CHAPTER 7

Theoretical Distributions Including the Normal Distribution

Summary

Empirical distributions are frequency distributions of observed scores. ***Theoretical distributions*** are distributions based on logic or mathematical formulas and they are "best estimates" of how events would occur. Chapter 7 focuses on finding probabilities by using theoretical distributions and then using those probabilities to answer questions about empirical distributions.

The concept of probability is central to understanding inferential statistics. To find the empirical probability of an event, form a ratio with successes in the numerator and the total of successes and failures in the denominator. To find the theoretical probability of an event, determine the proportion of a theoretical curve that corresponds to the events in question. With the theoretical curves in this chapter, the probability of an event is equal to the proportion of the curve that corresponds to that event. Remember probabilities can range from 0.0000 to 1.0000 and to express probabilities out to three or four decimals (this differs from other values in the text that are expressed out to only two decimals).

Three different theoretical distributions are described - ***rectangular distributions*** (for example, playing cards), ***binomial distributions*** (for example, flipping a coin) and the ***normal distribution***. Each distribution allows us to know the probability of events (or sets of events).

The normal curve is a bell-shaped distribution (sometimes called the "bell curve"). For a normal distribution, the mean, median, and mode are the same number, which corresponds to the peak of the distribution. The unit of measurement on the *x*-axis is the *z* score, which extends from approximately -3 on the left side to 0 in the middle (corresponding to the mean) to approximately +3 on the right side.

Any empirical distribution that is not normally distributed can be converted to a normal distribution using *z* scores $\left(\frac{X-\mu}{\sigma}\right)$. Using *z* scores and the table of proportions of the normal curve (Table C), you can find:

1. the proportion of the empirical distribution that is beyond a particular score
2. the proportion of the empirical distribution that is between two particular scores
3. a particular score that separates out a proportion of the curve
4. two particular scores that encompass a given proportion of the curve
5. the number of cases with particular scores
6. the number of cases within a particular proportion of the curve

The best way to start these problems is to draw a picture of the normal curve, identify the mean of the population (μ), draw a line corresponding to the value identified in the problem (with values above the mean on the right side of the curve and values below the mean on the left side of the curve), and shade the area identified in the problem. Some of the later problems in the chapter will require you to find the probabilities of two areas under the curve, so it can be helpful to have pens or highlighters in multiple colors to help you keep track of the areas you are adding or subtracting.

Multiple-Choice Questions

1. The difference between an empirical distribution and a theoretical distribution is that a theoretical distribution ______.

   a. is based on many more observations
   b. is theory and cannot be used
   c. is based on mathematics and logic
   d. is based solely on observations

2. Which of the following is an empirical distribution?

a. the given names and their frequencies of all high school graduates in the United States for the year 2017
b. the scores expected from an infinite number of throws of one die
c. the normal distribution in Table C in your text
d. all of the above

3. To use the theoretical normal curve as described in this chapter, which of the following things must be known about the population?

a. mean
b. standard deviation
c. the form of the distribution
d. all of the above

4. The area under the curve of a normal distribution is ______.

a. dependent on the number of frequencies
b. dependent on the size of the mean
c. 1.00
d. none of the above

5. When the unit of measurement is *z* scores, the theoretical normal curve has a mean equal to ______ and a standard deviation equal to ______.

a. 1.00; 0.00
b. 0.00; 1.00
c. 1.00; 1.00
d. 1.00; the standard deviation of the population

6. If you were given one *z* score from a population of measurements and nothing else, you could determine ______.

a. the mean of the population
b. the standard deviation of the population
c. both a. and b.
d. neither a. nor b.

7. .4332 of the normal curve lies between μ and 1.5σ. The proportion between μ and $.75\sigma$ is ______.

a. .8664
b. .2166
c. .0668
d. none of the above

8. If a normal distribution of empirical scores is converted to a distribution of *z* scores, ______.

a. the new mean will be 0.00
b. the new standard deviation will be 1.00
c. both a. and b.
d. neither a. nor b.

9. Which of the following is a theoretical distribution?

a. a professor flipped a coin 10,000 times and the number of heads and tails was recorded
b. the price of every house sold in the last five years in Hampden County was obtained from courthouse records
c. the number of students who arrived late was recorded every time a statistics course met during the semester
d. none of the above

10. Kyle and James were discussing scores on the admission test for the Gifted and Talented program at their child's elementary school. The scores are distributed normally with a mean of 50. They agreed that 10% of the population had scores of 60 or better (and they were correct on this). Kyle also claimed that 10% of the population had scores of 40 or below.

a. Kyle is correct
b. Kyle is correct, but only because each score point is worth one percentage point
c. Kyle is mistaken
d. cannot be determined from the information given

11. Continuing the example of the admission test for the Gifted and Talented program, Kyle claimed that, because 10% of the population had scores of 60 or better, which is 10 points from the mean, 5% must have had scores of 70 or better, because doubling the score distance always halves the percentage.

a. Kyle is correct
b. Kyle is mistaken
c. Kyle is correct, but only because each score point is worth one percentage point
d. cannot be determined from the information given

12. Suppose the mean of a particular normal distribution is 3.95. The median of this distribution will be ______.

a. larger than 3.95
b. smaller than 3.95
c. 3.95
d. cannot be determined from the information given

13. Suppose that if *k* should occur, it will be called a success. If *j* should occur it will be called a failure. The ratio $\frac{k}{k+j}$ is ______.

a. the empirical probability of *k*
b. the empirical probability of *j*
c. the theoretical probability of *k*
d. the theoretical probability of *j*

14. The advantage of calculating *z* scores is that they allow you to ______.

a. know the sample mean
b. know the sample standard deviation
c. compare scores that come from different distributions
d. none of the above

15. If you flip a coin a thousand times and plot the results, the distribution will be ______.

a. a normal distribution
b. a rectangular distribution
c. a relative distribution
d. a binomial distribution

16. Suppose a student has a *z* score of -1.00. That means she ______.

a. scored above the mean
b. had a negative raw score
c. scored 1 standard deviation below the mean
d. had the lowest score in the group

17. The area under part of a theoretical curve is ______.

a. the mean of the curve
b. the standard deviation of the curve
c. the probability of events covered by that area of the curve
d. cannot be determined from the information given

18. Rectangular, binomial, and normal curves have the following in common: ______.

a. the most frequent score is in the middle of the distribution
b. Table C gives the probabilities for all three curves
c. the area under the curve is 1.00
d. all of the above

19. The normal curve, as it approaches the *x*-axis, is reaching ______.

a. a binomial
b. an inflection point
c. an asymptote
d. the mean

20. If you knew that scores were distributed normally and you wanted to find the number of people in a population who scored between 150 and 175, you would need to know

a. the number of people in the population
b. the mean of the population
c. the standard deviation of the population
d. all of the above

Short-Answer Questions

1. Distinguish between theoretical and empirical distributions.

2. Suppose a small college has 500 freshmen, 400 sophomores, 300 juniors, and 200 seniors and that you randomly pick one student. What is the probability that the student will be ______?
a. a freshman
b. a senior
c. a sophomore or a junior
d. not a senior

3. What is the difference between a binomial distribution and a normal curve?

4. Write a paragraph describing the normal curve.

Problems

1. Identify each of the following distributions as theoretical or empirical.
    a. Students determined the hypothetical probability of tossing a quarter ten times and getting ten heads in a row.
    b. At Collegiate University, the Registrar recorded the grade point average of every freshman for the years 1985, 1995, 2005, and 2015.
    c. For all Saturdays since the college began, the proportion of rainy days was determined from official weather records.
    d. Each offspring of a single fruit fly was classified as red-eyed or white-eyed.
    e. The length of pine tree needles is normally distributed and a researcher identified the proportion that would be expected to be more than two standard deviations shorter than the mean.
    f. Researchers recorded the weight of the first ten trucks at a weigh-in station on a weekday morning and a weekend morning.

2. Suppose you live in a state with a lottery system that selects 6 numbers from 36 balls (labeled 1-36) to determine a winner. The order of selection does not influence the outcome. To win, you simply have to match the 6 numbers that the lottery commission selects. What are the odds of you winning that lottery by purchasing one ticket?

3. Women in the U.S. Air Force have hands that are 6.9 inches long on average (from the tip of the middle finger to the heel of the hand). The standard deviation is .34 inches. Assume that hand length is normally distributed.
    a. What proportion would have hands 7 inches or longer?
    b. What proportion would have hands 6 inches or shorter?
    c. What proportion would have hands between 6.5 and 7.2 inches long?
    d. Of 3000 women in the U.S. Air Force, how many would have hands 7.5 inches (the mean for male Air Force personnel) or longer?
    e. How long would a woman's hand have to be to put her among the 10% with the longest hands in the U.S. Air Force?

4. Psychological and educational tests, including the MMPI (Minnesota Multiphasic Personality Inventory), are often designed so they have a mean of 50 and a standard deviation of 10.
    a. The MMPI says that scores that are two standard deviations or more above the mean may indicate pathology. Among a thousand persons, how many would be expected to have scores two standard deviations or more above the mean?
    b. What proportion would be expected to score 68 or above?
    c. What proportion would be expected to score between 53 and 63?
    d. What score separates out the lowest one-fifth from the rest?

5. Graduate Record Exam (GRE) scores for seniors taking the biology subject test are approximately normally distributed with a mean of 473 and a standard deviation of 85.
    a. What proportion of the population of biology seniors would have scores between 450 and 650?
    b. What proportion would have scores of 700 or above?
    c. If 10,000 biology majors took the test, how many would be expected to have scores of 670 (the highest score yet made by a recent State U. graduate) or above?
    d. Find the scores that separate the middle 95 percent of biology seniors from the lower 2 ½ percent and the upper 2 ½ percent.

6. On average it takes 27 throws to complete the College Frisbee Golf Course and the standard deviation is 4. Assume that the number of throws to complete the course is normally distributed.

a. What proportion of the population would be expected to score 22 or less?
b. What proportion would score between 25 and 30?
c. If 950 students played and a prize was given to those who scored 20 or less, how many would get prizes?
d. What proportion would be expected to score between 28 and 32?
e. What proportion would be expected to score 29 or less?

ANSWERS

Multiple-Choice Questions

1. c
2. a
3. d
4. c
5. b
6. d
   Explanation: While the population mean and standard deviation are both required to determine a *z* score, you have no way of knowing what they were by just looking at the *z* score.
7. d
   Explanation: It is tempting to assume that the proportion between μ and $.75\sigma$ would be half of the proportion between μ and 1.5σ. But, the shape of the normal curve shows that equal distances along the horizontal axis do not correspond to equal areas of the curve.
8. c
9. d
   Explanation: All of the distributions given are empirical distributions because they are derived from actual observations of scores.
10. a
   Explanation: Because the normal distribution is symmetrical, if 10% of the population had scores of 60 (which is 10 above the mean) or better, there will also be 10% of the population with scores of 40 (which is 10 below the mean) or below.
11. b
   Explanation: Because 70 is a more extreme score than 60, it is much less likely to occur. Because the shape of the normal distribution has scores clustered around the center and then become much less common after the inflection points, it's likely that the proportion of scores 70 or better is much smaller than 5%. If you have a hard time imagining this, try drawing it out and you will see that the shaded area is much smaller for people with scores of 70 or better.
12. c
   Explanation: Remember the mean, median, and mode of the normal distribution are always equal.
13. c
14. c
15. d
16. c
17. c
18. c
19. c
20. d

Short-Answer Questions

1. Theoretical distributions are based on math and logic and they represent our "best guess" of outcomes based on probability. Empirical distributions come from actual observations of outcomes.
2.
   a. *p* (a freshman) = .3571
   To answer this question, we need to take the number of freshmen and divide by the total number of students.

   $$\frac{500}{500+400+300+200} = \frac{500}{1400} = .3571$$

   b. *p* (a senior) = .1429
   To answer this question, we need to take the number of seniors and divide by the total number of students.

   $$\frac{200}{500+400+300+200} = \frac{200}{1400} = .1429$$

   c. *p* (a sophomore or a junior) = .5000

To answer this question, we need to add the number of sophomores and juniors and divide by the total number of students.

$\frac{400+300}{500+400+300+200} = \frac{700}{1400} = .5000$

d. p (not a senior) = .8571
To answer this question, we need to add the number of freshmen, sophomores, and juniors (all of the "not seniors") divide by the total number of students.

$\frac{500+400+300}{500+400+300+200} = \frac{1200}{1400} = .8571$

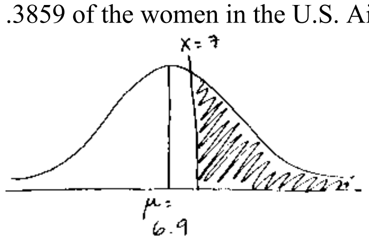
3. Binomial and normal distributions are both theoretical distributions. The primary difference between the two is that binomial distributions are based on the frequency of events when there are *only* two possible outcomes, while normal distributions are based on the frequency of events with multiple possible outcomes. There are other important properties of the normal distributions (for example, they are asymptotic) that are not shared by binomial distributions, but the primary difference is the number of possible outcomes.
4. The normal curve is a bell-shaped symmetrical distribution. The mean, median, and mode are all the same number. The points of inflection are at one standard deviation above and one standard deviation below the mean. The curve is asymptotic to the x-axis. You can use the normal curve to determine the probability of events.

Problems

1.
   a. theoretical
   b. empirical
   c. empirical
   d. empirical
   e. theoretical
   f. empirical

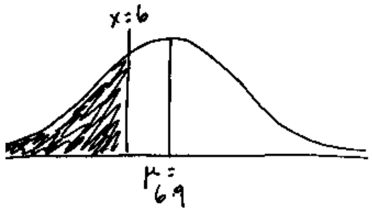
2. p (winning 6 number lottery with one ticket) = .000000513
When you want to know the probability of multiple events all occurring, multiply the probability of each event. The probability of matching the first number is 6/36 (you have chosen 6 numbers so any one of those could match the first number/the total number of choices), the probability of matching the next number is 5/35 (because there is one number already removed), and so forth until all six numbers are drawn.
$\frac{6}{36} \times \frac{5}{35} \times \frac{4}{34} \times \frac{3}{33} \times \frac{2}{32} \times \frac{1}{31} =$.000000513 (this is about one chance in two million)

3.
   a. .3859 of the women in the U.S. Air Force would have hands 7 inches or longer.



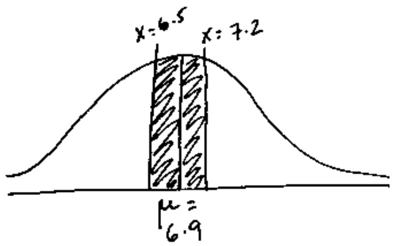
$z = \frac{7-6.9}{.34} = .29$
proportion from z to ∞ = .3859

b. .0040 of the women in the U.S. Air Force would have hands 6 inches or shorter.



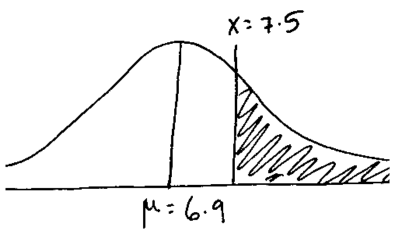
$z = \frac{6-6.9}{.34}$ = -2.65
proportion from z to ∞ = .0040

c. .6916 of the women in the U.S. Air Force would have hands between 6.5 and 7.2 inches long.



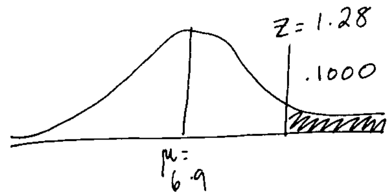
$z = \frac{6.5-6.9}{.34}$ = -1.18
proportion from μ to z = .3810
$z = \frac{7.2-6.9}{.34}$ = .88
proportion from μ to z = .3106
proportion from μ to z (z = -1.18) + proportion from μ to z (z = .88) = .3810 + .3106 = .6916

d. Of 3000 women in the U.S. Air Force, only 118 of them would have hands 7.5 inches (the mean for male Air Force personnel) or longer.



$z = \frac{7.5-6.9}{.34}$ = 1.76
proportion from z to ∞ = .0392
.0392 x 3000 = 117.6 = 118 women

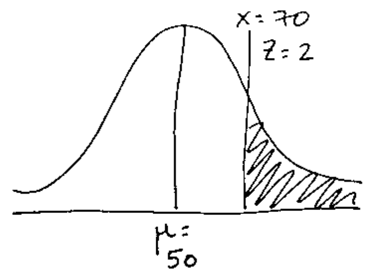
e. A woman's hand would have to be 7.34 inches long to put her among the 10% with the longest hands in the U.S. Air Force.



proportion from z to ∞ = .1000 is z = 1.28
$X = \mu + (z)(\sigma)$ = 6.9 + (1.28)(.34) = 6.9 + .44 = 7.34

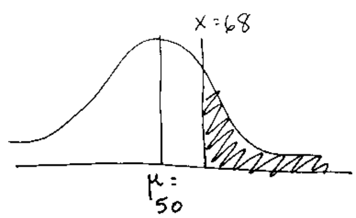
4.

a. Among a thousand persons, 23 would be expected to have scores two standard deviations or more above the mean.



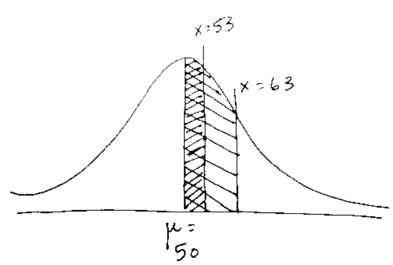
$z =$ 2 (this is because the score is two standard deviations above the mean; we also know the actual score is 70, because that is the mean plus two standard deviations)
proportion from z to ∞ = .0228
.0228 x 1000 = 22.8 = 23 people

b. .0359 would be expected to score 68 or above.



$z = \frac{68-50}{10} = 1.8$
proportion from z to ∞ = .0359

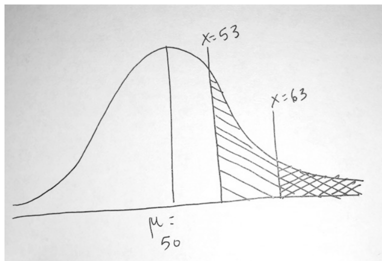
c. .2853 would be expected to score between 53 and 63.
There are actually two ways to complete this problem. The first is to find μ to z for each score and to subtract the smaller value from the larger value.



$z = \frac{53-50}{10} = .30$
proportion from μ to z = .1179

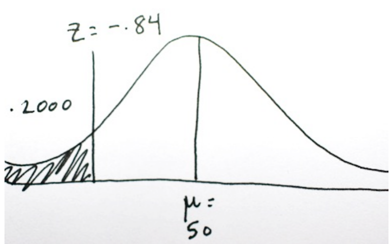
$z = \frac{63-50}{10}$ = 1.3
proportion from μ to z = .4032
proportion from μ to z (z = 1.3) - proportion from μ to z (z = .3) = .4032 - .1179 = .2853

The second way to solve this problem is to find z to ∞ for each score and to subtract the smaller value from the larger value.



$z = \frac{53-50}{10}$ = .30
proportion from z to ∞ = .3821
$z = \frac{63-50}{10}$ = 1.3
proportion from z to ∞ = .0968
proportion from z to ∞ (z = .3) - proportion from z to ∞ (z = 1.3) = .3821 - .0968 = .2853

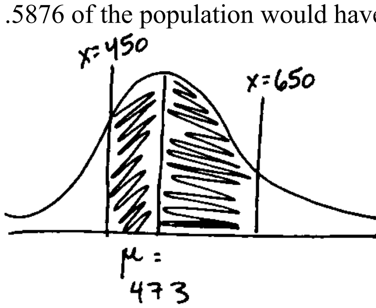
d. A score of 41.6 separates out the lowest one-fifth from the rest.



proportion from z to ∞ = .2000 is z = -.84 (note that this is negative because it is a score lower than the mean)
$X = \mu + (z)(\sigma)$ = 50 + (-.84)(10) = 50 – 8.4 = 41.6

5.

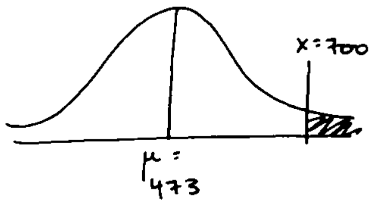
a. .5876 of the population would have scores between 450 and 650.



$z = \frac{450-473}{85}$ = -.27
proportion from μ to z = .1064
$z = \frac{650-473}{85}$ = 2.08

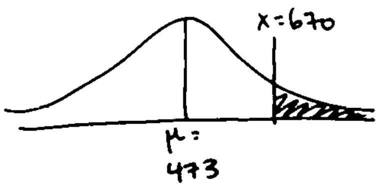
proportion from μ to z = .4812
proportion from μ to z (z = -.27) + proportion from μ to z (z = 2.08) = .1064 + .4812 = .5876

b. .0038 of the population would have scores of 700 or above.



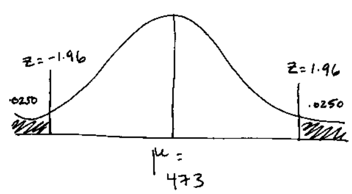
$z = \frac{700-473}{85}$ = 2.67
proportion from z to ∞ = .0038

c. If 10,000 psychology majors took the test, 102 people would be expected to have scores of 670 (the highest score yet made by a recent State U. graduate) or above.



$z = \frac{670-473}{85}$ = 2.32
proportion from z to ∞ = .0102
.0102 x 10,000 = 102 people

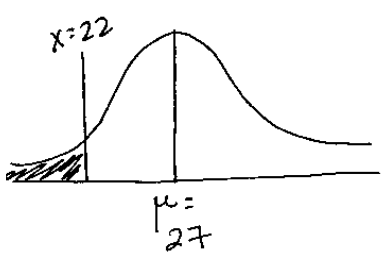
d. The scores that separate the middle 95 percent from the lower 2 ½ percent and the upper 2 ½ percent are 306 and 640.



proportion from z to ∞ = .0250 is z = 1.96 and z = -1.96
$X = \mu + (z)(\sigma)$ = 473 + (1.96)(85) = 473 + 166.6 = 639.6 = 640
$X = \mu + (z)(\sigma)$ = 6473 + (-1.96)(85) = 473 - 166.6 = 306.4 = 306

6.

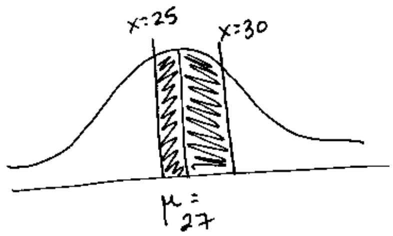
a. .1056 of the population would be expected to score 22 or less.



$z = \frac{22-27}{4}$ = -1.25

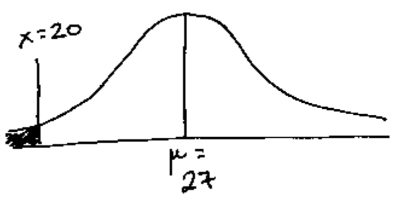
proportion from z to ∞ = .1056

b. .4649 of the population would score between 25 and 30.



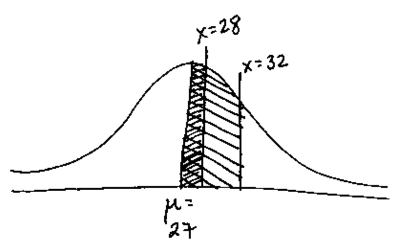
$z = \frac{25-27}{4}$ = -.5
proportion from μ to z = .1915
$z = \frac{30-27}{4}$ = .75
proportion from μ to z = .2734
proportion from μ to z (z = -.5) + proportion from μ to z (z = .75) = .1915 + .2734 = .4649

c. If 950 students played and a prize was given to those who scored 20 or less, 38 would get prizes.



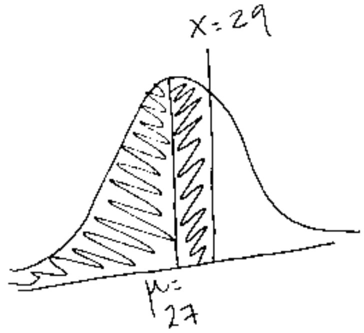
$z = \frac{20-27}{4}$ = -1.75
proportion from z to ∞ = .0401
.0401 x 950 = 38.10 = 38 people

d. .2957 of the population would be expected to score between 28 and 32.



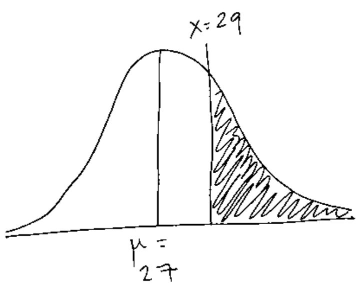
$z = \frac{28-27}{4}$ = .25
proportion from μ to z = .0987
$z = \frac{32-27}{4}$ = 1.25
proportion from μ to z = .3944
proportion from μ to z (z = 1.25) - proportion from μ to z (z = .25) = .3944 - .0987 = .2957

e. .6915 of the population would be expected to score 29 or less
There are two ways to solve this problem. First, you can find μ to z for a score of 29, and then add .5000, which represents all of the scores below 27.



$z = \frac{29-27}{4} = .50$
proportion from μ to z = .1915
proportion from μ to z (z = .50) + .5000 (proportion of all scores below the mean) = .1915 + .5000 = .6915

The second way to solve this problem is to find z to ∞ for scores 29 and above, and subtract that from the total proportion 1.000.



$z = \frac{29-27}{4} = .50$
proportion from z to ∞ = .3085
1.000 (proportion of all scores) - proportion from z to ∞ (z = .50) = 1.000 - .3085 = .6915