APPENDIX 3.3: CALCULATING SKEW AND KURTOSIS

We mentioned in Chapter 1 that parameters are simply numbers that characterize the scores of populations. The mean (μ) and variance (σ^2) are two of the most important parameters associated with statistical analysis in psychology. In subsequent chapters, we will use statistics to estimate these important parameters. Although μ and σ^2 will be our primary concern, skew and kurtosis are parameters in the same way as μ and σ^2. Skew and kurtosis can be computed from the scores in populations in a manner very similar to the computation of the mean and variance. Therefore, we will take a moment to comment on how skew and kurtosis are computed in populations and samples.

Moments

At the beginning of Chapter 3, we noted that the mean and variance had very similar definitions. The mean of a population is the sum of all scores divided by N. Statisticians sometimes call this the *first raw moment* of the distribution. The variance in a population is the sum of squared deviations from the mean, divided by N. Statisticians call this the *second central moment* of a distribution. Central moments are computed by subtracting the mean from all scores and then raising these deviation scores to some power. (Raw moments do not subtract the mean from all scores.) The following expressions define the second, third, and forth central moments:

$$\theta_2 = \frac{\sum(y-\mu)^2}{N}$$

$$\theta_3 = \frac{\sum(y-\mu)^3}{N}$$

$$\theta_4 = \frac{\sum(y-\mu)^4}{N}.$$

The symbol θ is pronounced theta. Therefore, θ_1, θ_2, θ_3, and θ_4 represent the first, second, third, and fourth central moments of a distribution, respectively. All are moments computed in the same way, and they differ only in the power to which the differences between y and μ are raised. You may not be familiar with exponents other than 2 (squaring), but they are really not complicated, as shown in the following examples:

$$y^1 = y$$

$$y^2 = y * y$$

$$y^3 = y * y * y$$

$$y^4 = y * y * y * y.$$

So, the exponents simply tell us how many times to multiply a number by itself.

Skew and Kurtosis

The second, third, and fourth central moments are related to skew and kurtosis. In a population, skew is defined as

$$\text{skew} = \frac{\theta_3}{\sigma^3} \qquad (3.A3.1)$$

where σ is the population standard deviation (or the square root of the second central moment, θ_2). Skew, as defined in equation 3.A3.1, can take on positive and negative values. Symmetrical distributions (such as in Figures 3.3a and 3.4a) have zero skew. Distributions that are skewed to the right yield positive skew values, and those that are skewed to the left yield negative skew values. The right-skewed distribution in Figure 3.3 has a skew of about 3.6, and the left-skewed distribution has a skew of about −3.6.

Kurtosis in a population is defined as

$$\text{kurtosis} = \frac{\theta_4}{\sigma^4}. \qquad (3.A3.2)$$

Kurtosis, as defined in equation 3.A3.2, can take on only positive values. The larger the values, the more leptokurtic the distribution. However, when statisticians talk about kurtosis, they often mean *excess kurtosis*. A normal distribution has a kurtosis of 3, when defined using equation 3.A3.2. Statisticians define 3 as *normal kurtosis*. Excess kurtosis is the difference between kurtosis and normal kurtosis. Therefore, excess kurtosis is defined as follows:

$$\text{excess kurtosis} = \frac{\theta_4}{\sigma^4} - 3. \qquad (3.A3.3)$$

A normal distribution has an excess kurtosis of 0. The leptokurtic distribution in Figure 3.4b has an excess kurtosis of 3, and the platykurtic distribution in Figure 3.4c has an excess kurtosis of −1. The flattest possible distribution (most platykurtic) is the *uniform*, or *rectangular*, *distribution*. A uniform distribution is one for which the densities are equal for all possible numbers between some minimum and maximum. For example, a uniform distribution may have **min** = 0 and **max** = 1, and all values between **min** and **max** are equally probable. Uniform distributions have excess kurtosis of −1.2.

Estimating Skew and Kurtosis

The definitions of skew and excess kurtosis in equations 3.A3.1 and 3.A3.3 are *parameters*. These formulas should not be applied to samples to estimate the population parameters. Just as the definition of the sample variance differs from that of the population variance, the definitions of the sample skew and sample kurtosis differ from those of the population skew and kurtosis. In all cases, the differences in the formulas have to do with making the statistics good estimators of the parameters. Once again, I promise this will be explained in Chapter 5.

The formula for skew for a sample is

$$\text{skew} = \frac{\hat{\theta}_3}{s^3} * \frac{n^2}{(n-1)(n-2)}, \tag{3.A3.4}$$

where n is sample size, s is the sample standard deviation, and $\hat{\theta}_3$ is computed exactly like θ_3 but from the scores in a sample rather than scores in a population; i.e., $\hat{\theta}_3 = \Sigma(y - m)^3/n$. Equation 3.A3.4 looks horrible. (If I'd seen something like this in my first statistics course, I would have had an anxiety attack. Sorry.) Notice, however, that the black term in equation 3.A3.4 looks exactly like equation 3.A3.1, except that $\hat{\theta}_3$ and s are computed from scores in the sample. That's not so bad. The blue term is called a correction factor, which we already discussed when talking about the sample variance. If you play around with the correction factor, you'll notice that it gets closer to 1 as n (sample size) gets larger, because n−1 and n−2 get closer and closer to n, making $(n-1)(n-2)$ closer and closer to n^2. This means that $\hat{\theta}_3/s^3$ needs less correction as sample size increases. In statistics, we like large samples!

The formula for excess kurtosis for a sample is

$$\text{excess kurtosis} = \frac{\hat{\theta}_4}{s^4} * \frac{n^2(n+1)}{(n-1)(n-2)(n-3)} - 3 * \frac{(n-1)^2}{(n-2)(n-3)}, \tag{3.A3.5}$$

where n is sample size, s is the sample standard deviation, and $\hat{\theta}_4$ is computed exactly like θ_4 but from the scores in a sample rather than scores in a population; i.e., $\hat{\theta}_4 = \Sigma(y - m)^4/n$. If equation 3.A3.4 looks horrible, then equation 3.A3.5 looks positively ghastly. But notice again that the black terms in equation 3.A3.5 look exactly like equation 3.A3.3, except that θ_4 and s are computed from scores in the sample. The correction factors in blue behave the same way as the correction factor in the sample variance and the sample skew. As sample size increases, the correction factors get closer to 1.

We are rarely interested in estimating skew and kurtosis for their own sake. Rather, we use skew and kurtosis computed from a sample to assess the normality of the population from which the sample was drawn. When a sample is drawn from a normal population, then skew and excess kurtosis should be close to 0. Large departures from 0 (positive or negative) would suggest that our sample was not drawn from a normal population. Just how large a departure from 0 would be cause for concern is something we will discuss in later chapters.

LEARNING CHECK 1

1. If y = {2, 2, 5, 5, 5, 11} is a small population of scores, calculate the following: (a) μ, (b) σ, (c) θ_2, (d) θ_3, (e) θ_4, (f) skew, (g) kurtosis, and (h) excess kurtosis.

Answers

1. (a) $\mu = 5$. (b) $\sigma = 3$. (c) $\theta_2 = 9$. (d) $\theta_3 = 27$. (e) $\theta_4 = 243$. (f) Skew = θ_3/σ^3 = 27/27 = 1. (g) Kurtosis = θ_4/σ^4 = 243/81 = 3. (h) Excess kurtosis = kurtosis − 3 = 3 − 3 = 0.

APPENDIX 3.4: THE IMPORTANCE OF THE MEAN AND STANDARD DEVIATION

In Chapter 2, we saw that probability distributions convey the relative standing of scores within a sample or population. That is, knowing how scores are distributed allows us to determine the proportion of the distribution at or below any score of interest. We can say that a score is extreme or unusual if there are very few scores above it (thus an extremely high score) or very few scores below it (thus an extremely low score).

In Chapter 3, we saw that the mean is an important measure of central tendency for a distribution, and the standard deviation is an important measure of dispersion. There is an important connection between the mean and standard deviation of a distribution, and the relative standing of scores in a distribution.

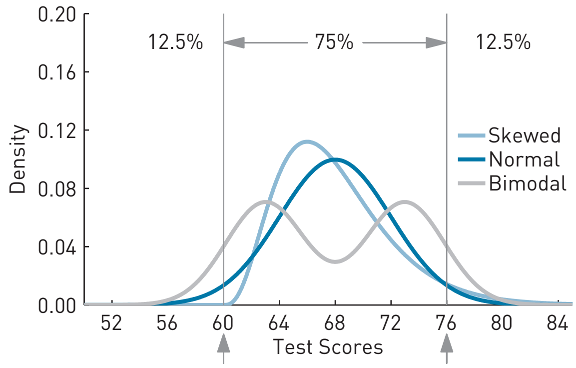
Looking back at the distributions in Figures 3.3, 3.4, and 3.5, you might notice that scores tend to fall close to the mean, on average. In fact, it is unusual for a score to fall a long way from the mean. In later chapters, we will use the standard deviation of a distribution as our measure of distance, and we will talk about how many standard deviations a score is from the mean of its distribution. In Chapter 4, we will make an important connection between how many standard deviations a score is from the mean of its distribution and the proportion of scores in the distribution above or below it.

Here is an interesting preview of where we are going. The Russian mathematician Pafnuty Chebyshev (1821–1894) proved that *at least* 75% of a distribution of scores falls within two standard deviations of the mean. That is, *at least* 75% of a distribution of scores falls in the interval $\mu - 2(\sigma)$ to $\mu + 2(\sigma)$. We can express this interval more compactly as $\mu \pm 2(\sigma)$. Some distributions will have more than 75% of their scores within two standard deviations of the mean, but no distribution will have less than 75% of scores within two standard deviations of the mean.

This point is illustrated in Figure 3.A4.1. Three distributions (skewed, normal, and bimodal) are shown. All three distributions have $\mu = 68$ and $\sigma = 4$. The two upward pointing arrows on the *x*-axis show $\mu \pm 2(\sigma)$, which means 68 ± 8, or 60 and 76. Thus, Chebyshev's theorem tells us that if we knew only that a distribution had $\mu = 68$ and $\sigma = 4$ but we knew nothing about the shape of the distribution (it could have been one of these three or infinitely many others), we would know that at least 75% of scores would fall between 60 and 76.

In Figure 3.A4.1, the skewed and normal distributions have much more than 75% of scores falling in the interval $\mu \pm 2(\sigma)$. There are many other distributions that also have more than 75% of scores falling in the interval $\mu \pm 2(\sigma)$. But even if we know nothing about the shape of a distribution, we can say that *at least* 75% of scores fall in the interval $\mu \pm 2(\sigma)$.

FIGURE 3.A4.1 ■ Chebyshev's Theorem



An illustration of the meaning of Chebyshev's theorem for the case of $k = 2$. The theorem says that no matter what shape the distribution has, *at least* 75% lies in the interval $\mu \pm 2(\sigma)$, which means that *at most* 12.5% of it lies above $\mu + 2(\sigma)$, and *at most* 12.5% lies below $\mu - 2(\sigma)$.

In fact, Chebyshev's result is more general than this because it states the minimum proportion of a distribution falling within *any* number (k) of standard deviations of the mean, as long as $k > 1$. That is, Chebyshev's theorem allows us to calculate the minimum proportion of a distribution of scores falling in the interval $\mu \pm k(\sigma)$, where k is the number of standard deviations from the mean. This proportion is given by the following, very simple expression:

$$1 - \frac{1}{k^2}. \tag{3.A4.1}$$

For our example with $k = 2$, we can see how we obtained 75%. By plugging numbers into this expression, we find

$$1 - \frac{1}{k^2} = 1 - \frac{1}{2^2} = 1 - \frac{1}{4} = 1 - .25 = .75.$$

So, the proportion of the distribution in the interval $\mu \pm 2(\sigma)$ is .75, which is 75%.

Equation 3.A4.1 shows us the *minimum* proportion of a distribution *within* the interval $\mu \pm k(\sigma)$. This means that the *maximum* proportion of the distribution *outside* the interval is simply

$$\frac{1}{k^2}. \tag{3.A4.2}$$

For our example with $k = 2$, we find that the maximum proportion of the distribution outside the interval $\mu \pm k(\sigma)$ is

$$\frac{1}{k^2} = \frac{1}{2^2} = \frac{1}{4} = .25,$$

or 25%. Figure 3.A4.1 shows that half of this percentage (12.5%) is above $\mu + 2(\sigma)$ and the other half is below $\mu - 2(\sigma)$. This means that if a score is 2 standard deviations *above* the mean, then at most 12.5% of scores are *higher* than it. If a score is 2 standard deviations *below* the mean, then at most 12.5% of scores are *lower* than it.

More generally, if a score is k standard deviations *above* the mean, then the maximum proportion of the distribution *above* it is

$$0.5\left(\frac{1}{k^2}\right) = \frac{0.5}{k^2}. \tag{3.A4.3}$$

If a score is k standard deviations *below* the mean, then the maximum proportion of the distribution *below* it is also $0.5/k^2$. For $k = 2$, we can see that

$$\frac{0.5}{k^2} = \frac{0.5}{2^2} = \frac{0.5}{4} = .125$$

or 12.5%, as shown in Figure 3.A4.1.

Chebyshev's theorem tells us about the *minimum* proportion of a distribution *within* the interval $\mu \pm k(\sigma)$ and the *maximum* proportion of a distribution *outside* the interval $\mu \pm k(\sigma)$. Once again, Figure 3.A4.1 shows that for many distributions, the proportion outside the interval $\mu \pm k(\sigma)$ will be much less than $1/k^2$.

The value of this theorem is that it gives us some guidance about what constitutes extreme or unusual scores when we know only the mean and standard deviation of a distribution. However, if we do know something about the shape of the distribution, then our judgments about extreme or unusual scores can be much more precise. In Chapter 4, we will see exactly this. If we happen to know that scores were drawn from a normal distribution, we can make far more accurate statements about scores falling k standard deviations from the mean. This will have huge practical consequences that will carry through the rest of this book.

LEARNING CHECK 1

Let's say we know the mean and standard deviation of a distribution but not its shape. Answer the following questions and round your answers to two decimal places.

1. What is the minimum proportion of the distribution falling in the intervals (a) $\mu \pm 2(\sigma)$, (b) $\mu \pm 3(\sigma)$, (c) $\mu \pm 1.25(\sigma)$, and (d) $\mu \pm 1.4142(\sigma)$?
2. What is the maximum proportion of the distribution falling (a) above $\mu + 1.25(\sigma)$ and (b) below $\mu - 1.25(\sigma)$?
3. What is the maximum proportion of the distribution falling (a) above $\mu + 1.4142(\sigma)$ and (b) below $\mu - 1.4142(\sigma)$?

Answers

1. (a) $1 - 1/2^2 = .75$. (b) $1 - 1/3^2 = .89$. (c) $1 - 1/1.25^2 = .36$. (d) $1 - 1/1.4142^2 = .50$.
2. (a) $0.5/1.25^2 = .32$. (b) $0.5/1.25^2 = .32$.
3. (a) $0.5/1.4142^2 = .25$. (b) $0.5/1.4142^2 = .25$.