APPENDIX M: NOTES ON MOMENTS

Every stats textbook covers the properties of the mean and variance in great detail, but the higher moments are often neglected. This is unfortunate, because they are often of important real-world significance. For example, returns from financial stocks are known to have larger-than-Normal skew and kurtosis: Mandelbrot [1963, footnote 3] traces awareness of this non-Normality as far back as 1915; see also, e.g., Blattberg and Gonedes [1974] or Kon [1984]. Models that assume Normal skew and kurtosis will underpredict extreme outcomes.

This technical appendix goes over a few properties and notes to help you better understand the skew and kurtosis.

As in the main textbook, $\mathbf{x}$ is a data vector of size n, and x_i is a representative element. The skew is defined as the third central moment,

$$\mathcal{S}(\mathbf{x}) = \sum_i (x_i - \mu)^3/n,$$

and the kurtosis is defined as the fourth central moment,[1]

$$\mathcal{K}(\mathbf{x}) = \sum_i (x_i - \mu)^4/n.$$

The sample moment is much like the population moment, but the population moment is based on $(x_i - \mu)^m$, where μ is the true mean of the data and m is two for variance, three for skew, or four for kurtosis; while the sample moment is based on $(x_i - \bar{\mathbf{x}})^m$, where $\bar{\mathbf{x}}$ is the sample mean. We can

[0]Although it is not bound with the book itself, this is a technical appendix to *Modeling with Data: Tools and Techniques for Scientific Computing*. It uses the same notation as the book, as on pp 12–13. Should you need to cite this appendix, simply cite the book and give a page number of the form *online appendix M, page* n.

[1]Notice that the definition of skew and kurtosis are not divided by σ^3 or σ^4, or otherwise modified from the definition of a central moment. Such renormalizations make comparison to the standard Normal distribution very easy, but make absolutely every other computation more difficult.

expect that $\bar{\mathbf{x}} \neq \mu$. Also, $\bar{\mathbf{x}}$ has a variance (and skew and kurtosis), while μ is a constant. On the other hand, we can calculate $\bar{\mathbf{x}}$ from the data we have, while μ may be entirely unknowable.

As on page 222, the unbiased sample estimator of the sample variance,

$$\hat{\sigma}^2(\mathbf{x}) = \sum_i (x_i - \bar{\mathbf{x}})^2/(n-1),$$

involves a sum divided by $n-1$, not n. There are similar adjustments to be made to achieve unbiased skew and kurtosis. The derivation is involved, but it gives us a chance to look over some other properties of the central moments.

To find the relationship between the sample moments based on $\bar{\mathbf{x}}$ and the population moments based on μ, we will first find a few intermediate expected values:

1. $\sigma^2(\bar{\mathbf{x}}), \mathcal{S}(\bar{\mathbf{x}}), \mathcal{K}(\bar{\mathbf{x}})$
2. $E(x_i), E(x_i^2), E(x_i^3), E(x_i^4)$
3. $E(\bar{\mathbf{x}}^2), E(\bar{\mathbf{x}}^3), E(\bar{\mathbf{x}}^4)$
4. $E(x_i\bar{\mathbf{x}}^2), E(x_i^2\bar{\mathbf{x}}), E(x_i^3\bar{\mathbf{x}}), \ldots$
5. $E((x_i - \bar{\mathbf{x}})^2), E((x_i - \bar{\mathbf{x}})^3), E((x_i - \bar{\mathbf{x}})^4)$
6. $\hat{\sigma}^2(\mu), \hat{\mathcal{S}}(\mu), \hat{\mathcal{K}}(\mu)$

Series (1) is relatively easy, and the same derivation works for all central moments. Series (2) and (3) are the *noncentral moments*, where the mean is not subtracted; the slight difference between these two series proves to be important later. Series (4) consists of hybrid terms that multiply a power of a representative element by a power of $\bar{\mathbf{x}}$, and require new techniques for estimation. Series (5) is an estimate of the sample moments using the prior results. Series (6) uses series (5) to produce an unbiased estimate of the true population moments, using sample data.

MOMENTS OF THE MEAN

First, we can calculate the variance, skew, and kurtosis of the mean.[2] The steps involved are to expand the expected value to the sum over n that it is (the ellipses indicate additional terms whose expected value is zero), take the $\frac{1}{n}$ out of the moment calculation, and then note that the sum of n iid elements can be broken down to n moments of any representative element (here, x_1). The second moment of a representative element is simply $\sigma^2(x)$, and similarly for the other moments.

[2] Following Casella and Berger [1990, p 208].

$$\begin{aligned}
\sigma^2(\bar{\mathbf{x}}) &= E\left(\frac{1}{n}\sum_i (x_i-\mu)\right)^2 \\
&= \frac{1}{n^2}E\left(\sum_i (x_i-\mu)\right)^2 \\
&= \frac{1}{n^2}E\left(\sum_i (x_i-\mu)^2 + \sum_i\sum_{j\neq i}(x_i-\mu)(x_j-\mu)\right) \\
&= \frac{1}{n^2}E\left(n(x_1-\mu)^2\right) \\
&= \frac{1}{n}\cdot\sigma^2\left(x_1\right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{S}(\bar{\mathbf{x}}) &= E\left(\frac{1}{n}\sum_i (x_i-\mu)\right)^3 \\
&= \frac{1}{n^3}E\left(\sum_i (x_i-\mu)^3 + 3\sum_i\sum_{j\neq i}(x_i-\mu)^2(x_j-\mu) + \dots\right) \\
&= \frac{1}{n^3}E\left(n(x_1-\mu)^3\right) \\
&= \frac{1}{n^2}\cdot\mathcal{S}\left(x_1\right)
\end{aligned}$$

$$\begin{aligned}
\mathcal{K}(\bar{\mathbf{x}}) &= E\left(\frac{1}{n}\sum_i (x_i-\mu)\right)^4 \\
&= \frac{1}{n^4}E\left(\sum_i (x_i-\mu)^4 + 6\sum_i\sum_{j>i}(x_i-\mu)^2(x_j-\mu)^2 + \dots\right) \\
&= \frac{1}{n^4}E\left(n(x_1-\mu)^4 + 6(n(n-1)/2)\sigma^2(x_i)\sigma^2(x_j)\right) \\
&= \frac{1}{n^3}\cdot\left(\mathcal{K}\left(x_1\right) + 3(n-1)\sigma^4(x_1)\right)
\end{aligned}$$

To summarize the above:

$$\boxed{\begin{aligned} \sigma^2(\bar{\mathbf{x}}) &= \tfrac{\sigma^2(x)}{n} \\ \mathcal{S}(\bar{\mathbf{x}}) &= \tfrac{\mathcal{S}(x)}{n^2} \\ \mathcal{K}(\bar{\mathbf{x}}) &= \tfrac{\mathcal{K}(x)+3(n-1)\sigma^4(x)}{n^3} \end{aligned}}$$

It is reasonably intuitive that the mean of several draws will have a smaller variance, skew, or kurtosis than a single draw, and these relations embody that intuition. They will drive much of the story to follow.

NONCENTRAL MOMENTS The expression $E(x_i^2)$ is not quite the variance, which is the expected squared distance to the mean: $E((x_i - \mu)^2)$. The further μ is from zero, the further $E(x_i^2)$ will be from the variance.

The trick to calculating $E(x_i^m)$ is to rewrite it as $E\left((x_i - \mu + \mu)^m\right)$. For the variance:

$$E\left[((x_i - \mu) + \mu)^2\right] \;= E\left[\; (x_i - \mu)^2 + 2(x_i - \mu)\mu + \mu^2 \;\right].$$

The expectation of a sum is the sum of expectations; that is, this expression can be broken down into its components:

$$\begin{aligned} E(x_i^2) &= E\left[(x_i - \mu)^2\right] & +E\left[2(x_i - \mu)\mu\right] & +E\left[\mu^2\right] \\ &= \sigma^2(\mathbf{x}) & +0 & +\mu^2 \end{aligned}$$

This can be rewritten to produce the familiar statement that $\sigma^2(\mathbf{x}) = E(\mathbf{x}^2) - E^2(\mathbf{x})$.

We can repeat the same steps for the skew and kurtosis: write $E(x_i^m)$ as $E\left((x_i - \mu + \mu)^m\right)$, expand, and find the expectation of each individual term. For the skew:

$$\begin{aligned} E\left[((x_i - \mu) + \mu)^3\right] &= E\left[(x_i - \mu)^3\right] & +E\left[3(x_i - \mu)^2\mu\right] & +E\left[3(x_i - \mu)\mu^2\right] & +E[\mu^3] \\ &= \mathcal{S}(\mathbf{x}) & +3\sigma^2(\mathbf{x})\mu & +0 & +\mu^3. \end{aligned}$$

For kurtosis:

$$\begin{aligned} E\left[x^4\right] &= E\left[(x_i - \mu)^4\right] & +E\left[4(x_i - \mu)^3\mu\right] & +E\left[6(x_i - \mu)^2\mu^2\right] & +E\left[4(x_i - \mu)\mu^3\right] & +E[\mu^4]. \\ &= \mathcal{K}(\mathbf{x}) & +4\mathcal{S}(\mathbf{x})\mu & +6\sigma^2(\mathbf{x})\mu^2 & +0 & +\mu^4. \end{aligned}$$

To summarize:

$$\boxed{\begin{aligned} E\left(\mathbf{x}^2\right) &= \sigma^2(\mathbf{x}) + \mu^2 \\ E\left(\mathbf{x}^3\right) &= \mathcal{S}(\mathbf{x}) + 3\sigma^2(\mathbf{x})\mu + \mu^3 \\ E\left(\mathbf{x}^4\right) &= \mathcal{K}(\mathbf{x}) + 4\mathcal{S}(\mathbf{x})\mu + 6\sigma^2(\mathbf{x})\mu^2 + \mu^4 \end{aligned}}$$

So the noncentral variance depends on the mean, the noncentral skew depends on the variance and mean, and the noncentral kurtosis depends on the skew, variance, and mean. In all cases, the relationship is positive; for example, a larger variance means a larger noncentral skew and kurtosis.

NONCENTRAL MOMENTS FOR $\bar{\mathbf{x}}$ We can reuse much of the above calculation regarding $E[x_i]$ for $E[\bar{\mathbf{x}}]$. The results from the series of central moments of the mean already come in handy, as we can write the variance of $\bar{\mathbf{x}}$ as $\sigma^2(\mathbf{x})/n$, for example. Expanding $E(\bar{\mathbf{x}}^2)$, we get:

$$\begin{array}{rclll} E\left[((\bar{\mathbf{x}}-\mu)+\mu)^2\right] & = & E\left[(\bar{\mathbf{x}}-\mu)^2\right] & +E\left[2(\bar{\mathbf{x}}-\mu)\mu\right] & +E\left[\mu^2\right] \\ & = & \sigma^2(\mathbf{x})/n & +0 & +\mu^2. \end{array}$$

For the skew:

$$\begin{array}{rllll} E\left[((\bar{\mathbf{x}}-\mu)+\mu)^3\right] & = E\left[(\bar{\mathbf{x}}-\mu)^3\right] & +E\left[3(\bar{\mathbf{x}}-\mu)^2\mu\right] & +E\left[3(\bar{\mathbf{x}}-\mu)\mu^2\right] & +E\left[\mu^3\right] \\ & = \mathcal{S}(\mathbf{x})/n^2 & +3\sigma^2(\mathbf{x})\mu/n & +0 & +\mu^3. \end{array}$$

For kurtosis:

$$\begin{array}{rlllll} E\left[x^4\right] & = E\left[(\bar{\mathbf{x}}-\mu)^4\right] & +E\left[4(\bar{\mathbf{x}}-\mu)^3\mu\right] & +E\left[6(\bar{\mathbf{x}}-\mu)^2\mu^2\right] & +E\left[4(\bar{\mathbf{x}}-\mu)\mu^3\right] & +E[\mu^4]. \\ & = \frac{\mathcal{K}(\mathbf{x})+3(n-1)\sigma^4(\mathbf{x})}{n^3} & +4\mathcal{S}(\mathbf{x})\mu/n^2 & +6\sigma^2(\mathbf{x})\mu^2/n & +0 & +\mu^4. \end{array}$$

$$\boxed{\begin{aligned} E\left(\bar{\mathbf{x}}^2\right) &= \sigma^2(\mathbf{x})/n + \mu^2 \\ E\left(\bar{\mathbf{x}}^3\right) &= \mathcal{S}(\mathbf{x})/n^2 + 3\sigma^2(\mathbf{x})\mu/n + \mu^3 \\ E\left(\bar{\mathbf{x}}^4\right) &= \mathcal{K}(\mathbf{x})/n^3 + 3(n-1)\sigma^4(\mathbf{x})/n^3 + 4\mathcal{S}(\mathbf{x})\mu/n^2 + 6\sigma^2(\mathbf{x})\mu^2/n + \mu^4 \end{aligned}}$$

You can see that this series is analogous to the last, save for the divisions by powers of n. As $n \to \infty$, the variations among the data points become less important, and the final value of $E(\bar{\mathbf{x}}^m)$ is simply μ^m. That is, these are asymptotically unbiased estimates of the powers of μ.

HYBRIDS So far, we have considered only $\bar{\mathbf{x}}$ or a single element x_i, but what about combinations, like $E(x_1\bar{\mathbf{x}})$? [As above, use x_1 or x_2 as representative elements.] $\bar{\mathbf{x}}$ is the sum of all of the elements of the data set, but elements of the form $x_1 \cdot x_1$ behave differently from elements of the form $x_1 \cdot x_2$, so we need to break down the sum. For $E(x_1\bar{\mathbf{x}})$:

$$\begin{aligned}
E[x_1\bar{\mathbf{x}}] &= E[x_1 \textstyle\sum_{i=1}^n x_i/n] \\
&= E[x_1^2/n \qquad + x_1 \textstyle\sum_{i=2}^n x_i/n] \\
&= E[x_1^2]/n \qquad + (n-1)E[x_1x_2] \\
&= \sigma^2(\mathbf{x})/n + \mu^2/n \quad + (n-1)\mu^2/n \\
&= \sigma^2(\mathbf{x})/n + \mu^2 \\
&= E(\bar{\mathbf{x}}^2)
\end{aligned}$$

In this case, things worked out well—in fact, $E(x_1\bar{\mathbf{x}}^m) = E(\bar{\mathbf{x}}^{m+1})$ for any m. But other cases are not so elegant. For example, $E(x_1^2\bar{\mathbf{x}})$ is a new form:

$$\begin{aligned}
E[x_1^2\bar{\mathbf{x}}] &= E[x_1^2 \textstyle\sum_{i=1}^n x_i/n] \\
&= E[x_1^3/n \qquad + x_1^2 \textstyle\sum_{i=2}^n x_i/n] \\
&= E[x_1^3]/n \qquad + (n-1)E[x_1^2x_2] \\
&= (\mathcal{S}(\mathbf{x}) + 3\sigma^2(\mathbf{x})\mu + \mu^3)/n \quad + (n-1)(\sigma^2(\mathbf{x}) + \mu^2)\mu/n \\
&= \mathcal{S}(\mathbf{x})/n + \tfrac{n+2}{n}\sigma^2(\mathbf{x})\mu + \mu^3
\end{aligned}$$

Here is a catalog of some of the forms that will be used below. The especially tedious derivations have not been shown, but take the same form as the derivations above, in expanding the mean or square-of-mean and counting instances of the form $x_1^2x_2$, $x_1^2x_2^2$, et cetera.

$$\begin{aligned}
E[x_1\bar{\mathbf{x}}] &= \sigma^2(\mathbf{x})/n + \mu^2 = E(\bar{\mathbf{x}}^2) \\
E[x_1^2\bar{\mathbf{x}}] &= \mathcal{S}(\mathbf{x})/n + \tfrac{n+2}{n}\sigma^2(\mathbf{x})\mu + \mu^3 \\
E[x_1\bar{\mathbf{x}}^2] &= \mathcal{S}(\mathbf{x})/n^2 + \tfrac{3}{n}\sigma^2(\mathbf{x})\mu + \mu^3 = E(\bar{\mathbf{x}}^3) \\
E[x_1^3\bar{\mathbf{x}}] &= \mathcal{K}(\mathbf{x})/n + \tfrac{n+3}{n}\mathcal{S}(\mathbf{x})\mu + \tfrac{3n+3}{n}\sigma^2(\mathbf{x})\mu^2 + \mu^4 \\
E[x_1\bar{\mathbf{x}}^3] &= \mathcal{K}(\mathbf{x})/n^3 + 3(n-1)\sigma^4(\mathbf{x})/n^3 + 4\mathcal{S}(\mathbf{x})\mu/n^2 + 6\sigma^2(\mathbf{x})\mu^2/n + \mu^4 \quad = E(\bar{\mathbf{x}}^4) \\
E[x_1^2\bar{\mathbf{x}}^2] &= \mathcal{K}(\mathbf{x})/n^2 + \tfrac{2(n+1)}{n^2}\mathcal{S}(\mathbf{x})\mu + \tfrac{n+5}{n}\sigma^2(\mathbf{x})\mu^2 + \tfrac{n-1}{n^2}\sigma^4 + \mu^4
\end{aligned}$$

THE SAMPLE CENTRAL MOMENTS We now have the components needed to find the sample central moments. A *sample central moment* is centered not around μ, where it would have a form like $\sum(x-\mu)^2/n$, but is centered around $\bar{\mathrm{x}}$, like $\sum(x-\bar{\mathbf{x}})^2/n$.

In all three cases, the sample central moment based on $\bar{\mathrm{x}}$ and the actual central moment based on μ are not equal in expectation. The sample mean $\bar{\mathbf{x}}$ itself has some variance, so the final sample variance will be a combination of the true population variance and the variance of the mean; similarly for skew and kurtosis.

Let the sample central moments based on $\bar{\mathbf{x}}$ be $\sigma^2_{\bar{\mathbf{x}}}(\mathbf{x})$, $\mathcal{S}_{\bar{\mathbf{x}}}(\mathbf{x})$, and $\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})$. The procedure for calculating this sample central moment is exactly as above: expand the square, cube, or quadratic, then take the expectation of each individual term.

$$\begin{aligned} E\left[(x_i - \bar{\mathbf{x}})^2\right] &= E[\bar{\mathbf{x}} - 2x_i\bar{\mathbf{x}} + \bar{\mathbf{x}}^2] \\ &= E[x_i^2 - \bar{\mathbf{x}}^2] \\ &= E\left[x_i^2\right] - E\left[\bar{\mathbf{x}}^2\right] \\ &= \sigma^2(\mathbf{x}) + \mu^2 - \sigma^2(\mathbf{x})/n - \mu^2 \\ &= \tfrac{n-1}{n}\sigma^2(\mathbf{x}) \end{aligned}$$

In the last step, the μ terms easily cancel out, but the remaining expected value of $\sigma^2(x)$ differs from the remaining expected value of $\bar{\mathbf{x}}^2$ by that fraction of n, and that means that there is the inconvenient $(n-1)/n$ term when they merge.

We can rewrite the final form to reveal a simple relationship among the various types of variance. Because $\sigma^2(\bar{\mathbf{x}}) = \sigma^2(\mathbf{x})/n$, this result becomes:

$$\sigma^2_{\bar{\mathbf{x}}}(\mathbf{x}) = \sigma^2(\mathbf{x}) - \sigma^2(\bar{\mathbf{x}}).$$

Now repeat the process for the skew. Here we begin to use the hybrid expressions above, such as that $E[x_i\bar{\mathbf{x}}^2] = E[\bar{\mathbf{x}}^3]$.

$$\begin{aligned} E\left[(x_i - \bar{\mathbf{x}})^3\right] &= E[x_i^3 - 3x_i^2\bar{\mathbf{x}} + 3x_i\bar{\mathbf{x}}^2 - \bar{\mathbf{x}}^3] \\ &= E[x_i^3] - 3E[x_i^2\bar{\mathbf{x}}] + 2E[\bar{\mathbf{x}}^3] \end{aligned}$$

The expansion for each term in the equation is given at some point above, and each consists of a sum of $\mathcal{S}$, $\mu\sigma^2$, and μ^3. To keep things organized, each row in the table below is the expansion of one of the three terms in the above expression, and each column gives the coefficient on $\mathcal{S}$, $\mu\sigma^2$, or μ^3 for the given expression. For example, the first row can be read as $E(x_i^3) = 1\mathcal{S} + 3\mu\sigma^2 + 1\mu^3$.

term	$\mathcal{S}$	$\mu\sigma^2$	μ^3
$E(x_i^3)$	1	3	1
$-3E(x_i^2\bar{\mathbf{x}})$	$-3/n$	$-3\frac{n+2}{n}$	-3
$+2E(\bar{\mathbf{x}}^3)$	$2/n^2$	$6/n$	2

With terms neatly in columns, you can quickly verify that the coefficients on $\mu\sigma^2$ and on μ^3 sum to zero. We are left with:

$$\mathcal{S}_{\bar{\mathbf{x}}}(\mathbf{x}) = \frac{(n-2)(n-1)}{n^2}\mathcal{S}(\mathbf{x})$$

Like the variance, this can be expressed in terms of $\mathcal{S}(\mathbf{x})$ and $\mathcal{S}(\bar{\mathbf{x}})$:

$$\mathcal{S}_{\bar{\mathbf{x}}}(\mathbf{x}) = \mathcal{S}(\mathbf{x}) - (3n-2)\mathcal{S}(\bar{\mathbf{x}}).$$

Tragically, any simple pattern established by the variance and skew does not carry forward, and the sample kurtosis is a mess. Expanding the quadratic gives us several terms to evaluate:

$$E\left[(x_i - \bar{\mathbf{x}})^4\right] \;=\; E\left[x_i^4\right] - E\left[4x_i^3\bar{\mathbf{x}}\right] + E\left[6x_i^2\bar{\mathbf{x}}^2\right] - E\left[4x_i\bar{\mathbf{x}}^3\right] + E\left[\bar{\mathbf{x}}^4\right] \; .$$

Again organizing the expansion of the expectations into a table of coefficients:

term	$\mathcal{K}$	$\mathcal{S}\mu$	$\mu^2\sigma^2$	σ^4	μ^4
$E(x_i^4)$	1	4	6		1
$-4E(x_i^3\bar{\mathbf{x}})$	$-4/n$	$-4(n+3)/n$	$-12(n+1)/n$		-4
$+6E(x_i^2\bar{\mathbf{x}}^2)$	$6/n^2$	$12(n+1)/n^2$	$6(n+5)/n$	$6(n-1)/n^2$	6
$-4E(x_i\bar{\mathbf{x}}^3) + E(\bar{\mathbf{x}}^4)$	$-3/n^3$	$-12/n^2$	$-18/n$	$-9(n-1)/n^3$	-3

You can again quickly verify that the coefficients in the $\mathcal{S}\mu$, $\mu^2\sigma^2$, and μ^4 columns sum to zero. The remaining columns give us two rather unpleasant terms in the final expected value of the sample kurtosis.

$$\begin{aligned} E\left[(x_i - \bar{\mathbf{x}})^4\right] &= \tfrac{(n-1)(n^2-3n+3)}{n^3}\mathcal{K}(\mathbf{x}) + \tfrac{(n-1)(6n-9)}{n^3}\sigma^4(\mathbf{x}) \\ &= \tfrac{(n-1)}{n^3}\left((n^2-3n+3)\mathcal{K}(\mathbf{x}) + (6n-9)\sigma^4(\mathbf{x})\right) \end{aligned}$$

Remarkably, this mess can also be expressed as a weighted sum of $\mathcal{K}(\mathbf{x})$ and $\mathcal{K}(\bar{\mathbf{x}})$:

$$\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x}) = \frac{(n-2)^2}{n^2}\mathcal{K}(\mathbf{x}) + (2n-3)\mathcal{K}(\bar{\mathbf{x}}).$$

$$\begin{aligned} \sigma^2_{\bar{\mathbf{x}}}(\mathbf{x}) &= \tfrac{n-1}{n}\sigma^2(\mathbf{x}) \\ &= \sigma^2(\mathbf{x}) - \sigma^2(\bar{\mathbf{x}}) \\[2em] \mathcal{S}_{\bar{\mathbf{x}}}(\mathbf{x}) &= \tfrac{(n-2)(n-1)}{n^2}\mathcal{S}(\mathbf{x}) \\ &= \mathcal{S}(\mathbf{x}) - (3n-2)\mathcal{S}(\bar{\mathbf{x}}) \\[2em] \mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x}) &= \tfrac{(n-1)}{n^3}\left((n^2-3n+3)\mathcal{K}(\mathbf{x}) + (6n-9)\sigma^4(\mathbf{x})\right) \\ &= \tfrac{(n-2)(n-2)}{n^2}\mathcal{K}(\mathbf{x}) + (2n-3)\mathcal{K}(\bar{\mathbf{x}}) \end{aligned}$$

Sample estimates of the true moments

We are interested in the true variance, based on μ, but μ is typically not knowable from finite data, so we must make do with estimates based on statistics of the data, like $\bar{\mathbf{x}}$. We just saw that $\sigma^2_{\bar{\mathbf{x}}}(\mathbf{x}), \mathcal{S}_{\bar{\mathbf{x}}}(\mathbf{x})$, and $\mathcal{K}_{\bar{\mathbf{x}}}(\mathbf{x})$ are biased estimators of their respective true moments, so they won't work without modification. Fortunately, we know exactly how large the bias is, and so can produce an unbiased estimate of the the true variance, skew, or kurtosis by simply moving the true moment to the left-hand side of above equations.

To make the computations clearer (and the concepts a little more opaque), I present the estimates of the central moments fully written out. Because we are moving from the expected value of a potentially incalculable value to an operational estimate, we put a hat on the central moment to indicate that this is a sample-based estimator.

For the variance and skew, we can easily produce an estimate that does not depend on knowledge of the true value of μ, but uses only information from the data set at hand. For the kurtosis, the unfortunate σ^4 term needs to be replaced with an unbiased estimate, so plug in the square of $\hat{\sigma}^2(\mathbf{x})$.

$$\boxed{\begin{aligned}
\hat{\sigma}^2(\mathbf{x}) &= \tfrac{1}{n-1} \textstyle\sum_i (x_i - \bar{\mathbf{x}})^2 \\
\hat{\mathcal{S}}(\mathbf{x}) &= \tfrac{n}{(n-1)(n-2)} \textstyle\sum_i (x_i - \bar{\mathbf{x}})^3 \\
\hat{\mathcal{K}}(\mathbf{x}) &= \tfrac{n^2}{(n-1)(n^2-3n+3)} \textstyle\sum_i (x_i - \bar{\mathbf{x}})^4 - \tfrac{6n-9}{n^2(n^2-3n+3)} \left(\sum_i (x_i - \bar{\mathbf{x}})^2\right)^2
\end{aligned}}$$

The sample estimate of the population variance should be familiar to you from textbooks, using the definition of variance but with $\bar{\mathbf{x}}$ replacing μ and dividing the sum by $n-1$ instead of n. You can see that it derives rather simply from the fact that the expected value of the $\bar{\mathbf{x}}$-based variance and the μ-based variance differ by $\sigma^2(\bar{\mathbf{x}}) = \sigma^2(\mathbf{x})/n$; from there it is simple algebra to get the common form.

Especially among ANOVA users, the variance formula is often explained using a degrees-of-freedom story:[3] given the average $\bar{\mathbf{x}}$, there are only $n-1$ free elements in the data, because the last can be deterministically calculated from the first $n-1$ and $\bar{\mathbf{x}}$. That story is a useful fiction, in that it makes a great deal of sense in the ANOVA context, and saves the trouble of the algebra above. However, if it were really just a story of counting degrees of freedom, then we are hard-pressed to explain the forms for the unbiased estimate of the sample skew or kurtosis. Fortunately, researchers make use only of analysis of variance and not analysis of skew or kurtosis (ANOSK and ANOKU?), and so the useful fiction of the degrees-of-freedom story is consistent and a help.

[3] See, e.g., Snedecor and Cochran [1976, p 45].

References

Robert C Blattberg and Nicholas J Gonedes. A comparison of the stable and student distributions as statistical models for stock prices. *The Journal of Business*, 47(2):244–280, 1974.

George Casella and Roger L Berger. *Statistical Inference*. Duxbury Press, 1990.

Stanley J Kon. Models of stock returns–a comparison. *The Journal of Finance*, 39(1):147–165, 1984.

Benoit Mandelbrot. The variation of certain speculative prices. *The Journal of Business*, 36(4): 394–419, 1963.

George W Snedecor and Willian G Cochran. *Statistical Methods*. Iowa State University Press, 6th edition, 1976.