November 4, 1999

Formulas for Skewness and Kurtosis

Karl G. Jöreskog[1]

PRELIS output gives a table called *Univariate Summary Statistics for Continuous Variables.* In addition to the usual statistics, means and standard deviations, this table gives measures of skewness and kurtosis of each variable. Another table gives test statistics for testing the hypotheses of zero skewness and zero kurtosis. The purpose of this note is to present the formulas we use for skewness and kurtosis. I do not explain the meaning of skewness and kurtosis and their uses in other contexts. DeCarlo (1997) discusses the meaning of kurtosis and its use in testing normality, and in issues of robustness and outliers.

To present the formulas used to calculate skewness and kurtosis, I must first define some population quantities. Let X be a continuous random variable with moments existing up through order four. Let $\mu = E(X)$ be the mean of X and denote

$$\mu_i = E(X - \mu)^i \, , \qquad i = 2, 3, 4 \, . \tag{1}$$

These are the population central moments of order 2, 3 , and 4. μ is a location parameter; it tells where the distribution is located. μ_2 is the variance; it tells something about the variation in the distribution. $\sqrt{\mu_2}$, the standard deviation, is a scale parameter; it can be used to define a unit of measurement for X.

Skewness and kurtosis (sometimes called excess) are defined as follows.

$$\gamma_1 = \frac{\mu_3}{{\mu_2}^{3/2}} = \frac{\mu_3}{\mu_2\sqrt{\mu_2}} \tag{2}$$

$$\gamma_2 = \frac{\mu_4}{\mu_2^2} - 3 \tag{3}$$

These are parameters describing the shape of the distribution independent of location and scale. Here I follow the notation of Cramer (1957, eqs. 15.8.1 and 15.8.2). In the literature, sometimes the notation $\beta_1 = \gamma_1^2$ and $\beta_2 = \gamma_2 + 3$ is used instead of γ_1 and γ_2, see, *e.g.*, Kendall & Stuart (1952, p. 85). A normal distribution has $\gamma_1 = 0$ and $\gamma_2 = 0$. All symmetric distributions have $\gamma_1 = 0$, but they vary in terms of γ_2 which can be positive or negative. Non-symmetric distributions have positive or negative γ_1.

γ_1 and γ_2 are population parameters. To decide whether the distribution is normal or non-normal, one can estimate γ_1 and γ_2 from the sample and decide whether the estimates differ significantly from zero. If they do, this is an indication of non-normality. There are several ways of doing this. I describe the one that is used in PRELIS.

Let $x_1, x_2, \ldots, x_n$ be a random sample of size n from the distribution of X. One can then use the sample quantities

$$m_i = (1/n)\sum_{a=1}^{n}(x_a - \bar{x})^i \, , \qquad i = 2, 3, 4 \, . \tag{4}$$

to estimate μ_i, where

$$\bar{x} = (1/n)\sum_{a=1}^{n} x_a \, , \tag{5}$$

[1] I thank Larry DeCarlo for leading me to the Fisher's statistics and for pointing out an error in the z-score for skewness and Ken Bollen for allowing me to use his data

is the sample mean.

The estimates m_i are consistent estimates of μ_i but they are not unbiased. This means that in a large sample they are close to the population parameter but in a small sample they differ on average from the population parameter. It is possible to construct unbiased estimates of μ_i. In fact, Cramer (1957, p. 352) gives the following unbiased estimates:

$$u_2 = \frac{n}{n-1} m_2 = c_2 m_2 \tag{6}$$

$$u_3 = \frac{n^2}{(n-1)(n-2)} m_3 = c_3 m_3 \tag{7}$$

$$u_4 = \frac{n(n^2-2n+3)}{(n-1)(n-2)(n-3)} m_4 - \frac{3n(2n-3)}{(n-1)(n-2)(n-3)} m_2^2 = c_4 m_4 - c_5 m_2^2 \tag{8}$$

The second term in (8) is of order n^{-1} and can be ignored in most cases.

I define sample skewness and kurtosis as

$$g_1 = \frac{m_3}{{m_2}^{3/2}} = \frac{m_3}{m_2\sqrt{m_2}} = \sqrt{b_1} \tag{9}$$

$$g_2 = \frac{m_4}{m_2^2} - 3 = b_2 - 3 \tag{10}$$

Here I follow Cramer (1957, p. 356) and use the notation g_1 and g_2 for skewness and kurtosis. In analogy with the use of β's instead of γ's, more recent literature commonly uses the notation[2] $\sqrt{b_1}$ for g_1 and b_2 for $g_2 + 3$, see, *e.g.*, D'Agostino (1970, 1971, 1982, 1986), D'Agostino, *et al.* (1990), Mardia (1980), Bollen (1989, eqs. 9.74 and 9.75), and DeCarlo (1997). The difference in notation seems to be one between the "old" and the "young" generation of statisticians. For those who are used to the notation $\sqrt{b_1}$ and b_2, I will write the formulas that follow by using the "old" notation on the left, the definition in the middle and the "new" notation on the right.

Using unbiased moments u_i instead of m_i in these formulas gives two other estimates $\hat{\gamma}_1$ and $\hat{\gamma}_2$ of γ_1 and γ_2, where

$$\hat{\gamma}_1 = \frac{u_3}{{u_2}^{3/2}} = \frac{u_3}{u_2\sqrt{u_2}} \tag{11}$$

$$\hat{\gamma}_2 = \frac{u_4}{u_2^2} - 3 \tag{12}$$

For a general distribution, $\hat{\gamma}_1$ and $\hat{\gamma}_2$ are not unbiased estimates of γ_1 and γ_2, but by using unbiased moments instead of biased moments, one should get estimates with a smaller bias.

Using the constants c_2, c_3, c_4, and c_5 in (6) – (8), the relationships between $\hat{\gamma}_1$, $\hat{\gamma}_2$ and g_1, g_2 are:

$$\hat{\gamma}_1 = (c_3/c_2^{3/2}) g_1 = (c_3/c_2^{3/2})\sqrt{b_1} \tag{13}$$

$$\hat{\gamma}_2 = (c_4/c_2^2)(g_2+3) - (c_5/c_2^2) - 3 = (c_4/c_2^2) b_2 - (c_5/c_2^2) - 3 \tag{14}$$

If we ignore the second term in (8) we get still another estimate $\hat{\hat{\gamma}}_2$ of γ_2:

$$\hat{\hat{\gamma}}_2 = (c_4/c_2^2)(g_2+3) - 3 = (c_4/c_2^2) b_2 - 3 \tag{15}$$

To test if skewness and kurtosis are zero in the population, one would like to know the mean and variance of these estimates and transform them to a z-statistic which can

[2] $\sqrt{b_1}$ is to be interpreted as a single entity. This notation is awkward. Mathematically $\sqrt{b_1}$ means the positive square root of b_1, with $b_1 = (m_3^2/m_2^3)$ positive. According to this definition, skewness cannot be negative. But skewness can be positive or negative as indicated by the sign of m_3. So the proper definition of skewness should $\sqrt{b_1} \times sign(m_3)$.

be used as a test statistic. In the general case, the exact mean and variance of g_1 and g_2 (or of $\hat{\gamma}_1$ and $\hat{\gamma}_2$) are not available. For a general distribution, Cramer (1957, eq. 27.7.8) gives expressions for the mean and variance accurate to the order of n^{-1}. However, to test normality, the normal distribution is the assumed distribution under the null hypothesis, and Cramer (1957, eq. 29.3.7) gives expressions for the mean and variance that are exact under normality. These expressions show that g_1 and $\hat{\gamma}_1$ are unbiased and that g_2 and $\hat{\gamma}_2$ both have bias $-6/(n+1)$. One can also show that $\hat{\hat{\gamma}}_2$ has a positive bias which is smaller in magnitude than that of $\hat{\gamma}_2$.

Cramer (1957, eq. 29.3.8) gives the following two alternative estimates due to Fisher (1930), see also Fisher (1973, p. 75)[3], both of which are unbiased under normality.

$$G_1 = \frac{\sqrt{n(n-1)}}{n-2} g_1 = \frac{\sqrt{n(n-1)}}{n-2} \sqrt{b_1} \tag{16}$$

$$G_2 = \frac{n-1}{(n-2)(n-3)}[(n+1)g_2 + 6] = \frac{n-1}{(n-2)(n-3)}[(n+1)b_2 - 3(n-1)] \tag{17}$$

G_1 is identical to $\hat{\gamma}_1$ but G_2 is slightly different from $\hat{\gamma}_2$. The exact variances of G_1 and G_2 under normality are given by Cramer (1957, eq. 29.3.9). These could be used to define z-statistics for testing zero skewness and kurtosis. However, for very small n the normality approximation is not sufficiently good.

In the general case, it is difficult to say which of the sets of estimates is best. In large samples differences among them are ignorable. In small samples G_1 and G_2 have the advantage of being unbiased if the distribution is normal. But as is often the case, a reduction in bias is accompanied by an increased sampling variance. One would have to do a simulation study to settle this.

Bollen (1989) provided some data that can be used to illustrate these different estimates. The variables are GNP per capita (x_1) and energy consumption per capita (x_2) in 75 countries, as well as their logarithms. Table 1 gives estimates of skewness and Table 2 gives estimates of kurtosis.

Table 1: *Different Estimates of Skewness*

Variable	g_1	$\hat{\gamma}_1$	G_1
x_1	1.761	1.797	1.797
$\ln x_1$	0.259	0.264	0.264
x_2	3.087	3.150	3.150
$\ln x_1$	–0.353	–0.390	–0.360

The values of g_1 ($\sqrt{b_1}$) and g_2 ($b_2 - 3$) agree with those given by Bollen (1989, p. 420). The values of $\hat{\gamma}_1$ were computed by (13) and those of G_1 were computed by (16). This confirms that these estimates are identical.

By making a logarithmic transformation of skewness D'Agostino (1986) and D'Agostino, *et al.* (1990) developed another z-statistic that can be used with n as small as 8. Another kind of transformation is used for kurtosis which works well for $n \geq 20$. These z-statistics are based directly on g_1 and g_2. The formulas to perform these tests are quite complicated.

[3]Fisher defines skewness and kurtosis in terms of cumulants as $g_1 = k_3/k_2^{3/2}$ and $g_2 = k_4/k_2^2$ but this is equivalent to our G_1 and G_2.

Table 2: *Different Estimates of Kurtosis*

Variable	g_2	$\hat{\gamma}_2$	$\hat{\hat{\gamma}}_2$	G_2
x_1	3.138	3.312	3.229	3.442
$\ln x_1$	–0.693	–0.627	–0.710	–0.657
x_2	11.638	12.053	11.970	12.537
$\ln x_1$	–0.504	–0.434	–0.517	–0.455

They are summarized in Bollen (1989, Table 9.2).[4] In PRELIS we follow these formulas exactly to compute the z-scores for skewness and kurtosis. There is also an omnibus test to test for skewness and kurtosis simultaneously. This is simply the sum of squares of the z-scores for skewness and kurtosis. Under normality this has a chi-square distribution with 2 degrees of freedom.

In PRELIS we compute skewness by $\hat{\gamma}_1 = G_1$. Prior to November 1999 we used $\hat{\hat{\gamma}}_2$ to compute kurtosis. In November 1999 (Patch 6)[5] we changed that to G_2. If $n < 100$ these may give slightly different results. G_1 and G_2 are used in SAS and SPSS to compute skewness and kurtosis.

Unfortunately, there was an error in PRELIS in the z-score for testing zero skewness. This error had the effect that the standard z-score (a_1 in Table 9.2 in Bollen, 1989):

$$a_1 = \sqrt{\frac{(n+1)(n+3)}{6(n-2)}} g_1 = \sqrt{\frac{(n+1)(n+3)}{6(n-2)}} \sqrt{b_1} \tag{18}$$

was used even for $n < 150$. Since this z-score is fairly accurate even for sample sizes less than 150, this error has not done much harm in terms of practical conclusions. In November 1999 (Patch 6) this error has been corrected and the z-score for skewness is now correct. The z-score for kurtosis has always been correct in PRELIS.

Using Bollen's data, I can illustrate the results given in PRELIS. I first give the results obtained with the previous version (prior to November 1999):

```
Univariate Summary Statistics for Continuous Variables

Variable     Mean  St. Dev.   T-Value  Skewness  Kurtosis
--------     ----  --------   -------  --------  --------
      X1  205.787   167.655    10.630     1.797     3.229
    lnX1    5.054     0.733    59.724     0.264    -0.710
      X2  297.360   435.072     5.919     3.150    11.970
    lnX2    4.792     1.511    27.472    -0.360    -0.517

Test of Univariate Normality for Continuous Variables

             Skewness          Kurtosis      Skewness and Kurtosis

Variable Z-Score P-Value   Z-Score P-Value   Chi-Square P-Value
```

[4]The formulas are correct in the first printing of the book but in subsequent printings there is an error in the formula for a_4; the numerator should be 2 not 1.

[5]Available at SSI's website for current users of LISREL 8.30 for Windows: http://www.ssicentral.com

X1	6.437	0.000	3.283	0.001	52.217	0.000
lnX1	0.947	0.344	-1.537	0.124	3.261	0.196
X2	11.282	0.000	5.289	0.000	155.263	0.000
lnX2	-1.290	0.197	-0.880	0.379	2.437	0.296

Next, I give the results obtained with the current version (November 1999, Patch 6):

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	T-Value	Skewness	Kurtosis
X1	205.787	167.655	10.630	1.797	3.442
lnX1	5.054	0.733	59.724	0.264	-0.657
X2	297.360	435.072	5.919	3.150	12.537
lnX2	4.792	1.511	27.472	-0.360	-0.456

Test of Univariate Normality for Continuous Variables

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
X1	4.978	0.000	3.283	0.001	35.554	0.000
lnX1	0.979	0.327	-1.537	0.124	3.323	0.190
X2	6.838	0.000	5.289	0.000	74.738	0.000
lnX2	-1.320	0.187	-0.880	0.379	2.517	0.284

The second results differ from the first for two reasons:

- The value of kurtosis differs because we now use G_2 instead of $\hat{\hat{\gamma}}_2$.
- The z-score for skewness differs because the previous version used the standard z-score (17) whereas the current version uses the final z-score in Bollen's Table 9.2.

The skewness and kurtosis estimates in the last results agree exactly with those obtained with SAS and SPSS. The z-scores for skewness and kurtosis agree exactly with the result reported by Bollen (1989, p. 422).[6] They can also be obtained by the SPSS macro of DeCarlo (1997, pp. 304–307).

Another data set that can be used to test the procedures in PRELIS is the data from the Framingham heart study. It consists of a sample of cholesterol values from 62 subjects. The individual values can be read off from the stem-and-leaf plot in Table 1 of D'Agostino, *et al.* (1990). PRELIS gives the following results.

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	T-Value	Skewness	Kurtosis
X	250.032	41.443	47.505	1.049	1.816

Test of Univariate Normality for Continuous Variables

[6] The z-values for skewness are incorrect in the first printing of the book but they are correct in subsequent printings.

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
X	3.139	0.002	2.213	0.027	14.752	0.001

D'Agostino, *et al.* (1990) report the value $\sqrt{b_1} = 1.02$ (our g_1) and $b_2 = 4.58$ (equivalent to our $g_2 = 1.58$). According to (15) and (16) these values are consistent with the values $G_1 = 1.049$ and $G_2 = 1.816$ that PRELIS gives. Furthermore, all values given by PRELIS for testing skewness and kurtosis agree exactly with those given by D'Agostino, *et al.* (1990).

Another data set that can be used to illustrate the testing procedure is the nine psychological tests provided with the distribution of LISREL 8.30, see Jöreskog, *et al.* (1999, p. 148). The data file is NPV.RAW and consists of 145 cases on nine variables. Running the command file NPV1.PR2 gives the following results:

```
Total Sample Size =    145
```

Univariate Summary Statistics for Continuous Variables

Variable	Mean	St. Dev.	T-Value	Skewness	Kurtosis
VIS PERC	29.579	6.914	51.517	-0.119	-0.046
CUBES	24.800	4.445	67.183	0.239	0.872
LOZENGES	15.966	8.317	23.115	0.623	-0.454
PAR COMP	9.952	3.375	35.502	0.405	0.252
SEN COMP	18.848	4.649	48.817	-0.550	0.221
WORDMEAN	17.283	7.947	26.186	0.729	0.233
ADDITION	90.179	23.782	45.660	0.163	-0.356
COUNTDOT	109.766	20.995	62.955	0.698	2.283
S-C CAPS	191.779	37.035	62.355	0.200	0.515

Test of Univariate Normality for Continuous Variables

	Skewness		Kurtosis		Skewness and Kurtosis	
Variable	Z-Score	P-Value	Z-Score	P-Value	Chi-Square	P-Value
VIS PERC	-0.604	0.546	0.045	0.964	0.367	0.833
CUBES	1.202	0.229	1.843	0.065	4.842	0.089
LOZENGES	2.958	0.003	-1.320	0.187	10.491	0.005
PAR COMP	1.995	0.046	0.761	0.447	4.559	0.102
SEN COMP	-2.646	0.008	0.693	0.489	7.483	0.024
WORDMEAN	3.385	0.001	0.720	0.472	11.977	0.003
ADDITION	0.826	0.409	-0.937	0.349	1.560	0.458
COUNTDOT	3.263	0.001	3.325	0.001	21.699	0.000
S-C CAPS	1.008	0.313	1.273	0.203	2.638	0.267

These results differ slightly from those reported on p. 163 in Jöreskog, *et al.* (1999) for reasons already stated.

Since skewness and kurtosis can be positive or negative, the P-value given for each z-score corresponds to a double-sided test. The P-value for chi-square is a one-sided (upper) test. To test each hypothesis on the 5% level, say, all one needs to to is to examine the P-value and reject the hypothesis if it is smaller than 0.05.

Which of the variables are skewed? Which variables have a high or low kurtosis? Which variables are non-normal? I leave it to the reader, as an exercise, to answer these questions.

Here I have not covered measures of multivariate skewness and kurtosis and the testing of multivariate normality. For these, PRELIS uses the formulas (9.78) – (9.80) and those in Table 9.3 of Bollen (1989) and we have established that PRELIS gives the same results as reported by Bollen (1989, p. 423)[7] on the same data (eight indicators of political democracy).

In this note I have only considered the problem of detecting non-normality. I have not considered the issue of treatment of non-normality. I shall consider this on another occasion in this corner. As the doctor would say: first comes diagnosis, then comes treatment.

References

Bollen, K.A. (1989) *Structural equations with latent variables.* New York: Wiley.

Cramer, H. (1957) *Mathematical methods of statistics.* Princeton University Press, Seventh Printing.

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

D'Agostino, R.B. (1971) An omnibus test of normality for moderate and large sample size. *Biometrika*, 58, 341–348.

D'Agostino, R.B. (1986) Tests for the normal distribution. Pp. 367–419 in R.B. D'Agostino & M.A. Stephens (Eds.) *Goodness-of-fit techniques.* New York: Marcel Dekker.

D'Agostino, R.B., Belanger, A., & D'Agostino, R.B., Jr (1990) A suggestion for using powerful and informative tests of normality. *The American Statistician*, 44, 316–321.

DeCarlo, L. (1997) On the meaning and use of kurtosis. *Psychological Methods*, 2, 292–397.

Fisher, R.A. (1930) The moments of the distribution for normal samples of measures of departures from normality. *Proceedings of the Royal Society, London, Series A*, 130, p. 16.

Fisher, R.A. (1973) *Statistical methods for research workers.* New York: Hafner Publishing, Fourteenth Printing.

Jöreskog, K., Sörbom, D., du Toit, S., & du Toit, M. (1999) *LISREL 8: New Statistical Features.* Chicago: Scientific Software International.

Kendall, M.G., & Stuart, A. (1952) *The advanced theory of statistics, Volume 1: Distribution theory.* London: Griffin & Co. Fifth Edition.

Mardia, K.V. (1980) Tests of univariate and multivariate normality. In P.R. Krishnaiah (Ed.), *Handbook of Statistics* (Vol. 1, Pp. 279–320. Amsterdam: North Holland.

[7] In the first printing of the book the z-score for skewness, $W(b_{1,p})$, is incorrect. The correct value, as given in subsequent printings, should be 3.99.