

2

Variation in Nature

No one supposes that all the individuals of the same species are cast in the very same mould. These individual differences are highly important for us, as they afford material for natural selection to accumulate, in the same manner as man can accumulate in any given direction individual differences in his domesticated productions.

Charles Darwin, Origin of Species, 1859, p. 45

Charles Darwin started the *Origin of Species* with two chapters on variation, the first on variation under domestication, and the second on variation in nature (Darwin, 1859). He followed this with publication of a full two-volume treatise on the *Variation of Animals and Plants under Domestication* (Darwin, 1868). The first volume of *Variation* is a descriptive species-by-species survey of domesticated animals and plants, but the second volume is more thematic in dealing with inheritance, cross-breeding, selection in domestication, variation, and finally Darwin's theory of pangenesis as a mechanism of inheritance.

Darwin (1868; 1875) summarized the subject of variation in terms of laws, but given what he knew concluded:

When we reflect on the several foregoing laws, imperfectly as we understand them, and when we bear in mind how much remains to be discovered, we need not be surprised at the intricate and to us unintelligible manner in which our domestic productions have varied, and still go on varying.

Darwin, 1875, v. 2, p. 348

Darwin was concerned with the origin of variants and varieties as a step toward the origin of species. He emphasized that individual differences are essential for natural selection but never made much progress finding patterns in the variation that he studied so carefully.

2.1 Quantification of Human Variation

Some of Darwin's contemporaries were more gifted quantitatively. These included the Belgian polymath Adolphe Quetelet, the American statistician Ezekiel Elliott, and Darwin's cousin Francis Galton. All three focused their studies on humans, a large species living in large populations that are both accessible and easy to measure.

2.1.1 Adolphe Quetelet

Adolphe Quetelet (1846) combined measurements of chest circumference for a number of local militias in Scotland, taken from the *Edinburgh Medical and Surgical Journal* for April 1817. The journal reported observations for 5,758 men, measured to the nearest inch (imperial inch, 25.4 mm), and Quetelet grouped all of the measurements by inch. He then constructed a table showing the number (frequency) of men in each group, and calculated the weighted mean of the measurements. The mean value for chest circumference based on Quetelet's table is 39.84 inches (1.012 m), with a probable error or median deviation of 1.381 inches (35.08 mm). The mean falls within the most frequent or modal group, and observations differing from the mean decrease in number as their difference increases (Figure 2.1a). In Quetelet's time, "probable error" was the value used to characterize the dispersion of observations about the mean value. The probable error is equal to 0.6745 times the standard deviation (see below).

Quetelet also studied measurements of stature for 100,000 military conscripts from France compiled by Antoine-Audet Hargenvilliers in 1817 and published by Louis Villermé (1829). Here the original measurements were in French inches (*pouce*, 27.07 mm), but this time the groups were centered on midpoints of the successive one-inch ranges. No increments were reported for statures less than 58 inches nor for statures greater than 65 inches, reducing the sample size from 100,000 to 68,890 and making calculation of relevant statistics problematical. Villermé reported a mean of 59.67 inches (1.615 m), to which Quetelet added a probable error of 1/33 times the mean. Graphing frequencies for measurements reported to the nearest inch yields the distribution shown in Figure 2.1b, where it appears that a mean of 60.1 inches (1.626 m) fits the distribution better than that reported by Villermé. Quetelet's probable error should then be 1.821 inches (49.29 mm). Quetelet did not graph either of the distributions shown in Figure 2.1 but showed equivalents in tabular form.

Quetelet (1846) found variation in the empirical measurements of chest circumference and stature satisfying because their patterns mimicked the distributions of error recorded in astronomy and physics. He believed in natural laws, among them

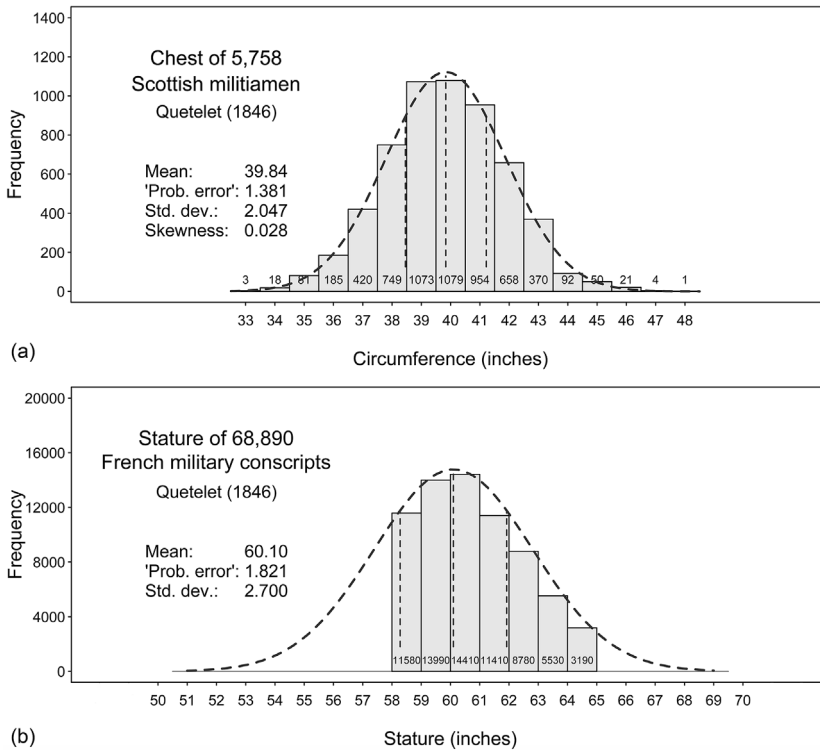


Figure 2.1 Histograms of the measurements compiled by Quetelet (1846). (a) Chest circumference of Scottish militiamen, based on measurements in imperial inches published anonymously in Edinburgh in 1817. (b) Standing height or stature of French military conscripts in French inches published by Villerme (1829; censored by failure to report shorter and taller statures). Mean stature for the French recruits would be 64.1 inches on an imperial scale. Numbers of men in each category are shown near the base of the corresponding column. Note the bell-shaped distributions of variation in both sets of measurements: variation considered to mimic distributions of error in astronomical measurements. Probable error was an early measure of dispersion, superseded by the standard deviation.

conservation of a human “type,” and he went so far as to call such laws divine (Quetelet, 1870, p. 21). Quetelet is famous for his interest in social science or “social physics” (*physique sociale*) and for his concept of the “average man” (*l’homme moyen*; Quetelet, 1835). He accepted distributions of “error” in human measurements as mathematical confirmation of the applicability of physics-like laws to both the human type and the average man. However, Quetelet’s idea of an “average man” or human “type” was misguided and the antithesis of Darwin’s subsequent emphasis that individual differences are important for natural selection.

Before leaving Quetelet (1846), it is interesting to note his preoccupation with human giants and dwarfs. He went so far as to calculate, from the sample of French conscripts, that a population with a mean stature of 1.62 meters might be expected to include Frenchmen ranging in stature from 1.21 to 2.03 meters (deviations of ± 0.41 meters from the mean). The smaller men Quetelet considered to be dwarfs, and larger men to be giants. Herschel (1850, p. 27) challenged this, citing examples, and argued that “the ‘probable’ deviation of nature’s workmanship from her universal human type cannot possibly be less than double that resulting from the French measurements.” If we follow Herschel, then Quetelet’s range of human stature should reach 0.80 meters for dwarfs and 2.44 meters for giants. We need not worry about Quetelet’s or Herschel’s numbers here but will return to the subject of dwarfs and giants.

2.1.2 Ezekiel Elliott

Ezekiel Elliott was an American statistician and a government delegate to the International Statistical Congress that met in Berlin in 1863. There he presented a study of American military statistics (Elliott, 1863; 1865). Elliott included measurements of the stature of a large sample of Civil War volunteers. Here again measurements were in inches, and the measurements were grouped by the inch (imperial inch, 25.4 mm). Following Quetelet (1846), Elliott constructed a table showing the number of men in each group. Observations differing from the mean decreased in number as the difference increased (Figure 2.2a), as Quetelet had shown for chest circumference and, less well, for stature. Elliott did not graph the distribution of variation for his large sample, but he did construct a graph like that in Figure 2.2a for a smaller and more manageable set of measurements.

Elliott was impressed by Quetelet, with his emphasis on the human type, and by the regularity of variation about the type:

Statistical researches, conducted by M. Quetelet of Belgium, have established the fact, previously contested, of the existence of a human type, and that the casual variations from it are subject to the same symmetrical law in their distribution as that, which the doctrine of probabilities assigns to the distribution of errors of observation. In the accompanying tables, showing the distribution of heights and of measurements of the circumference of chests of American soldiers, the conclusions of this eminent statist and mathematician are strikingly confirmed.

Elliott, 1863, p. 14

The distributions are the same for variation within a species and for errors of astronomical observation, but it is a mistake to conflate variation and error. We shall return to this below.

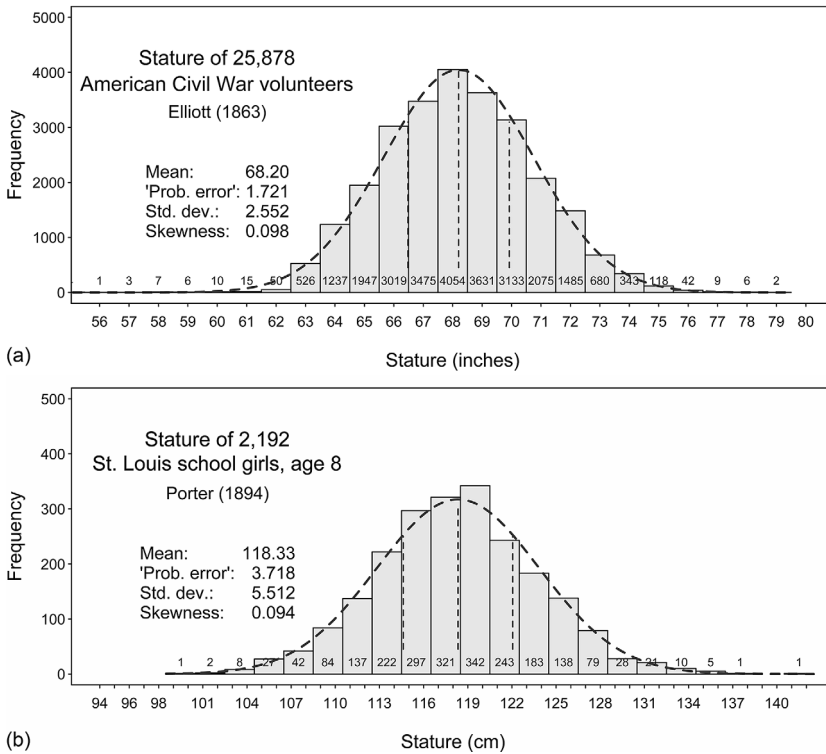


Figure 2.2 Histograms of standing heights in (a) American Civil War recruits (Elliott, 1863); and (b) eight-year-old St. Louis school girls (Porter, 1894). Measurements were reported to the nearest inch or centimeter, respectively, and numbers in each category are shown near the base of the corresponding column. Porter's school girl measurements were analyzed by Karl Pearson (1895), who was interested in their skewness. Probable error was an early measure of dispersion, superseded by the standard deviation.

Elliott analyzed both the physical measurements of Civil War volunteers and their ages. He was more perceptive than his contemporaries in distinguishing a law of error for ages based on *differences* from a law of error for human forms (“types”) based on *proportions*:

According to the law, already stated, which appears to obtain with the distribution of the ages of the volunteers, the differences of the numbers at consecutive equidistant ages are very nearly in equi-rational or [arithmetical] progression. — In the distribution of the representatives of a type, where the law assigned by the theory of probability strictly holds, the quotients (not the differences) of the proportionate numbers at consecutive equidistant points of measurement are in equi-rational progression.

Elliott, 1863, p. 15

This represents, in cryptic form, the kernel of a key insight of Francis Galton to be developed below.

2.2 Probability and the Law of Error

We need some background to understand Quetelet's fascination with what he called the "binomial law", "law of probability," or "mathematical law of size." Quetelet used all three names for what is essentially the same concept; others have called this the "law of error."

Start with a coin. It has two sides, two "faces," and a negligible edge separating these. Call one side a tail (T), and the other side a head (H). If you flip a coin in the air, it will land and fall flat, with a tail facing upward or a head facing upward. On a fair coin the probabilities of these alternative events are equal. We can note the possible outcomes of one coin flip as a "T" for tail or an "H" for head.

What happens if we flip two coins? There is one way we can get two tails: TT. There are two ways to get a tail and a head: TH if the tail comes up first and the head comes up second, or HT if the head comes up first and the tail second. And there is one way to get two heads: HH. We can summarize this by calling TT, TH, HT, and HH the four possible *permutations* resulting from flipping two coins. Two tails, a tail and a head, and two heads, are the corresponding three *combinations*. We may not know what permutations went into the combinations we see, but we know that for two coins we can expect the permutations to be related to the combinations in the proportions 1:2:1.

If we flip three coins, then we may find any of eight permutations, TTT, TTH, THT, HTT, THH, HTH, HHT, and HHH, in one of four combinations of three heads, two heads and a tail, one head and two tails, or three tails. Without knowing the history, again we don't know what the permutations were, but we can expect the permutations to be related to the combinations in the proportions 1:3:3:1.

Extending this logic, we can expect flipping ten coins to yield combinations of heads and tails in the proportions 1:10:45:120:210:252:210:120:45:10:1, for a total of 1,024 permutations (Figure 2.3). Each combination has a probability or relative likelihood or "density" equal to the number of permutations yielding the combination, divided by the total number of permutations. In Figure 2.3, the combination with the greatest probability is five tails and five heads, for which the probability is $252/1024 = 0.246$. This indicates too that the combination of five tails and five heads has the greatest likelihood in relation to other combinations.

The permutations and combinations of a coin-flipping exercise are interesting for many reasons, but the point to be made here is illustrated in Figure 2.4. Flipping a single coin yields two permutations and two combinations, one in each case a tail and one a head. Graphing yields a simple symmetrical histogram of one permutation for each combination. Graphs for the permutations and combinations of two and three coins are slightly more complicated, but the first inkling of the relationship of interest emerges in a graph of the permutations and combinations for four

Ten coins		Permutations	Probabilities
Combinations	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	1	0.001
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	10	0.010
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	45	0.044
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	120	0.117
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	210	0.205
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	252	0.246
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	210	0.205
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	120	0.117
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	45	0.044
	⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	10	0.010
⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙ ⊙	1	0.001	
	<hr/>	<hr/>	<hr/>
	1024	1.000	

Figure 2.3 Permutations and combinations of tails (T) and heads (H) for ten coins. There is one permutation yielding the combination “all tails”; ten permutations of “nine tails and one head” (the head can be in any of 10 positions in the row); etc. The numbers of permutations that can yield the observed proportion of tails and heads in each combination are listed to the right of the coins, along with the associated probability for each combination. The latter is calculated as the number of permutations in a particular combination (e.g., 1, 10, 45, etc.), divided by the total number of permutations in the experiment (1,024).

coins. The graph for four coins has vertical bars differing by small, then large, then small amounts flanking the modal value. This becomes increasingly clear for permutations representing larger numbers of coins, and the permutations for 9 or 10 coins follow the bell-shape of a “normal” distribution. In a binary coin-flipping experiment like this the permutations for each combination are known as the “binomial coefficients” for a given number of trials. Historically, the mathematical shape of what we today call a “normal distribution” was derived by finding the limiting shape of the continuous curve of permutations for an infinite number of coins.

The same distribution and curve can be derived by rolling dice (Figure 2.5). One die has six sides or faces, with negligible edges separating these. The sides are generally numbered from 1 to 6. If you roll a die on a table, it will land and fall flat, with one of the numbers facing upward. On a fair die the probabilities of the alternative events are equal. We can note the possible outcomes for rolling one die by recording the resulting number: There is one way to score a “1,” one way to score a “2,” etc. Each has the same probability, and graphing yields a simple symmetrical histogram of one permutation for each combination (Figure 2.5a).

A graph for the permutations and combinations of two dice is slightly more complicated because there is one way to achieve a score of two, by rolling a one on each die (“1 + 1”). A combination score of three can be achieved in two ways, by rolling “1 + 2,” or “2 + 1.” The modal combination score is seven, which can be achieved six ways, as “1 + 6,” “2 + 5,” “3 + 4,” “4 + 3,” “5 + 2,” or “6 + 1.” The

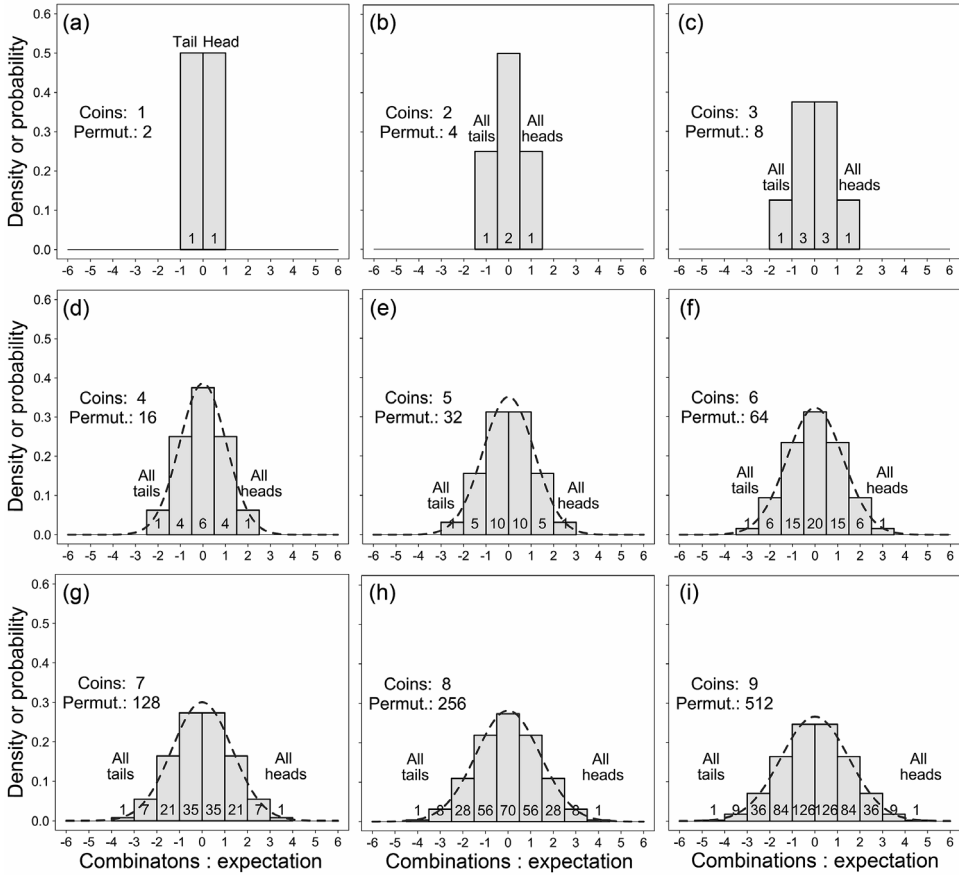


Figure 2.4 Histograms of permutations for increasing numbers of coins and combinations in a coin-tossing experiment. For n coins, the number of combinations of tails and heads increases as $n + 1$, but the number of permutations increases as 2^n . Note that the histogram of permutations converges to a normal curve (dashed line) as the number of coins increases.

sequence of successive permutations for each combination forms a symmetrical pyramid rising on one side to the mode and descending on the other (Figure 2.5b).

The relationship of interest emerges clearly in a graph of the permutations and combinations for three dice. Now the permutations for each combination form the series 1:3:6:10:15:21:25:27:27:25:21:15:10:6:3:1, for a total of 216 permutations. These are no longer pyramidal on a graph but now approximate a normal distribution (Figure 2.5c), with the heights of the vertical bars approximating a normal curve (dashed line). The fit to a normal curve for permutations of three dice is even better than the fit for permutations of 9 or 10 tosses of a coin. Here again, the mathematical shape of the normal curve can be derived by finding the limiting shape of the continuous curve for an infinite number of dice.

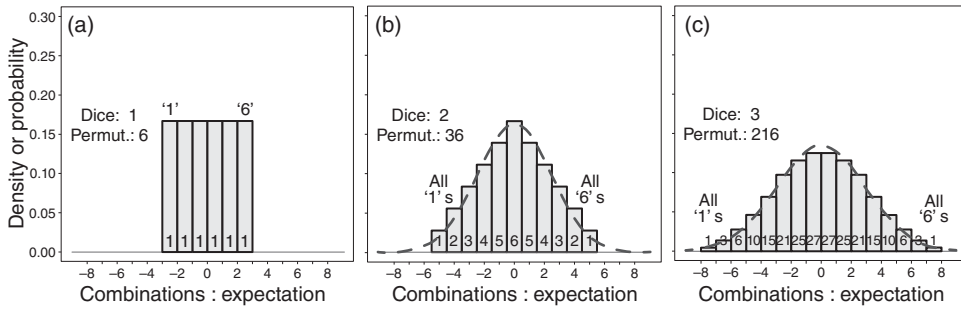


Figure 2.5 Histograms of permutations for increasing numbers of dice and combinations in a dice-rolling experiment. For n dice, the number of combinations of scores “1” through “6” increases as $5n + 1$, but the number of permutations increases as 6^n . Note that the histogram of permutations converges rapidly to a normal curve (dashed line) as the number of dice increases.

The importance of this exercise is that it shows how normal distributions of variation can be derived from the relative frequencies of small differences accumulating in the simple combinations expected by chance. The distributions of variation we see in biological populations reflect the contributions of many small genetic differences in constituent individuals.

2.3 The Normal Curve

“Normal” is an adjective, derived from Latin, that is commonly used in European languages to represent a “norm” or expected occurrence. As we have seen, this bland label is assigned to curves and distributions like those in the preceding figures. Common usage means it should come as no surprise that “normal” has no fixed point of origin in the history of statistics. Gustav Fechner (1860, p. 125) wrote of a “*normaler Fehlervertheilung*” (normal distribution of error). Quetelet (1869, p. 36) wrote of “*déviations d’une grandeur normale*” (deviations from a normal size; translated from Herschel’s 1850, “deviations from a standard”). Charles Peirce (1873, p. 206), who read Fechner, wrote of comparing an observed curve of errors to “the normal least-squares curve.” Wilhelm Lexis (1877, p. 34ff.), who read both Fechner and Quetelet, repeatedly compared “*normaler Dispersion*” (normal dispersion) to a dispersion greater or less than normal. Normal is an adjective that grew slowly into a name. By the time Francis Galton wrote *Natural Inheritance*, normal variability was the title of a chapter, the name of a curve, and the name of a distribution (Galton, 1889). In this he was followed by Karl Pearson (1894) and many others.

A normal distribution is sometimes called a Gaussian distribution, named for Carl Friedrich Gauss. Gauss (1809) is the one credited with inferring the form of

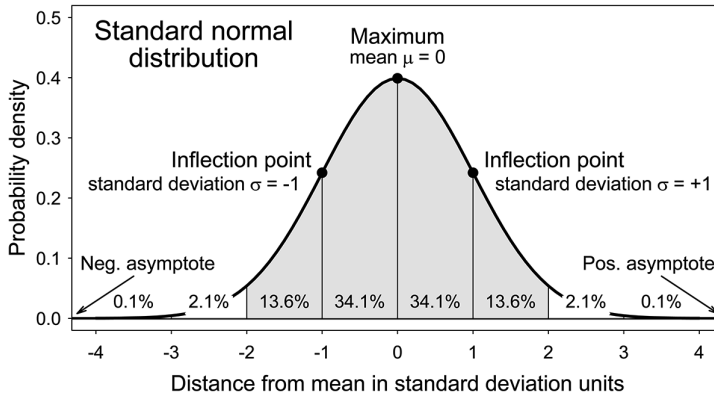


Figure 2.6 Landmarks of a normal distribution. The maximum value of the curve on the vertical axis gives the “location” of the distribution, which is the mean value of the probability density under the curve (here 0 on the horizontal axis). Left and right inflection points of the curve give the “dispersion” of the distribution. Each inflection point is one standard deviation from the mean (± 1 on the horizontal axis). The curve here is standardized to a unit integral or unit area under the curve (making the total probability density = 1). Note that 68.3% of the area under the curve lies within ± 1 standard deviation of the mean, 95.4% lies within ± 2 standard deviations, 99.7% lies within ± 3 standard deviations, and virtually all lies within ± 4 standard deviations. Reducing the standard deviation will narrow the dispersion, and increasing the standard deviation will broaden it, but the probability density within a given standard deviation range does not change.

the error curve, which he wrote as $\varphi(\Delta) = (h/\sqrt{\pi}) \cdot e^{-h \cdot h \cdot \Delta \cdot \Delta}$ (parentheses and multiplication symbols added), where *phi* denotes a function of delta, *h* is a measure of precision (or dispersion), *pi* is the familiar ratio of circumference to diameter of a circle, and *e* is the base of natural logarithms. Abraham de Moivre published something similar in 1733 as an approximation to the binomial distribution, as did Laplace in 1810 in the form of the central limit theorem (Stigler, 1986).

2.3.1 Landmarks of a Normal Curve

If we look at any continuous normal curve (Figure 2.6), we see that it is symmetrical, with three landmarks familiar from calculus. The most prominent landmark is the highest point on the distribution. This point, the maximum value of the curve, is also, from symmetry, the *mean* value of the distribution on the horizontal axis, a parameter represented by μ . The mean specifies the location of the curve. Secondary landmarks are the two inflection points equidistant from the mean. These differ from the mean on the horizontal axis by minus one

and plus one *standard deviation*, a parameter represented by σ , and the square root of the *variance* σ^2 .

For a distribution of constant size (standard area under the curve), the standard deviation specifies the dispersion of the distribution. There is no negative or positive limit to the distribution because the extreme values are asymptotic to the horizontal axis. Note that 68.2% of the area under the curve lies within ± 1 standard deviation, 95.5% lies within ± 2 standard deviations, 99.7% lies within ± 3 standard deviations, and virtually all of the area under the curve lies within ± 4 standard deviations. These proportions hold for any normal curve or distribution.

The general form of the probability density function comprising the normal curve and normal distribution is:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (2.1)$$

The curve is fit to an empirical sample by substituting sample values \bar{x} and s for the parametric mean μ and standard deviation σ , respectively. For $\mu = 0$ and $\sigma = 1$, Equation 2.1 simplifies to:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} \quad (2.2)$$

Both equations yield normal distributions standardized to a unit integral or a unit area under the curve, making the total probability density equal to one as well.

2.4 Logarithms and Coefficients of Variation

Logarithms and exponentials of the same base are inverse or mirror transformations. That is, $\log_{10} x$, or simply “log x ” here, is the inverse of 10^x . $\text{Log}_e x$, or “ln x ” here, is the inverse of e^x . $\text{Log}_2 x$ is the inverse of 2^x . The choice of bases is not completely arbitrary but depends on the range of numbers being compared. Log_{10} is most useful when the range spans several orders of magnitude. Log_2 is most useful when the range spans several doublings but lies within an order of magnitude.

Napierian or natural logarithms, \log_e or \ln , are “natural” because they have special properties. First, the natural logarithm for a number $a > 0$ can be defined as the area under the curve $y = 1/x$ as x increases (or decreases) from 1 to a (with the area being negative for $a < 1$). Second, the slope of the curve $y = \ln x$ is $1/x$, and the slope is 1 for $x = 1$ when and only when the base is e (y itself is 0 at this point, as it is for logarithms of all bases). Natural logarithms are “natural” too with respect to the population or sample variation that interests us here.

For normal distributions of different sizes and shapes, the standard deviation s increases in proportion to the mean \bar{x} as well as the dispersion. The ratio of the two, $V = s/\bar{x}$, is commonly called the coefficient of variation (and sometimes multiplied by 100 to represent s as a percentage of \bar{x}). The ratio is considered a measure of dispersion independent of the mean. It is a special property of natural logarithms that the variance of ln transformed measurements approximates the squared coefficient of variation V^2 (Lewontin, 1966), from which it follows that the standard deviation of ln-transformed measurements is a close approximation to the coefficient of variation V . The opposite is also true: The coefficient of variation V is a close approximation to the standard deviation of ln-transformed measurements. For this reason, natural logarithms of base e are the preferred transformation for measurements in biology rather than \log_2 or \log_{10} .

Empirically, linear measurements of organisms commonly have $V \approx 0.05$, meaning that for linear measurements a standard deviation will be about 0.05 units on a natural log scale. Measurements of area commonly have $V \approx 0.10$, and a standard deviation will be about 0.10 units on a natural log scale. And finally, volumetric measurements of organisms such as weight commonly have $V \approx 0.15$, and a standard deviation will be about 0.15 units on a natural log scale. Yablokov (1974) provides extensive documentation. Note that for organisms that vary in size but have similar shapes, the coefficients of variation 0.05, 0.10, and 0.15 are proportional to the dimensions of the measurement: 1, 2, and 3 (Schmalhausen, 1935; Yablokov, 1974; Lande, 1977).

2.5 Arithmetic Normality versus Geometric Normality

Arithmos is the Greek word for number, and *arithmetic* is the science of counting and calculation that involves addition and subtraction (and multiplication and division), moving forward and backward on what is commonly called a “number line.” An arithmetic progression is one that involves successive numbers differing by equal amounts, such as 1, 2, 3, 4; or $-10, -20, -30, -40$.

Ge or *gaia* is the Greek word for Earth, and *geometry* is the science of measurement, shape, and proportion. It started as measurement of land on the Earth’s surface but progressed to measurement and comparison of sizes and shapes of all kinds. The emphasis on proportion is evident in geometric progressions, such as 1, 2, 4, 8; or 1, 10, 100, 1000; or 0.050, 0.135, 0.368, 1.000.

A number line and arithmetic operations on it are relatively easy to relate to our everyday experience of addition and subtraction, and slightly more complicated versions of addition and subtraction in the form of multiplication and division. Most of our measuring devices, such as rulers and weighing balances, are calibrated arithmetically and we read them by *counting* in units of convenience.

This works well enough for comparisons that are inherently linear, yielding numbers in arithmetic progression, but it does not work well for comparisons that are inherently areal or volumetric, yielding numbers in geometric progression. Geometry is called geometry because it is different from arithmetic. It involves arithmetic, but it involves more and is an extension of arithmetic.

2.5.1 Francis Galton's Giants and Dwarfs

The extraordinary limits [to the height of man], beyond which are found monstrosities, seem to me difficult to fix . . . When we suppose the number of observations infinite, we may carry the differences to equally infinite distances from the mean, and find the corresponding probabilities. *This mathematical conception evidently cannot agree with that which is in nature.*

Adolphe Quetelet 1846 [1849], p. 102; italics added

The ordinary law of Frequency of Error, based on the arithmetic mean, . . . asserts that the existence of giants, whose height is more than double the mean . . . implies . . . *the existence of dwarfs, whose stature is less than nothing at all.*

Francis Galton 1879, p. 367; italics added

Quetelet took 5 feet 4 inches (1.62 m) as an average human height, and he accepted 1 foot 5 inches (0.43 m), exaggerated or not, as the height of the smallest dwarf. This is a difference of 3 feet 11 inches. Quetelet then added 3 feet 11 inches to the average height and predicted the limit to the size of giants to be 9 feet 3 inches (2.82 m). Quetelet recognized, intuitively at least, that there was some disagreement between the symmetry of his mathematical expectation and the asymmetry of limits actually observed in nature.

Galton exaggerated slightly in claiming giants to exist that are twice as tall as the average person, but he was clever in making an important point. Quetelet's "average man" was 1.62 meters or 162 cm tall. If a giant could be more than double this mean, say $162 + 164 = 326$ cm, then symmetry of the normal curve would imply that a dwarf could be $162 - 164 = -2$ cm tall: a Galtonian dwarf "whose stature is less than nothing at all." The conundrum is illustrated graphically in Figure 2.7a. However, experience now interferes — because we do not see people of negative stature, nor do we see people that we might consider to be approaching zero or negative stature. Galton's exercise demonstrates that the "normal" curve of human stature cannot be symmetrical. In arithmetic terms, there are fewer standard deviations to the lower limit of human stature than there are to the upper limit.

The geometric equivalent of Figure 2.7a is illustrated in Figure 2.7b. The two graphs have the same arithmetic axes, but vertical-axis values in Figure 2.7b are

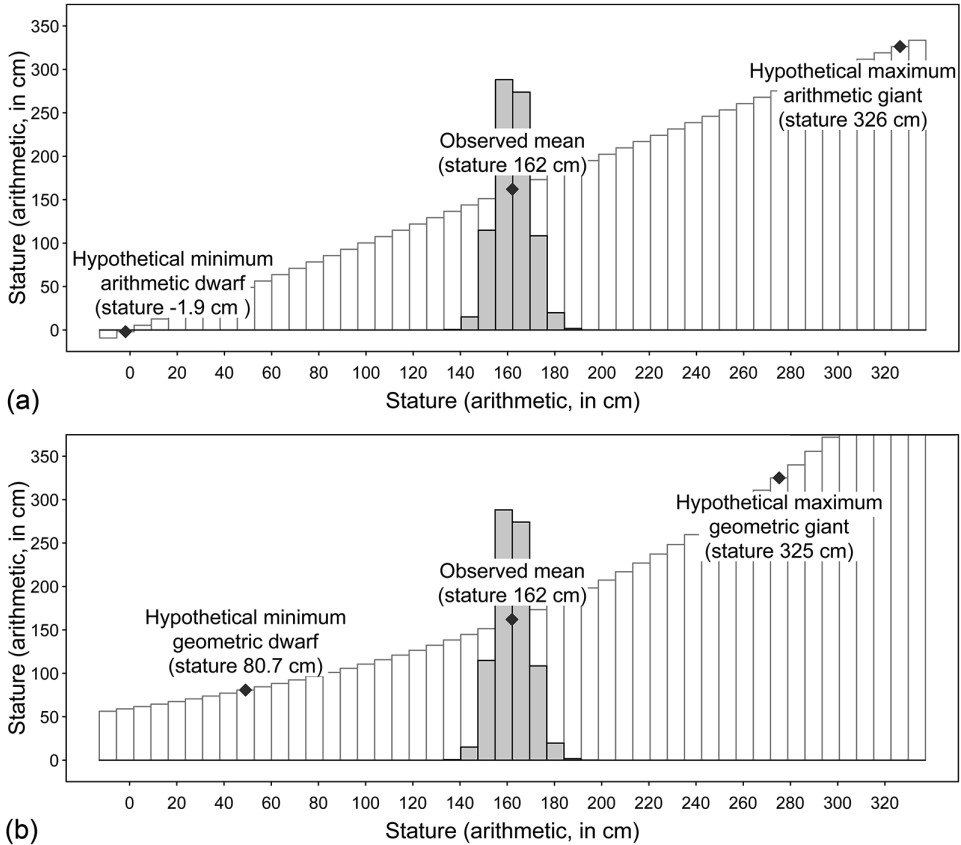


Figure 2.7 Comparisons of human stature in hypothetical human giants and dwarfs. On an arithmetic scale (a), where stature is added and subtracted in equal amounts, the postulated existence of giants more than double the mean implies, as Francis Galton argued, the existence of dwarfs whose stature is negative or “less than nothing at all.” In contrast, on a geometric scale (b), when stature is added and subtracted in equal proportions, the existence of giants more than double the mean implies, more plausibly, the existence of dwarfs whose stature is less than half the mean. Vertical bar widths and step heights are standard deviations, with heights in (b) calculated geometrically. The observed mean and standard deviation are those of Quetelet’s French military conscripts, 162 cm and 7.28 cm, and extreme statures assuming arithmetic normality are -1.8 and 326 cm. Equivalents assuming geometric normality, expressed on a proportional (natural-log) scale, are 5.09 and 0.045, with extreme statures of 4.39 and 5.78. Note that the existence of a geometric giant is more likely than the existence of an arithmetic giant because the doubling defining a geometric giant begins fewer standard deviations from the mean (15.5 versus 22.5; neither is really likely).

scaled geometrically. The vertical steps increase and decrease not by equal amounts but by equal proportions. There are fewer standard-deviation steps to the upper limit of 325 or 326 cm, and hence fewer steps to the lower limit when both are scaled proportionally. Fewer standard-deviation steps to these limits, 15.5 in Figure 2.9b versus 22.5 in Figure 2.7a, means that the geometric limits are more likely. Importantly, in the geometric case a lower limit of one halving matches an upper limit of one doubling.

Galton (1879) and the Cambridge mathematician Donald McAlister (1879) recognized that distributions of error and variation in terms of proportion are just as plausible as distributions of error and variation in terms of amount. Both men recognized that the appropriate measure of central tendency in such a case is the geometric mean (exponentiated mean of the logarithms of measurements) rather than the arithmetic mean of the raw observations. McAlister (1879) then showed that the ordinary law of error applies to the logarithms of measurements in the geometric case just as it does to raw measurements in the arithmetic case.

Neither Galton nor McAlister used the term “lognormal,” but their work implicitly introduced the concept. Both surely recognized that geometric normality leads to an expectation of asymmetry, not symmetry, on an arithmetic scale. However Galton, oddly, in his subsequent work followed a facile path and ignored lognormality, writing, for example, “it was found that the distribution of stature was sufficiently normal to justify our ignoring any shortcomings in that respect” (Galton, 1889, p. 117). And “had I possessed better data, I should have tried the geometric mean throughout” (Galton, 1889, p. 119).

Empirical distributions of biological variation like those in Figures 2.1 and 2.2 here, and similar distributions published by Weldon (1893; 1895) and others, motivated Karl Pearson to develop an application of moments, borrowed from physics and mechanics, to investigate normality and departures from normality. He hoped, optimistically, to factor complex curves into normal components. The first moment of a normal curve is the mean, and the second moment is the variance. Pearson focused on the standard deviation, the positive square root of the second moment, as the most appropriate measure of dispersion (Pearson, 1894), and then on a standardized third moment as a measure of asymmetry or skewness (Pearson, 1895). Right or positive skewness was common if not ubiquitous in the empirical distributions studied by Quetelet, Galton, Weldon, Pearson, and others.

It may seem surprising that little attempt was made to analyze the distributions as lognormal rather than normal, but this requires substantial samples structured appropriately; it requires goodness-of-fit tests that were not available at the end of the nineteenth century; and it also requires computational power that was not yet

available. Pearson's final word on lognormality, published in his biography of Galton, reads as an epitaph:

I am unaware of any comprehensive investigation being ever undertaken to test the "goodness of fit" of [Galton and McAlister's] geometric mean curve to actual observations. McAlister gives no numerical illustration, and I do not think Galton ever returned to the topic. It would still form the subject of an interesting research, but *I fear the Galton-McAlister curve would be found wanting.*

Karl Pearson 1924, pp. 227–228; italics added

2.5.2 Empirical Support for Lognormality

One comparison of the normality and lognormality of biological variation started as an attempt to use a large set of measurements to reject one or the other hypothesis. Gingerich (2000) analyzed an extraordinary set of human measurements published in a series of monographs resulting from the mid-twentieth-century All India Anthropometric Survey. This professional government survey involved measurement of numerous homogeneous sets of 50 adult human males of the same caste and village, from villages broadly distributed across political states and geographic regions of India. These were compiled and finally published, state by state, in volumes issued by the Anthropometric Survey of India. Gingerich (2000) chose to analyze two of the larger samples from the states of Maharashtra (Basu et al., 1989) and Uttar Pradesh (Banerjee and Basu, 1991). The former provides 14 measurements and 14 indices for 6,869 individuals and the latter provides the same measurements and indices for 7,766 individuals.

The usual approach to the problem of normality is to consider arithmetic normality as a null hypothesis, H_0 , test this against a set of measurements, fail to reject the null hypothesis, and then proceed as if variation is arithmetically normal because the hypothesis was not rejected. However, failure to reject normality as a null hypothesis rarely means anything because the samples employed are usually too small to provide the statistical power required for rejection (Gingerich, 1995).

Empirical distribution function (or EDF) tests are among the most powerful non-parametric goodness-of-fit tests for normality (D'Agostino, 1986). Each test is based on the fit of a stepped cumulative empirical curve to a model cumulative normal curve with parameters estimated from the empirical sample. Lilliefors' version of the Kolmogorov–Smirnov goodness-of-fit test involves a supremum statistic representing the maximum deviation from expectation for all steps of an EDF. Cramer–von Mises and Anderson–Darling tests employ quadratic statistics representing sums of differently weighted squared differences. A full set of original measurements or indices is required for computation of these statistics, and the

goodness-of-fit is calculated first for the original measurements or indices, and then for the logarithmically transformed measurements or indices. Details are given in Gingerich (2000).

Goodness-of-fit tests of normality that treat arithmetic normality or geometric normality (lognormality) as a null hypothesis often fail because: (1) as mentioned above, small sample sizes mean neither hypothesis can be rejected or (2) when sample sizes are large, both hypotheses can be rejected. The appeal of a model is its generality, and empirical distributions rarely fit a general model exactly. Thus, large samples often provide the power to reject all models. In Gingerich (2000) measurements of low variability behaved differently from measurements of higher variability. Stature is a low-variability measure with relatively small coefficients of variation on an arithmetic scale and small standard deviations on a geometric scale. The shapes of the distributions on the two scales are similar (Figure 2.8), and the large Maharashtra and Uttar Pradesh samples taken as a whole generally failed to reject either of the alternatives as a null hypothesis. For measurements of higher variability such as body weight, involving larger coefficients of variation and larger standard deviations (Figure 2.9), the large Maharashtra and Uttar Pradesh samples generally forced rejection of both hypotheses. Hence the large-sample tests failed in different ways, depending at least in part on the variability of the measurement or index being examined. Whatever the reason, most large-sample tests failed to distinguish normality from lognormality.

An alternative approach to the normality versus lognormality problem is to compare the two as alternative hypotheses and ask which is better supported by the empirical information at hand. This is a classic likelihood solution (Edwards, 1972; 1992) to a problem where ordinary hypothesis testing fails. Alternative hypotheses (H_1 , H_2 , etc.) are tested not by comparison to some statistical-model critical value, but by comparison of the hypotheses to each other to see which has greater relative likelihood or support. Support is the difference in the natural logarithms of the probabilities associated with each hypothesis (arithmetic minus geometric) or, equivalently, the natural logarithm of the likelihood ratio of probabilities favoring one hypothesis over the other (arithmetic over geometric). Support scores are additive. A positive support score favors arithmetic normality, and a negative support score favors geometric normality.

Large-sample support scores for the Maharashtra measurements total -11.63 (for 12 of 14 measurement scores) and for the Uttar Pradesh measurements total -25.56 (12 of 14 scores). Large-sample support scores for the Maharashtra indices total -51.60 (6 of 14 index scores) and for the Uttar Pradesh indices total -55.24 (6 of 14 scores). Some measurement and index scores are missing because the probabilities required for their calculation are too small to be computed.

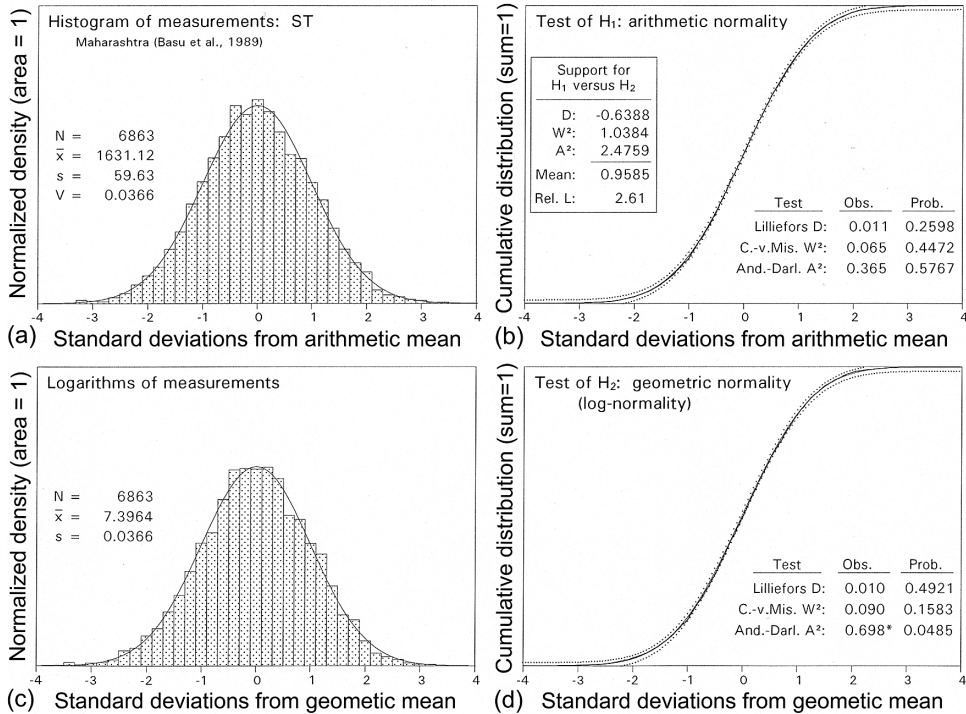


Figure 2.8 Likelihood comparison of arithmetic and geometric normality of human stature, based on a sample of 6,863 adult men from Maharashtra (Basu et al., 1989; Gingerich, 2000). (a and b) Normalized density histogram and cumulative empirical distribution function (EDF; stepped line) for raw measurements. (c and d) Normalized density histogram and cumulative EDF (stepped line) for \ln -transformed measurements. Goodness-of-fit statistics are Lilliefors D , Cramer-von Mises W^2 , and Anderson–Darling A^2 (D’Agostino, 1986). Support for hypothesis H_1 (arithmetic normality) relative to H_2 (geometric normality or lognormality) is given by the log-likelihood ratio, the natural logarithm of the ratio of probabilities for the corresponding test statistic (e.g., $-0.6388 = \ln [0.2598/0.4921]$). Mean support of 0.9585 for all three tests suggests that H_1 (arithmetic normality) is about 2.61 times more likely than H_2 (geometric normality) for these measurements. The only goodness-of-fit statistic with a probability less than the critical value for significance ($\alpha < 0.05$, asterisk) is Anderson–Darling A^2 . Subsamples are more homogeneous and their relative likelihoods are more tractable computationally (see text). Figure is modified from Gingerich (2000), reproduced by permission of Elsevier Publishing

Fortunately, both the Maharashtra and the Uttar Pradesh surveys were carried out caste by caste and village by village, and then published as a collection of smaller and more homogeneous subsamples preserving this information. Most subsamples include measurements and indices for 50 individuals. The best way to take advantage of all of the information is to calculate support scores for each

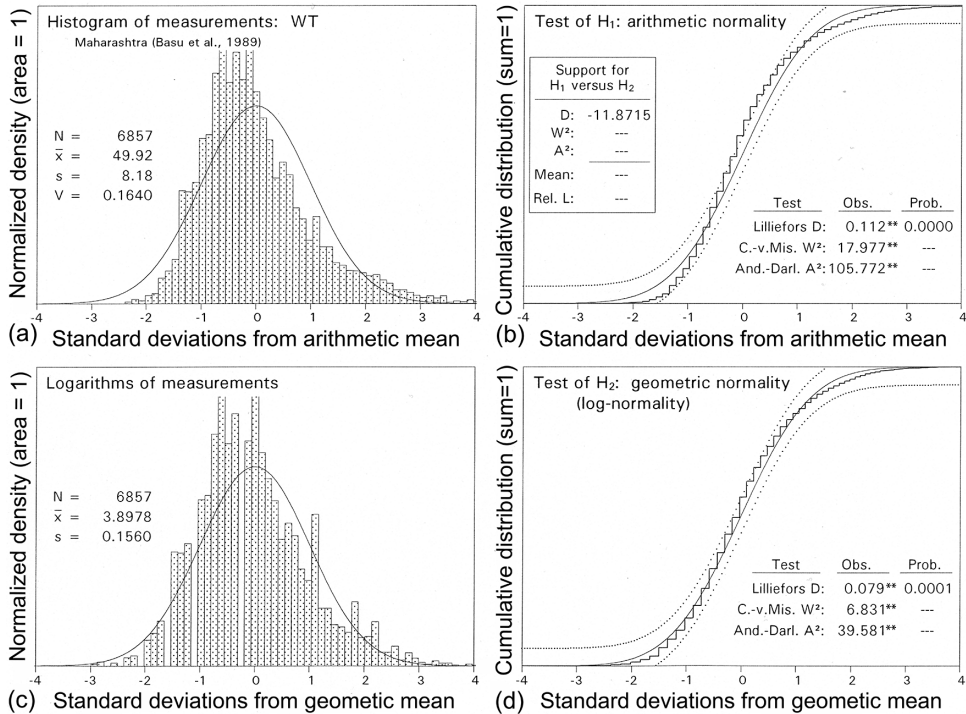


Figure 2.9 Likelihood comparison of arithmetic and geometric normality of human body weight, based on a large sample of 6,857 adult men from Maharashtra (Basu et al., 1989; Gingerich, 2000). (a and b) Normalized density histogram and cumulative empirical distribution function (EDF; stepped line) for raw measurements. (c and d) Normalized density histogram and cumulative EDF (stepped line) for ln-transformed measurements. Goodness-of-fit statistics are Lilliefors D , Cramer–von Mises W^2 , and Anderson–Darling A^2 (D’Agostino 1986), as in Figure 2.8. All test statistics for H_1 and H_2 have probabilities much less than the critical values for significance ($\alpha < 0.05$ and $\alpha < 0.01$, double asterisk), indicating that the empirical distributions do not fit either model. Probabilities are too small to compute in four of the six tests, compromising calculation of mean support. Subsamples are more homogeneous and their relative likelihoods are more tractable computationally (see text). Figure is modified from Gingerich (2000), reproduced by permission of Elsevier Publishing

measurement or index for each subsample and then add these together. Pooled support scores for 143 Maharashtra samples of measurements total -288.90 (14 of 14 scores) and for 153 Uttar Pradesh samples of measurements total -136.40 (14 of 14 scores). Pooled support scores for 143 Maharashtra samples of indices total -539.08 (14 of 14 scores) and for 153 Uttar Pradesh samples of indices total -338.50 (14 of 14 scores).

For Maharashtra 5 of 14 subsample support sums for measurements are positive and 9 are negative. Two subsample support sums for indices are positive and

12 are negative. For Uttar Pradesh 4 of 14 subsample support sums for measurements are positive and 10 are negative. One subsample support sum for indices is positive and 13 are negative.

There is some variation in likelihood support scores, but for large samples studied to date, whether analyzed as a whole or divided into homogeneous subsamples, virtually all support scores are negative. Thus, empirically, geometric normality is favored over arithmetic normality. This supports the Elliott (1863), Galton (1879), and McAlister (1879) claims that distributions of variation should be studied in terms of proportion, and supports the Galton–McAlister application of the “law of error” to the logarithms of measurements. Deficiencies of the arithmetic methods of measurement we use to study geometric phenomena are easily compensated by transforming measurements to logarithms.

2.6 Applications of Normality and Lognormality

The normality of biological variation has a number of important consequences, following first from normality itself, and then from the geometric nature of normality.

2.6.1 Phenotypes and Genotypes

The classical Greek words *phaino* and *phaneros*, meaning manifest and evident, are the roots of common English words such as phenomenon and phenomenal. Phenomena are *observable*, perceived through the senses rather than inference or intuition. *Phaenotypus* was introduced in biology by the Danish botanist Wilhelm Johannsen (1909, p. 123), and “phenotype” is the name commonly given to the observable form and behavior of a single organism or a statistical population of organisms. The word was introduced to distinguish what is seen (the phenotype) from what is unseen (Johannsen’s *Genotypus* or genotype) in an organism or population, reflecting how genetic inheritance and development were understood at the time.

Johannsen (1909, p. 130) was most impressed by the divergent phenotypes of sexually dimorphic organisms, and he reasoned that the way the phenotypes are manifest says nothing (“*absolut nichts*”) about the underlying genotype. Johannsen argued that phenotypic differences can be seen where no genotypic differences exist, and genotypic differences exist where no phenotypic differences can be seen. More is known today, and one is reluctant to criticize someone writing a century ago, but Johannsen was wrong to think phenotypes tell us nothing about underlying genotypes. Normal distributions of variation like those in Figures 2.1–2.2 and 2.8–2.9 are constructed from the addition of many genetic differences of small

effect. Additive genetic variance underlies virtually all of the polygenic, normally distributed traits of interest in quantitative evolutionary studies.

2.6.2 *Quetelet's "Average Man"*

Measurement error and biological variation are both normally distributed, and hence "error" and "variation" are commonly conflated. However, they are not the same. One reflects the error inherent in repeated observation of a system (possibly combined with variation in behavior of the system itself). The other reflects true variation in a system (possibly combined with error in observation of the system). This is important in considering the meaning of constructs like Quetelet's (1835) "average man." There is an average value for any characteristic we can measure, but what is its meaning?

The mean value for a normal distribution of error is taken to represent the true value being measured, and there is only one expected or "normal" value. On the other hand, the mean for a normal distribution of natural variation is just one of many expected values, and the whole of the observed distribution is what is "normal." If the chest circumference in Figure 2.1a were an error distribution, the expected value would be the mean value of 39.84 inches. However, it is not an error distribution but a distribution of natural variation, and the expected value is the whole bell-shaped exponential curve centered on the mean.

2.6.3 *Species Comparisons*

It is sometimes challenging to compare the variability of traits in one biological species or population with the variability of traits in another because the variability of the traits so often depends on the size of the organisms involved: standard deviations depend on their associated means. In the example of Figure 2.10a the white-tailed deer *Odocoileus virginianus* on the right side of the chart has a much broader range of variation in cranial length than does the deer mouse *Peromyscus maniculatus* on the left side of the chart. The two distributions of variation are very different in shape, with standard deviations of 1.53 and 20.45, respectively, and no consistent expectation. The problem is more difficult of course when attempting to understand how questionably identified organisms might group into species. One commonly accepted solution is to consider variability in relation to the mean by calculating a coefficient of variation: the standard deviation divided by the mean. When we do this for the deer mouse and the deer we see that the coefficients of variation for cranial length, 0.078 and 0.073, are closely comparable, and we can safely expect other species to have similar coefficients of variation.

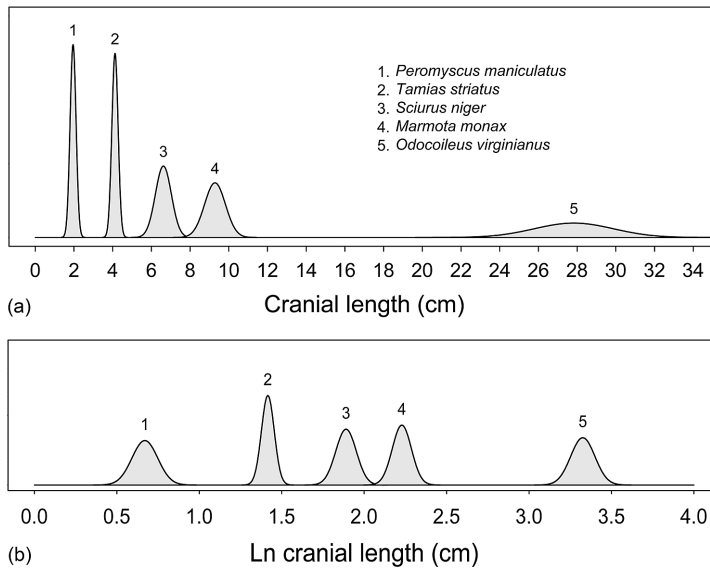


Figure 2.10 Comparison of empirical cranial length distributions for five mammalian species commonly found in Michigan. (a) Cranial length on the arithmetic scale of measurement, where small species have a relatively narrow range of dispersion and large species have a much greater range. (b) Cranial length on a geometric scale following transformation of measurements to base- e natural logarithms. Note that \ln -transformed measurements have distributions that are similar across species, facilitating interpretation of species differences and species boundaries. The ranges of \ln -transformed linear measurements like those shown here average about 0.3–0.4 units on an \ln scale ($\pm 3 = 6$ standard deviation units). The same standardization can be achieved with base-2 and base-10 logarithms, but base- e is preferred because one standard deviation closely approximates the ordinary coefficient of variation. (Lewontin, 1966)

An equivalent and more powerful approach to standardization is to compare the species on a logarithmic geometric scale rather than the arithmetic scale of measurement. This comparison is shown in Figure 2.10b, where now the distributions for all five species are much more similar. Natural-log transformation of measurements is preferred because the resulting standard deviations approximate the coefficients of variation just calculated (0.078 and 0.073, respectively; see Lewontin, 1966, for an analytical explanation). Base-2 and base-10 logarithms are equally effective in standardizing variation and making species comparable, but they do not have the advantage of practical equivalence to the coefficient of variation.

Why is the variability of a trait proportional to the size of the trait? This is an open question that may be related to the generation of variation, or to functional limits governing interactions within a population, or both. It is important to acknowledge,

first, that variability is proportional to size and, second, that logarithms provide a simple way to standardize this proportionality.

2.6.4 Allometry

The paleontologist Henry Fairfield Osborn (1925) introduced what he called the “principle of allometry” to emphasize the importance of “change of proportion” in vertebrate evolution. Julian Huxley (1924, 1932) first called this “heterogony,” following Albert Pézard, but later adopted the word “allometry” (Huxley and Teissier, 1936). Allometry is a biological equivalent of geometry in mathematics — each is given a name to distinguish it from simple additive arithmetic.

In a study of fiddler crabs, Huxley (1924) found that the variable y , representing the weight of the large chela of males, was related to x , body weight less the weight of the chela, by the relation:

$$\log y = k \log x + \log b \quad (2.3)$$

Huxley then expressed this in what he called its “simplest” form as:

$$y = bx^k \quad (2.4)$$

It is debatable whether a power function is simpler than a linear equation, although it may have seemed so in Huxley’s day when logarithmic transformation required book-length tables. Equations 2.3 and 2.4 are alternative representations of the same relation, one geometric and one arithmetic. One involves logarithms and the other does not.

Logarithmic transformation converts ranges of equivalent proportion to ranges of equivalent size. Exponentiation does the opposite, transforming ranges of equal size to ranges of equal proportion. The mouse-to-elephant simulations in Figure 2.11 show the effects of logarithmic and exponential transformation. The curved progression of small-to-large species in Figure 2.11a, each of equivalent coefficient of variation, becomes straight and uniform when measurements are transformed to proportions, logarithmically, in Figure 2.11b. The straight and uniform series of species in Figure 2.11b, each of equivalent range, becomes a curved progression of small-to-large species when proportions are transformed to measurements, exponentially, in Figure 2.11a.

The ellipses representing species in the simulation of Figure 2.11b have 95% confidence ranges of about 0.20 units on both the length (x) and width (y) axes. A 95% confidence range corresponds to ± 2 standard deviations (Figure 2.6), for a full range of four standard deviation units. The standard deviations of tooth length and width in the simulation are each 0.05, and the coefficients of variation V are 0.05 as expected for a linear measurement. We would expect a standard deviation

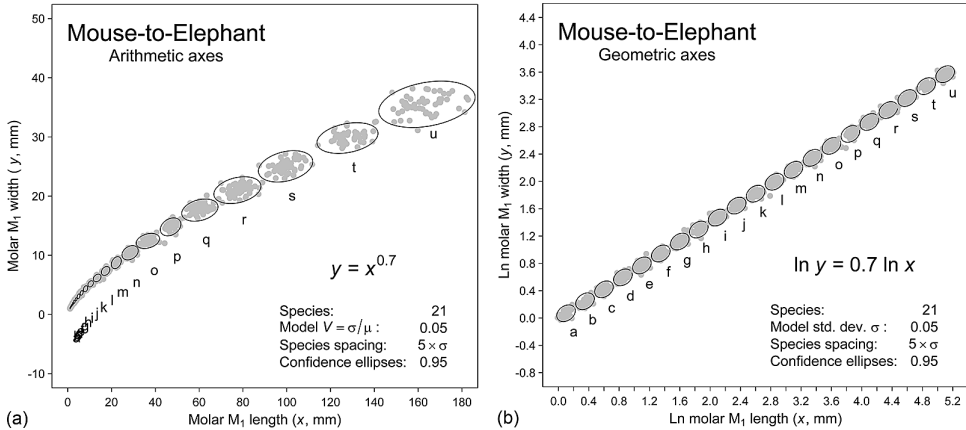


Figure 2.11 Relative sizes of 21 simulated “mouse-to-elephant” model species, here labeled $a - u$, represented by measurements of first lower molars. (a) Species on arithmetic axes representing the scales of measurement. (b) Species on geometric or allometric axes representing the scales of functional relationship. Ellipses enclose 95% confidence regions for 50 individuals drawn at random from bivariate normal populations. Means μ of successive species increase by the Hutchinsonian ratio 1.28, and standard deviations in A and B are $\sigma = 0.05 \cdot \mu$ and $\sigma = 0.05$, respectively. The within-species correlation between x and y is constant at $\rho = 0.3$. Note the consistent size ranges and uniform spacing of species when plotted on geometric axes. This consistency simplifies comparison and interpretation, and is itself an indication that the underlying functional relationships of molar length and width are geometric. Figure is modified from Gingerich (2014), reproduced by permission of John Wiley and Sons Publishing

of tooth crown area (length \times width) to be about 0.10, and a standard deviation of tooth crown volume (length \times width \times height) to be about 0.15. It would be difficult to see such regularity and consistency in the very same tooth sizes when plotted on arithmetic axes, as they are in Figure 2.11a.

2.6.5 Limiting Similarity

In 1958, G. Evelyn Hutchinson delivered a classic presidential address to the American Society of Naturalists. His title, “Homage to Santa Rosalia,” recalled a happy day spent collecting water bugs on Monte Pellegrino in Sicily (Hutchinson, 1959). Two species were present in a pond, one small and one a little larger. This started Hutchinson thinking about why there are so few, and so many, species in nature. He then asked himself about “limiting similarity.” What morphological difference is required for closely related species to occupy adjacent niches at the same level in a food web? Hutchinson found, empirically, that closely related species living sympatrically differ in linear measurements by a factor averaging

about 1.28, which he considered the minimum difference necessary to fill different niches. Hutchinson expressed this as 1.28/1.00, or 1:1.28, and 1.28 is now known as a “Hutchinsonian ratio.”

But what if the measurement comparing species is two dimensional instead of linear? What if it is three dimensional? We can investigate this as Hutchinson (1959) investigated the limiting similarity for linear measurements. Hutchinson’s first example, comparison of males of two sympatric mustelid species found in Great Britain, employed measurements published by Miller (1912). The smaller of the species, *Mustela nivalis*, is the British weasel, and the larger of the species, *M. erminea*, is the stoat or short-tailed weasel. Miller (1912) provided measurements for crania of 12 male *M. nivalis* and 12 male *M. erminea*, including condylobasal length, zygomatic breadth, and occipital depth (Miller’s measurements are in tables starting on his pages 392 and 408). The male *M. nivalis* crania average 39.5 mm in length, and the male *M. erminea* crania average 50.5 mm in length, yielding the 1.28 ratio that Hutchinson reported.

If we multiply Miller’s measurements of cranial length by zygomatic breadth, we have a measure of cranial area. This averages 859.1 mm² for crania of male *M. nivalis*, and 1474.1 mm² for crania of male *M. erminea*, yielding a Hutchinsonian ratio for area of 1.72. If we multiply Miller’s measurements of cranial length by zygomatic breadth by occipital depth (height), we have a measure of cranial volume. This averages 9,046 mm³ for crania of male *M. nivalis*, and 19,530 mm³ for crania of male *M. erminea*, yielding a Hutchinsonian ratio for volume of 2.17. Results are shown graphically in Figure 2.12. Hutchinsonian ratios for lengths, areas, and volumes are clearly sensitive to the dimension of the form being represented.

The Hutchinsonian ratios of 1.28 for lengths, 1.72 for areas, and 2.17 for volumes are sensitive to dimension. If we take the natural logarithms of these, we find that a ratio of 1.28 for length measurements is equivalent to separation by 0.25 units on a natural log scale; a ratio of 1.72 for areas is equivalent to separation by 0.54 units on a natural log scale; and a ratio of 2.17 for volumes is equivalent to separation by 0.77 units on a natural log scale. These separations of 0.25:0.54:0.77 are almost exactly in the proportions 1:2:3, showing again their sensitivity to the dimensions being measured.

If a standard deviation is equivalent to 0.05 units on a natural log scale, as modeled above, then a separation of 0.25 for limiting similarity of cranial length would be equivalent to a separation of 5 standard deviations. If a standard deviation is equivalent to 0.10 units on a natural log scale, then a separation of 0.54 units for limiting similarity of cranial area would be equivalent to a separation of 5 standard deviations. And finally, if a standard deviation is equivalent to 0.15 units on a natural

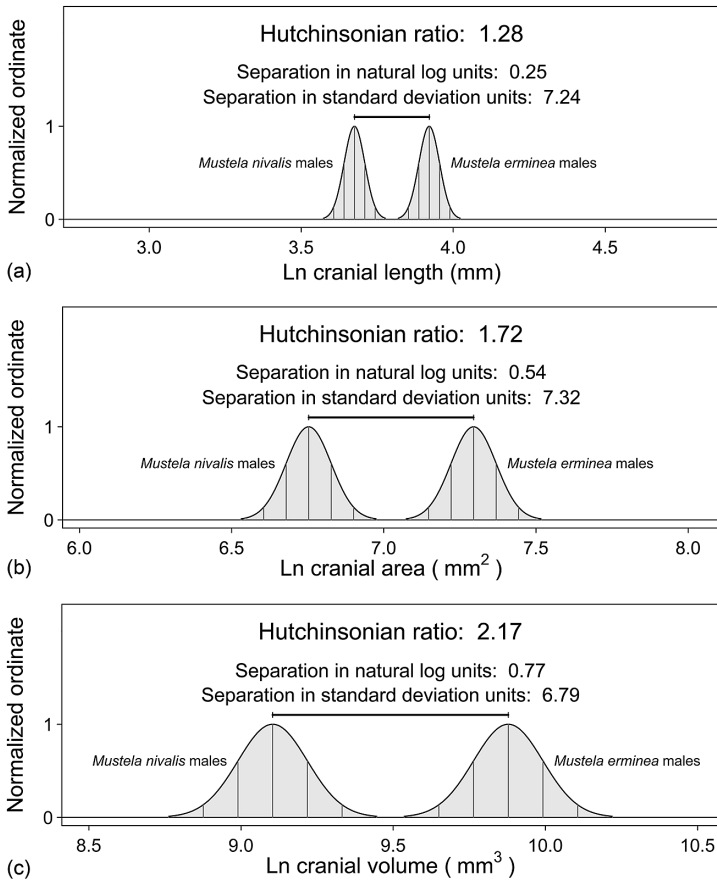


Figure 2.12 Quantification of limiting similarity for closely related species occupying adjacent niches at the same level in a food web. Example is from Hutchinson (1959), extended to include differences in (a) cranial length, (b) cranial area, and (c) cranial volume. Measurements from Miller (1912) compare male weasels, *Mustela nivalis*, and male stoats, *M. erminea*, sympatric in Great Britain. Differences between species quantified as Hutchinsonian ratios (1.28, 1.72, and 2.17) or in natural log units (0.25, 0.54, and 0.77) remain proportional to the dimensions (1, 2, and 3) of the lengths, widths, and volumes involved. Vertical lines within normal curves show standard deviations (see Figure 2.6; here averaged across the two species). Note that quantification in standard deviation units removes the effect of dimension and yields a consistent measure for limiting similarity of about seven standard deviation units in all three cases.

log scale, then a separation of 0.77 units for limiting similarity of cranial area would again be equivalent to a separation of 5 standard deviations. Empirically, the weasels and stoats studied by Miller and Hutchinson are a little less variable than modeled above, the standard deviations average 0.03, 0.07, and 0.11 rather than 0.05, 0.10, and 0.15, and the limiting similarity for weasels and stoats is close to seven standard

deviations for lengths, areas, and volumes. This example is important because it illustrates how standard deviation units incorporate dimensionality and eliminate its effect, making standard deviations the preferred units for expressing limiting similarity. For this reason standard deviation units represent differences between variable populations more effectively than any ratios or differences on a proportional scale. This comes up again later when we consider how to quantify evolutionary rates.

2.7 Summary

1. Darwin emphasized variation, individual differences within species, as “highly important” for natural selection. Variation is indeed essential for the process to work.
2. Attempts to quantify biological variation in the nineteenth century focused on resemblance of this variation to distributions of measurement error in astronomy and physics.
3. Measurement error is generated independently in astronomical and other physical observations, but in biology variable populations evolve from variable populations. Quetelet’s “average man” was misguided: The expectation in biology is not an average over and over, but a full distribution of variation in each successive generation.
4. Normal distributions can be generated by summing permutations across combinations of flipped coins or rolled dice. By analogy, normal distributions of variation in biological populations reflect “additive genetic variance” and the sum of many small differences in constituent individuals.
5. Arithmetic is mathematics based on counting, and geometry is mathematics based on proportion. Theoretically and empirically the normality of biological variation is geometric rather than arithmetic: biological variation is lognormal rather than normal, and individual differences are differences of proportion. Logarithms employed to transform counts to proportions can be chosen to reflect halving and doubling (\log_2), standardized deviations (\ln or \log_e), or orders of magnitude (\log_{10}).
6. Allometry is the biological equivalent of geometry, and the allometric equation in simplest form is a linear equation relating dependent $\log Y$ to independent $\log X$, or dependent $\ln Y$ to independent $\ln X$.
7. Finally, comparisons of populations in standard deviation units incorporate dimension and remove its effect, making standard deviations the preferred units for expressing the similarities and differences of variable populations.