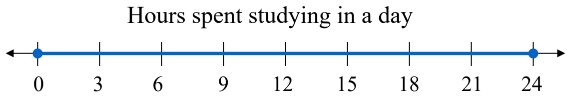
Properties of a Normal Distribution

Continuous random variable

- Has an infinite number of possible values that can be represented by an interval on the number line.



The time spent studying can be any number between 0 and 24.

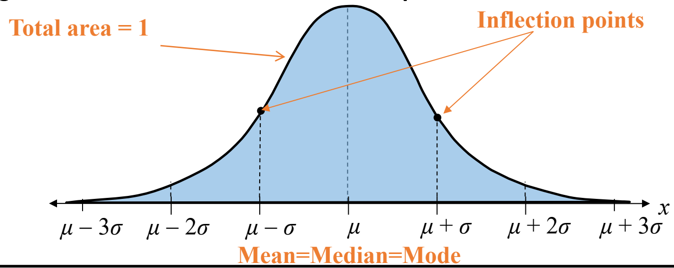
Continuous probability distribution

- The probability distribution of a continuous random variable.

Properties of a Normal Distribution

- A **normal distribution** is a continuous probability distribution for a random variable x.

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 one.
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 the **inflection points**.



Probability Density Function (PDF)

- A continuous probability distribution can be graphed with a **probability density function (pdf).**
- A probability density function has two requirements:
  1. The total area under the curve is equal to 1
  2. The function can never be negative.

  - The mathematical equation for the normal distribution

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

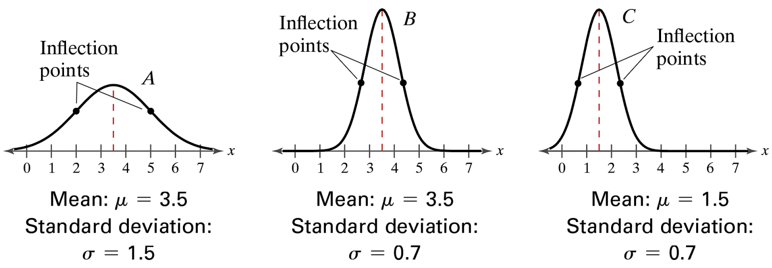
A normal curve depends completely on the two parameters μ and σ because $e \approx 2.718$ and $\pi \approx 3.14$ are constants.

e ≈ 2.718
π = 3.14
μ = population mean
σ^2 = population variance
σ = population standard deviation

3

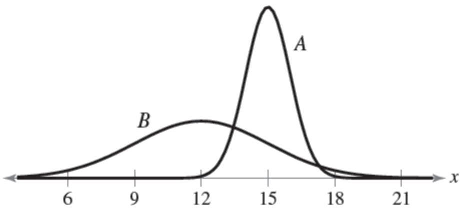
Means and Standard Deviations

- A normal distribution can have any mean and any positive standard deviation.
- The mean gives the location of the line of symmetry.
- The standard deviation describes the spread of the data.



4

Ex1. Which curve has the greater mean? Which curve has the greater standard deviation?



Solution:

Curve *A* has the greater mean (The line of symmetry of curve *A* occurs at $x = 15$. The line of symmetry of curve *B* occurs at $x = 12$.)

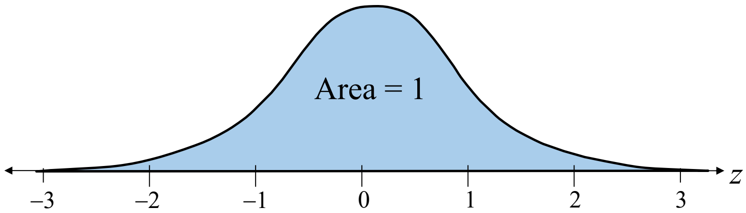
Solution:

Curve *B* has the greater standard deviation (Curve *B* is more spread out than curve *A*.)

The Standard Normal Distribution

Standard normal distribution

- The **standard normal distribution** is a normal distribution with a mean of 0 and a standard deviation of 1. The total area under its normal curve is 1.

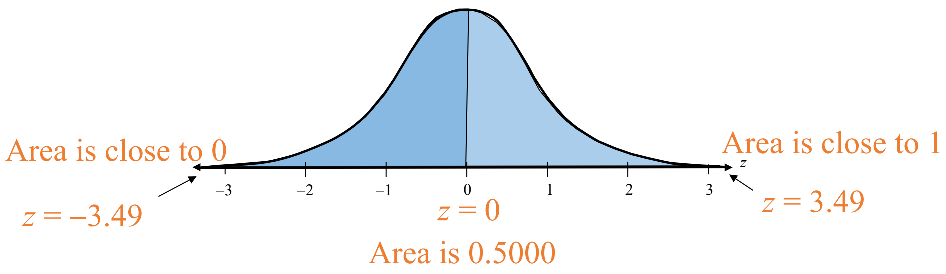


- Any x-value can be transformed into a z-score by using the formula

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

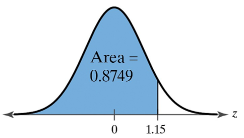
Properties of the Standard Normal Distribution

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 $z = 3.49$.



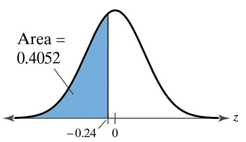
7

Ex1. Find the cumulative area that corresponds to a *z*-score of 1.15.



The area to the left of $z = 1.15$ is 0.8749.

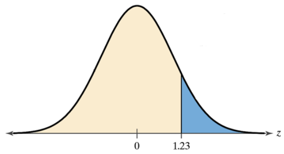
Ex2. Find the cumulative area that corresponds to a *z*-score of –0.24.



The area to the left of $z = -0.24$ is 0.4052.

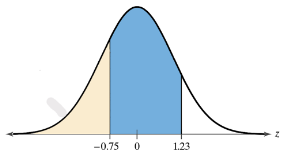
8

Ex3. Find the area that corresponds to a *z*-score of 1.23.



The area to the left of $z = 1.23$ is 0.8907.
Subtract to find the area to the right of z=1.23: 1–0.8907 = 0.1093.

Ex4. Find the cumulative area that corresponds to *z*-scores of –0.75 and 1.23.



The area to the left of $z = 1.23$ is 0.8907.
The area to the left of $z = –0.75$ is 0.2266.
Subtract to find the area of the region between the two *z*-scores:
0.8907 – 0.2266 = 0.6641.

9

Ex5. Find the area under the standard normal curve to the left of z = –0.99.

Ex6. Find the area under the standard normal curve to the right of *z* = 1.00.

Ex7. Find the area under the standard normal curve between *z* = –1.5 and *z* = 1.25.

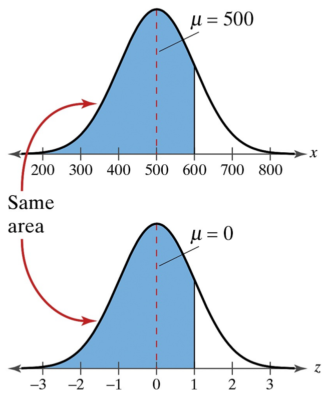
Ex8. Find the area under the standard normal curve between *z* = 2.00 and *z* = 2.47.

Ex8. Find the area under the standard normal curve between *z* = – 2.48 and *z* = – 0.83.

10

Probability and Normal Distributions

- If a random variable x is normally distributed, you can find the probability that x will fall in a given interval by calculating the area under the normal curve for that interval.
- Consider a normal curve with μ = 500 and σ = 100 (upper right).
- The value of x one standard deviation above the mean is $\mu + \sigma = 500 + 100 = 600$.
- Now consider the standard normal curve (lower right).
- The value of z one standard deviation above the mean is $\mu + \sigma = 0 + 1 = 1$.
- The z-score of 1 corresponds to an x-value of 600, and areas are not changed with a transformation to a standard normal curve, the shaded areas at the right are equal.
- $P(X < 600) = P(z < 1) = 0.8413$



11

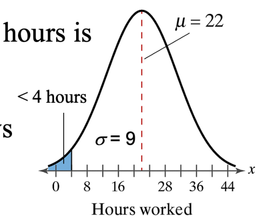
Ex1. A national study found that college students with jobs worked an average of 22 hours per week. The standard deviation is 9 hours. A college student with a job is selected at random. Find the probability that the student works for less than 4 hours per week. Assume that the lengths of time college students work are normally distributed and are represented by the variable x.

Solution

- The z-score that corresponds to 4 hours is

$$z = \frac{x - \mu}{\sigma} = \frac{4 - 22}{9} = -2.$$

- The Standard Normal Table shows that $P(z < -2) = 0.0228$.
- The probability that the student works for less that 4 hours per week is 0.0228.



So, 2.28% of college students with jobs worked for less than 4 hours per week. Because 2.28% is less than 5%, this is an unusual event.

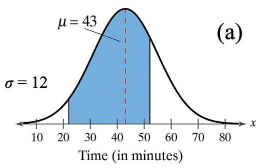
12

Ex2. A survey indicates that for each trip to a supermarket, a shopper spends an average of 43 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 between 22 and 52 minutes.

b. Find the probability that the shopper will be in the store for more than 37 minutes.

c. When 200 shoppers enter the store, how many shoppers would you expect to be in the store for each interval of time listed parts a and b?



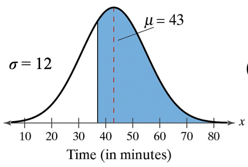
(a) $z_1 = \dfrac{x-\mu}{\sigma} = \dfrac{22-43}{12} = -1.75$

$z_2 = \dfrac{x-\mu}{\sigma} = \dfrac{52-43}{12} = 0.75$

$P(22 < x < 52) = P(-1.75 < z < 0.75)$

$= 0.7734 - 0.0401 = 0.7333$

(c) When 200 shoppers enter the store, you would expect about 200(0.7333) = 146.66 ≈ 147 shoppers to be in the store between 22 and 52 minutes.



(b) $z = \dfrac{x-\mu}{\sigma} = \dfrac{37-43}{12} = -0.5$

$P(x > 37) = P(z > -0.5) = 1 - 0.3085 = 0.6915$

(c) When 200 shoppers enter the store, you would expect about 200(0.6915) = 138.3 ≈ 138 shoppers to be in the store more than 37 minutes.

13

Ex3. The mean number of hours an American worker spends on the computer is 3.1 hours per workday. Assume that the standard deviation is 0.5 hour. Find the percentage of workers who spend less than 3.5 hours on the computer. Assume the variable is normally distributed.

Ex4. Each month, an American household generates an average of 28 pounds of newspaper for recycling. Assume the standard deviation is 2 pounds. If a household is selected at random, find the probability of generating newspaper for recycling is

a) Between 27 and 31 pounds per month.

b) More than 30.2 pounds per month.

14

Ex5. An American Automobile Association reports that the average time it takes to respond to an emergency call is 25 minutes. Assume the variable is approximately normally distributed and the standard deviation is 4.5 minutes. If 80 calls are randomly selected, approximately how many of them will be responded to in less than 15 minutes?

Ex6. To qualify for a police academy, candidates must score in the top 10% on a general ability test. The test has a mean of 200 and a standard deviation of 20. Find the lowest possible score to qualify. Assume the test scores are normally distributed.

15

Ex7. For a medical study, a researcher wishes to select people in the middle 60% of the population based on blood pressure. If the mean systolic blood pressure is 120 and the standard deviation is 8, find the upper and lower values that would qualify people to participate the study. Assume that blood pressure values are normally distributed.

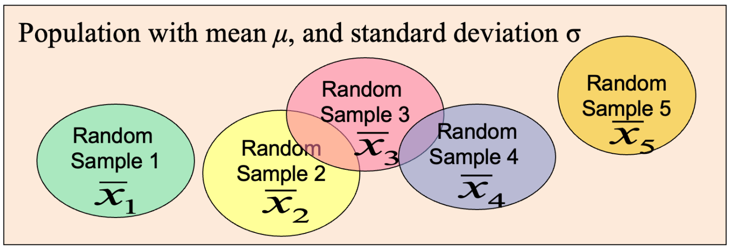
Ex8. In a randomly selected sample of women ages 20 –34, the mean total cholesterol level is 179 milligrams per deciliter with a standard deviation of 38.9 milligrams per deciliter. Assume the total cholesterol levels are normally distributed. Find the highest total cholesterol level a woman in this 20 –34 age group can have and still be in the bottom 1%.

16

Sampling Distributions

Sampling distribution

- The probability distribution of a sample statistic that is formed when random 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**



- **Sampling error** is the difference between the sample measure and the corresponding population measure due to the fact that the sample is not a perfect representation of the population.

17

Properties of Sampling Distributions of Sample Means

1. The mean of the sample means, μ_X, is equal to the population mean μ.

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

2. The standard deviation of the sample means, σ_X, is equal to the population standard deviation, σ divided by the square root of the sample size, n.

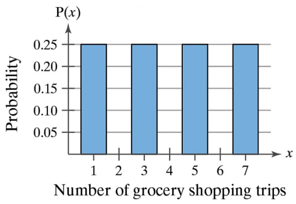
$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$$

- Called the **standard error of the mean.**

18

Ex. The number of times four people go grocery shopping in a month is given by the population values {1, 3, 5, 7}. A probability histogram for the data is shown.

You randomly choose two of the four people, 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 μ= 4, variance σ^2 = 5, and standard deviation $\sigma = \sqrt{5} \approx 2.2$ of the population.



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

The mean, variance, and standard deviation of the 16 sample means are

Mean of the sample means

$$\mu_X = \frac{1+2+\cdots+6+7}{16} = 4$$

Standard deviation of the sample means

$$\sigma_X^2 = \frac{5}{2} = 2.5 \qquad \sigma_X = \frac{\sqrt{5}}{\sqrt{2}} = 1.581$$

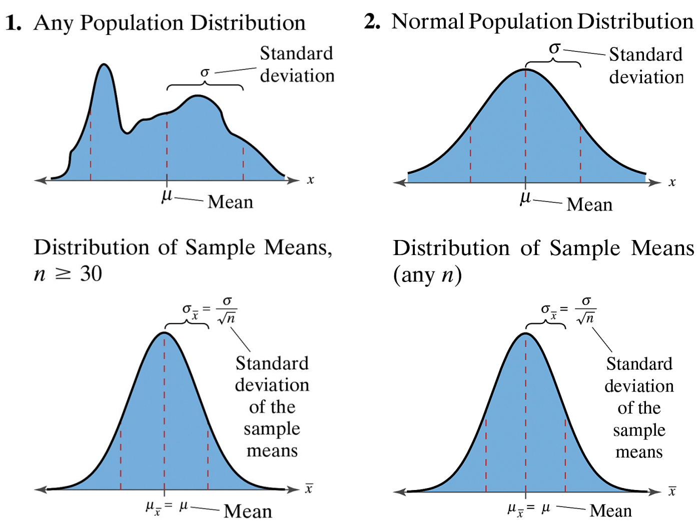
These results satisfy the properties of sampling distributions because

$$\mu_{\bar{x}} = \mu = 4 \quad \text{and} \quad \sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{\sqrt{5}}{\sqrt{2}} \approx 1.581.$$

19

The Central Limit Theorem

1. If samples of size $n \geq 30$, are drawn from any population with mean = μ and standard deviation = σ, then the sampling distribution of the 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 the sample means is normally distribution for ***any*** sample size n.



20

The Central Limit Theorem

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

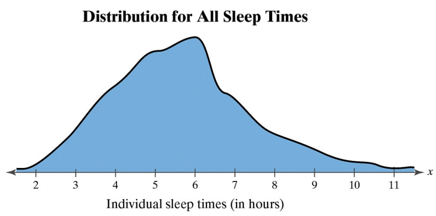
$$\mu_{\bar{x}} = \mu \qquad \text{Mean of the sample means}$$

- 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.*

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

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} \qquad \text{Standard deviation of the sample means (Standard error of the mean)}$$

Ex1. A study analyzed the sleep habits of college students. The study found that the mean sleep time was 6.8 hours, with a standard deviation of 1.4 hours. Random samples of 100 sleep times 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. *(Adapted from The Journal of American College Health)*



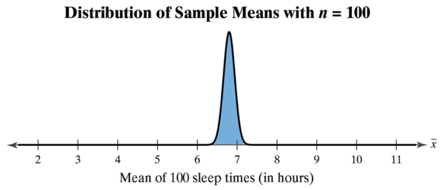
- The mean of the sampling distribution is equal to the population mean

$$\mu_x = \mu = 6.8$$

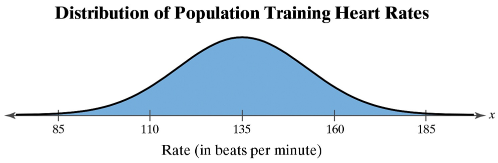
- The standard error of the mean is equal to the population standard deviation divided by $\sqrt{n}$.

$$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{1.4}{\sqrt{100}} = 0.14$$

- Since the sample size is greater than 30, the sampling distribution can be approximated by a normal distribution with a mean of 6.8 hours and a standard deviation of 0.14 hour.



Ex2. The training heart rates of all 20-years old athletes are normally distributed, with a mean of 135 beats per minute and standard deviation of 18 beats per minute. Random samples of size 4 are drawn from this population, and the mean of each sample is determined. Find the mean and standard error of the mean of the sampling distribution. Then sketch a graph of the sampling distribution of sample means.



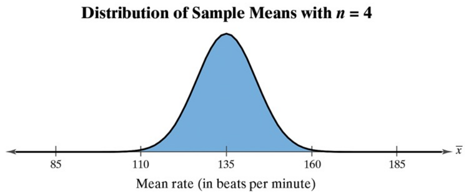
- The mean of the sample means

$$\mu_{\bar{x}} = \mu = 135 \text{ beats per minute}$$

- The standard deviation of the sample means

$$\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}} = \frac{18}{\sqrt{4}} = 9 \text{ beats per minute}$$

- Since the population is normally distributed, the sampling distribution of the sample means is also normally distributed.



23

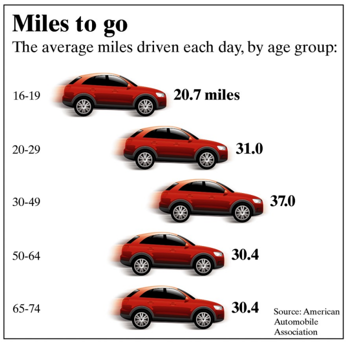
Probability and the Central Limit Theorem

- To transform $\bar{x}$ to a z-score

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

24

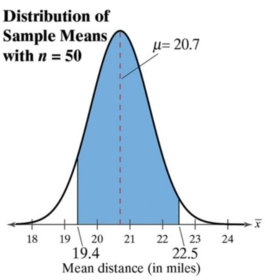
Ex1. The figure shows the mean distances traveled by drivers each day. You randomly select 50 drivers ages 16 to 19. What is the probability that the mean distance traveled each day is between 19.4 and 22.5 miles? Assume σ = 6.5 miles.



From the Central Limit Theorem (sample size is greater than 30), the sampling distribution of sample means is approximately normal with

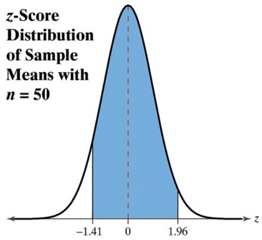
$\mu_x = \mu = 20.7$ miles and

$\sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{6.5}{\sqrt{50}} \approx 0.9$ miles



- The probability that the mean distance driven each day by the sample of 50 people is between 19.4 and 22.5 miles is

$$P(19.4 < \bar{x} < 22.5) = P(-1.41 < z < 1.96)$$
$$= P(z < 1.96) - P(z < -1.41)$$
$$= 0.9750 - 0.0793$$
$$= 0.8957$$



Of all samples of 50 drivers ages 16 to 19, about 90% will drive a mean distance each day between 19.4 and 22.5 miles,

25

Ex2. The mean room and board expense per year at four-year colleges is \$10,453. You randomly select 9 four-year colleges. What is the probability that the mean room and board is less than \$10,750? Assume that the room and board expenses are normally distributed with a standard deviation of \$1650.

- Because the population is normally distributed, you can use the Central Limit Theorem to conclude that the distribution of sample means is normally distributed, with a mean and a standard deviation of

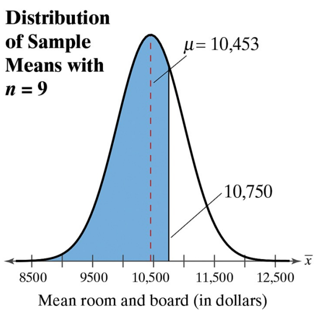
$$\mu_x = \mu = \$10{,}453 \text{ and } \sigma_x = \frac{\sigma}{\sqrt{n}} = \frac{\$1650}{\sqrt{9}} = = \$550.$$

- The area to the left of \$10,750 is shaded. The z-score that corresponds to \$10,750 is

$$z = \frac{10{,}750 - 10{,}453}{1650/\sqrt{9}} = \frac{297}{550} = 0.54.$$

- So, the probability that the mean room and board expense is less than \$10,750 is

$P(x < 10{,}750) = P(z < 0.54) = 0.7054.$



26

Ex3. A survey showed that children between the ages 2 and 5 watch television 25 hours per week. Assume the variable is normally distributed with the standard deviation of 3 hours. If 20 children between the ages of 2 and 5 are randomly selected, find the probability that the mean of the number of hours they watch television will be greater than 26.3.

Ex4. The average age of a vehicle registered in the United States is 8 years, or 96 months. Assume the standard deviation is 16 months. If a random sample of 36 vehicles is selected, find the probability that the mean of their age is in between 90 and 100 months.

27

Normal Approximation to a Binomial

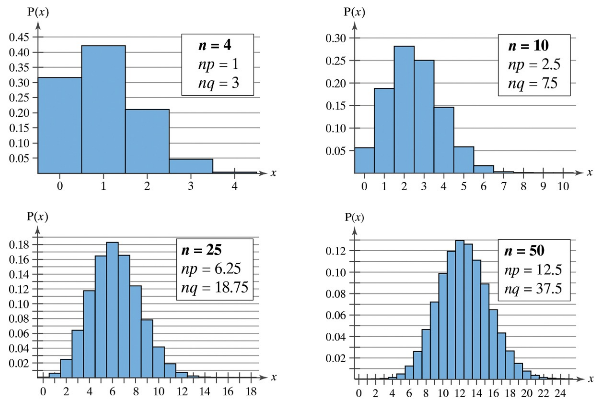
Normal Approximation to a Binomial Distribution

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

28

Normal Approximation to a Binomial

- Binomial distribution: p = 0.25, q =1 – 0.25, and n = 4, n = 10, n = 25 and n = 50.



- As n increases the histogram approaches a normal curve.

29

Ex1. Two binomial experiments are listed. Decide whether you can use the normal distribution to approximate x, the number of people who reply yes. If you can, find the mean and standard deviation. If you cannot, explain why.

1. Sixty-two percent of adults in the United States have an HDTV in their home. You randomly select 45 adults in the United States and ask them if they have an HDTV in their home.
2. Twelve percent of adults in the United States who do not have an HDTV in their home are planning to purchase one in the next two years. You randomly select 30 adults in the United States who do not have an HDTV and ask them if they are planning to purchase one in the next two years.

1. In this binomial experiment, $n = 45$, $p = 0.62$, and $q = 0.38$. So,

$$np = 45(0.62) = 27.9$$

and

$$nq = 45(0.38) = 17.1.$$

Because np and nq are greater than 5, you can use a normal distribution with

$$\mu = np = 27.9$$

and

$$\sigma = \sqrt{npq} = \sqrt{45 \cdot 0.62 \cdot 0.38} \approx 3.26$$

to approximate the distribution of x.

2. In this binomial experiment, $n = 30$, $p = 0.12$, and $q = 0.88$. So,

$$np = 30(0.12) = 3.6$$

and

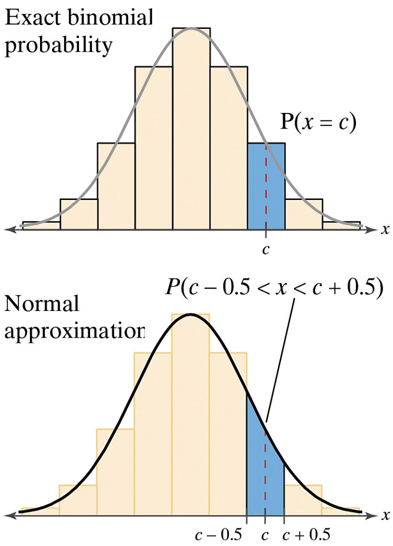
$$nq = 30(0.88) = 26.4.$$

Because $np < 5$, you cannot use a normal distribution to approximate the distribution of x.

30

Correction for Continuity

- A binomial distribution is discrete and can be represented by a probability histogram.
- To calculate *exact* binomial probabilities, the binomial formula is used for each value of *x* and the results are added.
- Geometrically this corresponds to adding the areas of bars in the probability histogram.
- When you use a *continuous* normal distribution to approximate a binomial probability, you need to move 0.5 unit to the left and right of the midpoint to include all possible *x*-values in the interval (**continuity correction**).



31

Ex1. Use a continuity correction to convert each binomial probability to a normal distribution probability.

1. The probability of getting between 270 and 310 successes, inclusive.
2. The probability of getting at least 158 successes.
3. The probability of getting fewer than 63 successes.

Solution1:

- The discrete midpoint values are 270, 271, …, 310.
- The corresponding interval for the continuous normal distribution is $269.5 < x < 310.5$.
- The normal distribution probability is $P(269.5 < x < 310.5)$.

Solution:

- The discrete midpoint values are 158, 159, 160, ….
- The corresponding interval for the continuous normal distribution is $x > 157.5$.
- The normal distribution probability is $P(x > 157.5)$.

Solution:

- The discrete midpoint values are …,60, 61, 62.
- The corresponding interval for the continuous normal distribution is $x < 62.5$.
- The normal distribution probability is $P(x < 62.5)$.

32

Correction for Continuity

Table1. Summary of the Normal Approximation to the Binomial Distribution

Binomial	**Normal**
when finding:	*use:*
P(X = a)	P(a – 0.5 < X < a + 0.5)
P(X ≥ a)	P(X > a – 0.5)
P(X > a)	P(X > a + 0.5)
P(X ≤ a)	P(X < a + 0.5)
P(X < a)	P(X < a – 0.5)

33

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 if 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 or subtract 0.5 from endpoints.
5. Find the corresponding z-score(s).	$z = \dfrac{x - \mu}{\sigma}$
6. Find the probability.	Use the Standard Normal Table.

34

Ex1. Sixty-two percent of adults in the United States have an HDTV in their home. You randomly select 45 adults in the United States and ask them if they have an HDTV in their home. What is the probability that fewer than 20 of them respond yes?

Solution

From Example 1, you know that you can use a normal distribution with $\mu = 27.9$ and $\sigma \approx 3.26$ to approximate the binomial distribution. Remember to apply the continuity correction for the value of x. In the binomial distribution, the possible midpoint values for "fewer than 20" are

$\ldots 17, 18, 19.$

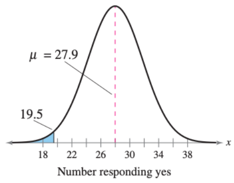
To use a normal distribution, add 0.5 to the right-hand boundary 19 to get $x = 19.5$. The graph at the left shows a normal curve with $\mu = 27.9$ and $\sigma \approx 3.26$ and a shaded area to the left of 19.5. The z-score that corresponds to $x = 19.5$ is

$$z = \frac{19.5 - 27.9}{3.26}$$
$$\approx -2.58.$$

Using the Standard Normal Table,

$$P(z < -2.58) = 0.0049.$$

Interpretation The probability that fewer than 20 people respond yes is approximately 0.0049, or about 0.49%.



35

Ex2. A study on aggressive driving found that 47% of drivers say they have yelled at another driver. You randomly select 200 drivers in the United States and ask them whether they have yelled at another driver. What is the probability that at least 100 drivers will say yes, they have yelled at another driver?

Ex3. 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?

36

Ex4. 58% of adults say that they never wear a helmet when riding a bicycle. You randomly select 200 adults in the United States and ask them if they wear a helmet when riding a bicycle. What is the probability that at most 95 adults will say they never wear a helmet when riding a bicycle?

Ex5. A magazine reported that 6% of the American drivers read the newspaper while driving. If 300 drivers are selected at random, find the probability that exactly 25 say they read the newspaper while driving.