	0.54	1.17	2.09	3.08	6.82
	500	ML	0.49	1.32	2.43	3.39	6.94
		GRO	0.35	0.97	1.96	3.11	6.08
		MOM	0.51	1.37	2.53	3.42	6.59
	1000	ML	0.46	1.30	2.61	3.68	6.60
		GRO	0.36	1.06	2.30	3.41	6.30
		MOM	0.47	1.35	2.70	3.77	6.59
	∞	ML	0.45	1.32	2.71	3.84	6.64

Table 2(b). Critical values for X^2, V_3^2 and V_4^2 based on 20,000 simulations and categories as specified in the text for K = 20 and α = 0.5, 0.25, 0.10, 0.05 and 0.01

Statistic	n	Estimator	0.50	0.25	0.10	0.05	0.01
X^2		ML	16.42	20.35	24.62	27.44	34.00
	50	GRO	16.87	20.87	25.12	28.00	34.52
		MOM	16.44	20.35	24.55	27.40	33.72
		ML	16.37	20.40	24.33	27.52	33.36
	75	GRO	16.88	21.00	25.94	28.15	34.08
		MOM	16.39	20.42	24.71	22.44	33.34
		ML	16.33	20.43	24.73	27.65	33.33
	100	GRO	16.90	21.09	25.45	28.35	34.03
		MOM	16.37	20.45	24.75	27.60	33.28
		ML	16.39	20.50	24.68	27.38	33.44
	200	GRO	17.13	21.31	25.56	28.33	34.32
		MOM	16.40	20.52	24.68	27.38	33.39
		ML	16.34	20.51	24.81	27.73	33.58
	500	GRO	17.58	21.83	26.23	29.23	35.14
		MOM	16.37	20.55	24.86	27.73	33.58
		ML	16.38	20.50	24.74	27.62	33.45
	1000	GRO	18.42	22.65	27.05	29.93	36.02
		MOM	16.40	20.52	24.80	27.66	33.51
	∞	ML	16.34	20.49	24.77	27.59	33.41
V_3^2		ML	0.46	1.32	2.65	3.73	6.42
	50	GRO	0.49	1.43	2.91	4.08	6.94
		MOM	0.47	1.37	2.74	3.83	6.64
		ML	0.46	1.32	2.69	3.84	6.63
	75	GRO	0.49	1.41	2.89	4.13	7.15
		MOM	0.46	1.32	2.71	3.85	6.74
		ML	0.46	1.32	2.73	3.82	6.41
	100	GRO	0.50	1.42	2.92	4.10	6.92
		MOM	0.46	1.33	2.74	3.85	6.50
		ML	0.47	1.37	2.73	3.89	6.82
	200	GRO	0.50	1.47	2.92	4.18	7.28
		MOM	0.47	1.38	2.74	3.91	6.84
		ML	0.45	1.34	2.75	3.90	6.77
	500	GRO	0.48	1.43	2.95	4.18	7.33
		MOM	0.45	1.34	2.76	3.92	6.86
		ML	0.47	1.35	2.71	3.84	6.48
	1000	GRO	0.50	1.44	2.90	4.11	6.96
		MOM	0.47	1.35	2.72	3.85	6.51
	∞	ML	0.45	1.32	2.71	3.84	6.64
V_4^2		ML	0.49	1.37	2.72	3.80	6.59
	50	GRO	0.52	1.46	2.92	4.08	6.99
		MOM	0.50	1.41	2.80	3.90	6.79
		ML	0.46	1.34	2.74	3.82	6.40
	75	GRO	0.49	1.44	2.92	4.11	7.04
		MOM	0.47	1.38	2.82	3.93	6.68
		ML	0.45	1.31	2.62	3.72	6.44
	100	GRO	0.48	1.40	2.81	3.99	6.97
		MOM	0.46	1.35	2.69	3.82	6.63

	ML	0.45	1.29	2.65	3.76	6.31
200	GRO	0.49	1.40	2.85	4.02	7.00
	MOM	0.46	1.34	2.73	3.88	6.57
	ML	0.46	1.32	2.69	3.77	6.28
500	GRO	0.49	1.44	2.92	4.16	7.03
	MOM	0.47	1.37	2.77	3.90	6.48
	ML	0.45	1.30	2.66	3.76	6.65
1000	GRO	0.50	1.44	2.93	4.19	7.35
	MOM	0.46	1.34	2.73	3.87	6.84
∞	ML	0.45	1.32	2.71	3.84	6.64

It is often suggested that the classes be pooled so that the χ^2 approximation can be used to give p-values for X^2 when testing for grouped normality. We recommend that pooling not be done and that when the 0.5 criterion is not met, p-values should be found by parametric bootstrap. Recent discussions concerning finding p-values via parametric bootstrap, in a goodness of fit context, are given in Gurtler and Henze (2000) and Gulati and Neus (2001).

It appears that the only goodness of fit tests for grouped normality in the literature are those based on the coefficients of skewness and kurtosis, and X^2. However other tests can easily be constructed. Since the Anderson-Darling A^2 test is very competitive for testing normality with ungrouped data, we will now compare powers of tests based on X^2, V_3^2, V_4^2 and a grouped version of the Anderson-Darling A^2. Put $H_j = \hat{p}_1 + ... + \hat{p}_j$, $j = 1, 2, \dots, K$, and

$$A^2 = n\sum_{j=1}^{K-1} R_j^2 \hat{p}_j / \{H_j(1 - H_j)\},$$

where for $j = 1, 2, \dots, K$, $R_j = N_1 + \dots + N_j - n(\hat{p}_1 + ... + \hat{p}_j)$.

Table 3(a) gives powers found using the parametric bootstrap technique as recently used by Gurtler and Henze (2000) and by Gulati and Neus (2001). Simulation runs of 1000 were used for both the inner and outer loops of the bootstrap. Given the results in Table 2 we would expect powers of X^2, V_3^2 and V_4^2 could also be found using appropriate χ^2 critical values and this is done in Table 3(b). The powers in Tables 3(a) and 3(b) are in very good agreement, verifying the accuracy of the χ^2 approximations for these alternatives and showing that the tests based on X^2 and A^2 generally have similar power. No pooling was done in the power calculations for X^2, V_3^2 and V_4^2. Use of V_3^2 and V_4^2 to amplify the X^2 test would seem sensible as V_3^2 does well for skewed alternatives and V_4^2 does well for alternatives with kurtosis different to that of the normal distribution. The Anderson-Darling test also provides a good omnibus test for grouped normality. However p–values for the Anderson-Darling test cannot be found using a convenient χ^2 approximation. The powers shown are for a $N(0, 1)$ null but seem very similar to those for general $N(\mu, \sigma^2)$ null hypotheses.

The first alternative considered in Table 3(a) is the uniform with $K = 8$, 10 and 12, $k_j = j/K$ for $j = 1, \dots, K - 1$ and with $x_1 = k_1/2$, $x_K = k_{K-1} + x_1$. Other x_j are class midpoints. The second alternative is the logistic distribution with $K = 11$, 13 and 15, $k_j = (j - 1)/2 - (K - 2)/4$ for $j = 1, \dots, K - 1$, $x_1 = (3k_1 - k_2)/2$, $x_K = (3k_{K-1} - k_{K-2})/2$ and with other x_j equal to the class midpoints. Classes for the Normal, Laplace and Student's t_3 alternatives were as defined for the logistic. The gamma (5) and gamma (6) distributions give skewed alternatives for which we take $K = 4$, 5, and 6, $k_j = (7 - K) + 2(j - 1)$, $x_1 = k_1/2$, $x_K = k_{K-1} + x_1$ and other x_j equal to the class midpoints. Another skewed alternative is the extreme value distribution with $K = 6$, 8 and 10, $k_j = (j - 1)/2 - (K - 2)/4$ for $j = 1, \dots, K - 1$ as for the logistic alternative. Here $x_1 =$

$(3k_1 - k_2)/2$, $x_K = (3k_{K-1} - k_{K-2})/2$ and other x_j are equal to the class midpoints. The first four alternatives are symmetric and the last three are skewed. Random values from these distributions were obtained using the IMSL (1995) software package.

Table 3(a). Parametric bootstrap powers against specified alternatives, for test size 0.05, n = 100 and various numbers of classes K

		MOM				ML				GRO		
Alternative	K	X^2	V_3^2	V_4^2	A^2	X^2	V_3^2	V_4^2	A^2	X^2	V_3^2	V_4^2
Uniform (0, 1)	8	0.21	0.06	0.31	0.21	0.20	0.05	0.34	0.19	0.11	0.05	0.13
	10	0.32	0.06	0.47	0.32	0.34	0.05	0.51	0.30	0.23	0.05	0.25
	12	0.38	0.07	0.56	0.36	0.36	0.04	0.56	0.35	0.27	0.04	0.29
Logistic	11	0.06	0.07	0.08	0.06	0.06	0.06	0.07	0.06	0.06	0.06	0.12
	13	0.07	0.07	0.10	0.08	0.05	0.06	0.11	0.08	0.07	0.06	0.13
	15	0.08	0.08	0.16	0.11	0.06	0.07	0.15	0.11	0.07	0.07	0.18
Laplace	11	0.40	0.08	0.62	0.46	0.36	0.07	0.64	0.46	0.38	0.08	0.66
	13	0.53	0.09	0.72	0.59	0.46	0.09	0.73	0.60	0.48	0.10	0.75
	15	0.57	0.12	0.83	0.72	0.57	0.14	0.80	0.69	0.59	0.14	0.81
Student's t_3	11	0.21	0.07	0.42	0.27	0.18	0.08	0.41	0.25	0.21	0.08	0.46
	13	0.30	0.10	0.64	0.39	0.26	0.08	0.62	0.38	0.30	0.09	0.64
	15	0.43	0.15	0.78	0.51	0.44	0.15	0.78	0.54	0.49	0.16	0.79
Normal	11	0.06	0.06	0.06	0.04	0.05	0.06	0.06	0.06	0.05	0.06	0.06
	13	0.06	0.05	0.06	0.04	0.06	0.05	0.05	0.06	0.06	0.05	0.05
	15	0.05	0.06	0.06	0.05	0.04	0.04	0.06	0.06	0.05	0.06	0.04
Gamma (5)	4	0.24	0.25	0.05	0.23	0.23	0.23	0.01	0.23	0.21	0.27	0.02
	5	0.51	0.60	0.07	0.49	0.52	0.61	0.07	0.53	0.52	0.61	0.07
	6	0.64	0.78	0.07	0.59	0.64	0.77	0.05	0.62	0.63	0.77	0.05
Gamma (6)	4	0.22	0.22	0.07	0.22	0.20	0.23	0.06	0.23	0.20	0.23	0.01
	5	0.39	0.51	0.06	0.41	0.39	0.48	0.06	0.38	0.39	0.49	0.06
	6	0.36	0.62	0.02	0.47	0.36	0.62	0.02	0.44	0.37	0.64	0.02
Extreme Value	6	0.21	0.32	0.07	0.25	0.23	0.32	0.06	0.27	0.10	0.36	0.04
	8	0.33	0.61	0.05	0.47	0.31	0.61	0.03	0.45	0.31	0.63	0.03
	10	0.39	0.81	0.05	0.61	0.36	0.79	0.03	0.64	0.40	0.81	0.04

Table 3(b). Powers against specified alternatives, for test size 0.05 and n = 100 using χ^2 critical values and 10,000 Monte Carlo simulations

		MOM			ML			GRO		
Alternative	K	X^2	V_3^2	V_4^2	X^2	V_3^2	V_4^2	X^2	V_3^2	V_4^2
Uniform (0, 1)	10	0.32	0.05	0.49	0.29	0.05	0.45	0.45	0.06	0.28
Logistic	13	0.06	0.06	0.11	0.06	0.06	0.09	0.10	0.07	0.15
Laplace	13	0.46	0.10	0.74	0.43	0.10	0.72	0.48	0.11	0.75
Student's t_3	13	0.28	0.09	0.62	0.26	0.09	0.60	0.31	0.11	0.64
Gamma (5)	5	0.51	0.60	0.07	0.52	0.61	0.07	0.52	0.61	0.06
Gamma (6)	5	0.39	0.51	0.05	0.40	0.51	0.05	0.45	0.53	0.05
Extreme Value	8	0.33	0.62	0.04	0.32	0.61	0.04	0.59	0.71	0.04

Table 3 compares the performance of MOM, ML and GRO estimators. Powers based on the ML and MOM estimators were very similar. However use of GRO estimators and χ^2 critical values often gave powers different to what the parametric bootstrap suggested they should be. For example, see the uniform, logistic and extreme value alternatives. Again we suggest that the GRO approach with χ^2 p-values is only an approximation to the better p-values available if ML or MOM estimators are used. In our experience ML and MOM estimators are very similar but if a data set occurred when V_1 and V_2 were not very close to zero with ML estimation then we suggest the use of MOM estimators. Klar (2000) gives reasons why MOM estimators should be used if tests involving higher order moments, such as tests based on V_3 and V_4, are of interest.

For the alternatives considered here, powers usually increase with K, but in practice we assume K is given and so we will not investigate this effect here.

4. An Improved Model and Additional Example

In this section we will emphasise the desirability of using the significant components of X^2 to construct an improved model. For the Pearson and Lee (1903) mothers' heights data, Table 1 gives expected counts, $E_j = n\hat{p}_j$, j = 1, … , 9, using the grouped normal model. The analysis in section 2 suggested that possibly the data differ from normality in regard to kurtosis, and the component V_4 is significantly large, using either MOM or ML estimation. We may therefore expect an appropriate kurtosis correction will significantly improve the model. Consider the kurtosis corrected model

$$p_j^* = C\,p_j\{1 + \theta_4 g_4(x_j)\},\ j = 1, \ldots, K.$$

where $p_j^* = 0$ if $p_j\{1 + \theta_4 g_4(x_j)\} < 0$ and where C is such that $\sum_j p_j^* = 1$. Barton (1955) suggested a similar model, although he did not suggest avoiding the negative frequencies.
It is routine to show that V_4 is the score test statistic for testing H: $\theta_4 = 0$ against K: $\theta_4 \neq 0$ for the model $\{p_j^*\}$, as it is for models of the form $\{C(\theta_4) \exp[\theta_4 g_4(x_j)]\, p_j\}$; see Rayner and Best (1989, p. 72). Moreover it is routine to show that for the model $\{p_j^*\}$, $E[V_4] = \theta_4\sqrt{n}$, so that V_4 properly normalized is a good 'indicator' of θ_4.

Table 4. Comparison of observed frequencies (N_j) of heights of mothers with E_j and E_j^*

Class (j)	Class limits	N_j	E_j	E_j^*	$(N_j - E_j)^2/E_j$	$(N_j - E_j^*)^2/E_j^*$
1	$(-\infty, 55]$	3	0.8	2.0	5.50	0.45
2	(55, 57]	8	10.1	12.0	0.42	1.36
3	(57, 59]	53	63.5	56.9	1.73	0.26
4	(59, 61]	215	204.3	196.8	0.56	1.68
5	(61, 63]	346	336.6	353.4	0.26	0.15
6	(63, 65]	277	284.3	287.4	0.19	0.38
7	(65, 67]	120	123.0	111.5	0.07	0.64
8	(67, 69]	24	27.2	27.2	0.38	0.38
9	$(69, \infty)$	7	3.2	5.7	4.34	0.29

Table 5. Bohemian income data

Class interval	$(-\infty, 1.53]$	(1.53, 2.15]	(2.15, 2.71]	(2.71, 3.32]	(3.32, 3.74]
Frequency	14	16	29	28	9
Class interval	(3.74, 4.18]	(4.18, 4.53]	(4.53, 4.70]	$(4.70, \infty)$	
Frequency	1	1	1	1	

With the $\{p_j^*\}$ model expected counts are given by

$$E_j^* = np_j^* = nCp_j\{1 + V_4 g_4(x_j)/\sqrt{n}\},$$

and Table 4 compares the observed frequencies, N_j, with E_j and E_j^*. In terms of the differences $|N_j - E_j^*|$, as seen, for example, on a histogram, it may not appear that the $\{E_j^*\}$ are an improvement on the $\{E_j\}$. However if we compare $X_{CN}^2 = \sum_j \left(N_j - E_j\right)^2 / E_j$ with $X_{ICN}^2 = \sum_j \left(N_j - E_j^*\right)^2 / E_j^*$, and the contributions to these metrics from each class, it is clear the kurtosis corrected $\{E_j^*\}$ give a better fit in the tails of the distribution. Using the improved model reduces the X^2 metric from $X_{CN}^2 = 13.45$ to $X_{ICN}^2 = 5.59$.

It is interesting to note that using the $\{p_j^*\}$ model, straightforward calculations show that

$$\sum_{j=1}^{K} \left(N_j - E_j^*\right)^2 / E_j = X_{CN}^2 - V_4^2 ,$$

from which it easily follows that $X_{ICN}^2 = X_{CN}^2 - V_4^2 + O(n^{-0.5})$. Roughly speaking, since these arguments generalise to corrections for other components, the reduction in X^2 from using a model that corrects the significant components is the sum of the squares of those components.

D'Agostino and Massaro (1992, p. 332) fit a logistic distribution to the grouped Bohemian income data shown in Table 5. We now test this grouped data for normality. It is readily found that $X^2 = 8.07$ with p-value 0.23 using MOM or ML estimation and the approximating χ_6^2 distribution. On this evidence alone we would conclude that the data are consistent with normality. However it is instructive to calculate the components of X^2. We find $\hat{V}_3^2 = 0.04$ and $\hat{V}_4^2 = 6.45$, the latter being highly significant and suggesting that non-normality is due to an excessive peak towards the centre of the data. This example demonstrates the value of looking at not only X^2 but its components as well. As with the mothers' heights data we could also obtain an improved model using V_4.

5. Conclusion

On the basis of the power study in section 3 it appears that X^2 provides a good omnibus test of normality with grouped data, while V_3^2 and V_4^2 are useful for suggesting whether or not the alternative is respectively symmetric or, relative to the normal, unusually peaked. From our simulations here we suggest that χ^2 approximations to the null distributions of the test statistics X^2, V_3^2 and V_4^2 will be reasonable for testing grouped normality if all class expectations are greater than 0.5 and method of moments or maximum likelihood estimation is used.

The suggestion to use X^2, V_3^2 and V_4^2 to test for grouped normality is hardly new, although previously skewness and kurtosis coefficients may not have been calculated as components of X^2. However with modern computing capabilities and bearing in mind the results above, we make four suggestions on how to improve on the classical approach.

1. Use method of moments or maximum likelihood estimation rather than the grouped frequency estimation of the normal mean and variance.
2. Do not pool the data so that p-values can be found using the χ^2 approximation to the distribution of X^2. Instead find p-values using the parametric bootstrap when the smallest expectation is less than 0.5.

3. If appropriate give an improved model based on V_3 and/or V_4, as illustrated in section 4 above.
4. Check that higher order moment differences do not exist by finding a p-value for $X^2 - V_3^2 - V_4^2$ using either a χ^2_{K-5} or a parametric bootstrap approximation.

References

Barton, D.E. (1955). A form of Neyman's Ψ^2 test of goodness of fit applicable to grouped and discrete data. *Skand. Aktuarietidskr*. *38*, 1-16.

Cochran, W.G. (1952). The χ^2 test of goodness of fit . *Ann. Math. Statist*., 23, 315-345.

D'Agostino, R.B. and Massaro, J.M. (1992). Goodness of fit tests. Chapter 13 in *Handbook of the Logistic Distribution*. Ed. N. Balakrishnan, New York: Dekker.

Emerson, P.L. (1968). Numerical construction of orthogonal polynomials from a general recurrence formula. *Biometrics*, 24, 695-701.

Fisher, R.A. (1924). The conditions under which χ^2 measures the discrepancy between observation and hypothesis. *Journal of the Royal Statistical Society*, 27, 442-450.

Freund, J.E. (2004). *Modern Elementary Statistics* (11th ed.). New Jersey: Pearson Education.

Gulati, S. and Neus, J. (2001). Goodness-of-fit statistics for the exponential distribution when the data are grouped. Chapter 9 in *Goodness-of-Fit Tests and Validity of Models*, eds. Huber-Carol, C., Balakrishnan, N., Nikulin, M. S. and Mesbah, M., Boston: Birkhauser, 113-123.

Gurtler, N. and Henze, N. (2000). Recent and classical goodness-of-fit tests for the Poisson distribution. *Journal of Statistical Planning and Inference*, 90, 207-225.

Hoel, P.G. (1984). *Introduction to Mathematical Statistics* (5th ed.). New York: Wiley.

IMSL (1995). *Fortran Numerical Libraries.* Houston: Visual Numerics.

Klar, B. (2000). Diagnostic smooth tests of fit. *Metrika*, 52, 237-252.

Lancaster, H.O. (1953). A reconciliation of χ^2 from metrical and enumerative aspects. *Sankhya*, 13, 1-10.

Pearson, K. and Lee, A. (1903). On the laws of inheritance in man. I Inheritance of physical characters. *Biometrika*, 2, 357-462.

Rayner, J.C.W. and Best, D.J. (1989). *Smooth Tests of Goodness of Fit*. New York: Oxford University Press.

Selvanathan, A., Selvanathan, S., Keller, G. and Warrack, B. (2000). *Australian Business Statistics*. Melbourne: Nelson.

Snedecor, G.W. and Cochran, W.G. (1989). *Statistical Methods* (8th ed.). Ames: Iowa State University Press.

Appendix A: Bivariate Newton-Raphson Estimation

(i) *Maximum Likelihood Estimators*

Suppose initial estimates (μ_0, σ_0) have been found as in the Introduction. Calculate the corresponding p_j, $j = 1, \ldots , K$ and based on these calculate the logarithm of the likelihood: $\ln L = constant + \sum_{j=1}^{k} N_j \ln p_j$. The constant is a multinomial coefficient that does not depend on (μ, σ). The maximum of ln L is found by simultaneously solving $\partial \ln L/\partial\mu = 0$ and $\partial \ln L/\partial\sigma = 0$. To solve these equations first find

$$r_{11} = \frac{\partial^2 \ln L}{\partial\mu^2},\ r_{12} = \frac{\partial^2 \ln L}{\partial\mu\partial\sigma},\ r_{21} = \frac{\partial^2 \ln L}{\partial\sigma\partial\mu} = r_{12} \text{ and } r_{22} = \frac{\partial^2 \ln L}{\partial\sigma^2},$$

and form the 2 × 2 matrix $R = \begin{pmatrix} r_{11} & r_{12} \\ r_{21} & r_{22} \end{pmatrix}$. Let f_0 be the vector $f = (\partial \ln L/\partial\mu,\ \partial \ln L/\partial\sigma)^T$ evaluated at $e = (\mu, \sigma)^T = (\mu_0, \sigma_0)^T = e_0$. If R_0 is just R evaluated at $e = e_0$ the Newton-Raphson method gives a new vector $e_1 = (\mu_1, \sigma_1)^T = e_0 - R_0^{-1} f_0$. In the same way calculate f_1 and R_1 using e_1 and then find $e_2 = e_1 - R_1^{-1} f_1$ and further e_i until a convergence criterion is satisfied, usually in five or fewer iterations. We now give the quantities f and R analytically. For $j = 1, \ldots, K$ write $y_j = (k_j - \mu)/\sigma$ and $z_j = \exp(-y_j^2/2)/\sqrt{(2\pi)}$. Then

$$\frac{\partial p_j}{\partial\mu} = -(z_j - z_{j-1})/\sigma,\ \frac{\partial p_j}{\partial\sigma} = -(y_j z_j - y_{j-1} z_{j-1})/\sigma,$$

$$\frac{\partial^2 p_j}{\partial\mu^2} = \frac{\partial p_j}{\partial\sigma}/\sigma,\ \frac{\partial^2 p_j}{\partial\mu\partial\sigma} = -\{(y_j^2 - 1)z_j - (y_{j-1}^2 - 1)z_{j-1}\}/\sigma^2 \text{ and}$$

$$\frac{\partial^2 p_j}{\partial\sigma^2} = \frac{-y_j^3 z_j + y_{j-1}^3 z_{j-1}}{\sigma^2} + 2\frac{\partial p_j}{\partial\sigma}/\sigma.$$

It follows that

$$\frac{\partial \ln L}{\partial\mu} = \sum_{j=1}^{K}\left(\frac{N_j}{p_j}\right)\left(\frac{\partial p_j}{\partial\mu}\right),\ \frac{\partial \ln L}{\partial\sigma} = \sum_{j=1}^{K}\left(\frac{N_j}{p_j}\right)\left(\frac{\partial p_j}{\partial\sigma}\right),$$

$$\frac{\partial^2 \ln L}{\partial\mu^2} = \sum_{j=1}^{K}\left(\frac{N_j}{p_j^2}\right)\left\{p_j \frac{\partial^2 p_j}{\partial\mu^2} - \left(\frac{\partial p_j}{\partial\mu}\right)^2\right\},$$

$$\frac{\partial^2 \ln L}{\partial\mu\partial\sigma} = \sum_{j=1}^{K}\left(\frac{N_j}{p_j^2}\right)\left\{p_j \frac{\partial^2 p_j}{\partial\mu\partial\sigma} - \left(\frac{\partial p_j}{\partial\mu}\right)\left(\frac{\partial p_j}{\partial\sigma}\right)\right\} \text{ and}$$

$$\frac{\partial^2 \ln L}{\partial\sigma^2} = \sum_{j=1}^{K}\left(\frac{N_j}{p_j^2}\right)\left\{p_j \frac{\partial^2 p_j}{\partial\sigma^2} - \left(\frac{\partial p_j}{\partial\sigma}\right)^2\right\}.$$

(ii) *Method of Moments Estimators*
To find these estimators we need to simultaneously solve the two non-linear equations $V_1 = 0$ and $V_2 = 0$, that is, solve

$$F = \sum_{j=1}^{k} x_j p_j - \mu = 0 \text{ and } G = \sum_{j=1}^{k} (x_j - \mu)^2 p_j - \sigma^2 = 0.$$

We proceed as above, but now with $f = (F, G)^T$ and

$$r_{11} = \frac{\partial F}{\partial\mu},\ r_{12} = \frac{\partial F}{\partial\sigma},\ r_{21} = \frac{\partial G}{\partial\mu} \text{ and } r_{22} = \frac{\partial G}{\partial\sigma}.$$

This time only first order partial derivatives are needed.

Appendix B: Explicit formula for $g_4(x_j)$

Define, as previously,

$$\mu = \sum_{j=1}^{K} x_j p_j \text{ and } \mu_r = \sum_{j=1}^{K} (x_j - \mu)^r p_j \text{ for } r = 2, 3, 4, \ldots,$$

and then put

$$c_1 = \mu_2\mu_4 - \mu_3^2 - \mu_2^3,\ c_2 = \mu_3\mu_4 - \mu_2^2\mu_3 - \mu_2\mu_5,$$
$$c_3 = \mu_3\mu_5 - \mu_2^2\mu_4 - \mu_4^2 - \mu_2\mu_3^2, \text{ and } c_4 = \mu_2^2\mu_5 - 2\mu_2\mu_3\mu_4 + \mu_3^3.$$

Now put $e = c_1(c_1\mu_6 + c_2\mu_5 + c_3\mu_4 + c_4\mu_5)$ and define a, b, c and d by

$$-ae = c_1^2\mu_7 + c_1c_2\mu_6 + c_1c_3\mu_5 + c_1c_4\mu_4,$$
$$-be = c_1c_2\mu_7 + c_2^2\mu_6 + c_2c_3\mu_5 + c_1c_4\mu_4 + \left(\mu_2\mu_6 - \mu_3\mu_5 - \mu_2^2\mu_4\right)e/c_1,$$
$$-ce = c_1c_3\mu_7 + c_2c_3\mu_6 + c_3^2\mu_5 + c_3c_4\mu_4 + \left(\mu_2\mu_3\mu_4 - \mu_3\mu_6 - \mu_2^2\mu_5 + \mu_4\mu_5\right)e/c_1,$$
$$-de = c_1c_4\mu_7 + c_2c_4\mu_6 + c_3c_4\mu_5 + c_4^2\mu_4 + \left(\mu_2\mu_4^2 - \mu_3^2\mu_4 + \mu_2\mu_3\mu_5 - \mu_2^2\mu_6\right)e/c_1.$$

It can be shown that $g_4(x_j)$ is

$$\frac{(x_j - \mu)^4 + a(x_j - \mu)^3 + b(x_j - \mu)^2 + c(x_j - \mu) + d}{\sqrt{\mu_8 + 2a\mu_7 + (a^2 + 2b)\mu_6 + 2(ab + c)\mu_5 + (b^2 + 2ac + 2d)\mu_4 + 2(ad + bc)\mu_3 + (c^2 + 2bd)\mu_2 + d^2}}.$$