6

Probability, Normal Distributions, and *z* Scores

LEARNING OBJECTIVES

After reading this chapter, you should be able to:

1. Identify eight characteristics of the normal distribution.
2. Define the standard normal distribution and the standard normal transformation.
3. Locate proportions of area under any normal curve above the mean, below the mean, and between two scores.
4. Locate scores in a normal distribution with a given probability.
5. Compute the normal approximation to the binomial distribution.
6. Convert raw scores to standard *z* scores using SPSS.

CHAPTER OUTLINE

6.1 The Normal Distribution in Behavioral Science

Behavioral data in a population tend to be normally distributed, meaning the data are symmetrically distributed around the mean, the median, and the mode, which are values at the center of the distribution. When data in the population are normally distributed, we use the *empirical rule* to determine the probability of an outcome in a research study.

6.2 Characteristics of the Normal Distribution

There are eight key characteristics that define a normal distribution.

1. *The normal distribution is mathematically defined:* An exact normal distribution is defined by a mathematical formula—however, rarely (if ever) do data fall exactly within the limits of the formula.
2. *The normal distribution is theoretical:* Because behavioral data rarely conform exactly to a normal distribution, the idea that data are normally distributed is theoretical.
3. *The mean, median, and mode are all located at the 50th percentile:* In a normal distribution, the mean, median, and mode are all located in the middle of the distribution at the 50th percentile, such that half the data fall above these measures and half fall below.
4. *The normal distribution is symmetrical:* The data in a normal distribution are equally distributed around the mean, median, and mode, making the distribution symmetrical.
5. *The mean can equal any value:* The mean of a normal distribution can be any number from positive to negative infinity.
6. *The standard deviation can equal any positive value:* The standard deviation of a normal distribution can be any positive number greater than 0.
7. *The total area under the curve of a normal distribution is equal to 1.00:* Half the area under the normal curve (or 50% of scores) falls above the mean, and half the area under the normal curve (or 50% of scores) falls below the mean. Hence, the total area under the normal curve is 1.00.
8. *The tails of a normal distribution are asymptotic:* The tails of the normal distribution are always approaching the *x*-axis but never touch it, allowing for the possibility of outliers in a normal distribution.

6.3 Research in Focus: The Statistical Norm

Approximately 95% of scores in a normal distribution will fall within 2 standard deviations (*SD*) of the mean. This percentage tends to be described in behavioral research as *the statistical norm*. The 5% of scores that fall outside of 2 *SD* of the mean are considered "not normal" because they fall far from the mean.

6.4 The Standard Normal Distribution

To determine the likelihood of obtaining a score in any normal distribution, researchers defined the standard normal distribution. All other normal distributions can then be converted to the standard normal distribution using the following *z* transformation formulas:

$$\text{Population } z \text{ transformation: } z = \frac{x - \mu}{\sigma}.$$

$$\text{Sample } z \text{ transformation: } z = \frac{x - M}{SD}.$$

- The standard normal distribution has a mean of 0 and a standard deviation of 1. The *z* distribution is distributed in *z*-score units along the *x*-axis.
- A *z* score is a value on the *x*-axis of the *z* distribution that specifies the distance from the mean in standard deviations.

6.5 The Unit Normal Table: A Brief Introduction

We can find the probability of obtaining a score when we know its *z* score in the standard normal distribution. The proportion of area or probability of *z* scores in the standard normal distribution is listed in the unit normal table (see Table B.1 in Appendix B of the book; also reprinted in the appendix of this study guide).

To read the *z* table, we use three columns:

- *Column A lists the* z *scores:* Column A in the *z* table lists only positive *z* scores. Because the *z* distribution is symmetrical, we know that a proportion associated with a *z* score above the mean is identical to that below the mean.
- *Column B lists the area between a* z *score and the mean:* As a *z* score is farther from the mean, the area under the curve between that *z* score and the mean increases.
- *Column C lists the area from a* z *score toward the tail:* As a *z* score is farther from the mean, the area under the curve between that *z* score and the tail decreases.

6.6 Locating Proportions

To locate the proportion, or probability, of a score in any normal distribution, we follow two steps:

- Step 1: Transform a raw score into a *z* score.
- Step 2: Locate the corresponding proportion for the *z* score in the unit normal table.

6.7 Locating Scores

To locate the raw score in a normal distribution based on the probability of that score occurring, we follow two steps:

- Step 1: Locate a *z* score associated with a given proportion in the unit normal table.
- Step 2: Transform the *z* score into a raw score (x).

6.8 SPSS in Focus: Converting Raw Scores to Standard *z* Scores

SPSS can be used to convert raw scores to standard *z* scores. After entering the data for a variable, we can convert raw scores to standard *z* scores using the Analyze, Descriptive Statistics, and Descriptives options in the menu bar. The *z* scores will appear in the Data View tab.

6.9 Going From Binomial to Normal

The standard normal distribution was derived from the binomial distribution. Thus, for a binomial distribution, as long as the sample size (n) is greater than or equal to 2, and as long as both np and nq (the sample size times the probability of the first outcome, p, and the sample size times the probability of the complementary outcome, q) are greater than 10, the binomial distribution will approximate a normal distribution.

6.10 The Normal Approximation to the Binomial Distribution

There are five steps to find the probability of a bivariate outcome.

- *Step 1: Check for normality:* If both np and nq are greater than 10, then the binomial distribution approximates a normal distribution.

- *Step 2: Compute the mean and standard deviation:* Compute the mean and the standard deviation of a binomial distribution using the following formulas, respectively:

$$\mu = np.$$

$$\sigma = \sqrt{npq}\ .$$

- *Step 3: Find the real limits:* The *real limits* for any binomial variable are the upper and lower values within which the probability of obtaining an outcome is contained. This can be defined as $x \pm 0.5$.
- *Step 4: Locate the* z *score for each real limit:* Calculate the z score for each real limit computed in Step 3.
- *Step 5: Find the proportion located within the real limits:* Find the proportion within the real limits using the unit normal table.

CHAPTER FORMULAS

z Transformations

$$z = \frac{x - \mu}{\sigma} \quad (z \text{ transformation for a population of scores})$$

$$z = \frac{x - M}{SD} \quad (z \text{ transformation for a sample of scores})$$

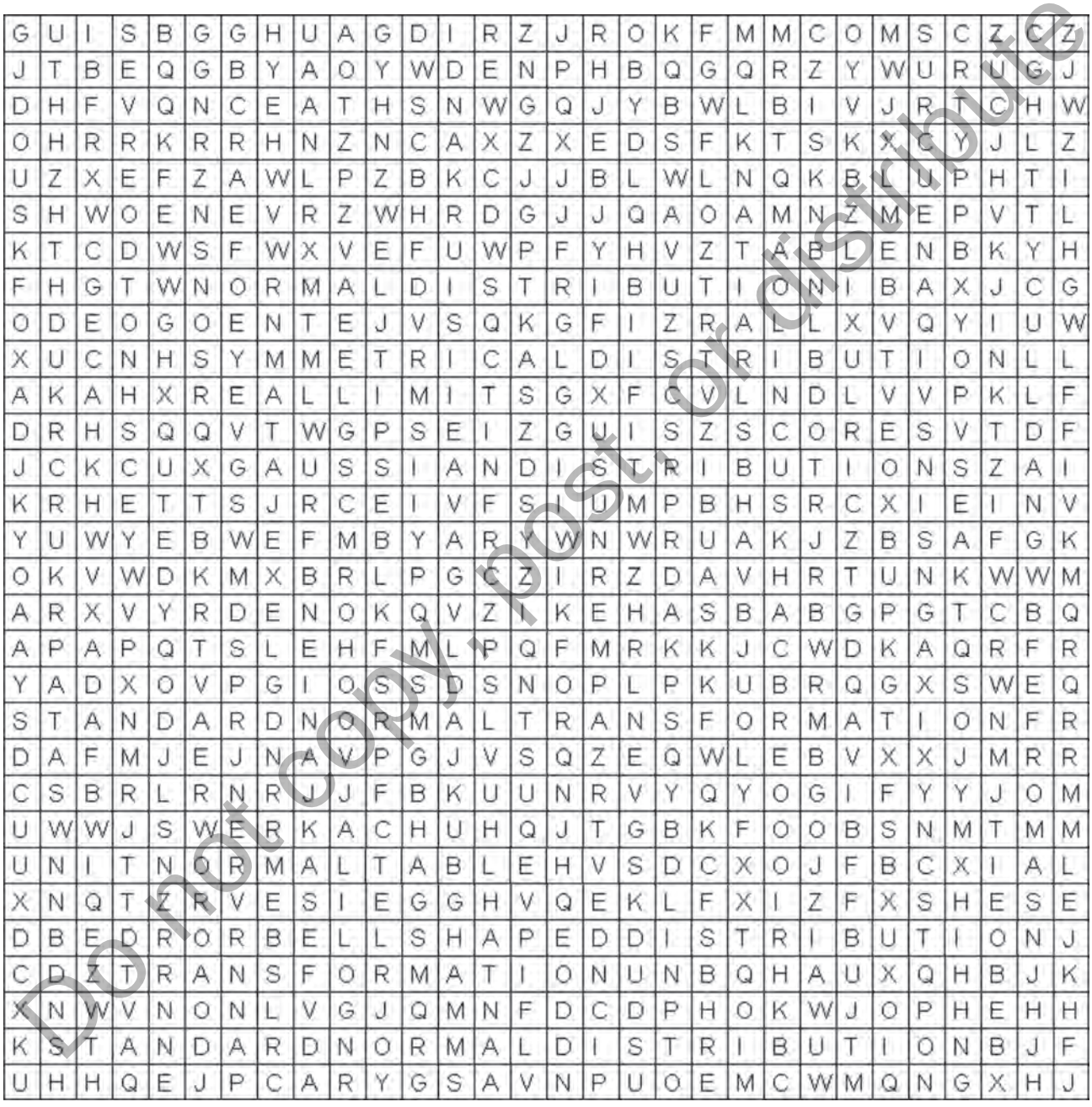
TIPS AND CAUTIONS FOR STUDENTS

- *Characteristics of the normal distribution:* Remember that the normal distribution is a theoretical distribution. In truth, most behavior does not exactly fit within the limits of the equation used to define the normal distribution. Instead, the behaviors we observe approximate the shape of a normal distribution. When we use inferential statistics, we will make an assumption that the populations we are interested in studying have data that are approximately normally distributed. As a general rule, as long as data in the population approximate a normal distribution, we can safely make this assumption.

- z *distribution:* Converting a score in any normal distribution to a *z* score will give you the position of that score in the standard normal distribution. The *z* score, which is a standard deviation, tells you the number of standard deviations that a score deviates from the mean in that distribution. Larger *z* scores are associated with a smaller probability of occurrence. Hence, we can use *z* scores to determine the distance that a score deviates from the mean and the probability of its occurrence.

- *Locating proportions:* When using the unit normal table to look up proportions, do not worry about whether the *z* score is negative or positive. Because the *z* distribution is symmetrical, any proportion in the positive tail of the distribution will be mirrored in the negative tail. This is why only positive values are listed in the unit normal table. Keep in mind that if the *z* score is negative, then the proportion, or area under the curve, will fall in the left or bottom half of the distribution.

KEY TERM WORD SEARCHES



G	U	I	S	B	G	G	H	U	A	G	D	I	R	Z	J	R	O	K	F	M	M	C	O	M	S	C	Z	C	Z
J	T	B	E	Q	G	B	Y	A	O	Y	W	D	E	N	P	H	B	Q	G	Q	R	Z	Y	W	U	R	U	G	J
D	H	F	V	Q	N	C	E	A	T	H	S	N	W	G	Q	J	Y	B	W	L	B	I	V	J	R	T	C	H	W
O	H	R	R	K	R	R	H	N	Z	N	C	A	X	Z	X	E	D	S	F	K	T	S	K	X	C	Y	J	L	Z
U	Z	X	E	F	Z	A	W	L	P	Z	B	K	C	J	J	B	L	W	L	N	Q	K	B	L	U	P	H	T	I
S	H	W	O	E	N	E	V	R	Z	W	H	R	D	G	J	J	Q	A	O	A	M	N	Z	M	E	P	V	T	L
K	T	C	D	W	S	F	W	X	V	E	F	U	W	P	F	Y	H	V	Z	T	A	B	L	E	N	B	K	Y	H
F	H	G	T	W	N	O	R	M	A	L	D	I	S	T	R	I	B	U	T	I	O	N	I	B	A	X	J	C	G
O	D	E	O	G	O	E	N	T	E	J	V	S	Q	K	G	F	I	Z	R	A	L	L	X	V	Q	Y	I	U	W
X	U	C	N	H	S	Y	M	M	E	T	R	I	C	A	L	D	I	S	T	R	I	B	U	T	I	O	N	L	L
A	K	A	H	X	R	E	A	L	L	I	M	I	T	S	G	X	F	C	V	L	N	D	L	V	V	P	K	L	F
D	R	H	S	Q	Q	V	T	W	G	P	S	E	I	Z	G	U	I	S	Z	S	C	O	R	E	S	V	T	D	F
J	C	K	C	U	X	G	A	U	S	S	I	A	N	D	I	S	T	R	I	B	U	T	I	O	N	S	Z	A	I
K	R	H	E	T	T	S	J	R	C	E	I	V	F	S	I	U	M	P	B	H	S	R	C	X	I	E	I	N	V
Y	U	W	Y	E	B	W	E	F	M	B	Y	A	R	Y	W	N	W	R	U	A	K	J	Z	B	S	A	F	G	K
O	K	V	W	D	K	M	X	B	R	L	P	G	C	Z	I	R	Z	D	A	V	H	R	T	U	N	K	W	W	M
A	R	X	V	Y	R	D	E	N	O	K	Q	V	Z	I	K	E	H	A	S	B	A	B	G	P	G	T	C	B	Q
A	P	A	P	Q	T	S	L	E	H	F	M	L	P	Q	F	M	R	K	K	J	C	W	D	K	A	Q	R	F	R
Y	A	D	X	O	V	P	G	I	O	S	S	D	S	N	O	P	L	P	K	U	B	R	Q	G	X	S	W	E	Q
S	T	A	N	D	A	R	D	N	O	R	M	A	L	T	R	A	N	S	F	O	R	M	A	T	I	O	N	F	R
D	A	F	M	J	E	J	N	A	V	P	G	J	V	S	Q	Z	E	Q	W	L	E	B	V	X	X	J	M	R	R
C	S	B	R	L	R	N	R	J	J	F	B	K	U	U	N	R	V	Y	Q	Y	O	G	I	F	Y	Y	J	O	M
U	W	W	J	S	W	E	R	K	A	C	H	U	H	Q	J	T	G	B	K	F	O	O	B	S	N	M	T	M	M
U	N	I	T	N	O	R	M	A	L	T	A	B	L	E	H	V	S	D	C	X	O	J	F	B	C	X	I	A	L
X	N	Q	T	Z	R	V	E	S	I	E	G	G	H	V	Q	E	K	L	F	X	I	Z	F	X	S	H	E	S	E
D	B	E	D	R	O	R	B	E	L	L	S	H	A	P	E	D	D	I	S	T	R	I	B	U	T	I	O	N	J
C	D	Z	T	R	A	N	S	F	O	R	M	A	T	I	O	N	U	N	B	Q	H	A	U	X	Q	H	B	J	K
X	N	W	V	N	O	N	L	V	G	J	Q	M	N	F	D	C	D	P	H	O	K	W	J	O	P	H	E	H	H
K	S	T	A	N	D	A	R	D	N	O	R	M	A	L	D	I	S	T	R	I	B	U	T	I	O	N	B	J	F
U	H	H	Q	E	J	P	C	A	R	Y	G	S	A	V	N	P	U	O	E	M	C	W	M	Q	N	G	X	H	J

BELL-SHAPED DISTRIBUTION

GAUSSIAN DISTRIBUTION

NORMAL DISTRIBUTION

REAL LIMITS

STANDARD NORMAL DISTRIBUTION

STANDARD NORMAL TRANSFORMATION

SYMMETRICAL DISTRIBUTION

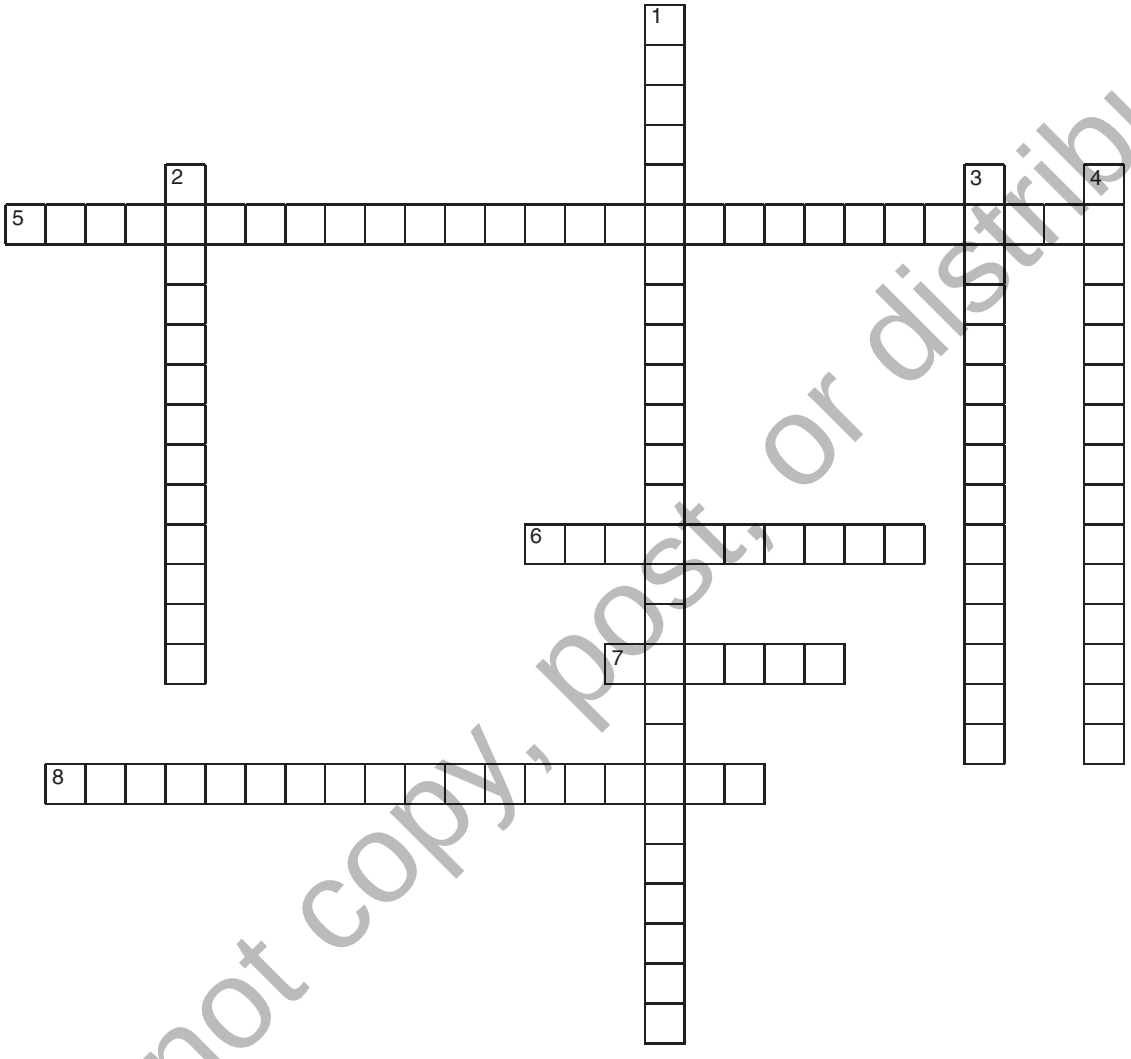
UNIT NORMAL TABLE

Z SCORES

Z TABLE

Z TRANSFORMATION

CROSSWORD PUZZLES



ACROSS

5 A formula used to convert any normal distribution with any mean and any variance to a standard normal distribution with a mean equal to 0 and a standard deviation equal to 1.

6 The upper and lower values within which the probability for obtaining the outcome of a binomial variable is contained.

7 A value on the *x*-axis of a standard normal distribution.

8 A theoretical distribution with data that are symmetrically distributed around the mean, the median, and the mode.

DOWN

1 A normal distribution with a mean equal to 0 and a standard deviation equal to 1.

2 A different name for the standard normal distribution.

3 A different name for the standard normal transformation.

4 A type of probability distribution table displaying a list of *z* scores and the corresponding probabilities associated with each *z* score listed.

PRACTICE QUIZZES

LO 1: Identify eight characteristics of the normal distribution.

1. A theoretical distribution in which scores are symmetrically distributed above and below the mean, the median, and the mode at the center of the distribution is called a:
   a. normal distribution
   b. fixed distribution
   c. skewed distribution
   d. bimodal distribution

2. The mean, the median, and the mode are located _____ of a normal distribution.
   a. at the extremes
   b. at the 50th percentile
   c. in the upper tail
   d. in different locations

3. The distribution of data above the mean in a normal distribution is the same as the distribution of data below the mean. Hence, the normal distribution is:
   a. disproportional
   b. skewed
   c. asymptotic
   d. symmetrical

4. The tails of the normal distribution are always approaching the *x*-axis but never touch it. Hence, the normal distribution is:
   a. disproportional
   b. symmetrical
   c. asymptotic
   d. skewed

5. Which of the following statements about the normal distribution is true?
   a. the mean can equal any value
   b. the standard deviation can equal any positive value
   c. the total area under the normal curve is equal to 1.0
   d. all of the above

LO 2: Define the standard normal distribution and the standard normal transformation.

6. A normal distribution with a mean of 0 and a standard deviation of 1 is called a:
   a. small normal distribution
   b. minimal normal distribution
   c. standard normal distribution
   d. centered normal distribution

7. A normal distribution has a mean of 10 and a standard deviation of 2. What is the *z* score for $x = 13$?
   a. 0.7
   b. 1.5
   c. 3.0
   d. 13

8. A normal distribution has a mean of 7.5 and a variance of 4. What is the *z* score for $x = 9.5$?
   a. 0.5
   b. 1.0
   c. 1.5
   d. 2.0

LO 3: Locate proportions of area under any normal curve above the mean, below the mean, and between two scores.

9. A normal distribution has a mean equal to 3.4. What is the probability of a score above the mean?
   a. 0
   b. 0.5
   c. 1.0
   d. not enough information

10. The unit normal table does not have negative *z* scores listed. Why?
    a. because it is not possible to observe negative *z* scores
    b. because negative *z* scores become positive when computed using a *z* transformation
    c. because a normal distribution is symmetrical, so proportions above and below the mean at any given *z* score are the same
    d. because probabilities are negative for negative *z* scores, so probabilities do not have meaning when *z* scores are negative

11. A normal distribution has a mean equal to 10 and a standard deviation equal to 2. Which outcome is more probable?
    a. obtaining a score between 10 and 12
    b. obtaining a score between 11 and 13
    c. none; they are equally probable
12. A normal distribution of quiz scores has a mean equal to 6 and a variance equal to 1.5. Which outcome is more probable?
    a. scoring above a 3.8 on this quiz
    b. scoring below an 8.2 on this quiz
    c. none; they are equally probable
13. A researcher gives two practice exams, with scores normally distributed on each exam. If Exam 1 is distributed as 71 ± 9.2 ($M \pm SD$) and Exam 2 is distributed as 75 ± 4.8 ($M \pm SD$), then on which exam is a score of 80 more likely?
    a. Exam 1
    b. Exam 2
    c. none; a score of 80 is equally likely

LO 4: Locate scores in a normal distribution with a given probability.

14. A normal distribution has a mean equal to 4.6 and a standard deviation equal to 1.4. What score is the cutoff for the top 10% of this distribution?
    a. .10
    b. 1.28
    c. 6.00
    d. 6.39
15. When computing a *z* transformation to find a score at a given probability for a distribution with a given mean and standard deviation, which value in the *z*-transformation formula is a variable (i.e., which value is unknown)?
    a. z
    b. M
    c. x
    d. SD
16. A normal distribution of memory recall scores has a mean equal to 12 items recalled and a standard deviation equal to 2.3. If Participant A scores a 16.6 and Participant B scores a 15.2, then which participant scored in the top 10% on this memory recall test?
    a. Participant A
    b. Participant B
    c. none; they both failed to score in the top 10%
    d. both; they both scored in the top 10%

17. A normal distribution has a mean equal to 10 and a standard deviation equal to 1. What is the *z*-score cutoff for the top and bottom 2.5% of this distribution?

    a. ±1.28

    b. ±1.96

    c. ±1.645

    d. ±10.96

LO 5: Compute the normal approximation to the binomial distribution.

18. The normal distribution was derived from the _____ distribution.

    a. skewed

    b. bimodal

    c. binomial

    d. standard

19. Suppose we select a sample from a population of children in which an equal number of children live with or without their biological parents. We select a sample of 50 children. What is the probability that 30 children in this sample will be children who live with their biological parents? *Hint:* Use the normal approximation to the binomial distribution, where p and q equal .50.

    a. .0414

    b. .414

    c. .50

20. Suppose that 10% of students earn an A in a statistics course. Using this percentage, what is the probability that exactly 25 students in a random sample of 200 students will earn an A in a statistics course? *Hint:* Use the normal approximation to the binomial distribution, where p = .10 (the probability that a student earns an A) and q = .90 (the probability that a student does not earn an A).

    a. .48

    b. .048

    c. .10

    d. .90

SPSS IN FOCUS

Converting Raw Scores to Standard *z* Scores

Follow the General Instructions Guidebook to complete this exercise. Also, an example for following these steps is provided in the SPSS in Focus section (Section 6.8) of the book. Complete and submit the SPSS grading template and a printout of the output file.

Exercise 6.1: Standardizing Participant Reactions

A researcher has participants read a vignette that describes a person making inappropriate remarks toward a colleague at work, and then they are asked to complete a survey. Lower total scores on the survey indicate more negative impressions of the person described in the vignette; higher total scores indicate more positive impressions. Convert the following data to *z* scores to determine who scored above, below, and at the mean in this sample.

18	19	8	20
12	14	12	25
17	14	8	14
8	14	20	14
23	9	16	15
7	17	14	11
10	21	14	19
10	7	18	8
12	11	5	10
25	8	19	14

With regard to the SPSS exercise, answer the following questions:

Based on the table shown in SPSS, state the following values for the original data:

Sample size _____

Minimum score _____

Maximum score _____

Mean score _____

Standard deviation _____

Based on the *z* scores listed in the *z*-scores column (Data View), state the following values:

Minimum *z* score _____

Maximum *z* score _____

Mean *z* score _____

State how many scores were:

Above the mean _____

Below the mean _____

Equal to the mean _____

CHAPTER SUMMARY ORGANIZED BY LEARNING OBJECTIVE

LO 1: Identify eight characteristics of the normal distribution.

- The normal distribution is a theoretical distribution in which scores are symmetrically distributed above and below the mean, the median, and the mode at the center of the distribution.
- Eight characteristics of a normal distribution are as follows:

  1. The normal distribution is mathematically defined.
  2. The normal distribution is theoretical.
  3. The mean, the median, and the mode are all located at the 50th percentile.
  4. The normal distribution is symmetrical.
  5. The mean can equal any value.
  6. The standard deviation can equal any positive value.
  7. The total area under the curve of a normal distribution is equal to 1.00.
  8. The tails of a normal distribution are asymptotic.

LO 2: Define the standard normal distribution and the standard normal transformation.

- The standard normal distribution, or z distribution, is a normal distribution with a mean equal to 0 and a standard deviation equal to 1. The standard normal distribution is distributed in z score units along the x-axis.
- The standard normal transformation, or z transformation, is an equation that converts any normal distribution with any mean and any positive standard deviation into a standard normal distribution with a mean equal to 0 and a standard deviation equal to 1:

  For a population: $z = \frac{x - \mu}{\sigma}$.

  For a sample: $z = \frac{x - M}{SD}$.

- The probabilities of z scores in a standard normal distribution are listed in the unit normal table in Table B.1 in Appendix B.

LO 3: Locate proportions of area under any normal curve above the mean, below the mean, and between two scores.

- To locate the proportion, and therefore probabilities, for scores in any normal distribution, we follow two steps:

  Step 1: Transform a raw score (x) into a z score.

  Step 2: Locate the corresponding proportion for the z score in the unit normal table.

LO 4: Locate scores in a normal distribution with a given probability.

- To locate scores that fall within a given proportion, or probability, we follow two steps:

  Step 1: Locate a z score associated with a given proportion in the unit normal table.

  Step 2: Transform the z score into a raw score (x).

LO 5: Compute the normal approximation to the binomial distribution.

- It is appropriate to use the normal distribution to approximate or estimate binomial probabilities when $np > 10$ and $nq > 10$.
- To use the normal approximation to estimate probabilities in a binomial distribution, follow five steps:

  Step 1: Check for normality.

  Step 2: Compute the mean and standard deviation.

  Step 3: Find the real limits.

  Step 4: Locate the *z* score for each real limit.

  Step 5: Find the proportion located within the real limits.

LO 6: Convert raw scores to standard *z* scores using SPSS.

- SPSS can be used to convert raw scores to standard *z* scores. After entering the data for each variable, raw scores are converted to standard *z* scores using the Analyze, Descriptive Statistics, and Descriptives options in the menu bar. These actions will allow you to select the "Save standardized values as variables" option to convert raw scores to standard *z* scores.