We will see in later chapters that many classical statistical methods are specifically designed for, and function best with, data that have been sampled from normal populations. We will further see that in many practical situations these methods also work very well for samples from nonnormal populations.

The normal distribution is of central importance in spite of the fact that many, perhaps most, naturally occurring biological distributions could be described better by a skewed curve than by a normal curve. A major use of the normal distribution is not to describe natural distributions, but rather to describe certain theoretical distributions, called sampling distributions, that are used in the statistical analysis of data. We will see in Chapter 5 that many sampling distributions are approximately normal even when the underlying data are not; it is this property that makes the normal distribution so important in the study of statistics.

Supplementary Exercises 4.S.1–4.S.21

4.S.1 The activity of a certain enzyme is measured by counting emissions from a radioactively labeled molecule. For a given tissue specimen, the counts in consecutive 10-second time periods may be regarded (approximately) as repeated independent observations from a normal distribution.[15] Suppose the mean 10-second count for a certain tissue specimen is 1,200 and the standard deviation is 35. Let Y denote the count in a randomly chosen 10-second time period. Find

(a) $\Pr\{Y \geq 1{,}250\}$

(b) $\Pr\{Y \leq 1{,}175\}$

(c) $\Pr\{1{,}150 \leq Y \leq 1{,}250\}$

(d) $\Pr\{1{,}150 \leq Y \leq 1{,}175\}$

4.S.2 The bill lengths of a population of male Blue Jays follow approximately a normal distribution with mean equal to 25.4 mm and standard deviation equal to 0.8 mm (as in Example 4.1.2). Find the 95th percentile of the bill length distribution.

4.S.3 Refer to the bill length distribution of Exercise 4.S.2.

(a) What percentage of bill lengths are greater than 26.2 mm?

(b) What percentage of bill lengths are less than 24 mm?

4.S.4 The heights of a certain population of corn plants follow a normal distribution with mean 145 cm and standard deviation 22 cm.[16] What percentage of the plant heights are

(a) 100 cm or more?

(b) 120 cm or less?

(c) between 120 and 150 cm?

(d) between 100 and 120 cm?

(e) between 150 and 180 cm?

(f) 180 cm or more?

(g) 150 cm or less?

4.S.5 Suppose four plants are to be chosen at random from the corn plant population of Exercise 4.S.4. Find the probability that none of the four plants will be more than 150 cm tall.

4.S.6 Refer to the corn plant population of Exercise 4.S.4. Find the 90th percentile of the height distribution.

4.S.7 For the corn plant population described in Exercise 4.S.4, find the quartiles and the interquartile range.

4.S.8 Suppose a certain population of observations is normally distributed.

(a) Find the value of z^* such that 95% of the observations in the population are between $-z^*$ and $+z^*$ on the Z scale.

(b) Find the value of z^* such that 99% of the observations in the population are between $-z^*$ and $+z^*$ on the Z scale.

4.S.9 In the nerve-cell activity of a certain individual fly, the time intervals between "spike" discharges follow approximately a normal distribution with mean 15.6 ms and standard deviation 0.4 ms (as in Example 4.1.3). Let Y denote a randomly selected interspike interval. Find

(a) $\Pr\{Y > 15\}$

(b) $\Pr\{Y > 16.5\}$

(c) $\Pr\{15 < Y < 16.5\}$

(d) $\Pr\{15 < Y < 15.5\}$

4.S.10 For the distribution of interspike-time intervals described in Exercise 4.S.9, find the quartiles and the interquartile range.

4.S.11 Among American women aged 20 to 29 years, 10% are less than 60.8 inches tall, 80% are between 60.8 and 67.6 inches tall, and 10% are more than 67.6 inches tall.[17] Assuming that the height distribution can adequately be approximated by a normal curve, find the mean and standard deviation of the distribution.

4.S.12 The intelligence quotient (IQ) score, as measured by the Stanford-Binet IQ test, is normally distributed in a certain population of children. The mean IQ score is 100 points, and the standard deviation is 16 points.[18] What percentage of children in the population have IQ scores

(a) 140 or more?

(b) 80 or less?

(c) between 80 and 120?

(d) between 80 and 140?

(e) between 120 and 140?

4.S.13 Refer to the IQ distribution of Exercise 4.S.12. Let Y be the IQ score of a child chosen at random from the population. Find $\Pr\{80 \le Y \le 140\}$.

4.S.14 Refer to the IQ distribution of Exercise 4.S.12. Suppose five children are to be chosen at random from the population. Find the probability that exactly one of them will have an IQ score of 80 or less and four will have scores higher than 80. (*Hint*: First find the probability that a randomly chosen child will have an IQ score of 80 or less.)

4.S.15 A certain assay for serum alanine aminotransferase (ALT) is rather imprecise. The results of repeated assays of a single specimen follow a normal distribution with mean equal to the true ALT concentration for that specimen and standard deviation equal to 4 U/l (see Example 2.2.12). Suppose that a certain hospital lab measures many specimens every day, performing one assay for each specimen, and that specimens with ALT readings of 40 U/l or more are flagged as "unusually high." If a patient's true ALT concentration is 35 U/l, what is the probability that his specimen will be flagged as "unusually high"?

4.S.16 Resting heart rate was measured for a group of subjects; the subjects then drank 6 ounces of coffee. Ten minutes later their heart rates were measured again. The change in heart rate followed a normal distribution, with a mean increase of 7.3 beats per minute and a standard deviation of 11.1.[19] Let Y denote the change in heart rate for a randomly selected person. Find

(a) $\Pr\{Y > 10\}$ (b) $\Pr\{Y > 20\}$
(c) $\Pr\{5 < Y < 15\}$

4.S.17 Refer to the heart rate distribution of Exercise 4.S.16. The fact that the standard deviation is greater than the average and that the distribution is normal tells us that some of the data values are negative, meaning that the person's heart rate went down, rather than up. Find the probability that a randomly chosen person's heart rate will go down. That is, find $\Pr\{Y < 0\}$.

4.S.18 Refer to the heart rate distribution of Exercise 4.S.16. Suppose we take a random sample of size 400 from this distribution. How many observations do we expect to obtain that fall between 0 and 15?

4.S.19 Refer to the heart rate distribution of Exercise 4.S.16. If we use the $1.5 \times \text{IQR}$ rule, from Chapter 2, to identify outliers, how large would an observation need to be in order to be labeled an outlier on the upper end?

4.S.20 It is claimed that the heart rates of Exercise 4.S.16 follow a normal distribution. If this is true, which of the following Shapiro–Wilk's test P-values for a random sample of 15 subjects are consistent with this claim?

(a) P-value $= 0.0149$ (b) P-value $= 0.1345$
(c) P-value $= 0.0498$ (d) P-value $= 0.0042$

4.S.21 The following four normal quantile plots, (a), (b), (c), and (d), were generated from the distributions shown by histograms I, II, and III and another histogram that is not shown. Which normal quantile plot goes with which histogram? How do you know? (There will be one normal quantile plot that is not used.)