Speaking Stata: The limits of sample skewness and kurtosis

Nicholas J. Cox
Department of Geography
Durham University
Durham, UK
n.j.cox@durham.ac.uk

Abstract. Sample skewness and kurtosis are limited by functions of sample size. The limits, or approximations to them, have repeatedly been rediscovered over the last several decades, but nevertheless seem to remain only poorly known. The limits impart bias to estimation and, in extreme cases, imply that no sample could bear exact witness to its parent distribution. The main results are explained in a tutorial review, and it is shown how Stata and Mata may be used to confirm and explore their consequences.

Keywords: st0204, descriptive statistics, distribution shape, moments, sample size, skewness, kurtosis, lognormal distribution

1   Introduction

The use of moment-based measures for summarizing univariate distributions is long established. Although there are yet longer roots, Thorvald Nicolai Thiele (1889) used mean, standard deviation, variance, skewness, and kurtosis in recognizably modern form. Appreciation of his work on moments remains limited, for all too understandable reasons. Thiele wrote mostly in Danish, he did not much repeat himself, and he tended to assume that his readers were just about as smart as he was. None of these habits could possibly ensure rapid worldwide dissemination of his ideas. Indeed, it was not until the 1980s that much of Thiele's work was reviewed in or translated into English (Hald 1981; Lauritzen 2002).

Thiele did not use all the now-standard terminology. The names *standard deviation*, *skewness*, and *kurtosis* we owe to Karl Pearson, and the name *variance* we owe to Ronald Aylmer Fisher (David 2001). Much of the impact of moments can be traced to these two statisticians. Pearson was a vigorous proponent of using moments in distribution curve fitting. His own system of probability distributions pivots on varying skewness, measured relative to the mode. Fisher's advocacy of maximum likelihood as a superior estimation method was combined with his exposition of variance as central to statistical thinking. The many editions of Fisher's 1925 text *Statistical Methods for Research Workers*, and of texts that in turn drew upon its approach, have introduced several generations to the ideas of skewness and kurtosis. Much more detail on this history is given by Walker (1929), Hald (1998, 2007), and Fiori and Zenga (2009).

st0204

Whatever the history and the terminology, a simple but fundamental point deserves emphasis. A name like skewness has a very broad interpretation as a vague concept of distribution symmetry or asymmetry, which can be made precise in a variety of ways (compare with Mosteller and Tukey [1977]). Kurtosis is even more enigmatic: some authors write of kurtosis as *peakedness* and some write of it as *tail weight*, but the skeptical interpretation that kurtosis is whatever kurtosis measures is the only totally safe story. Numerical examples given by Irving Kaplansky (1945) alone suffice to show that kurtosis bears neither interpretation unequivocally.[1]

To the present, moments have been much disapproved, and even disproved, by mathematical statisticians who show that in principle moments may not even exist, and by data analysts who know that in practice moments may not be robust. Nevertheless in many quarters, they survive, and they even thrive. One of several lively fields making much use of skewness and kurtosis measures is the analysis of financial time series (for example, Taylor [2005]).

In this column, I will publicize one limitation of certain moment-based measures, in a double sense. Sample skewness and sample kurtosis are necessarily bounded by functions of sample size, imparting bias to the extent that small samples from skewed distributions may even deny their own parentage. This limitation has been well established and discussed in several papers and a few texts, but it still appears less widely known than it should be. Presumably, it presents a complication too far for most textbook accounts. The presentation here will include only minor novelties but will bring the key details together in a coherent story and give examples of the use of Stata and Mata to confirm and explore for oneself the consequences of a statistical artifact.

2 Deductions

2.1 Limits on skewness and kurtosis

Given a sample of n values $y_1, \ldots, y_n$ and sample mean $\overline{y} = \sum_{i=1}^{n} y_i / n$, sample moments measured about the mean are at their simplest defined as averages of powered deviations

$$m_r = \frac{\sum_{i=1}^{n} (y - \overline{y})^r}{n}$$

so that m_2 and $s = \sqrt{m_2}$ are versions of, respectively, the sample variance and sample standard deviation.

Here sample skewness is defined as

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

---

1. Kaplansky's paper is one of a few that he wrote in the mid-1940s on probability and statistics. He is much better known as a distinguished algebraist (Bass and Lam 2007; Kadison 2008).

while sample kurtosis is defined as

$$\frac{m_4}{m_2^2} = \frac{m_4}{s^4} = b_2 = g_2 + 3$$

Hence, both of the last two measures are scaled or dimensionless: Whatever units of measurement were used appear raised to the same powers in both numerator and denominator, and so cancel out. The commonly used m, s, b, and g notation corresponds to a longstanding μ, σ, β, and γ notation for the corresponding theoretical or population quantities. If 3 appears to be an arbitrary constant in the last equation, one explanation starts with the fact that normal or Gaussian distributions have $\beta_1 = 0$ and $\beta_2 = 3$; hence, $\gamma_2 = 0$.

Naturally, if y is constant, then m_2 is zero; thus skewness and kurtosis are not defined. This includes the case of $n = 1$. The stipulations that y is genuinely variable and that $n \geq 2$ underlie what follows.

Newcomers to this territory are warned that usages in the statistical literature vary considerably, even among entirely competent authors. This variation means that different formulas may be found for the same terms—skewness and kurtosis—and different terms for the same formulas. To start at the beginning: Although Karl Pearson introduced the term *skewness*, and also made much use of β_1, he used skewness to refer to (mean − mode) / standard deviation, a quantity that is well defined in his system of distributions. In more recent literature, some differences reflect the use of divisors other than n, usually with the intention of reducing bias, and so resembling in spirit the common use of $n - 1$ as an alternative divisor for sample variance. Some authors call γ_2 (or g_2) the kurtosis, while yet other variations may be found.

The key results for this column were extensively discussed by Wilkins (1944) and Dalén (1987). Clearly, g_1 may be positive, zero, or negative, reflecting the sign of m_3. Wilkins (1944) showed that there is an upper limit to its absolute value,

$$|g_1| \leq \frac{n-2}{\sqrt{n-1}} \tag{1}$$

as was also independently shown by Kirby (1974). In contrast, b_2 must be positive and indeed (as may be shown, for example, using the Cauchy–Schwarz inequality) must be at least 1. More pointedly, Dalén (1987) showed that there is also an upper limit to its value:

$$b_2 \leq \frac{n^2 - 3n + 3}{n - 1} \tag{2}$$

The proofs of these inequalities are a little too long, and not quite interesting enough, to reproduce here.

Both of these inequalities are sharp, meaning attainable. Test cases to explore the precise limits have all values equal to some constant, except for one value that is equal to another constant: $n = 2$, $y_1 = 0$, $y_2 = 1$ will do fine as a concrete example, for which skewness is $0/1 = 0$ and kurtosis is $(1 - 3 + 3)/1 = 1$.

For $n = 2$, we can rise above a mere example to show quickly that these results are indeed general. The mean of two distinct values is halfway between them so that the two deviations $y_i - \overline{y}$ have equal magnitude and opposite sign. Thus their cubes have sum 0, and m_3 and b_1 are both identically equal to 0. Alternatively, such values are geometrically two points on the real line, a configuration that is evidently symmetric around the mean in the middle, so skewness can be seen to be zero without any calculations. The squared deviations have an average equal to $\{(y_1 - y_2)/2\}^2$, and their fourth powers have an average equal to $\{(y_1 - y_2)/2\}^4$, so g_2 is identically equal to 1.

To see how the upper limit behaves numerically, we can rewrite (1) as

$$|g_1| \leq \sqrt{n-1} - \frac{1}{\sqrt{n-1}}$$

so that as sample size n increases, first $\sqrt{n-1}$ and then $\sqrt{n}$ become acceptable approximations. Similarly, we can rewrite (2) as

$$b_2 \leq n - 2 + \frac{1}{n-1}$$

from which, in large samples, first $n-2$ and then n become acceptable approximations.

As it happens, these limits established by Wilkins and Dalén sharpen up on the results of other workers. Limits of $\sqrt{n}$ and n (the latter when n is greater than 3) were established by Cramér (1946, 357). Limits of $\sqrt{n-1}$ and n were independently established by Johnson and Lowe (1979); Kirby (1981) advertised work earlier than theirs (although not earlier than that of Wilkins or Cramér). Similarly, Stuart and Ord (1994, 121–122) refer to the work of Johnson and Lowe (1979), but overlook the sharper limits.[2]

There is yet another twist in the tale. Pearson (1916, 440) refers to the limit (2), which he attributes to George Neville Watson, himself later a distinguished contributor to analysis (but not to be confused with the statistician Geoffrey Stuart Watson), and to a limit of $n-1$ on b_1, equivalent to a limit of $\sqrt{n-1}$ on g_1. Although Pearson was the author of the first word on this subject, his contribution appears to have been uniformly overlooked by later authors. However, he dismissed these limits as without practical importance, which may have led others to downplay the whole issue.

In practice, we are, at least at first sight, less likely to care much about these limits for large samples. It is the field of small samples in which limits are more likely to cause problems, and sometimes without data analysts even noticing.

---

2. The treatise of Stuart and Ord is in line of succession, with one offset, from Yule (1911). Despite that distinguished ancestry, it contains some surprising errors as well as the compendious collection of results that makes it so useful. To the statement that mean, median, and mode differ in a skewed distribution (p. 48), counterexamples are 0, 0, 1, 1, 1, 1, 3, and the binomial $\binom{10}{k}0.1^k 0.9^{10-k}, k = 0, \ldots, 10$. For both of these skewed counterexamples, mean, median, and mode coincide at 1. To the statement that they coincide in a symmetric distribution (p. 108), counterexamples are any symmetric distribution with an even number of modes.

2.2   An aside on coefficient of variation

The literature contains similar limits related to sample size on other sample statistics. For example, the coefficient of variation is the ratio of standard deviation to mean, or $s/\overline{y}$. Katsnelson and Kotz (1957) proved that so long as all $y_i \geq 0$, then the coefficient of variation cannot exceed $\sqrt{n-1}$, a result mentioned earlier by Longley (1952). Cramér (1946, 357) proved a less sharp result, and Kirby (1974) proved a less general result.

3   Confirmations

[R] **summarize** confirms that skewness b_1 and kurtosis g_2 are calculated in Stata precisely as above. There are no corresponding Mata functions at the time of this writing, but readers interested in these questions will want to start Mata to check their own understanding. One example to check is

```
. sysuse auto, clear
(1978 Automobile Data)

. summarize mpg, detail

                        Mileage (mpg)

      Percentiles      Smallest
 1%          12              12
 5%          14              12
10%          14              14       Obs                 74
25%          18              14       Sum of Wgt.         74

50%          20                       Mean           21.2973
                         Largest      Std. Dev.     5.785503
75%          25              34
90%          29              35       Variance      33.47205
95%          34              35       Skewness      .9487176
99%          41              41       Kurtosis      3.975005
```

The `detail` option is needed to get skewness and kurtosis results from `summarize`.

We will not try to write a bulletproof skewness or kurtosis function in Mata, but we will illustrate its use calculator-style. After entering Mata, a variable can be read into a vector. It is helpful to have a vector of deviations from the mean to work on.

```
. mata :
                                            ─── mata (type end to exit) ───
: y = st_data(., "mpg")

: dev = y :- mean(y)

: mean(dev:^3) / (mean(dev:^2)):^(3/2)
  .9487175965

: mean(dev:^4) / (mean(dev:^2)):^2
  3.975004596
```

So those examples at least check out. Those unfamiliar with Mata might note that the colon prefix, as in `:-` or `:^`, merely flags an elementwise operation. Thus for example, `mean(y)` returns a constant, which we wish to subtract from every element of a data vector.

Mata may be used to check simple limiting cases. The minimal dataset $(0, 1)$ may be entered in deviation form. After doing so, we can just repeat earlier lines to calculate b_1 and g_2:

```
: dev = (.5 \ -.5)
: mean(dev:^3) / (mean(dev:^2)):^(3/2)
  0
: mean(dev:^4) / (mean(dev:^2)):^2
  1
```

Mata may also be used to see how the limits of skewness and kurtosis vary with sample size. We start out with a vector containing some sample sizes. We then calculate the corresponding upper limits for skewness and kurtosis and tabulate the results. The results are mapped to strings for tabulation with reasonable numbers of decimal places.

```
: n = (2::20\50\100\500\1000)
: skew = sqrt(n:-1) :- (1:/(n:-1))
: kurt = n :- 2 + (1:/(n:-1))
: strofreal(n), strofreal((skew, kurt), "%4.3f")
```

	1	2	3
1	2	0.000	1.000
2	3	0.914	1.500
3	4	1.399	2.333
4	5	1.750	3.250
5	6	2.036	4.200
6	7	2.283	5.167
7	8	2.503	6.143
8	9	2.703	7.125
9	10	2.889	8.111
10	11	3.062	9.100
11	12	3.226	10.091
12	13	3.381	11.083
13	14	3.529	12.077
14	15	3.670	13.071
15	16	3.806	14.067
16	17	3.938	15.062
17	18	4.064	16.059
18	19	4.187	17.056
19	20	4.306	18.053
20	50	6.980	48.020
21	100	9.940	98.010
22	500	22.336	498.002
23	1000	31.606	998.001

The second and smaller term in each expression for (1) and (2) is $1/(n-1)$. Although the calculation is, or should be, almost mental arithmetic, we can see how quickly this term shrinks so much that it can be neglected:

```
: strofreal(n), strofreal(1 :/ (n :- 1), "%4.3f")
             1       2

    1        2    1.000
    2        3    0.500
    3        4    0.333
    4        5    0.250
    5        6    0.200
    6        7    0.167
    7        8    0.143
    8        9    0.125
    9       10    0.111
   10       11    0.100
   11       12    0.091
   12       13    0.083
   13       14    0.077
   14       15    0.071
   15       16    0.067
   16       17    0.062
   17       18    0.059
   18       19    0.056
   19       20    0.053
   20       50    0.020
   21      100    0.010
   22      500    0.002
   23     1000    0.001

: end
```

These calculations are equally easy in Stata when you start with a variable containing sample sizes.

4   Explorations

In statistical science, we use an increasing variety of distributions. Even when closed-form expressions exist for their moments, which is far from being universal, the need to estimate parameters from sample data often arises. Thus the behavior of sample moments and derived measures remains of key interest. Even if you do not customarily use, for example, `summarize, detail` to get skewness and kurtosis, these measures may well underlie your favorite test for normality.

The limits on sample skewness and kurtosis impart the possibility of bias whenever the upper part of their sampling distributions is cut off by algebraic constraints. In extreme cases, a sample may even deny the distribution that underlies it, because it is impossible for any sample to reproduce the skewness and kurtosis of its parent.

These questions may be explored by simulation. Lognormal distributions offer simple but striking examples. We call a distribution for y lognormal if $\ln y$ is normally distributed. Those who prefer to call normal distributions by some other name (Gaussian, notably) have not noticeably affected this terminology. Similarly, for some people the terminology is backward, because a lognormal distribution is an exponentiated normal distribution. Protest is futile while the term *lognormal* remains entrenched.

If $\ln y$ has mean μ and standard deviation σ, its skewness and kurtosis may be defined in terms of $\exp(\sigma^2) = \omega$ (Johnson, Kotz, and Balakrishnan 1994, 212):

$$\gamma_1 = \sqrt{\omega - 1}(\omega + 2); \qquad \beta_2 = \omega^4 + 2\omega^3 + 3\omega^2 - 3$$

Differently put, skewness and kurtosis depend on σ alone; μ is a location parameter for the lognormal as well as the normal.

[R] **simulate** already has a worked example of the simulation of lognormals, which we can adapt slightly for the present purpose. The program there called `lnsim` merely needs to be modified by adding results for skewness and kurtosis. As before, `summarize, detail` is now the appropriate call. Before simulation, we (randomly, capriciously, or otherwise) choose a seed for random-number generation:

```
. clear all
. program define lnsim, rclass
  1.          version 11.1
  2.          syntax [, obs(integer 1) mu(real 0) sigma(real 1)]
  3.          drop _all
  4.          set obs `obs'
  5.          tempvar z
  6.          gen `z' = exp(rnormal(`mu',`sigma'))
  7.          summarize `z', detail
  8.          return scalar mean = r(mean)
  9.          return scalar var  = r(Var)
 10.          return scalar skew = r(skewness)
 11.          return scalar kurt = r(kurtosis)
 12. end
. set seed 2803
. simulate mean=r(mean) var=r(var) skew=r(skew) kurt=r(kurt), nodots
> reps(10000): lnsim, obs(50) mu(-3) sigma(7)

      command:  lnsim, obs(50) mu(-3) sigma(7)
         mean:  r(mean)
          var:  r(var)
         skew:  r(skew)
         kurt:  r(kurt)
```

We are copying here the last example from `help simulate`, a lognormal for which $\mu = -3, \sigma = 7$. While a lognormal may seem a fairly well-behaved distribution, a quick calculation shows that with these parameter choices, the skewness is about 8×10^{31} and the kurtosis about 10^{85}, which no sample result can possibly come near! The previously discussed limits are roughly 7 for skewness and 48 for kurtosis for this sample size. Here are the Mata results:

```
. mata
─────────────────────────────────────────────── mata (type end to exit) ───────
: omega = exp(49)
: sqrt(omega  - 1) * (omega + 2)
  8.32999e+31
: omega^4 + 2 * omega^3 + 3*omega^2 - 3
  1.32348e+85
: n = 50
```

```
: sqrt(n:-1) :- (1:/(n:-1)), n :- 2 + (1:/(n:-1))
                     1            2

    1   6.979591837    48.02040816

: end
```

Sure enough, calculations and a graph (shown as figure 1) show the limits of 7 and 48 are biting hard. Although many graph forms would work well, I here choose `qplot` (Cox 2005) for quantile plots.

```
. summarize
    Variable |       Obs        Mean    Std. Dev.       Min        Max
    ---------+------------------------------------------------------------
        mean |     10000    1.13e+09     1.11e+11    1.888205   1.11e+13
         var |     10000    6.20e+23     6.20e+25    42.43399   6.20e+27
        skew |     10000    6.118604    .9498364    2.382902   6.857143
        kurt |     10000    40.23354    10.06829    7.123528   48.02041
. qplot skew, yla(, ang(h)) name(g1, replace) ytitle(skewness) yli(6.98)
. qplot kurt, yla(, ang(h)) name(g2, replace) ytitle(kurtosis) yli(48.02)
. graph combine g1 g2
```
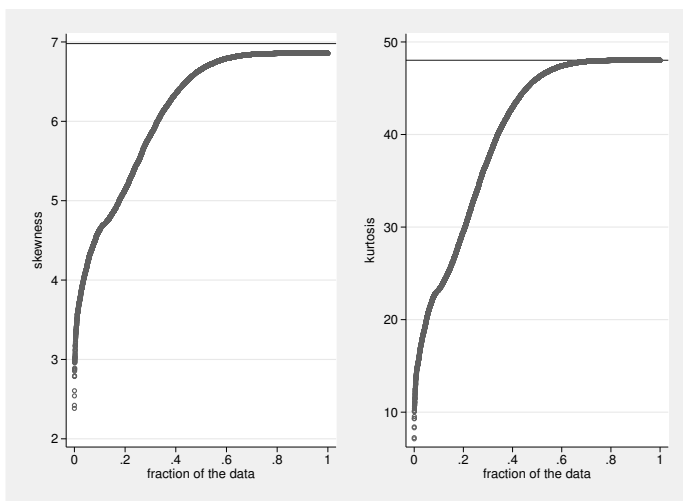

Figure 1. Sampling distributions of skewness and kurtosis for samples of size 50 from a lognormal with $\mu = -3, \sigma = 7$. Upper limits are shown by horizontal lines.

The natural comment is that the parameter choices in this example are a little extreme, but the same phenomenon occurs to some extent even with milder choices. With the default $\mu = 0, \sigma = 1$, the skewness and kurtosis are less explosively high—but still very high by many standards. We `clear` the data and repeat the simulation, but this time we use the default values.

```
. clear
. simulate mean=r(mean) var=r(var) skew=r(skew) kurt=r(kurt), nodots
> reps(10000): lnsim, obs(50)

      command:  lnsim, obs(50)
         mean:  r(mean)
          var:  r(var)
         skew:  r(skew)
         kurt:  r(kurt)
```

Within Mata, we can recalculate the theoretical skewness and kurtosis. The limits to sample skewness and kurtosis remain the same, given the same sample size $n = 50$.

```
. mata
———————————————————————————————————— mata (type end to exit) ———————
: omega = exp(1)
: sqrt(omega  - 1) * (omega + 2)
  6.184877139
: omega^4 + 2 * omega^3 + 3*omega^2 - 3
  113.9363922
: end
——————————————————————————————————————————————————————————————————
```

The problem is more insidious with these parameter values. The sampling distributions look distinctly skewed (shown in figure 2) but are not so obviously truncated. Only when the theoretical values for skewness and kurtosis are considered is it obvious that the estimations are seriously biased.

```
. summarize
    Variable |       Obs        Mean    Std. Dev.       Min        Max

        mean |     10000    1.657829    .3106537    .7871802   4.979507
         var |     10000    4.755659     7.43333    .3971136   457.0726
        skew |     10000    2.617803    1.092607     .467871   6.733598
        kurt |     10000    11.81865    7.996084    1.952879   46.89128
. qplot skew, yla(, ang(h)) name(g1, replace) ytitle(skewness) yli(6.98)
. qplot kurt, yla(, ang(h)) name(g2, replace) ytitle(kurtosis) yli(48.02)
. graph combine g1 g2
```
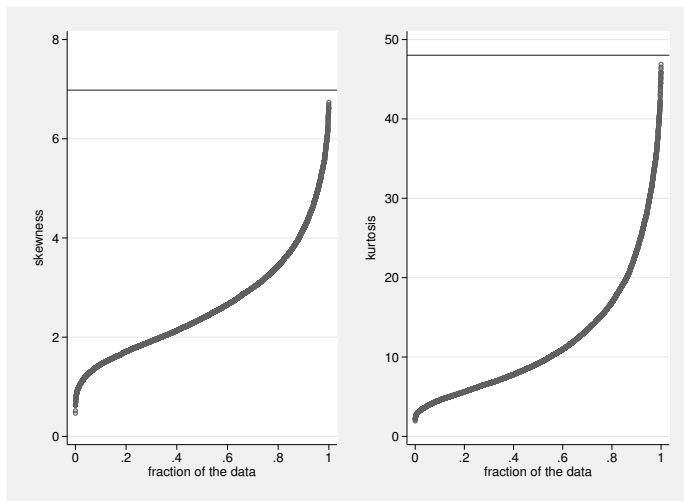
(*Continued on next page*)

Figure 2. Sampling distributions of skewness and kurtosis for samples of size 50 from a lognormal with $\mu = 0, \sigma = 1$. Upper limits are shown by horizontal lines.

Naturally, these are just token simulations, but a way ahead should be clear. If you are using skewness or kurtosis with small (or even large) samples, simulation with some parent distributions pertinent to your work is a good idea. The simulations of Wallis, Matalas, and Slack (1974) in particular pointed to empirical limits to skewness, which Kirby (1974) then established independently of previous work.[3]

5   Conclusions

This story, like any other, lies at the intersection of many larger stories. Many statistically minded people make little or no use of skewness or kurtosis, and this paper may have confirmed them in their prejudices. Some readers may prefer to see this as another argument for using quantiles or order statistics for summarization (Gilchrist 2000; David and Nagaraja 2003). Yet others may know that *L*-moments offer an alternative approach (Hosking 1990; Hosking and Wallis 1997).

Arguably, the art of statistical analysis lies in choosing a model successful enough to ensure that the exact form of the distribution of some response variable, conditional on the predictors, is a matter of secondary importance. For example, in the simplest regression situations, an error term for any really good model is likely to be fairly near normally distributed, and thus not a source of worry. But authorities and critics differ over how far that is a deductive consequence of some flavor of central limit theorem or a naïve article of faith that cries out for critical evaluation.

---

3. Connoisseurs of offbeat or irreverent titles might like to note some other papers by the same team: Mandelbrot and Wallis (1968), Matalas and Wallis (1973), and Slack (1973).

More prosaically, it is a truism—but one worthy of assent—that researchers using statistical methods should know the strengths and weaknesses of the various items in the toolbox. Skewness and kurtosis, over a century old, may yet offer surprises, which a wide range of Stata and Mata commands may help investigate.

6   Acknowledgments

7   References

Bass, H., and T. Y. Lam. 2007. Irving Kaplansky 1917–2006. *Notices of the American Mathematical Society* 54: 1477–1493.

Cox, N. J. 2005. Speaking Stata: The protean quantile plot. *Stata Journal* 5: 442–460.

Cramér, H. 1946. *Mathematical Methods of Statistics*. Princeton, NJ: Princeton University Press.

Dalén, J. 1987. Algebraic bounds on standardized sample moments. *Statistics & Probability Letters* 5: 329–331.

David, H. A. 2001. First (?) occurrence of common terms in statistics and probability. In *Annotated Readings in the History of Statistics*, ed. H. A. David and A. W. F. Edwards, 209–246. New York: Springer.

David, H. A., and H. N. Nagaraja. 2003. *Order Statistics*. Hoboken, NJ: Wiley.

Fiori, A. M., and M. Zenga. 2009. Karl Pearson and the origin of kurtosis. *International Statistical Review* 77: 40–50.

Fisher, R. A. 1925. *Statistical Methods for Research Workers*. Edinburgh: Oliver & Boyd.

Gilchrist, W. G. 2000. *Statistical Modelling with Quantile Functions*. Boca Raton, FL: Chapman & Hall/CRC.

Hald, A. 1981. T. N. Thiele's contribution to statistics. *International Statistical Review* 49: 1–20.

———. 1998. *A History of Mathematical Statistics from 1750 to 1930*. New York: Wiley.

———. 2007. *A History of Parametric Statistical Inference from Bernoulli to Fisher, 1713–1935*. New York: Springer.

Hosking, J. R. M. 1990. *L*-moments: Analysis and estimation of distributions using linear combinations of order statistics. *Journal of the Royal Statistical Society, Series B* 52: 105–124.

Hosking, J. R. M., and J. R. Wallis. 1997. *Regional Frequency Analysis: An Approach Based on L-Moments.* Cambridge: Cambridge University Press.

Johnson, M. E., and V. W. Lowe. 1979. Bounds on the sample skewness and kurtosis. *Technometrics* 21: 377–378.

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

Kadison, R. V. 2008. Irving Kaplansky's role in mid-twentieth century functional analysis. *Notices of the American Mathematical Society* 55: 216–225.

Kaplansky, I. 1945. A common error concerning kurtosis. *Journal of the American Statistical Association* 40: 259.

Katsnelson, J., and S. Kotz. 1957. On the upper limits of some measures of variability. *Archiv für Meteorologie, Geophysik und Bioklimatologie, Series B* 8: 103–107.

Kirby, W. 1974. Algebraic boundedness of sample statistics. *Water Resources Research* 10: 220–222.

———. 1981. Letter to the editor. *Technometrics* 23: 215.

Lauritzen, S. L. 2002. *Thiele: Pioneer in Statistics.* Oxford: Oxford University Press.

Longley, R. W. 1952. Measures of the variability of precipitation. *Monthly Weather Review* 80: 111–117.

Mandelbrot, B. B., and J. R. Wallis. 1968. Noah, Joseph, and operational hydrology. *Water Resources Research* 4: 909–918.

Matalas, N. C., and J. R. Wallis. 1973. Eureka! It fits a Pearson type 3 distribution. *Water Resources Research* 9: 281–289.

Mosteller, F., and J. W. Tukey. 1977. *Data Analysis and Regression: A Second Course in Statistics.* Reading, MA: Addison–Wesley.

Pearson, K. 1916. Mathematical contributions to the theory of evolution. XIX: Second supplement to a memoir on skew variation. *Philosophical Transactions of the Royal Society of London, Series A* 216: 429–457.

Slack, J. R. 1973. I would if I could (self-denial by conditional models). *Water Resources Research* 9: 247–249.

Stuart, A., and J. K. Ord. 1994. *Kendall's Advanced Theory of Statistics. Volume 1: Distribution Theory.* 6th ed. London: Arnold.

Taylor, S. J. 2005. *Asset Price Dynamics, Volatility, and Prediction*. Princeton, NJ: Princeton University Press.

Thiele, T. N. 1889. *Forlæsinger over Almindelig Iagttagelseslære: Sandsynlighedsregning og Mindste Kvadraters Methode*. Copenhagen: C. A. Reitzel. English translation included in Lauritzen 2002.

Walker, H. M. 1929. *Studies in the History of Statistical Method: With Special Reference to Certain Educational Problems*. Baltimore: Williams & Wilkins.

Wallis, J. R., N. C. Matalas, and J. R. Slack. 1974. Just a moment! *Water Resources Research* 10: 211–219.

Wilkins, J. E. 1944. A note on skewness and kurtosis. *Annals of Mathematical Statistics* 15: 333–335.

Yule, G. U. 1911. *An Introduction to the Theory of Statistics*. London: Griffin.

About the author

Nicholas Cox is a statistically minded geographer at Durham University. He contributes talks, postings, FAQs, and programs to the Stata user community. He has also coauthored 15 commands in official Stata. He wrote several inserts in the *Stata Technical Bulletin* and is an editor of the *Stata Journal*.