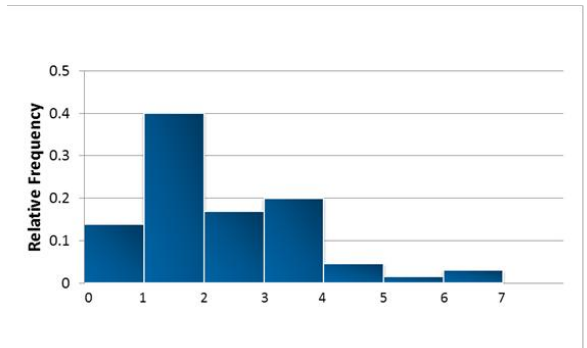
OBJECTIVES

1. Use a probability density curve to describe a population
2. Use a normal curve to describe a normal population
3. Convert values from a normal distribution to z-scores
4. Find areas under a normal curve
5. Find the value from a normal distribution corresponding to a given proportion

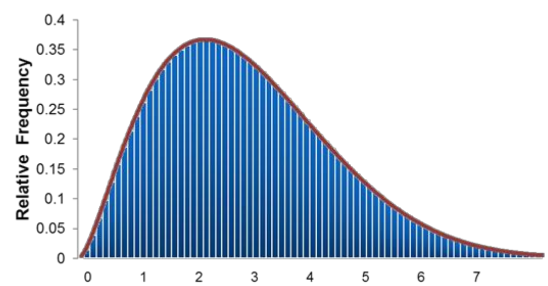
OBJECTIVE 1
USE A PROBABILITY DENSITY CURVE TO DESCRIBE A POPULATION

PROBABILITY DENSITY CURVES

The following figure presents a relative frequency histogram for the particulate emissions of a sample of 65 vehicles. If we had information on the entire population, containing millions of vehicles, we could make the rectangles extremely narrow.

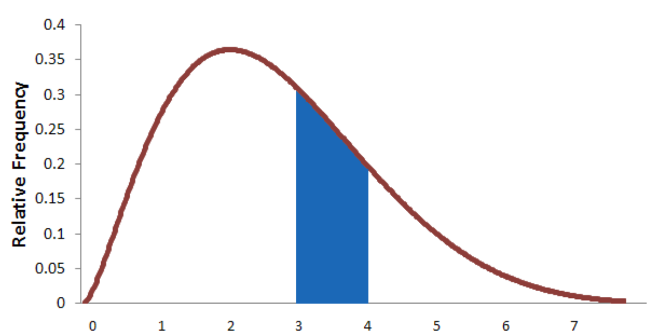


The histogram would then look smooth and could be approximated by a curve. The curve used to describe the distribution of this variable is called the **probability density curve** of the random variable. The probability density curve tells what proportion of the population falls within a given interval.



AREA AND PROBABILITY DENSITY CURVES

The area under a probability density curve between any two values a and b has two interpretations:

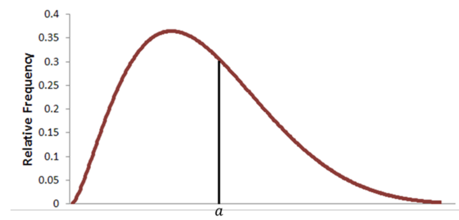


INTERPRETATIONS OF AREA UNDER A PROBABILITY DENSITY CURVE:

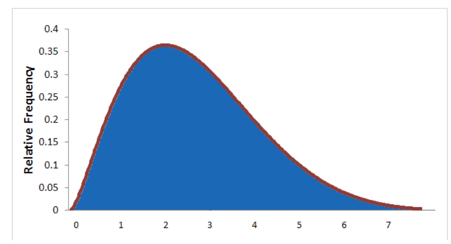
It represents the proportion of the population whose value are between a and b.

It represents the probability that a randomly selected value from the population will be between a and b

PROPERTIES OF PROBABILITY DENSITY CURVES



The region above a single point has no width, thus no area. Therefore, if X is a continuous random variable, $P(X = a) = 0$ for any number a. This means that $P(a < X < b) = P(a \le X \le b)$ for any numbers a and b.

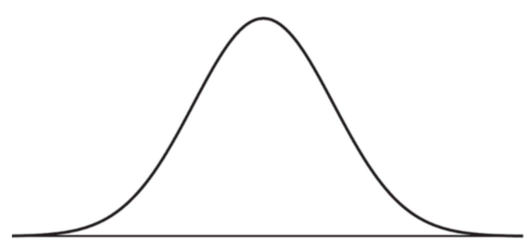


For any probability density curve, the area under the entire curve is 1, because this area represents the entire population.

OBJECTIVE 2
USE A NORMAL CURVE TO DESCRIBE A NORMAL POPULATION

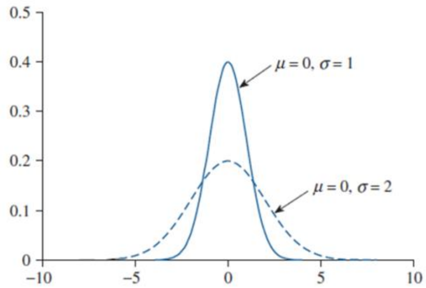
NORMAL CURVES

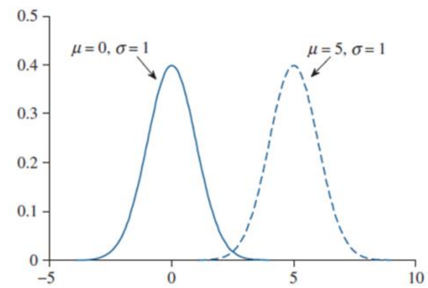
Probability density curves come in many varieties, depending on the characteristics of the populations they represent. Many important statistical procedures can be carried out using only one type of probability density curve, called a __normal curve__. A population that is represented by a normal curve is said to be **normally distributed**, or to have a **normal distribution**.



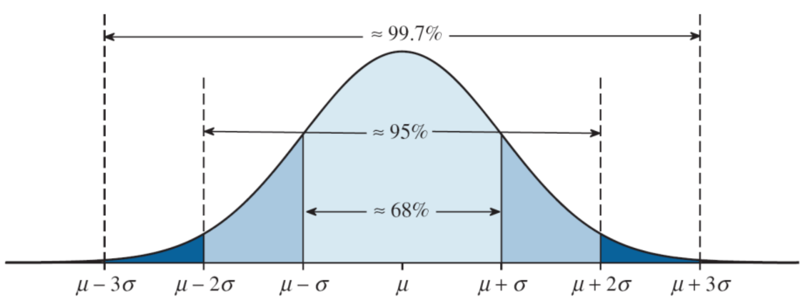
PROPERTIES OF A NORMAL CURVE

The **population mean** determines the **location of the peak**. The **population standard deviation** measures the **spread** of the population. Therefore, the normal curve is wide and flat when the population standard deviation is large, and tall and narrow when the population standard deviation is small. The **mean** and **median** of a normal distribution are both **equal to the mode**.





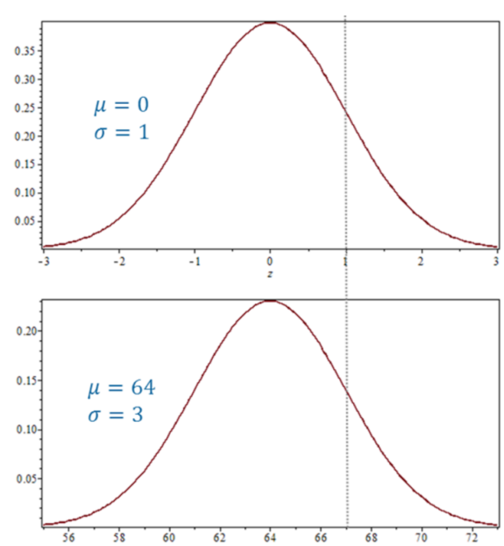
The normal distribution follows the Empirical Rule:



OBJECTIVE 3
CONVERT VALUES FROM A NORMAL DISTRIBUTION TO Z-SCORES

Recall that the z-score of a data value represents the number of standard deviations that data value is above or below the mean.

If x is a value from a normal distribution with mean μ and standard deviation σ, we can convert x to a z-score by using a method known as **standardization**. The z-score of x is $z = \frac{x-\mu}{\sigma}$.

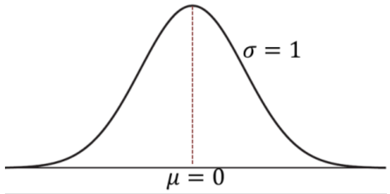


For example, consider a woman whose height is x = 67 inches from a normal population with mean μ = 64 inches and σ = 3 inches.

The z-score is:

$$z = \frac{x-\mu}{\sigma} = \frac{67-64}{3} = \frac{3}{3} = 1$$

A normal distribution can have any mean and any positive standard deviation. However, the normal distribution with a mean of ___0___ and standard deviation of ___1___ is known as the standard normal distribution.

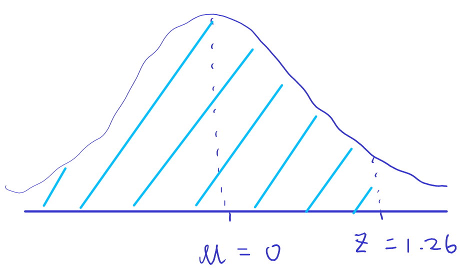


OBJECTIVE 4
FIND AREAS UNDER A NORMAL CURVE

TI-84 may be used to find the **area to the left** of a given z-score.

EXAMPLE 1: Find the area to the left of z = 1.26.

SOLUTION:



To find the area under the normal curve press $\boxed{2^{nd}}$ → $\boxed{\text{var}}$ → $\boxed{2}$ → normal cdf

TI-84
lower : -1E99
upper : 1.26
μ : 0
σ : 1

Area = $P(Z \le 1.26)$
= normalcdf(-1E99, 1.26, 0, 1)
= 0.8962

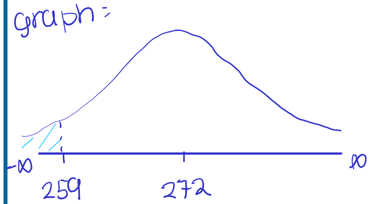
For a general normal curve, we first standardize to z-scores, then proceed using TI-84.

EXAMPLE 2: A study reported that the length of pregnancy from conception to birth is approximately normally distributed with mean μ = 272 days and standard deviation σ = 9 days. What proportion of pregnancies last less than 259 days?

SOLUTION: Let x = length of pregnancy in days. Compute the z-score of 259

$$z = \frac{x - \mu}{\sigma} = \frac{259 - 272}{9} = -1.444$$

graph:



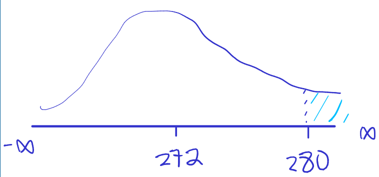
$P(x < 259) = P(Z < 1.444)$
= normalcdf(-1E99, -1.444, 0, 1)
= 0.07434

The proportion of pregnancies that last less than 259 days is 0.07434

EXAMPLE 3: A study reported that the length of pregnancy from conception to birth is approximately normally distributed with mean μ = 272 days and standard deviation σ = 9 days. What proportion of pregnancies last longer than 280 days?

SOLUTION: for $x = 280$, $z = \frac{280 - 272}{9} = \frac{8}{9} = 0.8889$

Graph:



$P(x > 280) = P(z > 0.8889)$

$= \text{normalcdf}(0.8889, 1E99, 0, 1)$

$= 0.1870$

EXAMPLE 4: The length of a pregnancy from conception to birth is approximately normally distributed with mean μ = 272 days and standard deviation σ = 9 days. A pregnancy is considered full-term if it lasts between 252 days and 298 days. What proportion of pregnancies are full-term?

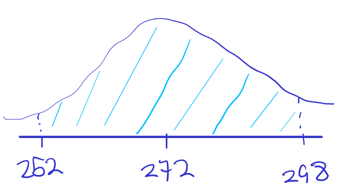
SOLUTION: for $x_1 = 252$ $\quad x_2 = 298$

$z_1 = \frac{252 - 272}{9} = -2.2222$ $\quad z_2 = \frac{298 - 272}{9} = \frac{26}{9} = 2.8889$

$P(252 < x < 298) = P(-2.2222 < x < 2.8889)$

$= \text{normalcdf}(-2.2222, 2.8889, 0, 1)$

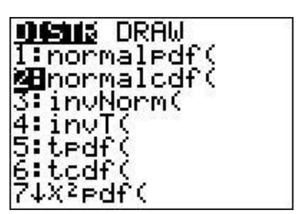
$= 0.9849$



The proportion of full-term pregnancies is 0.9849

FINDING AREAS WITH THE TI-84 PLUS

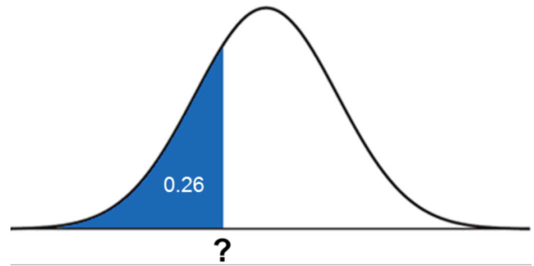
In the TI-84 PLUS calculator, the **normalcdf** command is used to find areas under a normal curve. Four numbers must be used as the input. The first entry is the **lower bound** of the area. The second entry is the **upper bound** of the area. The last two entries are the **mean** and **standard deviation**.



This command is accessed by pressing **2nd, Vars**.

OBJECTIVE 5
FIND THE VALUE FROM A NORMAL DISTRIBUTION CORRESPONDING TO A GIVEN PROPORTION

We have been finding areas under the normal curve from given z-scores. Many problems require us to go in the reverse direction. That is, if we are given an area, we need to find the z-score that corresponds to that area under the standard normal curve.



Suppose we want to find the value from a normal distribution that has a given z-score. To do this, we solve the standardization formula $z = \frac{x-\mu}{\sigma}$ for x.

The value of x that corresponds to a given z-score is: $x = \mu + z \cdot \sigma$

EXAMPLE: Heights in a group of men are normally distributed with mean μ = 69 inches and standard deviation σ = 3 inches. Find the height whose z-score is 0.6. Interpret the result.

SOLUTION: we want the height x with z-score of 0.6 then

$x = \mu + z \cdot \sigma = 69 + 0.6(3) = 70.8$

this means that a man 70.8 inches tall has height 0.6 standard deviation above the mean.

STEPS FOR FINDING NORMAL VALUES

The following procedure can be used to find the value from a normal distribution that has a given proportion above or below it using the TI-84:

Step 1: Sketch a normal curve, label the mean, label the value x to be found, and shade in and label the given area.

Step 2: Find the z-score corresponding to that area using the invNorm.

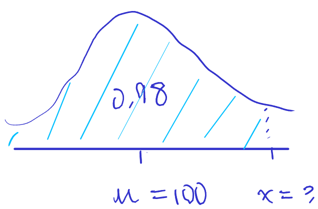
z = invNorm(area to the left, $\underset{\downarrow\,0}{\mu}$, $\underset{\downarrow\,1}{\sigma}$)

Step 3: Obtain the value from the normal distribution by computing $x = \mu + z \cdot \sigma$.

EXAMPLE: Mensa is an organization whose membership is limited to people whose IQ is in the top 2% of the population. Assume that scores on an IQ test are normally distributed with mean μ = 100 and standard deviation σ = 15. What is the minimum score needed to qualify for membership in Mensa?

SOLUTION: first, find the z-score that corresponds to the given area using invNorm. $\boxed{2^{nd}}$ → $\boxed{var}$ → 3

z = invNorm(0.98, 0, 1) = 2.0537



Now, $x = \mu + z \cdot \sigma$

$= 100 + 2.0537(15)$

$= 130.8 \approx 131$ (always round up)

The minimum score needed to qualify for membership in Mensa is 131

FINDING NORMAL VALUES GIVEN A PROPORTION WITH THE TI-84 PLUS

The **invNorm** command in the TI-84 PLUS calculator returns the z-score with a given **area to its left**. This command takes three values as its input. The first value is the area to the left, the second and third values are the mean and standard deviation.

This command is accessed by pressing **2nd, Vars**.

YOU SHOULD KNOW …

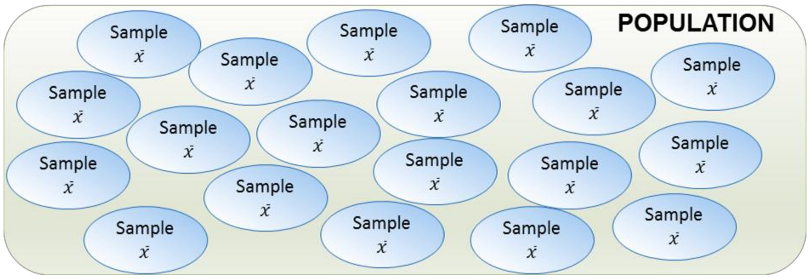
- How to use a probability density curve to describe a population
- How the shape of a normal curve is affected by the mean and standard deviation
- How to find areas under a normal curve
- How to find values from a population corresponding to areas under a normal curve

OBJECTIVES

1. Construct the sampling distribution of a sample mean
2. Use the Central Limit Theorem to compute probabilities involving sample means

OBJECTIVE 1
CONSTRUCT THE SAMPLING DISTRIBUTION OF A SAMPLE MEAN

In real situations, statistical studies involve sampling several individuals then computing numerical summaries of the samples. Most often ________________________ is computed.



If several samples are drawn from a population, they are likely to have different values for $\bar{x}$. Because the value of $\bar{x}$ varies each time a sample is drawn, $\bar{x}$ is a random variable. For each value of the random variable, $\bar{x}$, we can compute a probability. The probability distribution of $\bar{x}$ is called the
sampling distribution of $\bar{x}$.

AN EXAMPLE OF A SAMPLING DISTRIBUTION

Tetrahedral dice are shaped like a pyramid with four faces. Each face corresponds to a number between 1 and 4. Tossing a tetrahedral die is like sampling a value from the population $\{1, 2, 3, 4\}$. We can easily find the population mean, $\mu = 2.5$, and the population standard deviation $\sigma = 1.118$.

If a tetrahedral die is tossed three times, the sequence of three numbers observed is a sample of size 3 drawn with replacement. The table displays all possible samples of size 3 and their sample mean $\bar{x}$.

Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$	Sample	$\bar{x}$
1, 1, 1	1.00	2, 1, 1	1.33	3, 1, 1	1.67	4, 1, 1	2.00
1, 1, 2	1.33	2, 1, 2	1.67	3, 1, 2	2.00	4, 1, 2	2.33
1, 1, 3	1.67	2, 1, 3	2.00	3, 1, 3	2.33	4, 1, 3	2.67
1, 1, 4	2.00	2, 1, 4	2.33	3, 1, 4	2.67	4, 1, 4	3.00
1, 2, 1	1.33	2, 2, 1	1.67	3, 2, 1	2.00	4, 2, 1	2.33
1, 2, 2	1.67	2, 2, 2	2.00	3, 2, 2	2.33	4, 2, 2	2.67
1, 2, 3	2.00	2, 2, 3	2.33	3, 2, 3	2.67	4, 2, 3	3.00
1, 2, 4	2.33	2, 2, 4	2.67	3, 2, 4	3.00	4, 2, 4	3.33
1, 3, 1	1.67	2, 3, 1	2.00	3, 3, 1	2.33	4, 3, 1	2.67
1, 3, 2	2.00	2, 3, 2	2.33	3, 3, 2	2.67	4, 3, 2	3.00
1, 3, 3	2.33	2, 3, 3	2.67	3, 3, 3	3.00	4, 3, 3	3.33
1, 3, 4	2.67	2, 3, 4	3.00	3, 3, 4	3.33	4, 3, 4	3.67
1, 4, 1	2.00	2, 4, 1	2.33	3, 4, 1	2.67	4, 4, 1	3.00
1, 4, 2	2.33	2, 4, 2	2.67	3, 4, 2	3.00	4, 4, 2	3.33
1, 4, 3	2.67	2, 4, 3	3.00	3, 4, 3	3.33	4, 4, 3	3.67
1, 4, 4	3.00	2, 4, 4	3.33	3, 4, 4	3.67	4, 4, 4	4.00

The mean of all of values of $\bar{x}$ is $\mu_{\bar{x}} = 2.5$ and the standard deviation of all values of $\bar{x}$ is $\sigma_{\bar{x}} = 0.6455$. Next, we compare these values to the population mean (2.5) and population standard deviation (1.118).

Notice that we compute the population mean, variance, standard deviation as follows:

Population mean : $\mu = \frac{1+2+3+4}{4} = 2.5$

" variance : $\sigma^2 = \frac{\Sigma(x-\mu)^2}{n} = \frac{(1-2.5)^2+(2-2.5)^2+(3-2.5)^2+(4-2.5)^2}{4} = 1.25$

" standard deviation $= \sigma = \sqrt{\sigma^2} = \sqrt{1.25} = 1.118$

and the mean of sample is computed each sequence of three numbers then,
$\mu_{\bar{x}}$ is computed as the mean of $\bar{x}$'s

MEAN AND STANDARD DEVIATION OF A SAMPLING DISTRIBUTION

In the previous table, the mean of the sampling distribution is $\mu_{\bar{x}} = 2.5$, which is the same as the mean of the population, $\mu = 2.5$. This relation always holds.

The mean of the sampling distribution is denoted by $\mu_{\bar{x}}$ and equals the mean of the population:

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

The standard deviation of the sampling distribution is $\sigma_{\bar{x}} = 0.6455$, which is less than the population standard deviation $\sigma = 1.118$. It is not immediately obvious how these two quantities are related. Note, however, that $\sigma_{\bar{x}} = 0.6455 = \frac{1.118}{\sqrt{3}} = \frac{\sigma}{\sqrt{3}}$. Recall that the sample size is $n = 3$, which suggests that $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

The standard deviation of the sampling distribution, sometimes called the standard error, is denoted by $\sigma_{\bar{x}}$ and equals the standard deviation of the population divided by the square root of the sample size:

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

EXAMPLE: Among students at a certain college, the mean number of hours of television watched per week is μ = 10.5, and the standard deviation is σ = 3.6. A simple random sample of 16 students is chosen for a study of viewing habits. Let $\bar{x}$ be the mean number of hours of TV watched by the sampled students. Find the mean $\mu_{\bar{x}}$ and the standard deviation $\sigma_{\bar{x}}$ of $\bar{x}$.

SOLUTION: The mean of $\bar{x}$ is $\mu_{\bar{x}} = \mu = 10.5$

we know the sample size is $n = 16$, so the standard deviation of $\bar{x}$ is

$\sigma_{x} = \frac{\sigma}{\sqrt{n}} = \frac{3.6}{\sqrt{16}} = \frac{3.6}{4} = 0.9$

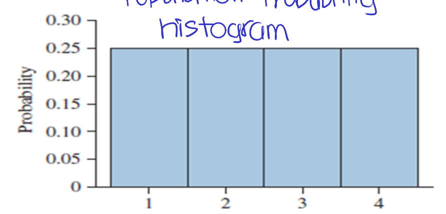
SAMPLING DISTRIBUTION FOR SAMPLE SIZE 3

Consider again the tetrahedral die example. The sampling distribution for $\bar{x}$ can be determined from the table of all possible values of $\bar{x}$ for a sample of size 3. The probability that the sample mean is 1.00 is $\frac{1}{64}$, or 0.015625, because out of the 64 possible samples, only 1 has a sample mean equal to 1.00. Similarly, the probability that $\bar{x}$ =1.33 is $\frac{3}{64}$, or 0.046875, because there are 3 samples whose sample mean is 1.33. The probability distribution is:

$\bar{x}$	$P(\bar{x})$
1.00	0.015625
1.33	0.046875
1.67	0.093750
2.00	0.156250
2.33	0.187500
2.67	0.187500
3.00	0.156250
3.33	0.093750
3.67	0.046875
4.00	0.015625

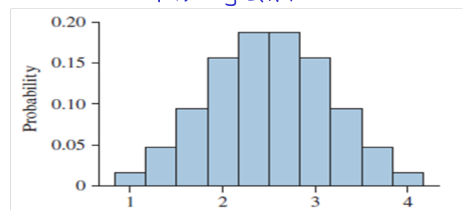
PROBABILITY HISTOGRAM FOR A SAMPLING DISTRIBUTION

Population Probability histogram



In the tetrahedral die example, the population is $\{1,\ 2,\ 3,\ 4\}$. When a die is rolled, each number has the same chance of appearing, $\frac{1}{4}$ or 0.25. The probability histogram for the sampling distribution of $\bar{x}$ with sample size 3 is obtained from the sampling distribution on the previous slide.

sampling distribution histogram



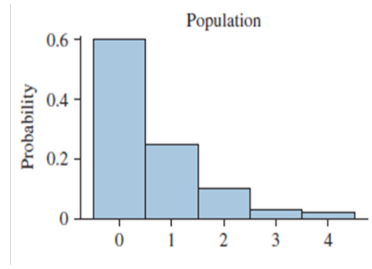
The probability histogram for the sampling distribution looks a lot like the normal curve, whereas the probability histogram for the population does not.

Remarkably, it is true that, for any population, if the sample size is large enough, the sample mean $\bar{x}$ will be approximately normally distributed. For a symmetric population like the tetrahedral die population, the sample mean is approximately normally distributed even for a small sample size like n = 3.

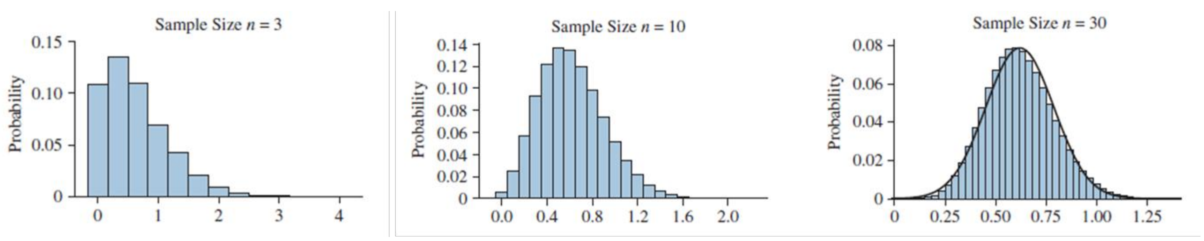
SAMPLING DISTRIBUTION OF A SKEWED POPULATION

If a population is skewed, a larger sample size is necessary for the sampling distribution of $\bar{x}$ to be approximately normal. Consider the following probability distribution

x	$P(x)$
0	0.60
1	0.25
2	0.10
3	0.03
4	0.02



Below are the probability histograms for the sampling distribution of $\bar{x}$ for samples of size 3, 10, and 30. Note that the shapes of the distributions begin to approximate a normal curve as the sample size increases.



The size of the sample needed to obtain approximate normality depends mostly on the skewness of the population. In practice, a sample of size n > 30 is large enough. The remarkable fact that the sampling distribution of $\bar{x}$ is approximately normal for a large sample from any distribution is part of one of the most used theorems in Statistics, the **Central Limit Theorem**.

THE CENTRAL LIMIT THEOREM: Let $\bar{x}$ be the mean of a large ($n > 30$) simple random sample from a population with mean μ and standard deviation σ, then $\bar{x}$ has an approximately normal distribution, with mean $\mu_{\bar{x}} = \mu$ and standard deviation $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

The Central Limit Theorem applies for all populations. However, for symmetric populations, a smaller sample size may suffice. If the population itself is normal, the sample mean $\bar{x}$ will be normal for any sample size.

EXAMPLE 1: A sample of size 45 will be drawn from a population with mean μ = 15 and standard deviation σ = 3.5. Is it appropriate to use the normal distribution to find probabilities for $\bar{x}$?

SOLUTION: Yes, by the central limit theorem (clt), since $n > 30$, $\bar{x}$ has an approximately normal distribution

EXAMPLE 2: A sample of size 8 will be drawn from a normal population with mean μ = –60 and standard deviation σ = 5. Is it appropriate to use the normal distribution to find probabilities for $\bar{x}$?

SOLUTION: Yes, since the population itself is normal, $\bar{x}$ has approximately normal distribution

EXAMPLE 3: A sample of size 24 will be drawn from a population with mean μ = 35 and standard deviation σ = 1.2. Is it appropriate to use the normal distribution to find probabilities for $\bar{x}$?

SOLUTION: No, since the population is not known to be normal and n is not greater than 30, we cannot be certain that $\bar{x}$ has an approximately normal distribution

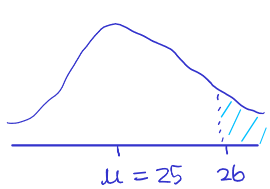
OBJECTIVE 2
USE THE CENTRAL LIMIT THEOREM TO COMPUTE PROBABILITIES INVOLVING SAMPLE MEANS

EXAMPLE 1: Based on data from the U.S. Census, the mean age of college students in 2011 was μ = 25 years, with a standard deviation of σ = 9.5 years. A simple random sample of 125 students is drawn. What is the probability that the sample mean age of the students is greater than 26 years?

SOLUTION: The sample size is $n = 125$, which is greater than 30; thus we use normal curve. (Apply clt)

Let's compute $\mu_{\bar{x}}$ and σ_x. $\mu_{\bar{x}} = \mu = 25$ and $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{9.5}{\sqrt{125}} = 0.85$

Find the area under the normal curve



$P(\bar{x} > 26) = \text{normalcdf}(\text{lower}, \text{upper}, \mu_{\bar{x}}, \sigma_{\bar{x}})$

$= \text{normalcdf}(26, 1E99, 25, 0.85)$

$= 0.1197$

The probability that the sample mean age of students is greater than 26 is 0.1197

EXAMPLE 2: Hereford cattle are one of the most popular breeds of cattle. Based on data from the Hereford Cattle Society, the mean weight of a one-year-old Hereford bull is 1135 pounds, with a standard deviation of 97 pounds. Would it be unusual for the mean weight of 100 head of cattle to be less than 1100 pounds?

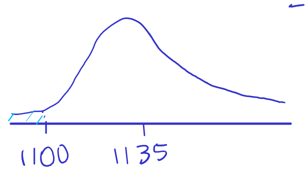
SOLUTION: $n = 100 > 30$, so we may use the normal curve

$$\mu_{\bar{x}} = \mu = 1135 \text{ and } \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{97}{\sqrt{100}} = 0.97$$

Find the area under the normal curve

$$P(\bar{x} < 100) = \text{normalcdf}(-1E99, 1100, 1135, 0.97)$$

$$= 0.0001542$$



since $P(\bar{x} < 100) = 0.000152 < 0.05$,
this event is unusual

YOU SHOULD KNOW …

- How to construct the sampling distribution of a sample mean
- How to find the mean and standard deviation of a sampling distribution of $\bar{x}$
- The Central Limit Theorem
- How to use the Central Limit Theorem to compute probabilities involving sample means

Sample proportion $\hat{p} = \frac{x}{n}$

SECTION 7.4: THE CENTRAL LIMIT THEOREM FOR PROPORTIONS

OBJECTIVES

1. Construct the sampling distribution for a sample proportion
2. Use the Central Limit Theorem to compute probabilities for sample proportions

OBJECTIVE 1
CONSTRUCT THE SAMPLING DISTRIBUTION FOR A SAMPLE PROPORTION

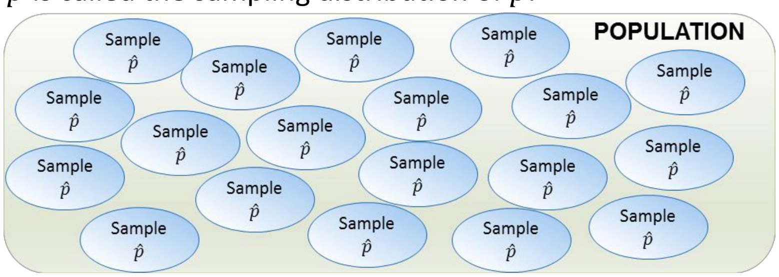
In a population, the proportion who have a certain characteristic is called the _population proportion_. The population proportion is denoted by p. In a simple random sample of n individuals, let x be the number in the sample who have the characteristic. The _sample proportion_ is $\hat{p} = \frac{x}{n}$.

EXAMPLE: A computer retailer wants to estimate the proportion of people in her city who own laptop computers. She cannot survey everyone in the city, so she draws a sample of 100 people and surveys them. It turns out that 35 out of the 100 people in the sample own laptops.

- The sample proportion is _$\hat{p} = \frac{35}{100}$_.
- The proportion of people in the entire city who own laptops is the population proportion, p.

SAMPLING DISTRIBUTION OF $\hat{p}$

If several samples are drawn from a population, they are likely to have different values for $\hat{p}$. Because the value of $\hat{p}$ varies each time a sample is drawn, $\hat{p}$ is a random variable, and it has a probability distribution. The probability distribution of $\hat{p}$ is called the sampling distribution of $\hat{p}$.



AN EXAMPLE OF A SAMPLING DISTRIBUTION

Consider tossing a fair coin five times. This produces a sample of size n = 5, where each item in the sample is either a head or a tail. The proportion of times the coin lands on heads is the sample proportion $\hat{p}$. The probability that it lands heads each time is 0.5. Therefore, the population proportion of heads is p = 0.5. All 32 possible samples are presented below. 2^n

Sample	$\hat{p}$	Sample	$\hat{p}$	Sample	$\hat{p}$
TTTTT	0.0	THTHH	0.6	HTHHT	0.6
TTTTH	0.2	THHTT	0.4	HTHHH	0.8
TTTHT	0.2	THHTH	0.6	HHTTT	0.4
TTTHH	0.4	THHHT	0.6	HHTTH	0.6
TTHTT	0.2	THHHH	0.8	HHTHT	0.6
TTHTH	0.4	HTTTT	0.2	HHTHH	0.8
TTHHT	0.4	HTTTH	0.4	HHHTT	0.6
TTHHH	0.6	HTTHT	0.4	HHHTH	0.8
THTTT	0.2	HTTHH	0.6	HHHHT	0.8
THTTH	0.4	HTHTT	0.4	HHHHH	1.0
THTHT	0.4	HTHTH	0.6		

The table displays all possible samples of size 5 and their sample proportion $\hat{p}$. The mean of all of values of $\hat{p}$ is $\mu_{\hat{p}} = 0.5$ and the standard deviation of all values of $\hat{p}$ is $\sigma_{\hat{p}} = 0.2236$. Next, we compare these values to the population proportion (0.5) and the sample size (5).

MEAN AND STANDARD DEVIATION OF A SAMPLING DISTRIBUTION

In the previous table, the mean of the sampling distribution is $\mu_{\hat{p}} = 0.5$, which is the same as the population proportion, $p = 0.5$. This relation always holds.

The mean of the sampling distribution of $\hat{p}$ is denoted by $\mu_{\hat{p}}$ and equals the population proportion:

$$\mu_{\hat{p}} = p$$

The standard deviation of the sampling distribution is $\sigma_{\hat{p}} = 0.2236$. It is not immediately obvious how this is related to the population proportion p and the sample size n. Note, however, that $\sigma_{\hat{p}} = 0.2236 = \sqrt{\frac{0.5(1-0.5)}{5}} = \sqrt{\frac{p(1-p)}{n}}$

The standard deviation of the sampling distribution of $\hat{p}$ is denoted by $\sigma_{\hat{p}}$, is:

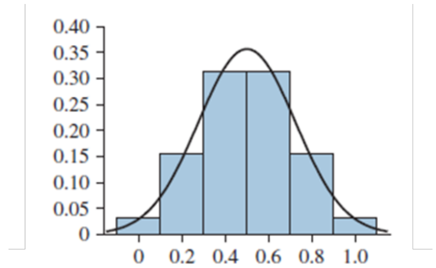
$$\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}}$$

EXAMPLE: The soft-drink cups at a certain fast-food restaurant have tickets attached to them. Customers peel off the tickets to see whether they win a prize. The proportion of tickets that are winners is p = 0.25. A total of n = 70 people purchase soft drinks between noon and 1:00 PM on a certain day. Let $\hat{p}$ be the proportion that win a prize. Find the mean and standard deviation of $\hat{p}$.

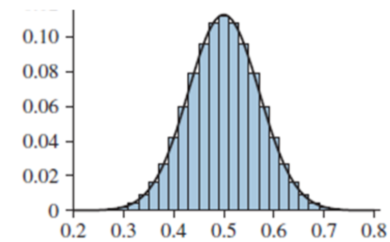
SOLUTION: The population proportion is $p = 0.5$
sample size $n = 70$
mean of $\hat{p}$: $\mu_{\hat{p}} = p = 0.25$
standard deviation of $\hat{p}$: $\sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.25(1-0.25)}{70}}$
$= 0.0518$

PROBABILITY HISTOGRAM FOR A SAMPLING DISTRIBUTION

The probability histogram for the sampling distribution of $\hat{p}$ for the proportion of heads in five tosses of a fair coin is presented. We can see that the distribution is reasonably well approximated by a normal curve.



The experiment is conducted again but with 50 tosses of a fair coin. As we increased the number of tosses, the sampling distribution of $\hat{p}$ is more closely approximated by a normal curve.



When p = 0.5, the sampling distribution of $\hat{p}$ is somewhat close to normal even for a small sample size like n = 5. When p is close to 0 or close to 1, a larger sample size is needed before the distribution of $\hat{p}$ is close to normal.

A common rule of thumb is that the distribution may be approximated with a normal curve whenever $n \cdot p$ and $n(1-p)$ are both at least 10.

THE CENTRAL LIMIT THEOREM FOR PROPORTIONS:
Let $\hat{p}$ be the sample proportion for a sample size n from population with population p. If $n \cdot p \geq 10$ and $n(1-p) \geq 10$, then the distribution of $\hat{p}$ is approximately normal, with mean $\mu_{\hat{p}} = p$ and standard deviation $\sigma_{\hat{p}} = \sqrt{\dfrac{p(1-p)}{n}}$

EXAMPLE 1: A sample of size 20 is drawn from a population with population proportion p = 0.7. Is it appropriate to use the normal distribution to find probabilities for $\hat{p}$?

SOLUTION: $n = 20$ and $p = 0.7$
since $n \cdot p = 20(0.7) = 14$ and $n(1-p) = 20(1-0.7) = 6$, we cannot be sure that the distribution of $\hat{p}$ is approximately normal

EXAMPLE 2: A sample of size 55 is drawn from a population with population proportion p = 0.8. Is it appropriate to use the normal distribution to find probabilities for $\hat{p}$?

SOLUTION: Since $n \cdot p = 55(0.8) = 44$ and $n(1-p) = 55(1-0.8) = 11$ are both at least 10, the distribution of $\hat{p}$ is approximately normal by CLT

OBJECTIVE 2
USE THE CENTRAL LIMIT THEOREM TO COMPUTE PROBABILITIES FOR SAMPLE PROPORTIONS

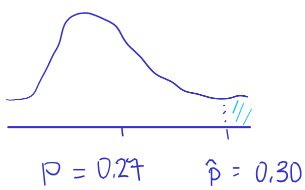
EXAMPLE 1: According to a Harris poll taken in September 2013, chocolate is the favorite ice cream flavor for 27% of Americans. If a sample of 100 Americans is taken, what is the probability that the sample proportion of those who prefer chocolate is greater than 0.30?

SOLUTION: Check assumption : $n \cdot p = 100(0.27) = 27 \geq 10$ ✓ and $n(1-p) = 100(1-0.27) = 73 \geq 10$ ✓

we may use normal curve

$$\mu_{\hat{p}} = p = 0.27 \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{(0.27(1-0.27))}{100}} = 0.044396$$

$$P(\hat{p} > 0.30) = \text{normalcdf}(\text{lower}, \text{upper}, \mu_{\hat{p}}, \sigma_{\hat{p}})$$
$$= \text{normalcdf}(0.30, 1, 0.27, 0.044396)$$
$$= 0.2496$$



EXAMPLE 2: In the 2012 U.S. presidential election, 51% of voters voted for Barack Obama. If a sample of 75 voters were polled, would it be unusual if less than 40% of them had voted for Barack Obama?

SOLUTION: We first check the assumptions, $n \cdot p = 75(0.51) = 38.25 \geq 0$
$n(1-p) = 75(1-0.51) = 36.75 \geq 0$

we may use the normal curve

$$\mu_{\hat{p}} = \mu = 0.51 \text{ and } \sigma_{\hat{p}} = \sqrt{\frac{p(1-p)}{n}} = \sqrt{\frac{0.51(1-0.51)}{75}} = 0.057723$$

$$P(\hat{p} < 0.4) = \text{normalcdf}(0, 0.40, 0.51, 0.057723)$$
$$= 0.0283$$

It would be unusual for the sample proportion to be less than 0.40.

YOU SHOULD KNOW …

- The notation for sample and population proportions
- How to construct the sampling distribution for a sample proportion
- How to find the mean and standard deviation of the sampling distribution of $\hat{p}$
- How to use the Central Limit Theorem for Proportions to compute probabilities for sample proportions

OBJECTIVES

1. Use dotplots to assess normality
2. Use boxplots to assess normality
3. Use histograms to assess normality
4. Use stem-and-leaf plots to assess normality
5. Use normal quantile plots to assess normality

OBJECTIVE 1
USE DOTPLOTS TO ASSESS NORMALITY

ASSESSING NORMALITY

Many statistical procedures require that we draw a sample from a population whose distribution is approximately normal. Often we don't know whether the population is approximately normal when we draw the sample. So the only way we have to assess whether the population is approximately normal is to **examine the sample**.

There are three important ideas to remember when assessing normality:
1. We are not trying to determine whether the population is exactly normal
2. Assessing normality is more important for small than
3. Hard and fast rules do not work well

We will **reject** the assumption that a population is approximately normal if a sample has any of the following features
1. The sample contains outlier
2. " " exhibits a large degree of skewness
3. " " has more than one distinct mode

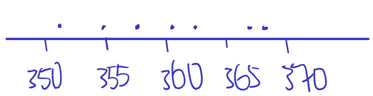
If the sample has none of the preceding features, we will treat the population as being approximately normal.

EXAMPLE: The accuracy of an oven thermostat is being tested. The oven is set to 360 degrees (F), and the temperature when the thermostat turns off is recorded. A sample of size 7 yields the following results:

358 363 361 355 367 352 368

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION: First construct a dotplot



The dotplot does not reveal any outliers, The dotplot does not exhibits a large degree of skewness, and there is no evidence that the population has more than one mode. Thus, we can treat this a sample

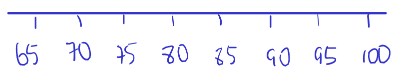
EXAMPLE: At a recent health fair, several hundred people had their pulse rates measured. A simple random sample of six records was drawn, and the pulse rates, in beats per minute, were

68 71 79 98 67 75

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION:

using the dotplot, it is clear that



OBJECTIVE 2
USE BOXPLOTS TO ASSESS NORMALITY

EXAMPLE: An insurance adjuster obtains a sample of 20 estimates, in hundreds of dollars, for repairs to cars damaged in collisions. Following are the data.

12.1	15.7	14.2	4.6	8.2	11.6	12.9	11.2	14.9	13.7
6.6	7.2	12.6	9.0	11.9	7.8	9.0	16.2	16.5	12.1

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION:

EXAMPLE: A recycler determines the amount of recycled newspaper, in cubic feet, collected each week. Following are the results for a sample of 18 weeks.

2129	2853	2530	2054	2075	2011	2162	2285	2668
3194	4834	2469	2380	2567	4117	2337	3179	3157

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION:

OBJECTIVE 3
USE HISTOGRAMS TO ASSESS NORMALITY

EXAMPLE: Diameters were measured, in millimeters, for a simple random sample of 20 grade A eggs from a certain farm. The results were

59	60	60	56	59	56	62	58	60	59
61	59	61	61	63	60	56	58	63	58

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION:

EXAMPLE: A shoe manufacturer is testing a new type of leather sole. A simple random sample of 22 people wore shoes with the new sole for a period of four months. The amount of wear on the right shoe was measured for each person. The results, in thousandths of an inch, were

24.1	2.2	11.8	2.7	4.1	13.9	33.6	2.4	36.2	16.8	5.4
4.6	4.5	4.1	6.1	6.3	22.6	29.1	12.2	4.6	15.8	7.7

Is it reasonable to treat this as a sample from an approximately normal population? Explain.

SOLUTION:

OBJECTIVE 4
USE STEM-AND-LEAF PLOTS TO ASSESS NORMALITY

EXAMPLE: A psychologist measures the time it takes for each of 20 rats to run a maze. The times, in seconds, are

54	48	49	54	63	54	66
32	45	52	41	37	56	56
52	53	41	45	48	43	

Construct a stem-and-leaf plot for these data. Is it reasonable to treat this as a random sample from an approximately normal population?

SOLUTION:

3	2 7
4	1 1 3 5 5 8 8 9
5	2 2 3 4 4 4 6 6
6	3 6

The stem-and-leaf plot reveals no outliers, strong skewness, or multimodality. We may treat this as a sample

OBJECTIVE 5
USE NORMAL QUANTILE PLOTS TO ASSESS NORMALITY

Normal quantile plots are somewhat more complex than dotplots, histograms, and stem-and-leaf plots. The following example illustrates how to construct a normal quantile plot.

A simple random sample of size n = 5 is drawn, and we want to determine whether the population it came from is approximately normal. The five sample values, in increasing order, are 3.0 3.3 4.8 5.9 7.8.

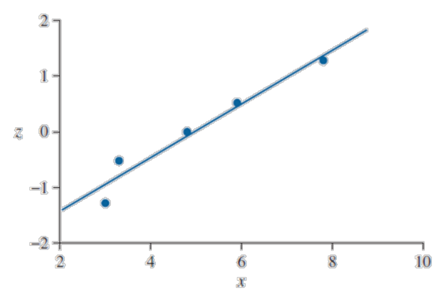
Step 1: Let n be the number of values in the data set. Spread the n values evenly over the interval from 0 to 1. This is done by assigning the value $\frac{1}{2n}$ to the first sample value, $\frac{3}{2n}$ to the second, and so forth. The last sample value will be assigned the value $\frac{2n-1}{2n}$. These values, denoted a_i , represent areas under the normal curve. For n = 5, the values are 0.1, 0.3, 0.5, 0.7, and 0.9.

i	1	2	3	4	5
x_i	3.0	3.3	4.8	5.9	7.8
a_i	0.1	0.3	0.5	0.7	0.9

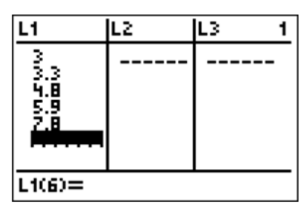
Step 2: The values assigned in Step 1 represent left-tail areas under the normal curve. We now find the z-scores corresponding to each of these areas. The results are shown in the following table.

i	1	2	3	4	5
x_i	3.0	3.3	4.8	5.9	7.8
a_i	0.1	0.3	0.5	0.7	0.9
z_i	–1.28	–0.52	0.00	0.52	128

Step 3: Plot the points (x_i , z_i). If the points approximately follow a straight line, then the population may be treated as being approximately normal. If the points deviate substantially from a straight line, the population should not be treated as normal. In this example, the points do approximately follow a straight line, so we may treat this population as approximately normal.

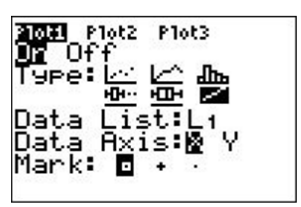


TI-84 PLUS PROCEDURE FOR QUANTILE PLOTS



Step 1: Enter the data into **L1** in the data editor.

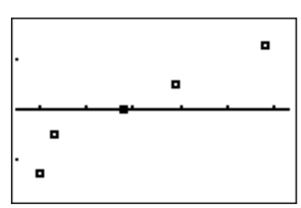
Step 2: Press **2nd, Y=** to access the STAT PLOTS menu and select Plot1 by pressing **1**.



Step 3: Select **On** and the normal quantileplot icon.

Step 4: For **Data List**, select **L1**, and for **Data Axis**, choose the **X** option.

Step 5: Press **ZOOM** and then **9: ZoomStat**.



EXAMPLE: A placement exam is given to each entering freshman at a large university. A simple random sample of 20 exam scores is drawn, with the following results.

61	60	60	68	63	63	94
66	65	98	61	71	74	63
66	61	61	65	72	85	

Construct a normal probability plot using technology. Is the distribution of exam scores approximately normal?

SOLUTION: The points on the quantile plot do not closely follows a striaght line.

The distribution is not approximately normal

YOU SHOULD KNOW …

- The conditions for rejecting the assumption that a population is approximately normal
- How to use dotplots to assess normality
- How to use boxplots to assess normality
- How to use histograms to assess normality
- How to use stem-and-leaf plots to assess normality
- How to use quantile plots to assess normality