Section 26 Probability and the Normal Curve

With the empirical approach to knowledge, we make observations and, based on them, make decisions. From previous observations, we can establish probabilities regarding the occurrence of specific events in the future. Weather forecasting is based on this approach. Events such as high- and low-pressure systems are observed, and predictions are based on previous observations of their effects on the weather.

Fortunately, for many problems, empirical probabilities are easy to determine because many distributions are normal.[1] Suppose that we conducted a large national survey to determine knowledge of basic math skills. Suppose we found that the distribution of math scores was normal, the mean was 50.00, and the standard deviation was 7.00. This information can be used to establish probabilities. To do so, we will need to use *z*-scores. (See Section 18 to review *z*-scores.) You may recall that the formula for them is

$$z = \frac{X - M}{S}$$

In this example, what is the probability of drawing an individual who has a score of 64 or higher at random from the population? To answer the question, first calculate the corresponding *z*-score (keeping in mind that the mean is 50.00 and the standard deviation is 7.00).

$$z = \frac{64 - 50.00}{7.00} = \frac{14}{7} = 2.00$$

Second, look up the *z*-score in Table 1 near the end of this book. There, we find that the proportion of scores at or less than 2 is .9772. Because we are interested in scores that are higher than 2, we can subtract .9772 from 1 to get .0228 (or 2.28%). Our probability of .0228 indicates that there are only slightly more than 2 chances in 100 of drawing a person with a score of 64 or higher from the population. This is referred to as a ***one-tailed probability*** because we asked the question about

[1]The normal curve was first introduced in Section 10 and was discussed in Sections 15 through 20.

only the upper tail of the normal distribution—the right-hand tail of the distribution in Figure 1.

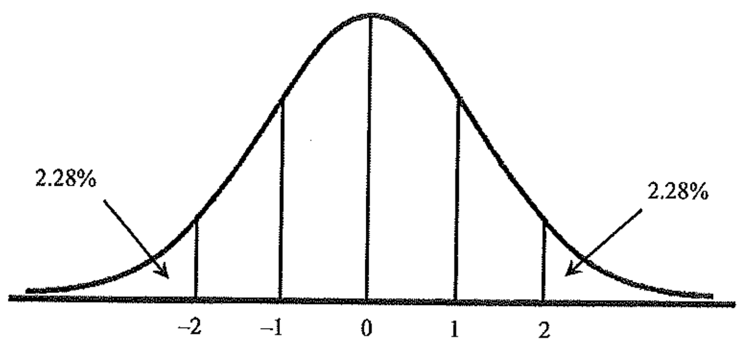

Figure 1. Normal distribution with selected *z*-scores.

Suppose, instead, that we asked this question: What is the probability of randomly drawing an individual with a *z*-score as high as 2.00 (or higher) *or* as low as –2.00 (or lower)? Obviously, the odds of doing so are double those of drawing just one of these. Thus, the odds are 2 × .0228 = .0456, which is a little more than 4 in 100. This is called a ***two-tailed probability*** because we are asking about the odds of drawing an individual at either tail of the normal distribution (see the percentages associated with each tail in Figure 1). The importance of distinguishing between one-tailed and two-tailed probabilities will become clear in later sections of this book.

The probabilities for both events described above represent ***unlikely events***. In most sciences, conventional wisdom indicates that any event that has a probability of occurrence of .05 or less is usually classified as unlikely to occur at random. A *z*-score of 1.96 has only a 2.5% chance of occurrence as a one-tailed probability. As a two-tailed probability, it has a 5.0% chance of occurrence (2 × 2.5% = 5.0%). Thus, an event with a *z*-score of 1.96 or greater or –1.96 or less (such as 1.97 or –1.97) is classified as an unlikely event.

A *z*-score of 2.58 or higher has only a .49% (or just less than 1/2 of 1%) chance of occurrence. The corresponding two-tailed probability is .98%, or almost 1%, for *z*-scores of 2.58 or –2.58. Some researchers only classify an event as unlikely if its likelihood is 1% or less.

There is no rule of nature that says at what point an event should be classified as unlikely. However, the 5% and 1% guidelines have evolved over time as the two most widely used. The only universally accepted rule is that a researcher must decide *in advance* of examining the data what guideline will be followed for declaring an event unlikely. Theoretically, any percentage may be specified, but to be accepted in most scientific circles, 5% or lower is generally used.

Identifying unlikely events is the basis of many of the tests of statistical significance presented in later sections of this book.

Terms to Review Before Attempting Worksheet 26

one-tailed probability, two-tailed probability, unlikely events

"I have a photographic memory—
I just seem to be out of film today."

Worksheet 26 Probability and the Normal Curve

***Riddle*: What did George Washington do when he was asked for his ID?**

DIRECTIONS: To find the answer to the riddle, write the answer to each question in the space immediately below it. In the solution section, the word in parentheses next to the answer to the first question is the first word in the answer to the riddle, the word beside the answer to the second question is the second word, and so on.

1. What is the one-tailed probability of drawing a subject with a z-score of 1.35 or higher at random from a normal distribution?

2. What is the one-tailed probability of drawing a subject with a z-score of -1.70 or lower at random from a normal distribution?

3. What is the two-tailed probability of drawing a subject with a z-score as extreme as 1.80 *or* -1.80 at random from a normal distribution?

4. For a normal distribution with a mean of 100.00 and a standard deviation of 16.00, what is the one-tailed probability of drawing a subject with a score of 124 or greater at random from a normal distribution?

5. For a normal distribution with a mean of 40.00 and a standard deviation of 8.00, what is the one-tailed probability of drawing a subject with a score of 30 or less at random from a normal distribution?

Worksheet 26 (continued)

6. According to the information in Question 5, what is the probability of drawing a subject with a score as extreme as 30 or 50 at random from a normal distribution?

7. According to the 5% guideline, should the answer to Question 2 be classified as an unlikely event?

8. According to the 1% guideline, should the answer to Question 5 be classified as an unlikely event?

Solution section:

.4554 (Congress)   no (it)   .0359 (public)   yes (showed)   .0718 (out)
.2112 (and)   .1336 (fooling)   .1056 (quarter)   .1770 (deficit)
.0668 (a)   .0885 (he)   .9554 (politicians)   .0446 (whipped)   .4115 (saying)

Write the answer to the riddle here, putting one word on each line: ________ ________ ________ ________

________ ________ ________ ________

Section 27 Percentiles and the Normal Curve

The major advantage of using standard scores is that you can use them to compare measures from different scales. For instance, suppose Hannah takes two college entrance exams—one has a total possible of 36 points (the ACT) and the other has a total possible of 2,400 points (the SAT). She can convert her raw scores to standard scores to compare the two. However, sometimes the comparison is more intuitive when using percentile ranks. A ***percentile rank*** is the percentage of scores at or below a given score. If Hannah's ACT score is in the 95th percentile, then 95% of the people who took the same test scored at or below Hannah. If her SAT score is in the 81st percentile, then we know that she did better on the ACT than the SAT.

Sometimes we need to convert from a percentile rank to a raw score. For instance, assume that Noah wants to apply to a college that only accepts students who score above the 90th percentile on the SAT. He has not taken the exam yet and wants to know what his target SAT score is. We can find that number by reversing the steps we used in the previous section.

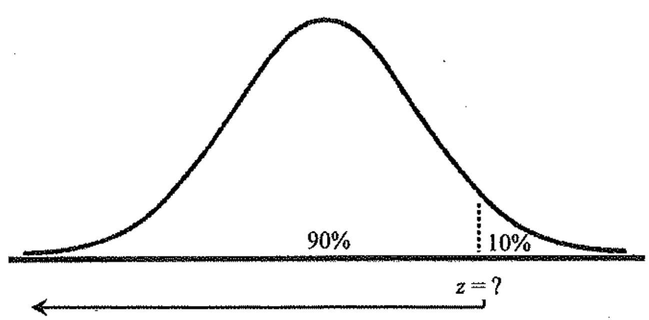

Figure 1. A score in the 90th percentile.

To review the use of Table 1 near the end of this book, we use both the left-most column and the top row as labels for z. For instance, the proportion of scores to the left of a z-value of 0.12 can be found in the row for .010 and the column for .02 (a proportion of .5478).

In Noah's case, the 90th percentile is above the mean (i.e., the 50th percentile), so we look in the positive values section of Table 1. Instead of using the row and column labels as we did when we knew the value of z, we will now search through the proportions in the middle of the table. The 90th percentile will have a value of .9000, and we want to find a value as close to that number as possible. There are two proportions that come very close: .8997 and .9015. Because .8997 is closest to our target, we use the row and column labels to identify the associated *z*-value of 1.28. At this point, we are not finished because we still want to know the SAT score (i.e., the raw score, or *X*-value) that Noah needs to obtain.

Recall that when calculating *z*-scores from *X*-values, we used the following formula:

$$z = \frac{X - M}{S}$$

To solve for *X*, the equation is changed to the following:

$$X = z(S) + M$$

If we know that the mean SAT score is 1,500 with a standard deviation of 100, then we can find the raw score necessary for Noah to earn on the SAT test:

$$X = 1.28(100) + 1{,}500 = 1{,}628$$

Term to Review Before Attempting Worksheet 27

percentile rank

Worksheet 27 Percentiles and the Normal Curve

***Riddle*: Why are math books always unhappy?**

DIRECTIONS: To find the answer to the riddle, write the answer to each question in the space immediately below it. In the solution section, the word in parentheses next to the answer to the first question is the first word in the answer to the riddle, the word beside the answer to the second question is the second word, and so on.

1. A percentile rank is the percentage of scores ______ a given score.

2. What equation do we use when finding the raw score?

Assume that the mean height is 70 inches with a standard deviation of 6 inches.

3. If Anne is 82 inches tall, what is her percentile rank?

4. If Kendra is 64 inches tall, what is her percentile rank?

5. If James is in the 77th percentile, what is his height?

6. If Alex is in the 25th percentile, what is his height?

7. What percentage of people is between 66 and 82 inches?

Worksheet 27 (continued)

8. What percentage of people is between 64 and 76 inches?

Solution section:

76 (puzzle)	98 (always)	$(X - M)/S$ (never)	16 (have)	near (rather)	73 (of)
68 (problems)	$z(S) + M$ (they)	2 (the)	64.5 (binder)	66 (lot)	
at or below (because)	50 (since)	74 (a)			

Write the answer to the riddle here, putting one word on each line: ________ ________ ________ ________

________ ________ ________ ________