IMPLEMENTING ALGORITHMS TO MEASURE COMMON STATISTICS

Berkeley Research Group

WHITE PAPER | NOVEMBER 2015

Prepared By:

Stuart McCrary
smccrary@thinkbrg.com
312.429.7902

Berkeley Research Group | 1800 M Street NW, Second floor | Washington, DC 20036

This manuscript is program documentation for three ways to calculate the mean, variance, skewness, kurtosis, covariance, correlation, regression parameters and other regression statistics. Although information contained in this manuscript is believed to be accurate, the documentation is offered without warranty, and users agree to assume all responsibilities and consequences from using this documentation.

Introduction

Formulas for common statistics are generally well known, and users have access to native routines in Microsoft Excel and most programming languages to calculate many statistics. Under most circumstances and with most data, these routines provide identical results. That is, they produce identical results within the mathematical precision available in that environment. However, these algorithms can be constructed in at least three ways, and sometimes the results differ because the algorithms exceed the precision of the environment. Stated differently, the three methods place unequal demands on the precision available for the calculations. Some data also put more demands on the precision available for calculations. For most data, the choice involves convenience; for some data, choosing the right algorithm is important.

Why One-Pass Statistics?

The standard definitions of the statistical formulas described below require two passes through the data. At times it is impossible or inconvenient to wait until all data is available to make the calculations. This might occur because it is necessary to calculate a statistic with all available data up to a point and recalculate after receiving each additional data point. Some data sets are large enough that retaining the data to make two passes is either impractical or impossible. These conditions argue for using one of the one-pass methods described below.

One example in which using one-pass statistics may be valuable involves Monte Carlo simulation, where the number of samples can quickly become very large. In this case, it is convenient to calculate distribution parameters such as the mean, standard deviation, sample skewness,[1] or kurtosis[2] using a one-pass method to avoid having to retain all of the data for ex post analysis. For the same reason, it is convenient to embed the statistical calculations in line in the same code that generates the Monte Carlo test rather than to rely on native statistical routines.

A second example in which one-pass statistics may be valuable also involves Monte Carlo simulation, where the tests are repeated until a certain level of statistical confidence is achieved. For example, the standard error of a Monte Carlo result generally declines proportionate with the square root of the number of trials. When the standard deviation of path results is known in advance, it is possible to also determine in advance how many trials are required. When the standard deviation of sample paths is not known in advance (for example, if this uncertainty depends on inputs to the simulation), it is convenient to run the test until the standard error of the estimate falls below a targeted level. A one-pass method that can be incrementally updated makes such a “smart” stop possible.

1 For the rest of this manuscript, “sample skewness” will just be called “skewness.”

2 In general, this manuscript will not assume that “kurtosis” will actually mean “excess kurtosis” unless labeled as such explicitly. In all cases, kurtosis will refer to the kurtosis of a sample.

Two-Pass Statistics

The standard definitions of variance, skewness, kurtosis, covariance, and simple linear regression begin by assuming that the mean of data to be analyzed is already known. An algorithm first calculates the mean. In the interest of completeness and to introduce the notation, that mean is shown in Equation 1:

$$\mu_X = \frac{\sum_{i=1}^{N} X_i}{N} \tag{1}$$

To calculate the mean, μ_X, of a vector X, add all the values for X and divide by the number of observations. The mean is sometimes called the first sample moment of a statistical distribution. The unit of measure that applies to μ_X is the same unit that applies to X. For example, if X is measured in feet, the mean produced by Equation 1 will be in feet.

The standard definition of sample variance appears in Equation 2:

$$\sigma_X^2 = \frac{\sum_{i=1}^{N} (X_i - \mu_X)^2}{N - 1} \tag{2}$$

An algorithm that first calculates the results of Equation 1 and then Equation 2 is called a two-pass algorithm for calculating variance. The variance is sometimes called the second sample moment of a statistical distribution and the numerator is called the sum of squares. The unit of measure that applies to σ_X^2 is the square of the unit that applies to X.

The definition of sample standard deviation appears in Equation 3:

$$\sigma_X = \sqrt{\frac{\sum_{i=1}^{N} (X_i - \mu_X)^2}{N - 1}} \tag{3}$$

Alternatively, the standard deviation can be described as the square root of variance, in which case the algorithm builder doesn't really need a separate formula as in Equation 3. Each of the one-pass methods described below follows that pattern: find the variance, then transform it into standard deviation if needed. The unit of measure that applies to σ_X is the same unit that applies to X. For example, if X is measured in feet, the standard deviation produced by Equation 3 will be in feet.

Equations 2 and 3 divide the sum of squares by the number of observations reduced by 1. This adjustment makes these sample statistics unbiased. A similar bias adjustment is required for skewness and kurtosis but often does not appear in published formulas. This manuscript will follow that convention and then discuss how to adjust the results to be unbiased.

The standard definition of skewness appears in Equation 4:

$$Skew_X = \frac{\sum_{i=1}^{N} (X_i - \mu_X)^3}{N * \sigma_X^3} \tag{4}$$

The denominator in Equation 4 can be described as either the standard deviation raised to the third power or the variance raised to the 1.5 power. Of course, that denominator requires a pass through the data, and the calculation of the denominator must be made before the skewness is calculated. However, the summations required to calculate the denominator using Equation 2 or Equation 3 can be tallied on the same pass through the data required to calculate the sum in the numerator of Equation 4. For this reason, it is still generally described as a two-pass formula.

The skewness is sometimes called the third moment of a statistical distribution. The unit of measure that applies to $Skew_X$ is independent of the unit that applies to X. For any data, a skewness of 0 is considered not skewed, while positive values are described as skewed right and negative values as skewed left.

The standard definition of kurtosis appears in Equation 5:

$$Kurtosis_X = \frac{\sum_{i=1}^{N}\left(X_i - \mu_X\right)^4}{N * \sigma_X^4} \tag{5}$$

This kurtosis formula could be described as a two-pass formula, because it relies on a prior step to calculate the mean, then a second step that sums values for both the numerator and the denominator.

The kurtosis is the fourth sample moment of a statistical distribution. The unit of measure that applies to $Kurtosis_x$ is independent of unit that applies to X. For any data, a kurtosis of about[3] 3 is considered typical of normally distributed values and described as mesokurtic. Kurtosis larger than about 3 is described as leptokurtic (fat tails), and kurtosis smaller than about 3 is platykurtic (thinner tails). A measure called excess kurtosis subtracts approximately 3[4] so that excess kurtosis is centered around 0.

The standard definition for covariance appears as Equation 6:

$$\sigma_{X,Y} = \frac{\sum_{i=1}^{N}\left(X_i - \mu_X\right)\left(Y_i - \mu_Y\right)}{N} \tag{6}$$

Equation 6 closely resembles the definition of variance in Equation 2. In fact, Equation 6 becomes Equation 2 (except for the minor difference in the denominators) when Equation 6 is used to measure the covariance between a variable and itself.

The covariance is not considered a moment. The units that apply to Equation 6 lack intuitive clarity. For this reason, correlation is calculated as a kind of standardized or normalized covariance. See Equation 7:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y} \tag{7}$$

3 See a longer description on how the adjustment differs from 3 the bias section.

4 See Appendix D for a more precise definition of the excess kurtosis adjustment.

Textbook Formulas for One-Pass Statistics

Each of the four moments described above and the covariance can be restated in a format that is conducive for building a one-pass algorithm. Each of the formulas is algebraically equivalent to the standard formulas. This means that if mathematical routines could produce the exact values called for in the equations above and below, the results would identically match the results from the equations above.

The equivalent one-pass formula for the variance[5] appears as equation 8:

$$\sigma_X^2 = \frac{\sum_{i=1}^{N} X_i^2 - N\overline{X}^2}{N-1} \tag{8}$$

Equation 8 is sometimes called the "textbook" formula because it is frequently included in statistical textbooks.[6] It permits construction of a one-pass algorithm because the mean is only needed at the end of a pass through the data. That algorithm sum both X and X^2.

The equivalent one-pass textbook formula for the skewness[7] appears as Equation 9:

$$\text{Skew}_X = \frac{\sum_{i=1}^{N} X_i^3 - 3\overline{X}\sum_{i=1}^{N} X_i^2 + 2N\overline{X}^3}{\sigma_X^3} \tag{9}$$

Although this author has not seen Equation 9 published, it is convenient to describe it as the textbook formula for skewness. It permits construction of a one-pass algorithm because the mean is only needed at the end of a pass through the data. That algorithm must sum X, X^2, and X^3.

The equivalent textbook one-pass formula for the kurtosis[8] appears as Equation 10:

$$\text{Kurt} = \frac{N\overline{X}^4 - 4\overline{X}^3\sum_{i=1}^{N} X_i + 6\overline{X}^2\sum_{i=1}^{N} X_i^2 - 4\overline{X}\sum_{i=1}^{N} X_i^3 + \sum_{i=1}^{N} X_i^4}{\left(\sum_{i=1}^{N} X_i^2 - N\overline{X}^2\right)^2} \tag{10}$$

Although this author has not seen Equation 10 published, it is convenient to describe it as the textbook formula for kurtosis. It permits construction of a one-pass algorithm because the mean is only needed at the end of a pass through the data. That algorithm must sum X, X^2, X^3, and X^4.

5 The derivation of Equation 8 appears in Appendix A.

6 Chan, Tony F., Gene H. Golub, and Randall J. LeVeque, "Algorithms for Computing Sample Variance, Analysis and Recommendations," *The American Statistician* 37:3 (August 1983), 242–247.

7 The derivation of Equation 9 appears in Appendix B.

8 The derivation of Equation 10 appears in Appendix C.

The textbook methodology lends itself to a one-pass method for calculating the covariance. Equation 11[9] follows the textbook strategy and requires the sums of X, Y, and XY.

$$\sigma_{X,Y} = \frac{\sum_{i=1}^{N} X_i Y_i - N\overline{X}\,\overline{Y}}{N-1} \tag{11}$$

Of course, a one-pass textbook algorithm to calculate the correlation coefficient follows, using Equation 11 and Equation 7. Calculate the covariance of two vectors using Equation 11 and the square root of Equation 8 (variance) to calculate the standard deviation of each vector with a single pass.

It is also possible to develop a similar one-pass formula for a regression slope, β, for a single[10] independent variable. Equation 11 requires the sums of X, X^2, Y, and XY and requires knowledge of the mean, but that mean is not required to complete the other calculations, so the terms needed to evaluate Equation 12 can be accumulated on a single pass through the data.

$$\beta = \frac{\sum_{i=1}^{N} X_i Y_i - \overline{X}\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i^2 - 2\overline{X}\sum_{i=1}^{N} X_i + N\overline{X}^2} \tag{12}$$

The intercept, α, in Equation 13 can be found using terms evaluated for Equation 12. Equation 13 relies on the knowledge that the means of the X and Y values represent a point on the regression line.

$$\overline{Y} = \beta\overline{X} + \alpha \tag{13}$$

$$\alpha = \frac{\sum_{i=1}^{N} Y_i - m\sum_{i=1}^{N} X_i}{N} \tag{14}$$

Therefore, the intercept can be determined from a single pass if the slope is known. By relying on Equation 12, which permits a one-pass algorithm, the regression line can be determined with a one-pass methodology similar to the textbook algorithms above.

Numerical Precision of Textbook One-Pass Algorithms

The textbook algorithms are vulnerable to computational errors for certain types of data. For example, if the magnitude of that data is large, requiring much of the available precision of a computer system, and if the variance is small relative to the underlying data, it is not difficult to construct a hypothetical data set where algorithms based on the textbook formulas for variance, skewness, and kurtosis produce unreliable results.[11] Some authors have advised against using the textbook algorithm because it is more prone to errors introduced by the computational limits of computer mathematical operations. Although two-pass methods are much less likely to exceed the computational precision of a computer, it is also possible to find data where the two-pass method can produce unreliable results. Some strategies can improve the accuracy of two-pass methods. For example, from all data, subtract a large number somewhat close to the expected mean. A third method introduced by Welford is generally least likely to require arithmetic operations that exceed the precision of the computer.

9 The derivation of Equation 11 appears in Appendix D.

10 The derivation of Equation 12 appears in Appendix E.

11 See for example, Cook, John D., "Comparing Three Methods of Computing Standard Deviation," John D. Cook blog (September 26, 2008), accessed at: http://www.johndcook.com/blog/2008/09/26/comparing-three-methods-of-computing-standard-deviation/

Online One-Pass Statistics

Welford[12] introduced a way to calculate variance without requiring the prior calculation of the mean. Knuth[13] designed a well-tested algorithm for calculating variance, relying on the Welford formulation. Welford defined the mean conventionally.

$$M_N = \frac{\sum_{i=1}^{N} X_i}{N} \tag{15}$$

A thoughtful algorithm accumulates the sum in the numerator. The mean is found by dividing this accumulation by the prevailing "N."

The Online variance accumulates the sum of squares and uses that sum to calculate variance. Equation 16 defines the sum of squares as an increment based on the previous sum of squares.

$$S_N = S_{N-1} + \frac{N-1}{N}(X_N - M_{N-1})^2 \tag{16}$$

Calculate variance of the N data points with the sum of squares using Equation 17:

$$\sigma_X^2 = \frac{S_N}{N-1} \tag{17}$$

Note that this formulation supports an incremental algorithm that works as long as the latest sum of the X_is and the sum of squares is preserved. As with Equation 8, the standard deviation is calculated using Equation 17 then taking the square root of the variance.

Chan et al.[14] extended the Online methodology to allow the accumulated analysis from one block of data to be merged with the accumulated analysis of a second block of data. Timothy B Terriberry[15] extended the Chan methodology to permit merging of data sets used to calculate skewness and kurtosis. The Terriberry equations reduce to the following methodology when one additional data point is added to a series:

$$\delta_N = X_N - M_{1,N-1} \tag{18}$$

Equation 18 calculates the deviation of the next data point from the mean prevailing before, updating the mean to reflect that data point. M_1 of course equals X_1 for N=1.

$$\overline{X}_N = M_{1,N} = M_{1,N-1} + \frac{\delta}{N} \tag{19}$$

Next, the mean is updated using Equation 19. Notice that Equations 18 through 22 introduce interim sums. M_1 is equal to the mean of X. M_2, M_3, and M_4 provide a convenient way to calculate standard deviation, skew, and kurtosis.

$$M_{2,N} = M_{2,N-1} + \delta_N^2 \frac{N-1}{N} \tag{20A}$$

$$\sigma_N = \sqrt{\frac{M_{2,N}}{N-1}} \tag{20B}$$

12 Welford, B. P., "Note on a Method for Calculating Corrected Sums of Squares and Products," *Technometrics* 4:3 (August 1962), 419–420.

13 Donald E. Knuth, *The Art of Computer Programming*, Volume 2: Seminumerical Algorithms, third ed. (1998), 232.

14 Chan et al. (1983).

15 Terriberry, Timothy, Computing Higher-Order Moments Online, 2008, https://people.xiph.org/~tterribe/notes/homs.html, accessed 8/14/15.

Use Equation 20A and 20B to update the standard deviation or to calculate variance. Notice that the value of M_2 is not altered and can be continually used to accumulate more data.

$$M_{3,N} = M_{3,N-1} + \delta^3 \frac{(N-1)(N-2)}{N^2} - \frac{3\delta * M_{2,N-1}}{N} \tag{21A}$$

$$\text{Skew}_N = \frac{\sqrt{N} M_{3,N}}{M_{2,N}^{3/2}} \tag{21B}$$

Update the skewness using Equations 21A and 21B.

$$M_{4,N} = M_{4,N-1} + \delta^4 \frac{(N-1)(N^2-3N+3)}{N^3} + \frac{6\delta^2 M_{2,N-1}}{N^2} - \frac{4\delta * M_{3,N-1}}{N} \tag{22A}$$

$$\text{Kurt}_N = \frac{NM_N}{M_2^2} \tag{22B}$$

Finally, update the kurtosis using Equations 22A and 22B.

It is also possible to derive a Welford-like formula for covariance.[16] The algorithm used herein relies on Equation 23[17]:

$$\sigma_{XY,N} = \frac{(N-1)}{N}\left(\sigma_{XY,N-1} + (X_N - \overline{X}_{N-1})(Y_N - \overline{Y}_{N-1})/N\right) \tag{23}$$

Online Regression Parameters

It is possible to calculate a regression beta using a formula similar to the Online variance formula. The algorithm is summarized in Equation 24[18]. Here, values for the numerator rely on previous values, which follow a now-familiar pattern because the algorithm adapts the previous sum to the new mean. The denominator is the sum of squared deviations accumulated for the independent valuation, shown as a variation on the sum of squares formulated in Equation 2.

$$\text{Beta}_N = \frac{\text{SumxY}_{N-1} + \frac{X_N - \overline{X}_{N-1}}{N}\left(Y_N(N-1) - \text{SumY}_{N-1}\right)}{(N-1)\sigma_{X,N}} \tag{24}$$

As described in the Appendix G, the "x" in "SumxY" refers to the data points X minus the mean of X but Equation 24 nevertheless presents a methodology that allows for incremental updating.

As before, calculate alpha using the means of both the independent and dependent variables. This relationship relies on the fact that a fit line passes through the coordinate equal to the means of X and Y in Equations 25 and 26.

$$\alpha = \mu_Y - \beta * \mu_X \tag{25}$$

$$= \frac{\sum_{1=1}^{N} Y_i}{N} - \beta * \frac{\sum_{1=1}^{N} X_i}{N} \tag{26}$$

16 Pebay, Phillip, *Formulas for Robust, One-Pass Parallel Computation of Covariances and Arbitrary-Order Statistical Moments*, Sandia Report, SAND2008-6212 (September 2008).

17 The derivation of Equation 23 appears in Appendix F.

18 The derivation of Equation 24 appears in Appendix G.

These means or summations can be calculated using the algorithm in Equation 15, because no other part of the calculations depend on the prevailing mean. The β is calculated with Equation 24.

Other Statistics

A large number of statistics involved with linear regression could potentially be calculated with a one-pass algorithm: total sum of squares, error sum of squares, regression sum of squares, R-square, F statistic, the standard error of estimate, the standard error of the slope, the standard error of the intercept, and t-tests of regression parameters. This manuscript will not seek to derive one-pass methods to calculate these additional statistical values.

Multiple regression is almost always conducted with the use of matrix operations: the inverse of a matrix, the transpose of a matrix, and matrix multiplication. The format does not appear to lend itself to one-pass algorithms. The formula for beta, for example, appears in Equation 27:

$$\beta_n = \left(X'X\right)^{-1} X' y \tag{27}$$

In Equation 27, {X} refers to a matrix that contains two or more independent variables, multiplication refers to matrix multiplication, the symbol {‘} refers to the transpose of a matrix, the symbol {–1} refers to the inverse of a matrix, and the vector {y} represents a vector of the deviations from the average of all the Y values.

This manuscript will not attempt to incrementally adapt to, for example, N data points following the analysis of N – 1 data points.

Exponential Smoothing

Exponentially smoothed data is inherently one-pass in nature. A weighted average of previous values is described in Equation 28, where an updated average equals a combination of the latest sample and the previous estimated average:

$$\hat{\mathrm{X}}_{\mathrm{i}} = \alpha \mathrm{X}_{\mathrm{i}} + \left(1-\alpha\right)\hat{\mathrm{X}}_{\mathrm{i}-1} \tag{28}$$

$$\text{Where } 0 \leq \alpha \geq 1$$

Exponential smoothing may offer computational efficiencies over other one-pass methods. By picking a relatively low value for α, the statistic should approximate an average of all data in the sample. Alternatively, by selecting a relatively high value for α, the statistic can be calibrated to match recent observations.

It follows that another way to create a one-pass method of calculating the variance is to adopt the method into the standard definition of variance. One example of such a hybrid is Equation 29:

$$\hat{\sigma}_i^2 = \alpha\left(X_i - \hat{\mathrm{X}}_i\right)^2 + \left(1-\alpha\right)\hat{\sigma}_{i-1}^2 \tag{29}$$

The use of α and 1 – α make the result an expected value of the sum of squares, which is also the intent in Equation 2, where the sum of squares is divided by N – 1. To calculate the analogue to the standard deviation, take the square root of the statistic in Equation 29.

Equation 30 applies the exponential weighting to the elements in the numerator of the skewness formula in Equation 4:

$$\hat{\mathrm{S}}\mathrm{kew}_{\mathrm{i}} = \frac{\mathrm{SumS}}{\hat{\sigma}_{\mathrm{i}}^3} \tag{30}$$

$$\text{where } \mathrm{SumS} = \alpha\left(\mathrm{X}_{\mathrm{i}} - \hat{\mathrm{X}}_{\mathrm{i}}\right)^3 + \left(1-\alpha\right)\mathrm{SumS}_{\mathrm{i}-1}$$

Here, the exponentially skewed average in Equation 28 is substituted for the sample mean, and a power of the statistic calculated in Equation 29 substitutes for the standard deviation.

Equation 31 applies the exponential weighting to elements in the numerator of the kurtosis formula in Equation 5:

$$\hat{\mathrm{K}}\mathrm{urt}_i = \frac{\mathrm{SumK}_i}{\hat{\sigma}_i^4} \tag{31}$$

$$\text{Where } \mathrm{SumK}_i = \alpha\left(X_i - \hat{X}_i\right)^4 + (1-\alpha)\mathrm{SumK}_{i-1}$$

As in Equation 30, the exponentially skewed average in Equation 28 is substituted for the sample mean, and the statistic calculated in Equation 29 substitutes for the standard deviation.

Bias adjustments

Taking the mean of a distribution removes a degree of freedom from a sample. This is why the formula for variance in Equation 2 and the formula for standard deviation in Equation 3 use N – 1 rather than N in the denominator. A similar adjustment[19] is necessary to make the formulas for skewness (Equation 30) unbiased for samples of data. Equation 32 matches the value of skewness as calculated by Minitab.

$$\mathrm{Skew}_{\mathrm{Unbiased}} = \mathrm{Skew}_{\mathrm{Biased}}\left(\frac{N-1}{N}\right)^{3/2} \tag{32}$$

Likewise, the sample kurtosis adjusted with Equation 33 should match the Minitab measure of unbiased excess kurtosis.

$$\text{Excess Kurt}_{\mathrm{Unbiased}} = \left(\frac{N-1}{N}\right)^2 \mathrm{Kurt}_{\mathrm{Biased}} - 3 \tag{33}$$

A slightly different adjustment is required to match the skewness calculated by SAS, SPSS, and Excel

$$\mathrm{Skew}_{\mathrm{Unbiased}} = \mathrm{Skew}_{\mathrm{Biased}} * \frac{\sqrt{N(N-1)}}{N-2} \tag{34}$$

Equation 35 shows the bias adjustment of kurtosis to match SAS, SPSS, and Excel:

$$\mathrm{Kurt}_{\mathrm{Unbiased}} = \mathrm{Kurt}_{\mathrm{Biased}} * \frac{N-1}{(N-2)(N-3)} \tag{35}$$

Finally, Equation 36 provides the adjustment needed to match the kurtosis calculated within SAS, SPSS, and Excel[20]:

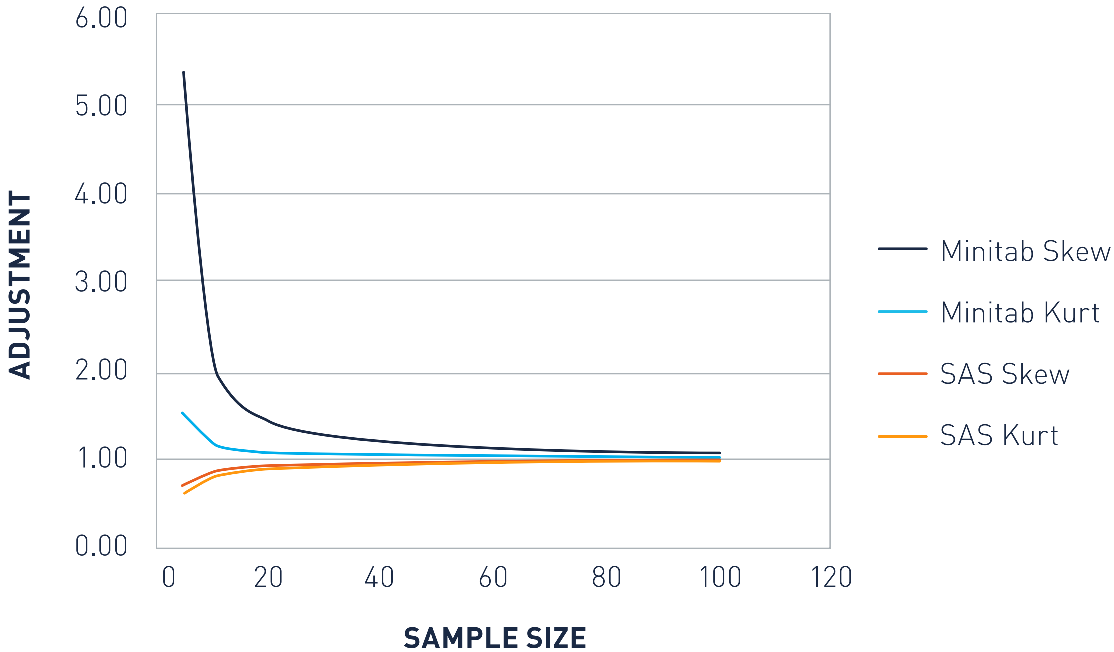
$$\mathrm{Kurt}_{\mathrm{Unbiased}} = \frac{N-1}{(N-2)(N-3)}\left((N+1) * \mathrm{Kurt}_{\mathrm{Biased}} + 6\right) \tag{36}$$

These adjustments matter primarily for small sample sizes. This adjustment factor would equal 1.00 if the preliminary estimate (that is, before adjusting for the bias) is essentially unbiased. Figure 1 shows that these adjustments are negligible for larger sample sizes.

19 Joanes, D.N., and C.A. Gill, "Comparing Measures of Sample Skewness and Kurtosis," *Journal of the Royal Statistical Society*, Series D (The Statistician), 47:1 (1998), 183–189.

20 Joanes, p 183-189

FIGURE 1. BIAS ADJUSTMENT



Standard Error of the Mean, Variance, Skewness, Kurtosis

Monte Carlo simulations frequently report the average of the outcomes. Because these are sampled estimates of the true mean, it is important to measure the standard error of the mean. The standard error is defined by Equation 37[21]:

$$\sigma_{\bar{X}} = \frac{\sigma_{X_i}}{\sqrt{N}} \tag{37}$$

The standard error can be derived from the standard deviation of the outcomes and is therefore available as a one-pass statistic.

The standard error of variance is defined in Equation 38[22]:

$$\sigma_{s^2} = s^2 \sqrt{\frac{2}{N-1}} \tag{38}$$

Where s^2 represents the variance of a sample

The standard error of variance can be derived from the variance of a sample and is therefore available as a one-pass statistic.

For large values of N, the standard error of the sample standard deviation is approximated by Equation 39[23]:

$$\sigma_s = s \frac{1}{\sqrt{2(N-1)}} \tag{39}$$

21 Ahn, Sangtae, and Jeffrey Fessler, *Standard Error of Mean, Variance, and Standard Deviation Estimators*, EECS Department, University of Michigan (2003). http://web.eecs.umich.edu/~fessler/papers/files/tr/stderr.pdf Accessed 8/14/15.

22 Ahn and Fessler (2003).

23 Ahn and Fessler (2003).

The standard error of the sample standard deviation can be approximated from the sample standard deviation and is therefore available as a one-pass statistic.

The standard error of the skewness is defined in Equation 40[24]:

$$\text{SES} = \sqrt{\frac{6\text{N}(\text{N}-1)}{(\text{N}-2)(\text{N}+1)(\text{N}+3)}} \tag{40}$$

The standard error of skewness is derived from the sample size and is therefore available as a one-pass statistic.

The standard error of the sample kurtosis is defined in Equation 41[25]:

$$\text{SEK} = 2SES\sqrt{\frac{\text{N}^2-1}{(\text{N}-3)(\text{N}+5)}} \tag{41}$$

The standard error of sample kurtosis is derived from the sample size and is therefore available as a one-pass statistic.

Conclusions

Statistical routines built into spreadsheets and statistical packages generally return numerically indistinguishable results for most data sets. Certain data sets create measurement problems using one or two methods described herein. For these data sets, the algorithms built around the methodology introduced by Welford may provide more accurate results.

This documentation primarily describes an application of one-pass methodologies to Monte Carlo trials. In these applications, a two-pass method may be impractical. Many such Monte Carlo samples are not problematic for either the textbook or Online method. Where the results are the same, it is difficult to argue that one method is better than the other.

While the textbook method can produce accurate results most of the time, a level of uncertainty remains that perhaps a particular trial pushes into an area where the textbook method is inaccurate. One way to be more confident about statistical measurements is to perform them with two or three different algorithms and confirm that the results are equivalent for whatever precision is required.

24 Cramer, Duncan, *Fundamental Statistics for Social Research*, (1997) p 85.

25 Cramer, p. 89

Appendix A – Derivation of Textbook Variance Formula

$$\sigma^2 = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^2}{N-1} = \frac{\left(X_1 - \overline{X}\right)^2 + \left(X_2 - \overline{X}\right)^2 + \ldots\left(X_N - \overline{X}\right)^2}{N-1} \qquad \text{(A1)}$$

$$\sigma^2 = \frac{X_1^2 + X_2^2 + \ldots X_N^2}{N-1} + \frac{N\left(\overline{X}^2\right)}{N-1} - \frac{2X_1\overline{X} + 2X_2\overline{X} + \ldots 2X_N\overline{X}}{N-1} \qquad \text{(A2)}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N-1} + \frac{N\overline{X}^2}{N-1} - \frac{2\overline{X}\sum_{i=1}^{N} X_i}{N-1} = \frac{\sum_{i=1}^{N} X_i^2}{N-1} + \frac{N\overline{X}^2}{N-1} - \frac{2\overline{X} * N\overline{X}}{N-1} \qquad \text{(A3)}$$

$$\sigma^2 = \frac{\sum_{i=1}^{N} X_i^2}{N-1} - \frac{N\overline{X}^2}{N-1} = \frac{\sum_{i=1}^{N} X_i^2}{N-1} - \frac{\left(\sum_{i=1}^{N} X_i\right)^2}{N(N-1)} = \frac{\sum_{i=1}^{N} X_i^2 - N\overline{X}^2}{N-1} \qquad \text{(A4)}$$

Appendix B – Derivation of Textbook Skewness Formula

$$\text{Skew} = \frac{N}{(N-1)(N-2)} \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)^3}{s^3} \tag{B1}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\left(X_1 - \overline{X}\right)\left(X_1 - \overline{X}\right)\left(X_1 - \overline{X}\right) + \left(X_2 - \overline{X}\right)\left(X_2 - \overline{X}\right)\left(X_2 - \overline{X}\right)\ldots \tag{B2}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\left[\left(X_1 - \overline{X}\right)\left(X_1^2 + \overline{X}^2 - 2X_1\overline{X}\right) + \left(X_2 - \overline{X}\right)\left(X_2^2 + \overline{X}^2 - 2X_2\overline{X}\right)\ldots\right] \tag{B3}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\begin{bmatrix} X_1^3 + X_1\overline{X}^2 - 2X_1^2\overline{X} - X_1^2\overline{X} - \overline{X}^3 + 2X_1\overline{X}^2 + \\ X_2^3 + X_2\overline{X}^2 - 2X_2^2\overline{X} - X_2^2\overline{X} - \overline{X}^3 + 2X_2\overline{X}^2 + \ldots \end{bmatrix} \tag{B4}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\left(\sum_{i=1}^{N} X_i^3 - N\overline{X}^3 + 3\overline{X}^2\sum_{i=1}^{N} X_i - 3\overline{X}\sum_{i=1}^{N} X_i^2\right) \tag{B5}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\left(\sum_{i=1}^{N} X_i^3 - N\overline{X}^3 + 3N\overline{X}^2\overline{X} - 3\overline{X}\sum_{i=1}^{N} X_i^2\right) \tag{B6}$$

$$\text{Skew} = \frac{N}{(N-1)(N-2)s^3}\left(\sum_{i=1}^{N} X_i^3 - 3\overline{X}\sum_{i=1}^{N} X_i^2 + 2N\overline{X}^3\right) \tag{B7}$$

Appendix C – Derivation of Textbook Kurtosis Formula

$$\text{Kurt} = \frac{\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^4}{\left(\sum_{i=1}^{N}\left(X_i - \bar{X}\right)^2\right)^2} \tag{C1}$$

$$\text{Num} = \left(X_1 - \bar{X}\right)\left(X_1 - \bar{X}\right)\left(X_1 - \bar{X}\right)\left(X_1 - \bar{X}\right) + \left(X_2 - \bar{X}\right)\left(X_2 - \bar{X}\right)\left(X_2 - \bar{X}\right)\left(X_2 - \bar{X}\right) + \ldots \tag{C2}$$

$$\begin{aligned}\text{Num} = {} & \left(X_1^2 + \bar{X}^2 - 2\bar{X}X_1\right)\left(X_1^2 + \bar{X}^2 - 2\bar{X}X_1\right) + \\ & \left(X_1^2 + \bar{X}^2 - 2\bar{X}X_2\right)\left(X_1^2 + \bar{X}^2 - 2\bar{X}X_2\right) + \ldots\end{aligned} \tag{C3}$$

$$\begin{aligned}\text{Num} = {} & X_1^4 + X_1^2\bar{X}^2 - 2X_1^3\bar{X} + X_1^2\bar{X}^2 + \bar{X}^4 - 2X_1\bar{X}^3 - 2X_1^3\bar{X} - 2X_1\bar{X}^3 + 4X_1^2\bar{X}^2 + \\ & X_2^4 + X_2^2\bar{X}^2 - 2X_2^3\bar{X} + X_2^2\bar{X}^2 + \bar{X}^4 - 2X_2\bar{X}^3 - 2X_2^3\bar{X} - 2X_2\bar{X}^3 + 4X_2^2\bar{X}^2 + \ldots\end{aligned} \tag{C4}$$

$$\text{Num} = \bar{X}^4 - 4X_1\bar{X}^3 + 6X_1^2\bar{X}^2 - 4X_1^3\bar{X} + X_1^4 + \bar{X}^4 - 4X_2\bar{X}^3 + 6X_2^2\bar{X}^2 - 4X_2^3\bar{X} + X_2^4 \tag{C5}$$

$$\text{Num} = N\bar{X}^4 - 4\bar{X}^3\sum_{i=1}^{N}X_i + 6\bar{X}^2\sum_{i=1}^{N}X_i^2 - 4\bar{X}\sum_{i=1}^{N}X_i^3 + \sum_{i=1}^{N}X_i^4 \tag{C6}$$

$$\text{Kurt} = \frac{N\bar{X}^4 - 4\bar{X}^3\sum_{i=1}^{N}X_i + 6\bar{X}^2\sum_{i=1}^{N}X_i^2 - 4\bar{X}\sum_{i=1}^{N}X_i^3 + \sum_{i=1}^{N}X_i^4}{\left(\sum_{i=1}^{N}X_i^2 - N\bar{X}^2\right)^2} \tag{C7}$$

Equation C6 replaces the numerator of Equation C1 and Equation A4 provides the denominator.

Appendix D – Derivation of Textbook Covariance Formula

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}\right)\left(Y_i - \overline{Y}\right)}{N-1} \tag{D1}$$

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N}\left(X_i Y_i - X_i\overline{Y} - Y_i\overline{X} + \overline{X}\overline{Y}\right)}{N-1} \tag{D2}$$

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N} X_i Y_i - \overline{Y}\sum_{i=1}^{N} X_i - \overline{X}\sum_{i=1}^{N} Y_i + N\overline{X}\overline{Y}}{N-1} \tag{D3}$$

$$\rho_{X,Y} = \frac{\sum_{i=1}^{N} X_i Y_i - N\overline{X}\overline{Y}}{N-1} \tag{D4}$$

Appendix E – Derivation of Textbook Regression Beta Formula

The slope of a linear regression line includes two terms.[26] The numerator equals the sum of the products of the Y observations times the amount that the X values deviate from their mean. The denominator is the sum of squared deviations.

$$\beta = \frac{\sum\left(X_i - \overline{X}_n\right)Y_i}{\sum\left(X_i - \overline{X}_n\right)^2} \tag{E1}$$

$$\beta = \frac{X_1Y_1 + X_2Y_2 \ldots + X_NY_N - \overline{X}Y_1 - \overline{X}Y_2 \ldots - \overline{X}Y_N}{X_1^2 + X_2^2 \ldots X_N^2 - 2\overline{X}X_1 - 2\overline{X}X_2 \ldots - 2\overline{X}X_N + N\overline{X}^2} \tag{E2}$$

$$\beta = \frac{\sum_{i=1}^{N} X_i Y_i - \overline{X}\sum_{i=1}^{N} Y_i}{\sum_{i=1}^{N} X_i^2 - 2\overline{X}\sum_{i=1}^{N} X_i + N\overline{X}^2} \tag{E3}$$

26 Wonnacott, Thomas H., and Ronald J. Wonnacott, *Introductory Statistics for Business and Economics*, second ed. (1977), 323, Equation 11-16.

Appendix F – Derivation of Online Covariance Formula

$$\sigma_{XY,N} = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X_N}\right)\left(Y_i - \overline{Y_N}\right)}{N} \tag{F1}$$

$$\sigma_{XY,N} = \frac{\sum_{i=1}^{N-1}\left(X_i - \overline{X_{N-1}} - \frac{X_N - \overline{X_{N-1}}}{N}\right)\left(Y_i - \overline{Y_{N-1}} - \frac{Y_N - \overline{Y_{N-1}}}{N}\right) + \left(\frac{N-1}{N}\right)^2\left(X_N - \overline{X_{N-1}}\right)\left(Y_N - \overline{Y_{N-1}}\right)}{N} \tag{F2}$$

$$\sigma_{X,Y} = \frac{\begin{array}{c}\sum_{i=1}^{N-1}\left(X_i - \overline{X_{N-1}}\right)\left(Y_i - \overline{Y_{N-1}}\right) - \frac{Y_N - \overline{Y_{N-1}}}{N}\sum_{i=1}^{N-1}\left(X_i - \overline{X_{N-1}}\right) - \frac{X_N - \overline{X_{N-1}}}{N}\sum_{i=1}^{N-1}\left(Y_i - \overline{Y_{N-1}}\right)\\ + (N-1)\frac{X_N - \overline{X_{N-1}}}{N} * \frac{Y_N - \overline{Y_{N-1}}}{N} + \left(\frac{N-1}{N}\right)^2\left(X_N - \overline{X_{N-1}}\right)\left(Y_N - \overline{Y_{N-1}}\right)\end{array}}{N} \tag{F3}$$

$$\sigma_{X,Y} = \frac{(N-1)}{N}\left(\sigma_{XY,N-1} + \frac{\left(X_N - \overline{X_{N-1}}\right)\left(Y_N - \overline{Y_{N-1}}\right)}{N}\right) \tag{F4}$$

Appendix G – Derivation of Welford Regression Beta Formula

$$\mathrm{SumxY}_N = \sum_{i=1}^{N}\left(X_i - \overline{X}_N\right)Y_N \qquad \text{(G1)}$$

$$\mathrm{SumxY}_N = \sum_{i=1}^{N-1}\left(X_i - \overline{X_{N-1}} - \frac{X_N - \overline{X}_{N-1}}{N}\right)Y_n + \frac{N-1}{N}\left(X_N - \overline{X}_{N-1}\right)Y_N \qquad \text{(G2)}$$

$$\mathrm{SumxY}_N = \mathrm{SumxY}_{N-1} - \frac{X_N - \overline{X}_{N-1}}{N}\mathrm{SumY}_{N-1} + \frac{N-1}{N}\left(X_N - \overline{X}_{N-1}\right)Y_N \qquad \text{(G3)}$$

$$\mathrm{SumxY}_N = \mathrm{SumxY}_{N-1} + \frac{X_N - \overline{X}_{N-1}}{N}\left(Y_N\left(N-1\right) - \mathrm{SumY}_{N-1}\right) \qquad \text{(G4)}$$

$$\beta_N = \frac{\sum_{i=1}^{N}\left(X_i - \overline{X}_N\right)Y_n}{\sum_{i=1}^{N}\left(X_i - \overline{X}_N\right)} \qquad \text{(G5)}$$

$$\beta_N = \frac{\mathrm{SumxY}_N}{\sigma_{X,N}} \qquad \text{(G6)}$$

$$\beta_N = \frac{\mathrm{SumxY}_{N-1} + \frac{X_N - \overline{X}_{N-1}}{N}\left(Y_N\left(N-1\right) - \mathrm{SumY}_{N-1}\right)}{\sigma_{X,N-1} + \frac{N-1}{N}\left(X_i - \overline{X}_{N-1}\right)^2} \qquad \text{(G7)}$$

About the Author

Stuart McCrary

Stuart McCrary is a trader and portfolio manager who specializes in traditional and alternative investments, quantitative valuation, risk management, and financial software. Before joining BRG, he spent 13 years consulting on a wide range of capital markets issues including litigation consulting, valuation, modeling, and risk management. Previously, he was president of Frontier Asset Management, a market-neutral hedge fund. He held positions with Fenchurch Capital Management as senior options trader and CS First Boston as vice president and market maker, where he traded OTC options and mortgage-backed securities. Prior to that, he was a vice president with the Securities Groups and a portfolio manager with Comerica Bank.

Mr. McCrary is the author of several books, including *How to Create and Manage a Hedge Fund: A Professional's Guide* (John Wiley & Sons, 2002).

About Berkeley Research Group

Berkeley Research Group, LLC (www.thinkbrg.com) is a leading global strategic advisory and expert consulting firm that provides independent advice, data analytics, authoritative studies, expert testimony, investigations, and regulatory and dispute consulting to Fortune 500 corporations, financial institutions, government agencies, major law firms, and regulatory bodies around the world. BRG experts and cons ultants combine intellectual rigor with practical, real-world experience and an in-depth understanding of industries and markets. Their expertise spans economics and finance, data analytics and statistics, and public policy in many of the major sectors of our economy, including healthcare, banking, information technology, energy, construction, and real estate. BRG is headquartered in Emeryville, California, with offices across the United States and in Asia, Australia, Canada, Latin America and the United Kingdom.