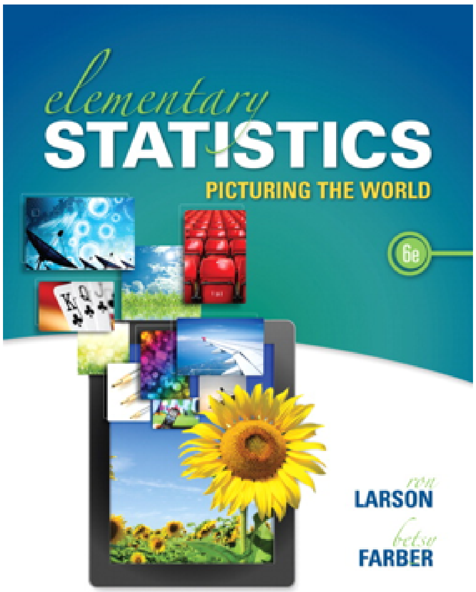
Elementary Statistics: Picturing The World

Sixth Edition



Chapter 5

Normal Probability Distributions

5-1 pg. 242 3,4,5, 17-37 EOO,39,47,50,53,56

5-2 pg. 249 9,10,13,14,17,18

5-3 pg. 257 1,5,9,13,17,19,21,22,25,30,31,32,34

5-4 pg.269 1,29,13,16,17,19,20,25,26,28,31,33,38

5-5 pg. 281 5-14,16,19,21,22,25,26,30

Chapter 5: Normal Probability Distribution

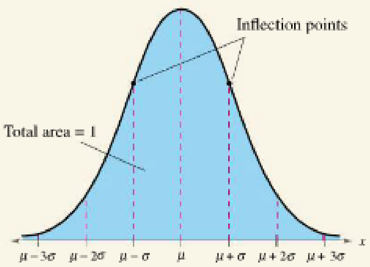
Section 5-1 Date: ______

Properties of a Normal Distribution

DEFINITION

A **normal distribution** is a continuous probability distribution for a random variable x. The graph of a normal distribution is called the **normal curve.** A normal distribution has these properties.

1. The mean, median, and mode are equal.
2. The normal curve is bell-shaped and is symmetric about the mean.
3. The total area under the normal curve is equal to 1.
4. The normal curve approaches, but never touches, the x-axis as it extends farther and farther away from the mean.
5. Between $\mu - \sigma$ and $\mu + \sigma$ (in the center of the curve), the graph curves downward. The graph curves upward to the left of $\mu - \sigma$ and to the right of $\mu + \sigma$. The points at which the curve changes from curving upward to curving downward are called **inflection points.**



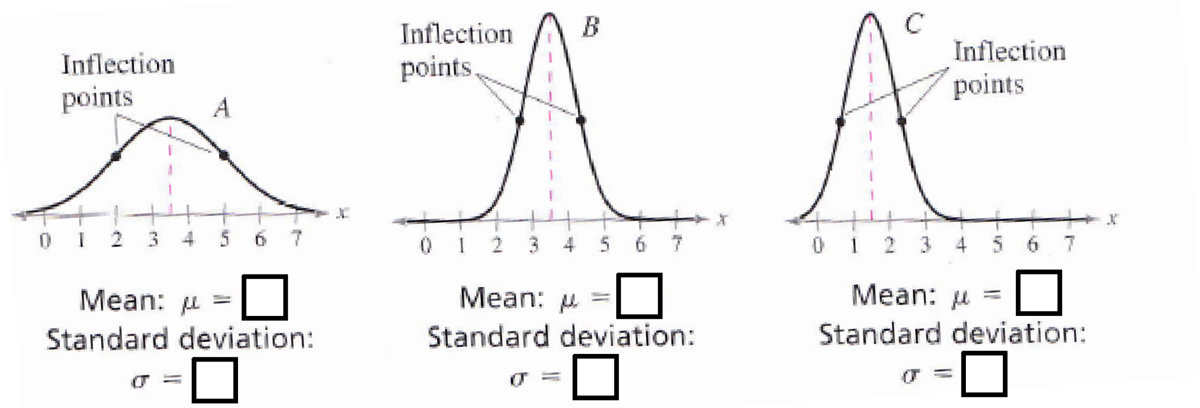
Something to never worry about!

$$y = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/2\sigma^2}$$

Just remember that the only parameters for the normal distribution are the mean and the standard deviation.

On a calculator:

2ND DISTR 1: normalpdf(χ, μ, σ)

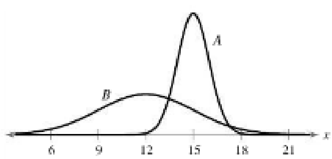


The area under each of these curves is 1 (100% of all possible outcomes)

EXAMPLE 1

Understanding Mean and Standard Deviation

1. Which normal curve has a greater mean?
2. Which normal curve has a greater standard deviation?



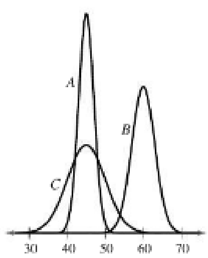
Solution

1. The line of symmetry of curve A occurs at $x = 15$. The line of symmetry of curve B occurs at $x = 12$. So, curve A has a greater mean.
2. Curve B is more spread out than curve A. So, curve B has a greater standard deviation.

Try It Yourself 1

Consider the normal curves shown at the right. Which normal curve has the greatest mean? Which normal curve has the greatest standard deviation?

a. Find the location of the line of symmetry of each curve. Make a conclusion about which mean is greatest.
b. Determine which normal curve is more spread out. Make a conclusion about which standard deviation is greatest.

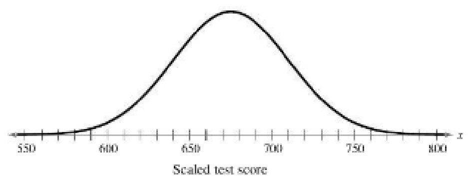


Section 5-1 (continued)

EXAMPLE 2

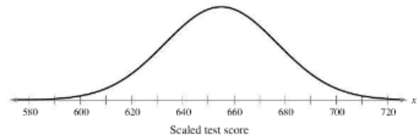
Interpreting Graphs of Normal Distributions

The scaled test scores for the New York State Grade 8 Mathematics Test are normally distributed. The normal curve shown below represents this distribution. What is the mean test score? Estimate the standard deviation of this normal distribution.



Try It Yourself 2

The scaled test scores for the New York State Grade 8 English Languag Arts Test are normally distributed. The normal curve shown belo represents this distribution. What is the mean test score? Estimate th standard deviation of this normal distribution.



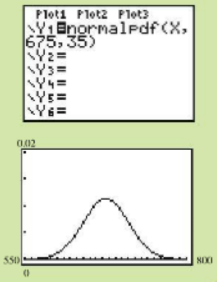
a. Find the line of symmetry and identify the mean.
b. Estimate the inflection points and identify the standard deviation.

1) Find the line of symmetry and identify the mean

2) Estimate the inflection points and identify the standard deviation

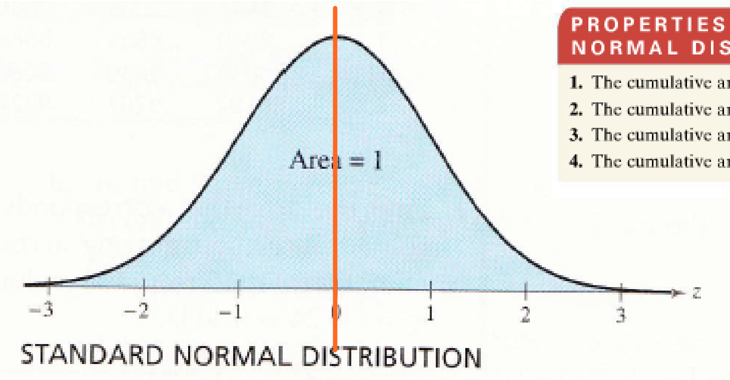
Study Tip

You can use technology to graph a normal curve. For instance, you can use a TI-84 Plus to graph the normal curve in Example 2.



The Standard Normal Distribution

Normal distribution with a MEAN of 0 and a STANDARD DEVIATION of 1.



PROPERTIES OF THE STANDARD NORMAL DISTRIBUTION

1. The cumulative area is close to 0 for z-scores close to $z = -3.49$.
2. The cumulative area increases as the z-scores increase.
3. The cumulative area for $z = 0$ is 0.5000.
4. The cumulative area is close to 1 for z-scores close to $z = 3.49$.

You can use "standardize" ANY x-value from a normal distribution by finding the z-score:

$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard deviation}} = \frac{x - \mu}{\sigma}.$$

Round to the nearest hundredth.

<u>Remember:</u> There is a difference between x and z. The random variable x is sometimes called a *raw score* and represents values in a nonstandard normal distribution. The z-score represents a value in the standard normal distribution.

Mar 4-3:10 PM

Section 5-1 (continued: using normal table)

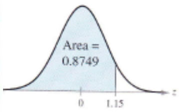
Open your book to page A-16 (appendix in back)

EXAMPLE 3

Using the Standard Normal Table

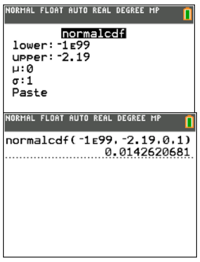
1. Find the cumulative area that corresponds to a z-score of 1.15.
2. Find the cumulative area that corresponds to a z-score of −0.24.

z	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
0.1	.5398	.5438	.5478	.5517	.5557	.5596	.5636
0.2	.5793	.5832	.5871	.5910	.5948	.5987	.6026
0.9	.8159	.8186	.8212	.8238	.8264	.8289	.8315
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131
1.4	.9192	.9207	.9222	.9236	.9251	.9265	.9279

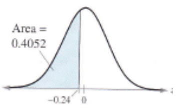


Try It Yourself 3

1. Find the cumulative area that corresponds to a z-score of −2.19.
2. Find the cumulative area that corresponds to a z-score of 2.17.



z	.09	.08	.07	.06	.05	.04	.03
−3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003
−3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004
−3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006
−0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981
−0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336
−0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707
−0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090
−0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483
−0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880

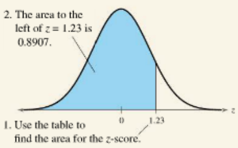


Left: number in table

GUIDELINES

Finding Areas Under the Standard Normal Curve

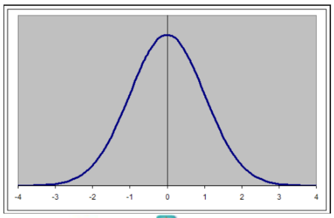
1. Sketch the standard normal curve and shade the appropriate area under the curve.
2. Find the area by following the directions for each case shown.
   a. To find the area to the *left* of z, find the area that corresponds to z in the Standard Normal Table.



EXAMPLE 4

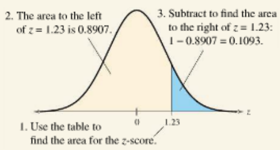
Finding Area Under the Standard Normal Curve

Find the area under the standard normal curve to the left of $z = -0.99$.



Right: 1- (number in table)

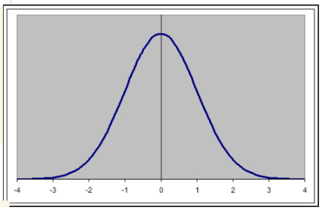
   b. To find the area to the *right* of z, use the Standard Normal Table to find the area that corresponds to z. Then subtract the area from 1.



EXAMPLE 5

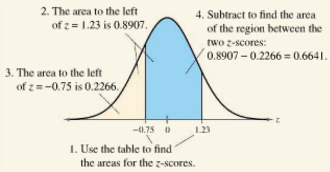
Finding Area Under the Standard Normal Curve

Find the area under the standard normal curve to the right of $z = 1.06$.



Between: (right number in table) - (left number in table)

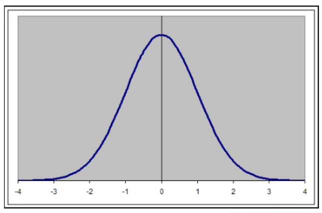
   c. To find the area *between* two z-scores, find the area corresponding to each z-score in the Standard Normal Table. Then subtract the smaller area from the larger area.

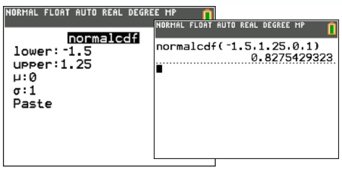


EXAMPLE 6

Finding Area Under the Standard Normal Curve

Find the area under the standard normal curve between $z = -1.5$ and $z = 1.25$.





Insight

Because the normal distribution is a continuous probability distribution, the area under the standard normal curve to the left of a z-score gives the probability that z is less than that z-score. For instance, in Example 4, the area to the left of $z = -0.99$ is 0.1611. So, $P(z < -0.99) = 0.1611$, which is read as "the probability that z is less than −0.99 is 0.1611."

Keep in mind that the Empirical Rule tells us that values more than 2 standard deviations away from the mean are considered unusual.

1. A z-score greater than 2 or less than -2 is unusual.

2. A z-score greater than 3 or less than -3 is VERY unusual.

Mar 4-3:25 PM

Section 5-1 (continued: normal on calculator)

2nd DISTR
2: normalcdf(

Notice this is cdf, not pdf like we have been using!

Area wanted	On calculator
to the left of z	normalcdf(-10000, z)
to the right of z	normalcdf(z,10000)
between a and b	normalcdf(a, b)

Study Tip

Here are instructions for finding the area that corresponds to $z = -0.24$ on a TI-83/84.

To specify the lower bound in this case, use −10,000.

2nd DISTR
2: normalcdf(
−10000, −.24)
ENTER

```
normalcdf(-10000
,-.24
      .405165175
```

These calculator commands will find the area to the LEFT of a z-score.

From the Standard Normal Table, this area is equal to 0.1611.

Try It Yourself 4

Find the area under the standard normal curve to the left of $z = 2.13$.

Try It Yourself 5

Find the area under the standard normal curve to the right of $z = -2.16$.

Try It Yourself 6

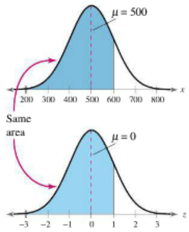
Find the area under the standard normal curve between $z = -2.165$ and $z = -1.35$.

pg. 242 3,4, 17-37 EOO 39, 47,50,53,56

Mar 5-8:17 AM

Section 5-2 Finding Probabilities Date: ______

To convert any normal distribution into a standard normal, let mean be equivalent to 0 and standard deviation be equivalent to 1.

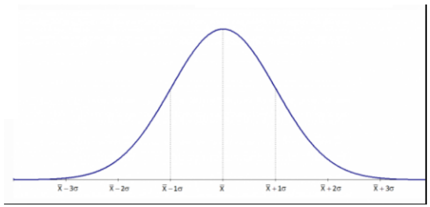


Reminder: $z = \dfrac{x - \mu}{\sigma}$

EXAMPLE 1

Finding Probabilities for Normal Distributions

A survey indicates that people keep their cell phone an average of 1.5 years before buying a new one. The standard deviation is 0.25 year. A cell phone user is selected at random. Find the probability that the user will keep his or her current phone for less than 1 year before buying a new one. Assume that the lengths of time people keep their phone are normally distributed and are represented by the variable x.

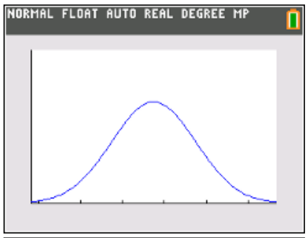


EXAMPLE 2

Finding Probabilities for Normal Distributions

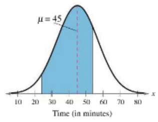
A survey indicates that for each trip to a supermarket, a shopper spends an average of 45 minutes with a standard deviation of 12 minutes in the store. The lengths of time spent in the store are normally distributed and are represented by the variable x. A shopper enters the store. (a) Find the probability that the shopper will be in the store for each interval of time listed below. (b) Interpret your answer when 200 shoppers enter the store. How many shoppers would you expect to be in the store for each interval of time listed below?

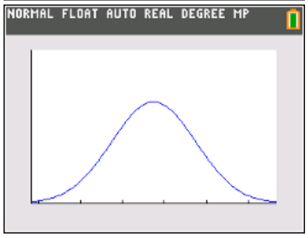
1. Between 24 and 54 minutes 2. More than 39 minutes



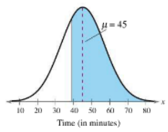
#1 Draw a picture find the z-score for both 24 and 54

Write a sentence





#2 Find the z-score for 39



The tricky part is usually deciding what portion of the graph you are looking for - DRAW A PICTURE!!!!

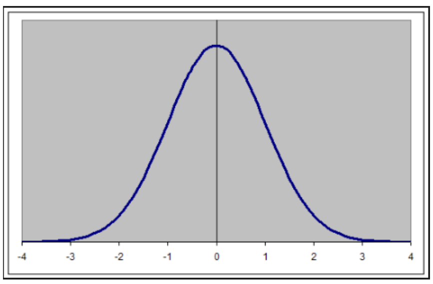
Try It Yourself 1

A Ford Focus manual transmission gets an average of 24 miles per gallon (mpg) in city driving with a standard deviation of 1.6 mpg. A Focus is selected at random. What is the probability that it will get more than 28 mpg? Assume that gas mileage is normally distributed. *(Adapted from U.S. Department of Energy)*

a. *Sketch* a graph.

b. *Find the z-score* that corresponds to 28 miles per gallon.

c. *Find the area* to the right of that z-score.



d. *Write* the result as a sentence.

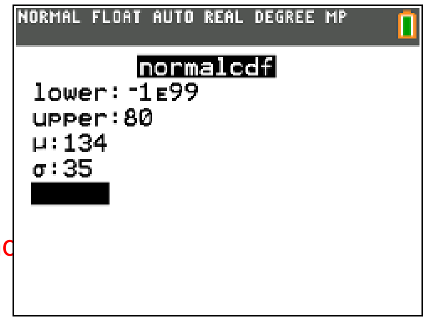
Section 5-2 Finding Probabilities (continued)

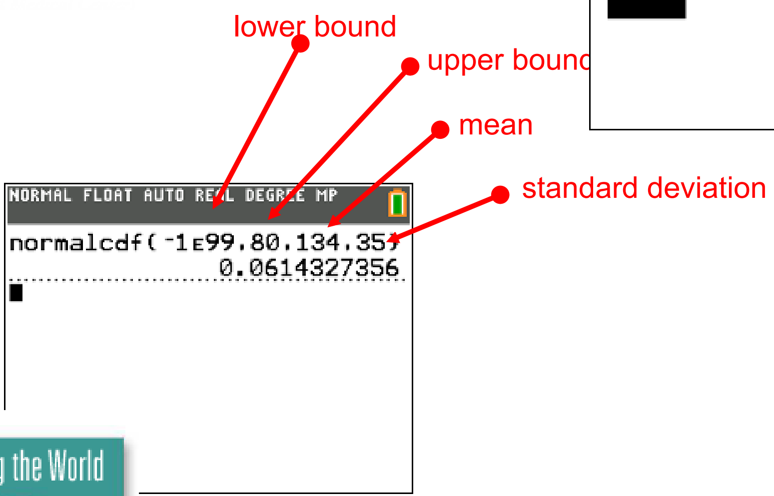
Many classes/businesses use MiniTab, Excel, or other statistical software.
For this class, look at the notes for TI 83/84.

EXAMPLE 3

Using Technology to Find Normal Probabilities

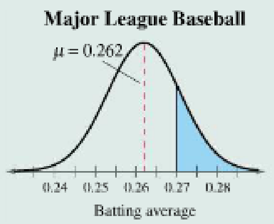
Triglycerides are a type of fat in the bloodstream. The mean triglyceride level in the United States is 134 milligrams per deciliter. Assume the triglyceride levels of the population of the United States are normally distributed, with a standard deviation of 35 milligrams per deciliter. You randomly select a person from the United States. What is the probability that the person's triglyceride level is less than 80? Use technology to find the probability.





Picturing the World

In baseball, a batting average is the number of hits divided by the number of at bats. The batting averages of all Major League Baseball players in a recent year can be approximated by a normal distribution, as shown in the figure. The mean of the batting averages is 0.262 and the standard deviation is 0.009. (Adapted from ESPN)



What percent of the players have a batting average of 0.270 or greater? Out of 40 players on a roster, how many would you expect to have a batting average of 0.270 or greater?

pg. 249: 9, 10, 13, 14,17,18

Mar 5-9:05 AM

Section 5-3 Finding Values Date: _______

We have been finding probability of getting above/below a certain value, or between two values.

What if, instead, we want to know the probability and want to find the corresponding value?

For example:

- University admissions may want to know the lowest SAT score a student can have and still be in the top 10% of their applicants

- Medical researcher may want to know the age range that would give them the middle 90% of patients by age.

TO USE A CALCULATOR TO FIND:

2ND DISTR

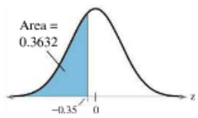
3: invNorm(given probability)

EXAMPLE 1

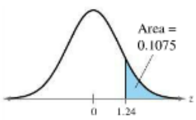
Finding a *z*-Score Given an Area

1. Find the z-score that corresponds to a cumulative area of 0.3632.
2. Find the z-score that has 10.75% of the distribution's area to its rig

z	.09	.08	.07	.06	.05	.04	.03
−3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003
−0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981
−0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336
−0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707
−0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090



z	.00	.01	.02	.03	.04	.05	.06
0.0	.5000	.5040	.5080	.5120	.5160	.5199	.5239
1.0	.8413	.8438	.8461	.8485	.8508	.8531	.8554
1.1	.8643	.8665	.8686	.8708	.8729	.8749	.8770
1.2	.8849	.8869	.8888	.8907	.8925	.8944	.8962
1.3	.9032	.9049	.9066	.9082	.9099	.9115	.9131



▸Try It Yourself 1

1. Find the z-score that has 96.16% of the distribution's area to the right.

2. Find the z-score for which 95% of the distribution's area lies between $-z$ and z.

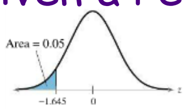
Finding a z score given a Percentile

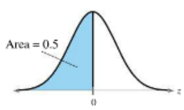
EXAMPLE 2

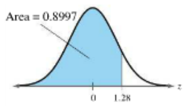
Finding a *z*-Score Given a Percentile

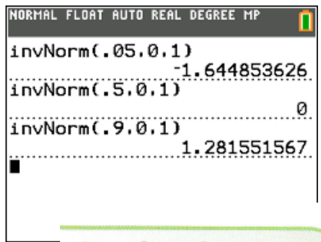
Find the z-score that corresponds to each percentile.

1. P_5 2. P_{50} 3. P_{90}









Study Tip

In most cases, the given area will not be found in the table, so use the entry closest to it. If the given area is halfway between two area entries, use the z-score halfway between the corresponding z-scores. For instance, in part 1 of Example 2, the z-score between −1.64 and −1.65 is −1.645.

Percentiles divide a data set into 100 equal parts.

- if a value x represents the 83rd percentile P_{83} then 83% of the data are below x and 17% of the data values are above x.

Section 5-3 continued (transforming a z-score to an x-value)

To turn a z-score back into an x-value: now just solve for x (get x by itself)

Recall that to transform an x-value to a z-score, you can use the formula

$$z = \frac{x - \mu}{\sigma}.$$

TRANSFORMING A z-SCORE TO AN x-VALUE

To transform a standard z-score to an x-value in a given population, use the formula

$$x = \mu + z\sigma.$$

EXAMPLE 3

Finding an x-Value Corresponding to a z-Score

A veterinarian records the weights of cats treated at a clinic. The weights are normally distributed, with a mean of 9 pounds and a standard deviation of 2 pounds. Find the weights x corresponding to z-scores of 1.96, −0.44, and 0. Interpret your results.

Study Tip

Here are instructions for finding a specific x-value for a given probability on a TI-84 Plus.

2nd DISTR

3: invNorm(

Enter the values for the area under the normal distribution, the mean, and the standard deviation.

invNorm(.9,50,10)
62.81551567

Finding a specific data value for a given probability

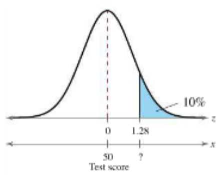
EXAMPLE 4

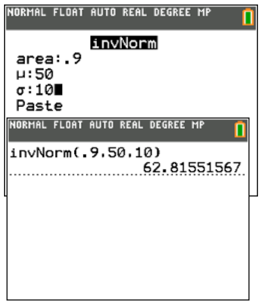
Finding a Specific Data Value

Scores for the California Peace Officer Standards and Training test are normally distributed, with a mean of 50 and a standard deviation of 10. An agency will only hire applicants with scores in the top 10%. What is the lowest score an applicant can earn and still be eligible to be hired by the agency?

Solution

Exam scores in the top 10% correspond to the shaded region shown.





Putting it all together!

Given a probability, find the x-value that goes with it.

- First, find the z-score that corresponds to the probability.
- Second, convert the z-score into an x-value.
- Lastly, interpret the results!

EXAMPLE 5

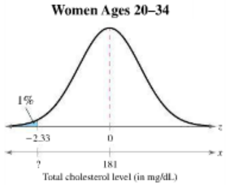
Finding a Specific Data Value

In a randomly selected sample of women ages 20–34, the mean total cholesterol level is 181 milligrams per deciliter with a standard deviation of 37.6 milligrams per deciliter. Assume the total cholesterol levels are normally distributed. Find the highest total cholesterol level a woman in this 20–34 age group can have and still be in the bottom 1%.

Solution

Total cholesterol levels in the lowest 1% correspond to the shaded region

Total Cholesterol Levels in Women Ages 20–34

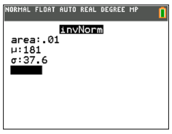


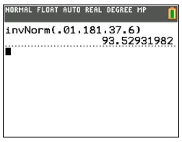
A total cholesterol level in the lowest 1% is any level below the 1st percentile. To find the level that represents the 1st percentile, you must first find the z-score that corresponds to a cumulative area of 0.01. In the Standard Normal Table, the area closest to 0.01 is 0.0099. So, the z-score that corresponds to an area of 0.01 is $z = -2.33$. To find the x-value, note that $\mu = 181$ and $\sigma = 37.6$, and use the formula $x = \mu + z\sigma$, as shown.

$$\begin{aligned} x &= \mu + z\sigma \\ &= 181 + (-2.33)(37.6) \\ &\approx 93.39 \end{aligned}$$

You can check this answer using technology. For instance, you can use a TI-84 Plus to find the x-value, as shown at the left.

Interpretation The value that separates the lowest 1% of total cholesterol levels for women in the 20–34 age group from the highest 99% is about 93 milligrams per deciliter.





⇨ pg. 257: 1,5,9,13,17,19,21,22,25,30, 31,32,34,

Mar 5-9:36 AM

Lesson 5-4: Sampling Distributions and Central Limit Theorem

Date: 4/2/14

We have been talking about population means, which hold true for an entire group/set of trials. What is the relationship between the population mean (the TRUE mean), and the mean we would get from a sample (the SAMPLE mean)?

Population mean: obtained from census - counts every single one

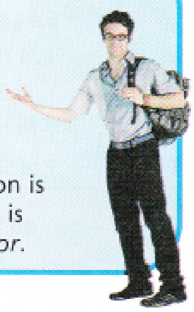
Sample mean: small group, assumed to represent the population (does it?)

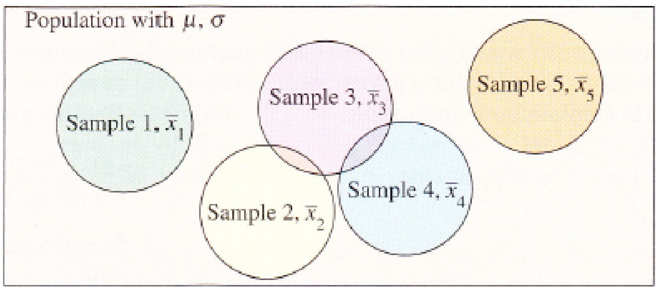
DEFINITION

A **sampling distribution** is the probability distribution of a sample statistic that is formed when samples of size n are repeatedly taken from a population. If the sample statistic is the sample mean, then the distribution is the **sampling distribution of sample means.** Every sample statistic has a sampling distribution.

Insight

Sample means can vary from one another and can also vary from the population mean. This type of variation is to be expected and is called *sampling error*.





Different samples have different means - ideally, if the sample has been selected to be representative of the population, all will be close the true population mean

PROPERTIES OF SAMPLING DISTRIBUTIONS OF SAMPLE MEANS

1. The mean of the sample means $\mu_{\bar{x}}$ is equal to the population mean μ.

$$\mu_{\bar{x}} = \mu$$

2. The standard deviation of the sample means $\sigma_{\bar{x}}$ is equal to the population standard deviation σ divided by the square root of n.

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

The standard deviation of the sampling distribution of the sample means is called the **standard error of the mean.**

Seems too good to be true? Let's verify that it really works (just once!) before we start using it blindly....

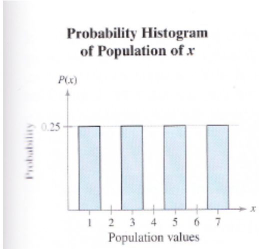
Mar 7-9:16 AM

Lesson 5-4 (continued) - verifying sample statistics

EXAMPLE 1

A Sampling Distribution of Sample Means

You write the population values {1, 3, 5, 7} on slips of paper and put them in a box. Then you randomly choose two slips of paper, with replacement. List all possible samples of size $n = 2$ and calculate the mean of each. These means form the sampling distribution of the sample means. Find the mean, variance, and standard deviation of the sample means. Compare your results with the mean $\mu = 4$, variance $\sigma^2 = 5$, and standard deviation $\sigma = \sqrt{5} \approx 2.236$ of the population.



True Population Stats:

$$\mu = \frac{1+3+5+7}{4} = 4$$

$$\sigma^2 = \sum P(x)(x-\mu)^2 = (.25)(1-4)^2 + (.25)(3-4)^2 + (.25)(5-4)^2 + (.25)(7-4)^2 = 5$$

$$\sigma = \sqrt{5} = 2.236$$

How would a sample compare if we randomly picked two of the four members of our population?

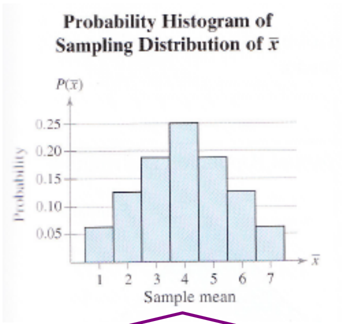
List all samples of 2 from our original population and the mean of each:

Sample	Sample mean, $\bar{x}$	Sample	Sample mean, $\bar{x}$
1, 1	1	5, 1	3
1, 3	2	5, 3	4
1, 5	3	5, 5	5
1, 7	4	5, 7	6
3, 1	2	7, 1	4
3, 3	3	7, 3	5
3, 5	4	7, 5	6
3, 7	5	7, 7	7

If we construct a probability distribution from these sample means, we can make this histogram:

Probability Distribution of Sample Means

$\bar{x}$	f	Probability
1	1	0.0625
2	2	0.1250
3	3	0.1875
4	4	0.2500
5	3	0.1875
6	2	0.1250
7	1	0.0625



$$\sum \bar{X} = 64$$

Let's find the mean of our sample means (the mean of means):

$$\mu_{\bar{X}} = \frac{\sum \bar{X} = 64}{16} = 4 \qquad \sigma^2_{\bar{X}} = \frac{5}{2}$$

and standard deviation is $\sigma_{\bar{X}} = \sqrt{\frac{5}{2}} = 1.581$ so $\mu_{\bar{X}} = 4 = \mu$

$$\sigma_{\bar{X}} = 1.581$$

Notice that $\frac{\sigma}{\sqrt{n}} = \frac{2.236}{\sqrt{2}} = 1.581$

This shows empirically, by examining each possible sample from a population that we had a complete view of already, that we can use $\mu_{\bar{x}} = \mu$

and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$ as the sample mean and sample standard deviation.

WHEW!

Mar 7-9:15 AM

Lesson 5-4 (continued) - Central Limit Theorem

This is really REALLY **REALLY** important!!!!!!

The CLT forms the basis for all the statistics we are going to do, and makes it possible to use samples to make assumptions about an entire population.

THE CENTRAL LIMIT THEOREM

1. If samples of size n, where $n \geq 30$, are drawn from any population with a mean μ and a standard deviation σ, then the sampling distribution of sample means approximates a normal distribution. The greater the sample size, the better the approximation.
2. If the population itself is normally distributed, the sampling distribution of sample means is normally distributed for *any* sample size n.

In either case, the sampling distribution of sample means has a mean equal to the population mean.

$$\mu_{\bar{x}} = \mu \qquad \text{Mean}$$

The sampling distribution of sample means has a variance equal to $1/n$ times the variance of the population and a standard deviation equal to the population standard deviation divided by the square root of n.

$$\sigma_{\bar{x}}^2 = \frac{\sigma^2}{n} \qquad \text{Variance}$$

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{Standard deviation}$$

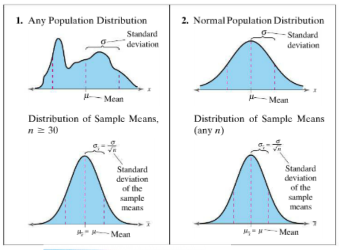
The standard deviation of the sampling distribution of the sample means, $\sigma_{\bar{x}}$, is also called the **standard error of the mean.**

Two major points

1. We sample at least 30 members of the population (bigger is better, but 30 is bare minimum)

2. If the original population was normally distributed, our distribution of the sample means will be roughly normal

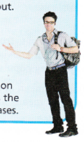
Worth remembering - since the sample standard deviation is the true s.d. DIVIDED by square root of n, it will be narrower



Insight

The distribution of sample means has the same mean as the population. But its standard deviation is less than the standard deviation of the population. This tells you that the distribution of sample means has the same center as the population, but it is not as spread out.

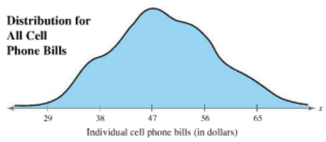
Moreover, the distribution of sample means becomes less and less spread out (tighter concentration about the mean) as the sample size n increases.



EXAMPLE 2

Interpreting the Central Limit Theorem

Cell phone bills for residents of a city have a mean of \$47 and a standard deviation of \$9, as shown in the figure. Random samples of 100 cell phone bills are drawn from this population, and the mean of each sample is determined. Find the mean and standard deviation of the sampling distribution of sample means. Then sketch a graph of the sampling distribution.



Solution

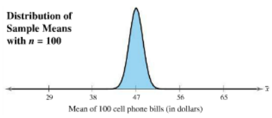
The mean of the sampling distribution is equal to the population mean, and the standard deviation of the sample means is equal to the population standard deviation divided by $\sqrt{n}$. So,

$$\mu_{\bar{x}} = \mu = 47 \qquad \text{Mean of the sample means}$$

and

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9}{\sqrt{100}} = 0.9. \qquad \text{Standard deviation of the sample means}$$

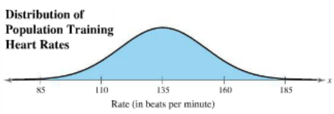
Interpretation From the Central Limit Theorem, because the sample size is greater than 30, the sampling distribution can be approximated by a normal distribution with a mean of \$47 and a standard deviation of \$0.90, as shown in the figure.



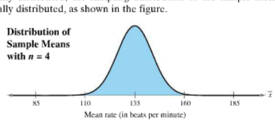
EXAMPLE 3

Interpreting the Central Limit Theorem

Assume the training heart rates of all 20-year-old athletes are normally distributed, with a mean of 135 beats per minute and a standard deviation of 18 beats per minute, as shown in the figure. Random samples of size 4 are drawn from this population, and the mean of each sample is determined. Find the mean and standard deviation of the sampling distribution of sample means. Then sketch a graph of the sampling distribution.



Interpretation From the Central Limit Theorem, because the population is normally distributed, the sampling distribution of the sample means is also normally distributed, as shown in the figure.



Mar 7-9:17 AM

Lesson 5-4 (continued) - Probability and the Central Limit Theorem

We saw in 5-2 how to find the probability that the random variable x would be in a range of values (z-score).

Same thing applies to finding the probability that a sample mean $\overline{x}$ will be in a range of values!

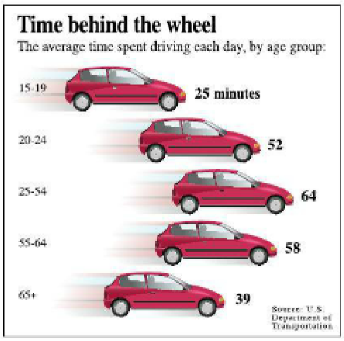
$$z = \frac{\text{Value} - \text{Mean}}{\text{Standard error}} = \frac{\overline{x} - \mu_{\overline{x}}}{\sigma_{\overline{x}}} = \frac{\overline{x} - \mu}{\sigma/\sqrt{n}}.$$

EXAMPLE 4

You can use the CLT because the sample size is greater than 30

Finding Probabilities for Sampling Distributions

The figure at the right shows the lengths of time people spend driving each day. You randomly select 50 drivers ages 15 to 19. What is the probability that the mean time they spend driving each day is between 24.7 and 25.5 minutes? Assume that $\sigma = 1.5$ minutes.



Solution

The sample size is greater than 30, so you can use the Central Limit Theorem to conclude that the distribution of sample means is approximately normal, with a mean and a standard deviation of

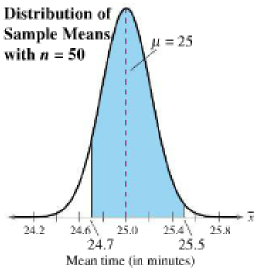
$$\mu_x = \mu =$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}} =$$

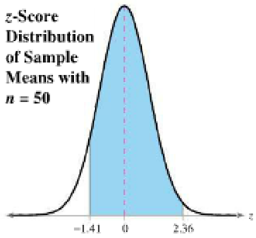
Study Tip

Before you find probabilities for intervals of the sample mean $\overline{x}$, use the Central Limit Theorem to determine the mean and the standard deviation of the sampling distribution of the sample means. That is, calculate $\mu_{\overline{x}}$ and $\sigma_{\overline{x}}$.

The graph of the distribution of sample means with n=50



X-score distribution of sample means with n=50

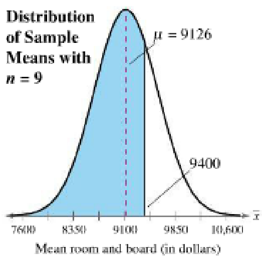


Interpretation

EXAMPLE 5

Finding Probabilities for Sampling Distributions

The mean room and board expense per year at four-year colleges is \$9126. You randomly select 9 four-year colleges. What is the probability that the mean room and board is less than \$9400? Assume that the room and board expenses are normally distributed with a standard deviation of \$1500.



$$\mu_x = \mu =$$

$$\sigma_x = \frac{\sigma}{\sqrt{n}} =$$

Interpretation So, about 71% of such samples with $n = 9$ will have a mean less than \$9400 and about 29% of these sample means will be greater than \$9400.

Mar 7-9:33 AM

Lesson 5-4 (continued) - Probability and the Central Limit Theorem

One of the most useful things about the CLT is that is makes a great lie detector.

Should we trust their claim?

Here's how:

Study Tip

To find probabilities for individual members of a population with a normally distributed random variable x, use the formula

$$z = \frac{x - \mu}{\sigma}.$$

To find probabilities for the mean $\bar{x}$ of a sample of size n, use the formula

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}}.$$

EXAMPLE 6

Finding Probabilities for x and $\bar{x}$

The average credit card debt carried by undergraduates is normally distributed, with a mean of \$3173 and a standard deviation of \$1120.

1. What is the probability that a randomly selected undergraduate, who is a credit card holder, has a credit card balance less than \$2700?
2. You randomly select 25 undergraduates who are credit card holders. What is the probability that their mean credit card balance is less than \$2700?
3. Compare the probabilities from (1) and (2).

1. In this case, you are asked to find the probability associated with a certain value of the random variable x. The z-score that corresponds to x =\$2700

2. Here, you are asked to find the probability associated with a sample mean

The z-score that corresponds to

3. Interpretation:

pg. 269 1,2,9,13,16,17,19,20, 22, 25, 26,28, 31, 33, 38

Mar 7-9:50 AM

Lesson 5-5: Normal Approximations to Binomial Distributions

Date:________

Binomial: DON'T FORGET 4.2

* n independent trials

*2 possible outcomes (success/failure)

*Probability of success (p) is same for each trial q = 1-p

* P is the same for each trial

The binomial formula was $P(x)=nCx(p)^{x}(q)^{(n-x)}$

This is great for a small number of x's, but a major headache to do for, say, P(0<x<50)

Normal Approximation to the rescue! DON'T forget 4.2

NORMAL APPROXIMATION TO A BINOMIAL DISTRIBUTION

If $np \geq 5$ and $nq \geq 5$, then the binomial random variable x is approximately normally distributed, with mean

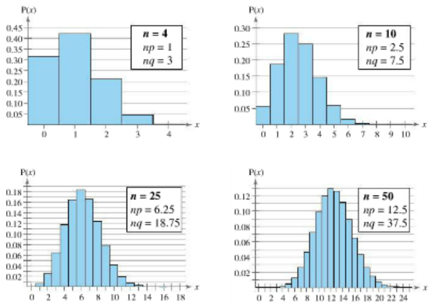
$$\mu = np$$

and standard deviation

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

where n is the number of independent trials, p is the probability of success in a single trial, and q is the probability of failure in a single trial.

The bigger n gets (the more we repeat the experiment and collect more data), the closer we get to a normal curve



Example 1:

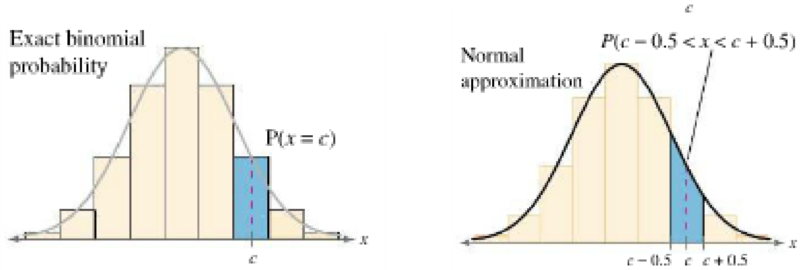
Decide if the normal approximation can be used. If so, find the mean and standard deviation. If not, why not?

1. Fifty-one percent of adults in the US whose New Year's resolution was to exercise more achieved their resolution. You randomly select 65 adults in the United States whose resolution was to exercise more and ask each if he or she achieved that resolution.

2. Fifteen percent of adults in the US do not make New Year's resolutions. You randomly select 15 adults in the US and ask each if he or she made a New Year's resolution.

Lesson 5-5 (continued) - Correction for continuity

Binomial is discrete, but normal is continuous. To fix that, we have to move out 0.5 units to the left and right of the region before calculating the normal approximation



Example 2:

Use a correction for continuity to convert each of the following binomial intervals to a normal distribution interval.

1. The probability of getting between 270 and 310 successes, inclusive

2. The probability of at least 158 successes

3. The probability of getting less than 63 successes

Shown below are several cases of binomial probabilities involving the number c and how to convert each to a normal distribution probability.

Binomial	Normal	Notes
Exactly c	$P(c - 0.5 < x < c + 0.5)$	Includes c
At most c	$P(x < c + 0.5)$	Includes c
Fewer than c	$P(x < c - 0.5)$	Does not include c
At least c	$P(x > c - 0.5)$	Includes c
More than c	$P(x > c + 0.5)$	Does not include c

Try It Yourself 2

Use a continuity correction to convert each binomial probability to a normal distribution probability.

1. The probability of getting between 57 and 83 successes, inclusive
2. The probability of getting at most 54 successes

a. List the midpoint values for the binomial probability.
b. Use a continuity correction to write the normal distribution probability.

Lesson 5-5 (continued) - Approximating

Putting it all together....

APPROXIMATING BINOMIAL PROBABILITIES

GUIDELINES

Using a Normal Distribution to Approximate Binomial Probabilities

IN WORDS	IN SYMBOLS
1. Verify that a binomial distribution applies.	Specify n, p, and q.
2. Determine whether you can use a normal distribution to approximate x, the binomial variable.	Is $np \geq 5$? Is $nq \geq 5$?
3. Find the mean μ and standard deviation σ for the distribution.	$\mu = np$ $\sigma = \sqrt{npq}$
4. Apply the appropriate continuity correction. Shade the corresponding area under the normal curve.	Add 0.5 to (or subtract 0.5 from) the binomial probability.
5. Find the corresponding z-score(s).	$z = \frac{x - \mu}{\sigma}$
6. Find the probability.	Use the Standard Normal Table.

7. Interpretation Write a sentence

EXAMPLE 3

Approximating a Binomial Probability

In a survey of 8- to 18-year-old heavy media users in the United States, 47% said they get fair or poor grades (C's or below). You randomly select forty-five 8- to 18-year-old heavy media users in the United States and ask them whether they get fair or poor grades. What is the probability that fewer than 20 of them respond yes?

Do the steps for example 3

1.

2.

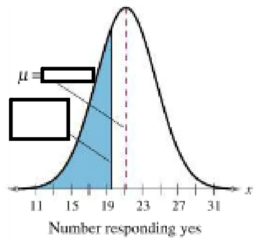
3.

4.

5.

6.

7 INTERPRETATION



Example 4:

Fifty-eight percent of Adults say they never wear a helmet when riding a bike. You randomly select 200 people in the US and ask each if they wear a helmet when riding a bike. What is the probability that at least 120 will say yes?

Do the steps for

1.

2.

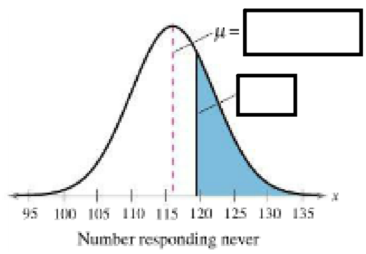
3.

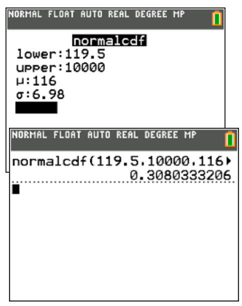
4.

5.

6.

7 INTERPRETATION





Try It Yourself 4

In Example 4, what is the probability that at most 100 adults will say they never wear a helmet when riding a bicycle?

a. Determine whether you can use a normal distribution to approximate the binomial variable (see Example 4).
b. Find the mean μ and the standard deviation σ for the normal distribution (see Example 4).
c. Apply a continuity correction to rewrite $P(x \leq 100)$ and sketch a graph.
d. Find the corresponding z-score.
e. Use the Standard Normal Table to find the area to the left of z and calculate the probability.

Lesson 5-5 (continued) - Approximating and making decisions

EXAMPLE 5

Approximating a Binomial Probability

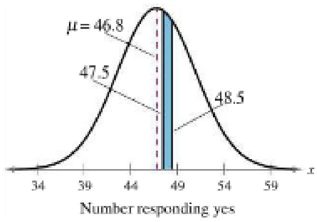
A study of National Football League (NFL) retirees, ages 50 and older, found that 62.4% have arthritis. You randomly select 75 NFL retirees who are at least 50 years old and ask them whether they have arthritis. What is the probability that exactly 48 will say yes?

Solution

Because $np = 75(0.624) = 46.8$ and $nq = 75(0.376) = 28.2$, the binomial variable x is approximately normally distributed, with

$$\mu = np = 46.8 \quad \text{and} \quad \sigma = \sqrt{npq} = \sqrt{75(0.624)(0.376)} \approx 4.19.$$

Using the continuity correction, you can rewrite the discrete probability $P(x = 48)$ as the continuous probability $P(47.5 < x < 48.5)$. The figure shows a normal curve with $\mu = 46.8$, $\sigma \approx 4.19$, and the shaded area under the curve between 47.5 and 48.5.



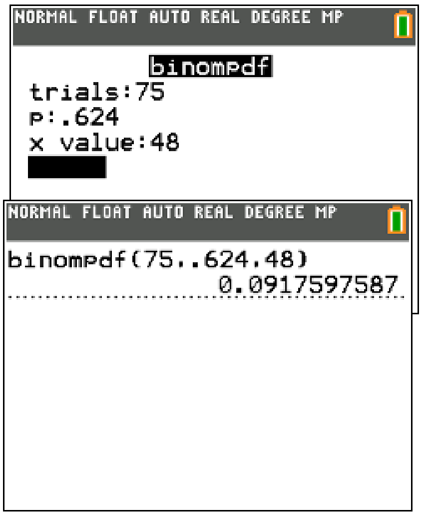
The z-scores that correspond to 47.5 and 48.5 are

$$z_1 = \frac{47.5 - 46.8}{\sqrt{75(0.624)(0.376)}} \approx 0.17 \quad \text{and} \quad z_2 = \frac{48.5 - 46.8}{\sqrt{75(0.624)(0.376)}} \approx 0.41.$$

So, the probability that exactly 48 NFL retirees will say they have arthritis is

$$\begin{aligned} P(47.5 < x < 48.5) &= P(0.17 < z < 0.41) \\ &= P(z < 0.41) - P(z < 0.17) \\ &= 0.6591 - 0.5675 \\ &= 0.0916. \end{aligned}$$

Interpretation The probability that exactly 48 NFL retirees will say they have arthritis is approximately 0.0916, or about 9.2%.



Try It Yourself 5

The study in Example 5 found that 32.0% of all men in the United States ages 50 and older have arthritis. You randomly select 75 men in the United States who are at least 50 years old and ask them whether they have arthritis. What is the probability that exactly 15 will say yes?

a. Determine whether you can use a normal distribution to approximate the binomial variable.
b. Find the mean μ and the standard deviation σ for the normal distribution.
c. Apply a continuity correction to rewrite $P(x = 15)$ and sketch a graph.
d. Find the corresponding z-scores.
e. Use the Standard Normal Table to find the area to the left of each z-score and calculate the probability.

pg. 281 5-14, 16, 19,21, 22, 25, 26,30

Mar 11-3:23 PM