

MS 204 In-class Problems

November 10, 2025

Table of Contents

Ch 1

Ch 1.1

Ch 1.2

Ch 1.4

Ch 2.1

Ch 2.2

Ch 2.3

Ch 3

Ch 3.1

Ch 3.2

Ch 3.3

Ch 4

Ch 4.1

Ch 4.4

Ch 5

Ch 5.1

Ch 5.2

Ch 6

Ch 6.1

Ch 6.2

Ch 6.3

Ch 6.4

Ch 7

Ch 7.1

Ch 7.2

Ch 8

Ch 8.1

Ch 8.2

Ch 8.3

Ch 10

Ch 10.1

Ch 10.2

Ch 10.3

Ch 12

Ch 12.1

Ch 12.2

Chapter 1 Section 1

DETERMINE WHETHER THE STATEMENT DESCRIBES A
POPULATION OR A SAMPLE.

The price of homes of all the employees at a software company.

- ▶ Population
- ▶ Sample

DETERMINE WHETHER THE STATEMENT DESCRIBES A
POPULATION OR A SAMPLE.

The heights of 5 out of the 32 eggplant plants at Mr. Lonardo's greenhouse.

- ▶ Population
- ▶ Sample

IDENTIFY THE **population** BEING STUDIED.

The number of times 10 out of 20 students on your floor order pizza in a week.

- ▶ The 20 students on your floor.
- ▶ All students who ordered pizza in a week.
- ▶ The 10 students on your floor.

DETERMINE WHETHER THE STATEMENT DESCRIBES A
DESCRIPTIVE OR INFERENTIAL STATISTIC.

A recent poll of 1443 luxury car owners in West Virginia showed that the average price of a luxury car in the U.S. is \$48,900.

- ▶ Descriptive Statistic
- ▶ Inferential Statistic

DETERMINE WHETHER THE STATEMENT DESCRIBES A
DESCRIPTIVE OR INFERENTIAL STATISTIC.

The average price of a car at the new car dealership in town is \$28,400.

- ▶ Descriptive Statistic
- ▶ Inferential Statistic

DETERMINE IF THE NUMERICAL VALUE DESCRIBES A POPULATION PARAMETER OR A SAMPLE STATISTIC.

A recent poll of 2935 corporate executives showed that the average price of their cars is \$27,100.

- ▶ Population Parameter
- ▶ Sample Statistic

DETERMINE IF THE NUMERICAL VALUE DESCRIBES A POPULATION PARAMETER OR A SAMPLE STATISTIC.

The average price of a house in the new subdivision is \$339,000.

- ▶ Population Parameter
- ▶ Sample Statistic

IDENTIFY THE SAMPLE CHOSEN FOR THE STUDY.

The number of times 4 out of 37 students on your floor order take-out in a week.

- ▶ The 4 students on your floor.
- ▶ All students who ordered take-out in a week.
- ▶ The 37 students on your floor.

Chapter 1 Section 2

Types of cars people own are an example of which type of data?

- ▶ Qualitative
- ▶ Quantitative
- ▶ Inferential
- ▶ Statistic

Football jersey numbers are an example of which type of data?

- ▶ Qualitative
- ▶ Quantitative
- ▶ Inferential
- ▶ Statistic

Goals scored during a soccer game are an example of which type of data?

- ▶ Qualitative
- ▶ Quantitative
- ▶ Inferential
- ▶ Statistic

INDICATE THE LEVEL OF MEASUREMENT FOR THE DATA SET DESCRIBED.

Monthly amounts of rain in Seattle over 10 years

- ▶ Interval
- ▶ Ratio
- ▶ Ordinal
- ▶ Nominal

INDICATE THE LEVEL OF MEASUREMENT FOR THE DATA SET DESCRIBED.

Categories of hurricanes that have hit the Atlantic coast

- ▶ Interval
- ▶ Ratio
- ▶ Ordinal
- ▶ Nominal

CLASSIFY DATA AS DISCRETE OR CONTINUOUS

Lengths of time it takes for new light bulbs to burn out are an example of which type of data?

- ▶ Discrete
- ▶ Continuous
- ▶ Neither

CLASSIFY DATA AS DISCRETE OR CONTINUOUS

Types of movies people go to see are an example of which type of data?

- ▶ Discrete
- ▶ Continuous
- ▶ Neither

CLASSIFY DATA AS DISCRETE OR CONTINUOUS

The numbers of each color of jelly beans in a jar (assuming they are all whole) are an example of which type of data?

- ▶ Discrete
- ▶ Continuous
- ▶ Neither

Chapter 1 Section 4



What is response bias and how can you avoid it?*

*This webpage seems to explain each type well, but I didn't read every sentence. I mainly put the link here for attributive purposes.

Chapter 2 Section 1

The following data describes grades of students in biology. Complete the frequency table for this data.

88.2, 94.9, 86.6, 80.0, 83.5, 96.1, 87.3, 89.7, 83.5, 93.1, 89.5, 88.6, 95.2, 96.7, 86.8, 96.8, 95.1, 89.0, 88.2, 94.9, 86.6, 80.0, 83.5, 96.1, 87.3, 89.7, 83.5, 93.1, 89.5, 88.6, 95.2, 96.7, 86.8, 96.8, 95.1, 89.0

Determine the frequency of each class in the table shown.

Grades of Students in Biology	
Class	Frequency
77.0–80.9	
81.0–84.9	
85.0–88.9	
89.0–92.9	
93.0–96.9	

Consider the following frequency table representing the distribution of hours students spend on homework in a week.

Hours Students Spend on Homework in a Week	
Class	Frequency
19–28	3
29–38	11
39–48	15
49–58	6
59–68	9

Determine the class width of each class.

Consider the following frequency table representing the distribution of hours students spend on homework in a week.

Price of a Newspaper (in Dollars)	
Class	Frequency
0.34–0.42	11
0.43–0.51	12
0.52–0.60	14
0.61–0.69	10
0.70–0.78	10

Determine the class width of each class.

Consider the following frequency table representing the distribution of cost of a paperback book (in dollars).

Cost of a Paperback Book (in Dollars)	
Class	Frequency
5.7–6.1	6
6.2–6.6	13
6.7–7.1	12
7.2–7.6	14
7.7–8.1	1

1. Determine the relative frequency for the second class as a simplified fraction.
2. Determine the relative frequency for the fourth class as a simplified fraction.






Consider the following frequency table representing the distribution of hourly wages for first jobs of a certain population.

Hourly Wage at First Job	
Class	Frequency
6.1–7.1	2
7.2–8.2	9
8.3–9.3	9
9.4–10.4	13
10.5–11.5	9

1. Determine the cumulative frequency for the fifth class.
2. Determine the cumulative frequency for the third class.

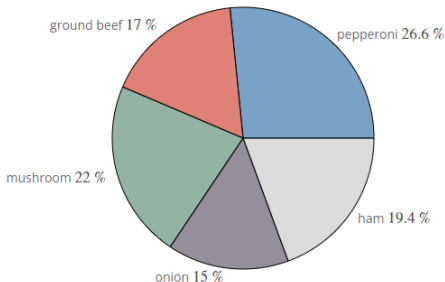
Chapter 2 Section 2

Qualitative Data

Type of Graph	Description
Pie Chart 	A pie chart shows how large each category is in relation to the whole; that is, it uses the relative frequencies from the frequency distribution to divide the "pie" into different-sized wedges. It can only be used to display qualitative data.
Bar Graph 	In a bar graph, bars are used to represent the amount of data in each category; one axis displays the categories of qualitative data and the other axis displays the frequencies.
Pareto Chart 	A Pareto chart is a bar graph with the bars in descending order of frequency. Pareto charts are typically used with nominal data.
Side-by-Side Bar Graph 	A side-by-side bar graph is a bar graph that compares the same categories for different groups.
Stacked Bar Graph 	A stacked bar graph is a bar graph that compares the same categories for different groups and shows category totals.

The Pizza Pie 'N Go sells about 2260 one-topping pizzas each month. The circle graph displays the most requested one-topping pizzas, by percentage, for one month.

Most Popular One-Topping Pizzas

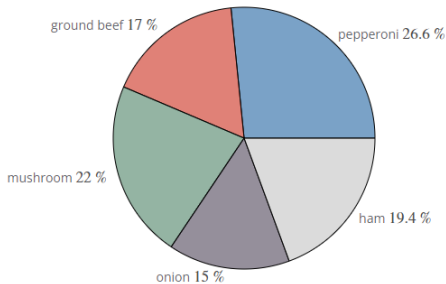


Step 1 of 5: Find the number of pepperoni pizzas sold each month. Round your answer to the nearest integer.

Step 2 of 5: Find the number of ground beef pizzas sold each month. Round your answer to the nearest integer.

The Pizza Pie 'N Go sells about 2260 one-topping pizzas each month.

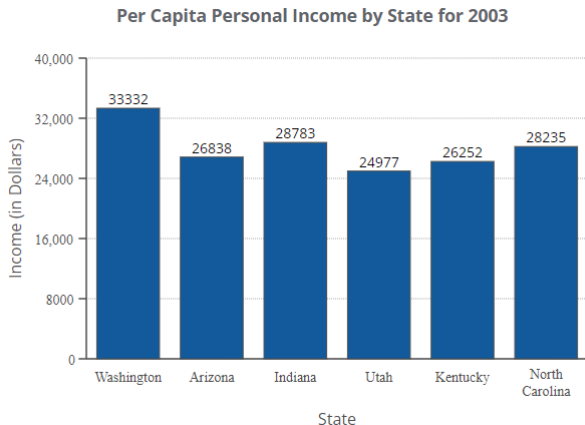
Most Popular One-Topping Pizzas



Step 3 of 5: Find the number of mushroom pizzas sold each month. Round your answer to the nearest integer.

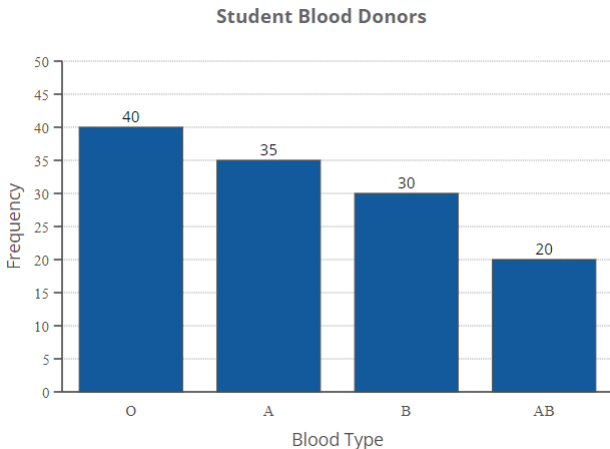
Step 4 of 5: Find the number of onion pizzas sold each month. Round your answer to the nearest integer.

Step 5 of 5: Find the number of ham pizzas sold each month. Round your answer to the nearest integer.



- Step 1 of 2:** Find the lowest per capita personal income for the six states shown.
- Step 2 of 2:** Find the highest per capita personal income for the six states shown.

Consider the Pareto chart, which shows the number of student blood donors by their type for one day of a campus blood drive. How many students donated blood on that day?



Quantitative Data

Type of Graph

Definition

Histogram



A histogram is a bar graph of a frequency distribution of quantitative data; the horizontal axis is a number line.

Stem-and-Leaf Plot

Stem	Leaves
32	0
33	7 7 7 8
34	0 0 0 0

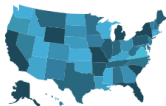
A stem-and-leaf plot retains the original data; the leaves are the last significant digit in each data value and the stems are the remaining digits.

Dot Plot



A dot plot retains the original data by plotting a dot above each data value on a number line.

Heat Map



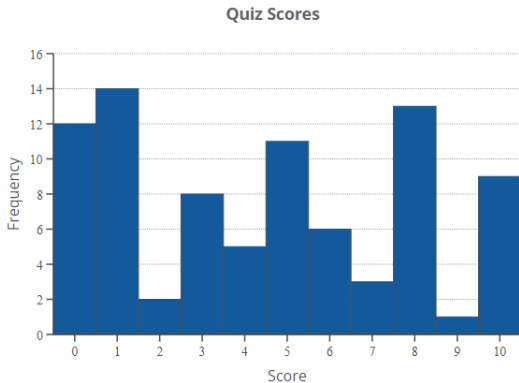
A heat map depicts relative values of the data using shades of color.

Line Graph



A line graph uses straight lines to connect points plotted at the value of each measurement above the time it was taken.

The following histogram represents the distribution of scores on a ten point quiz.



Step 1 of 3: Which score has the highest frequency?

Step 2 of 3: What is the frequency corresponding to a score of 6?

Step 3 of 3: What is the total number of people who made a score between 0 and 2 inclusive?

The following stem-and-leaf plot represents the distribution of weights for a group of people.

Stem	Leaves					
8	0	3	6	6		
9	1	3				
10	1	2	4	4	5	9
11	1	2	5	7	8	
12	1	2	3	3	6	8
13	1	2	2	7	7	
14	8	8				
15	2	4	5	8	9	
16	4	5	5	6	9	

Key: 8|0 = 80 pounds

Step 1 of 3: What is the weight of the lightest person in the group?

Step 2 of 3: How many people weigh in the range from 110 to 140 inclusive?

Step 3 of 3: What is the weight of the heaviest person in the range 80 to 89 inclusive?

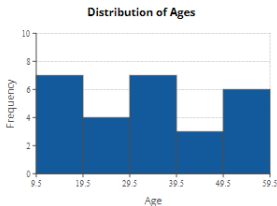
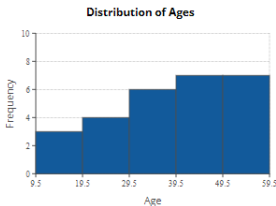
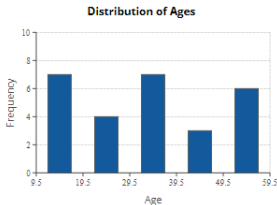
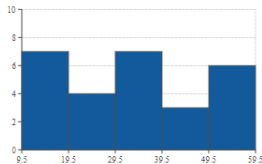
The following data represent the test scores for 18 students in a class on their most recent test. Use the given data to determine the stems for this stem-and-leaf plot.

87 84 69 85 73 58
 65 53 63 66 67 82
 66 82 79 89 52 60

Test Scores by Student						
Stem	Leaves					
_____	2	3	8			
_____	0	3	5	6	6	7 9
_____	2	9				
_____	2	2	4	5	7	9

The following data represents the distribution of ages of a group of people. Determine the graph that correctly represents the data.

Age	10-19	20-29	30-39	40-49	50-59
Frequency	7	4	7	3	6



Chapter 2 Section 3

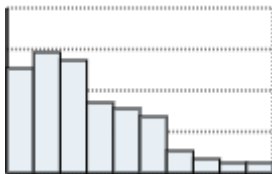
Select the graph that best illustrates the following distribution shape:

Uniform

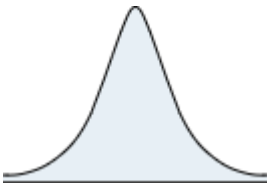
a)



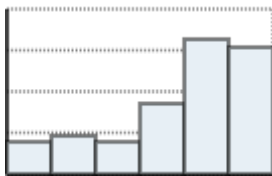
b)



c)

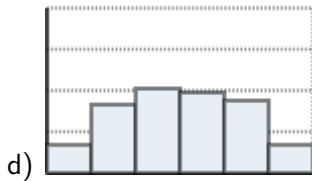
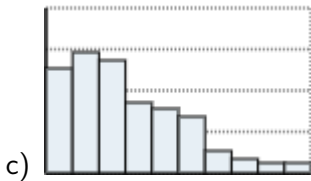
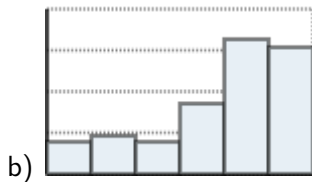
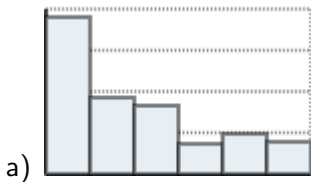


d)

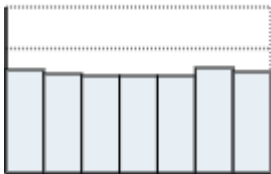


Select the graph that best illustrates the following distribution shape:

Symmetrical

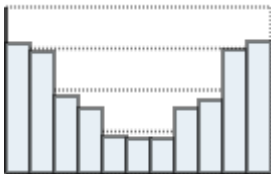


For the set of data displayed below, describe the most likely shape of its distribution.



-
- ▶ Skewed to the right
 - ▶ Symmetrical, but not uniform
 - ▶ Skewed to the left
 - ▶ Uniform

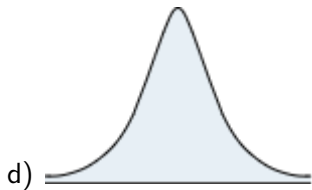
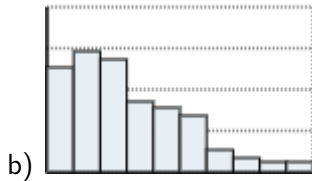
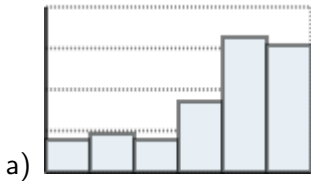
For the set of data displayed below, describe the most likely shape of its distribution.



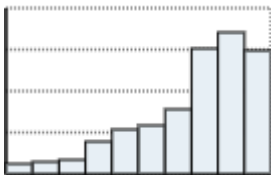
-
- ▶ Uniform
 - ▶ Skewed to the left
 - ▶ Symmetrical, but not uniform
 - ▶ Skewed to the right

Select the graph that best illustrates the following distribution shape:

Skewed to the right



For the set of data displayed below, describe the most likely shape of its distribution.



-
- ▶ Symmetrical, but not uniform
 - ▶ Uniform
 - ▶ Skewed to the left
 - ▶ Skewed to the right

Chapter 3 Section 1

For the data set shown in the table:

Data	
i	x_i
1	3
2	7
3	4

Find $\sum x_i$ written in two ways: as an unevaluated sum (it will have multiple terms) and then as an evaluated sum (a single term):

Unevaluated _____

Evaluated _____

Find the mean of:

1. 1, 2, 3

2. 1, 2, 5, 8, 9

3. 1, 1, 5, 9, 9

4. 1, 1, 2, 8, 9, 9

5. 1, 2, 5, 9, 13

6. 1, 2, 4, 8, 100

Find the mean of

50000, 30000, 45000, 33000, 47000, 51000, 6744000.

Consider the following data.

$14, -10, 7, 13, 3, -3$

Step 1 of 3: Determine the mean of the given data.

Step 2 of 3: Determine the median of the given data

Step 3 of 3: Determine if the data set is unimodal, bimodal, multimodal, or has no mode. Identify the mode(s), if any exist.

- ▶ No Mode
- ▶ Unimodal
- ▶ Bimodal
- ▶ Multimodal

Consider the following data.

$-9, 11, 7, 11, 7, -9$

Step 2 of 3: Determine the median of the given data.

Step 3 of 3: Determine if the data set is unimodal, bimodal, multimodal, or has no mode. Identify the mode(s), if any exist.

- ▶ No Mode
- ▶ Unimodal
- ▶ Bimodal
- ▶ Multimodal

Find the mode of

3, 2, 3, 1, 5, 1

For the following type of data set, would you be more interested in looking at the mean, median, or mode? State your reasoning.

The price for homes with similar floor plans in a new neighborhood

Correct measure of center:

a) mean

b) median

c) mode

Justification

- ▶ The prices for homes are quantitative data with outliers.
- ▶ The prices for homes are qualitative data.
- ▶ The prices for homes are quantitative data with no outliers.

A company has given you the task to research the storage cost for similarly sized spaces in downtown Houston. Would you be more interested in looking at the mean, median, or mode?

- ▶ mean
- ▶ median
- ▶ mode

A company has given you the task to research the color of a car preferred by the average male. Would you be more interested in looking at the mean, median, or mode?

- ▶ mean
- ▶ median
- ▶ mode

A company has given you the task to research the cost of cars at all the dealerships in town. Would you be more interested in looking at the mean, median, or mode?

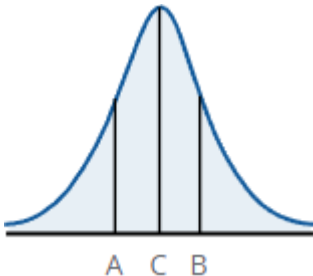
- ▶ mean
- ▶ median
- ▶ mode

Calculate the GPA of a student with the following grades:

A (17 hours), B (17 hours), F (10 hours).

Note that an A is equivalent to 4.0, a B is equivalent to a 3.0, a C is equivalent to a 2.0, a D is equivalent to a 1.0, and an F is equivalent to a 0. Round your answer to two decimal places.

For the graph shown, determine which letter represents the mean, the median, and the mode. Letters may be used more than once.

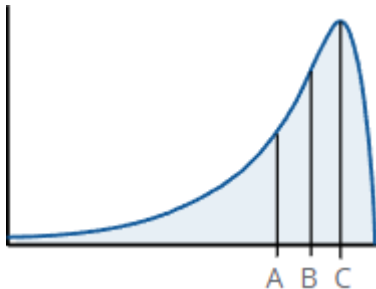


Mean =

Median =

Mode =

For the graph shown, determine which letter represents the mean, the median, and the mode. Letters may be used more than once.



Mean =

Median =

Mode =

Chapter 3 Section 2

Calculate the range, population variance, and population standard deviation for the following data set. If necessary, round to one more decimal place than the largest number of decimal places given in the data.

14, 18, 16

Range = _____

Population variance = _____

Population standard deviation = _____

Calculate the range, population variance, and population standard deviation for the following data set. If necessary, round to one more decimal place than the largest number of decimal places given in the data.

9, 9, 9, 9, 9, 9, 9, 9, 9, 9

Range = _____

Population variance = _____

Population standard deviation = _____

Use Excel:

Calculate the range, population variance, and population standard deviation for the following data set. If necessary, round to one more decimal place than the largest number of decimal places given in the data.

14, 18, 16, 5, 13, 9, 18, 16, 11, 17

Range = _____

Population variance = _____

Population standard deviation = _____

Donna is looking into investing a portion of her recent bonus into the stock market. While researching different companies, she discovers the following standard deviations of one year of daily stock closing prices.

Handy Prosthetics: Standard deviation of stock prices = \$1.12

El Lobo Malo Incorporated: Standard deviation of stock prices = \$9.63

Based on the data and assuming these trends continue, which company would give Donna a stable long-term investment?

- ▶ Handy Prosthetics; the smaller standard deviation indicates that Handy Prosthetics has a greater mean closing price than El Lobo Malo Incorporated.
- ▶ Handy Prosthetics; the smaller standard deviation indicates that Handy Prosthetics has a less variability in its closing prices than El Lobo Malo Incorporated.
- ▶ El Lobo Malo Incorporated the larger standard deviation indicates that El Lobo Malo Incorporated has a less variability in its closing prices than Handy Prosthetics.
- ▶ El Lobo Malo Incorporated the larger standard deviation indicates that El Lobo Malo Incorporated has a greater mean closing price than Handy Prosthetics.

Suppose that IQ scores have a bell-shaped distribution with a mean of 97 and a standard deviation of 17. Using the empirical rule, what percentage of IQ scores are between 46 and 148?

Suppose that grade point averages of undergraduate students at one university have a bell-shaped distribution with a mean of 2.52 and a standard deviation of 0.42. Using the empirical rule, what percentage of the students have grade point averages that are no more than 1.26? Please do not round your answer.

Chapter 3 Section 3

Ch 3.3

5 number summary

Calculate the 5 number summary for:

19
15
20
10
13
18
11
13
7
13
10
2
12
13
17
15
7
13
13
4

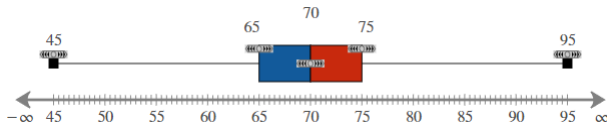
Copy Data

Goto problem in homework:

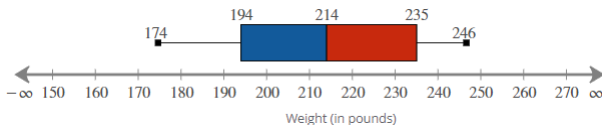
Construct a box plot from the given data. Use the approximation method.

Scores on a Statistics Test: 86, 79, 70, 91, 56, 48, 45, 81, 50, 89

Draw the box plot by selecting each of the five movable parts to the appropriate position.



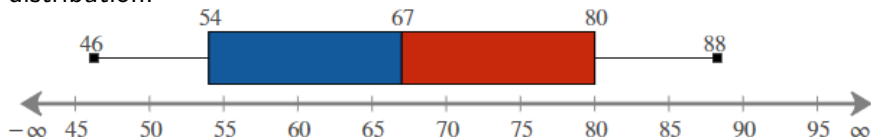
A high school has 52 players on the football team. The summary of the players' weights is given in the box plot. Approximately, what is the percentage of players weighing less than or equal to 194 pounds?



From:

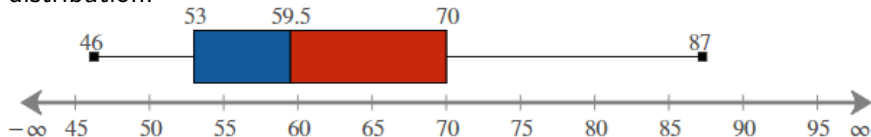
- ▶ 194-235?
- ▶ 194-246?
- ▶ 174-246?

Given the following box plot, choose the best description of the distribution.



- ▶ The distribution of the data is skewed left.
- ▶ The distribution of the data is skewed right.
- ▶ The distribution of the data is symmetric.

Given the following box plot, choose the best description of the distribution.

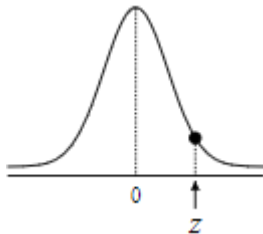


- ▶ The distribution of the data is skewed left.
- ▶ The distribution of the data is skewed right.
- ▶ The distribution of the data is symmetric.

Calculate the standard score of the given x value, $x = 59.6$, where $\mu = 65.5$, $\sigma = 3.7$. Round your answer to two decimal places.

Calculate the standard score of the given x value, $x = 22.8$, where $\bar{x} = 20.9$, $s = 3.6$. Round your answer to two decimal places.

Given the following graph, where the mean is marked, which value best represents the z-score shown?



- ▶ $z = 1.25$
- ▶ $z = 0$
- ▶ $z = -2.46$

Chapter 4 Section 1

WRITE OUT THE SAMPLE SPACE FOR THE GIVEN EXPERIMENT.
USE THE LETTER R TO INDICATE RED, G TO INDICATE GREEN,
AND B TO INDICATE BLUE.

A die shows 3 different colors on it. Give the sample space for the next
2 rolls.

Experimental probability example

	A	B	C	D	E	F	G
1	3		3		Pips	Freq	Rel Freq
2	2		3		1	0	0
3	6		3		2	1	0.1
4	4		3		3	7	0.7
5	3		4		4	1	0.1
6	2		6		5	0	0
7	5		3		6	1	0.1
8	4		2				
9	5		3		Sum	10	
10	1		3				

Code:

	A	B	C	D	E	F	G
1	=RANDBETWEEN(1,6)		3		Pips	Freq	Rel Freq
2	=RANDBETWEEN(1,6)		3		1	=COUNTIF(C:C, E2)	=F2/\$F\$9
3	=RANDBETWEEN(1,6)		3		2	=COUNTIF(C:C, E3)	=F3/\$F\$9
4	=RANDBETWEEN(1,6)		3		3	=COUNTIF(C:C, E4)	=F4/\$F\$9
5	=RANDBETWEEN(1,6)		4		4	=COUNTIF(C:C, E5)	=F5/\$F\$9
6	=RANDBETWEEN(1,6)		6		5	=COUNTIF(C:C, E6)	=F6/\$F\$9
7	=RANDBETWEEN(1,6)		3		6	=COUNTIF(C:C, E7)	=F7/\$F\$9
8	=RANDBETWEEN(1,6)		2				
9	=RANDBETWEEN(1,6)		3		Sum	=SUM(F2:F7)	
10	=RANDBETWEEN(1,6)		3				

There are 219 identical plastic chips numbered 1 through 219 in a box. What is the probability of reaching into the box and randomly drawing the chip numbered 170? Express your answer as a simplified fraction or a decimal rounded to four decimal places.

There are 756 identical plastic chips numbered 1 through 756 in a box. What is the probability of reaching into the box and randomly drawing the chip number that is smaller than 570? Express your answer as a simplified fraction or a decimal rounded to four decimal places.

There are 569 identical plastic chips numbered 1 through 569 in a box. What is the probability of reaching into the box and randomly drawing the chip number that is greater than 220? Express your answer as a simplified fraction or a decimal rounded to four decimal places.

You decide to record the hair colors of people leaving a lecture at your school. What is the probability that the next person who leaves the lecture will have gray hair? Express your answer as a simplified fraction or a decimal rounded to four decimal places.

Blonde	Red	Brown	Black	Gray
20	45	21	44	33

What is the probability that a randomly selected person will have a birthday in November? Assume that this person was not born in a leap year. Express your answer as a simplified fraction or a decimal rounded to four decimal places.

A coin is tossed 6 times.

What is the probability of getting all heads? Express your answer as a simplified fraction or a decimal rounded to four decimal places.

A standard six-sided die is rolled.

What is the probability of rolling a number less than or equal to 5?

Express your answer as a simplified fraction or a decimal rounded to four decimal places.

What is the probability of rolling a sum of 9 on a standard pair of six-sided dice? Express your answer as a fraction or a decimal number rounded to three decimal places, if necessary.

$(1, 1)$	$(1, 2)$	$(1, 3)$	$(1, 4)$	$(1, 5)$	$(1, 6)$
$(2, 1)$	$(2, 2)$	$(2, 3)$	$(2, 4)$	$(2, 5)$	$(2, 6)$
$(3, 1)$	$(3, 2)$	$(3, 3)$	$(3, 4)$	$(3, 5)$	$(3, 6)$
$(4, 1)$	$(4, 2)$	$(4, 3)$	$(4, 4)$	$(4, 5)$	$(4, 6)$
$(5, 1)$	$(5, 2)$	$(5, 3)$	$(5, 4)$	$(5, 5)$	$(5, 6)$
$(6, 1)$	$(6, 2)$	$(6, 3)$	$(6, 4)$	$(6, 5)$	$(6, 6)$

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

	$+$	$-$	
C Yes	1688	187	1875
C' No	32381	65744	98125
	34069	65931	100000

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

	+	-	
C Yes	1688	187	1875
C' No	32381	65744	98125
	34069	65931	100000

So

$$P(C) = \frac{1875}{100000} = \frac{3}{160} \approx .01875$$

$$P(+|C) = \frac{1688}{1875} = .9, \quad P(+|C') = \frac{32381}{98125} = .33$$

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

	$+$	$-$	
C Yes	.9	.1	1
C' No	.33	.67	1

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

	+	-	
C Yes	.9	.1	1
C' No	.33	.67	1

, $P(C) = \frac{1875}{100000} = \frac{3}{160}$

But what we really want to know is:

$$\begin{aligned}
 P(C|+) &= \frac{P(C)P(+|C)}{P(C)P(+|C) + P(C')P(+|C')} \\
 &= \frac{\frac{3}{160}(.9)}{\frac{3}{160}(.9) + \left(1 - \frac{3}{160}\right)(.33)} \\
 &= .0495
 \end{aligned}$$

So the probability is 4.95% that you have cancer given that the test came back positive.

Let C represent the event that you have cancer, C' that you do not have cancer, $+$ that the prostate cancer test came back positive, and $-$ that the prostate cancer test came back negative.

	$+$	$-$	
C Yes	.9	.1	1
C' No	.33	.67	1

So when $P(C) \approx .01875$

$$P(C|+) = .0495$$

and the probability is 4.95% that you have cancer given that the test came back positive.

Now when a male is 60 years old or older, then $P(C) = .4$. This results in

$$P(C|+) = .645$$

or as a percentage 64.5%.

Monty Hall Problem



Birthday Paradox

Birthday Paradox - How many people would you need in a room so that there is a 50% chance that two of them have the same birthday?

Related question - How many people would you need in a room so that there is a 50% chance that someone has your birthday?

Chapter 4 Section 4

Evaluate the following expression.

$$6!$$

Evaluate the following expression.

$${}_7P_3$$

Evaluate the following expression.

$${}_9C_7$$

Evaluate the following expression.

$$\frac{13!}{5!(13 - 5)!}$$

Pascal's triangle I

1

Pascal's triangle II

$$\begin{array}{ccc} & 1 & \\ 1 & & 1 \end{array}$$

Pascal's triangle III

$$\begin{array}{cccc} & & 1 & & \\ & 1 & & 1 & \\ 1 & & 2 & & 1 \end{array}$$

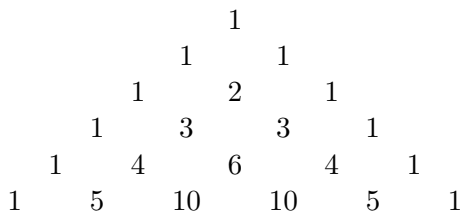
Pascal's triangle IV

			1			
		1		1		
	1		2		1	
1		3		3		1

Pascal's triangle V

			1					
		1		1				
	1		2		1			
1		3		3		1		
	1	4		6		4		1

Pascal's triangle VI



A Pascal's triangle with 6 rows. The numbers are arranged in a triangular shape, with each row containing one more number than the row above it. The numbers are: Row 1: 1; Row 2: 1, 1; Row 3: 1, 2, 1; Row 4: 1, 3, 3, 1; Row 5: 1, 4, 6, 4, 1; Row 6: 1, 5, 10, 10, 5, 1.

				1				
			1		1			
		1		2		1		
	1		3		3		1	
1	1	4		6		4	1	
1	5	10	10	5	1			

Pascal's triangle Combinations

n										
0					1					
1				1		1				
2			1		2		1			
3		1		3		3		1		
4		1	4		6		4		1	
5	1		5	10		10		5		1
0					${}_0C_0$					
1				${}_1C_0$		${}_1C_1$				
2			${}_2C_0$		${}_2C_1$		${}_2C_2$			
3		${}_3C_0$		${}_3C_1$		${}_3C_2$		${}_3C_3$		
4		${}_4C_0$	${}_4C_1$		${}_4C_2$	${}_4C_3$		${}_4C_4$		
5	${}_5C_0$	${}_5C_1$	${}_5C_2$	${}_5C_3$	${}_5C_4$	${}_5C_5$				

Chapter 5 Section 1

Expected value example here.

An important concept in probability is what is called a *Random Variable*.

There is nothing random nor variablely about a random variable. Let's flip a fair coin 3 times and use this to learn about random variables.

First thing is to determine what outcome are we interested in.

1. How many heads were flipped
2. How many tails were flipped
3. How many heads minus how many tails were flipped

We will look at how many heads were flipped.

Let us take a moment and look at an easier example and then we will return to the flip coin three times example.

Let us flip a fair coin once.

Sample space is $S = \{H, T\}$.

Let X *count* the number of times we have seen a head for each flip.

Outcome	Count
T	$X(T) = 0$
H	$X(H) = 1$

Back to three flips.

$$S =$$

Let's make a table of Outcomes, Count of heads, and the Random Variable of that outcome:

Outcome	# of H	RV

Why add this complexity?

Count	Outcomes	Probability
$X = 0$		$P(X = 0) =$
$X = 1$		$P(X = 1) =$
$X = 2$		$P(X = 2) =$
$X = 3$		$P(X = 3) =$

This is a probability distribution and one way of writing it is in table form:

x				
$P(X = x)$				

Definition 1 (Probability distribution, Discrete random variable)

A **probability distribution** is a function that gives the probabilities of occurrence of different possible outcomes for an experiment and must satisfy:

1. $0 \leq P(X = x) \leq 1$
2. $\sum P(X = x) = 1$

A **discrete random variable** has either a finite number of possible values or a countably infinite number.

Finite Number of Possible Values

Roll a 2D6 and let r.v. $X = R_1 + R_2$.

x	2	3	4	5	6	7	8	9	10	11	12
$P(X = x)$	$\frac{1}{36}$	$\frac{1}{18}$	$\frac{1}{12}$	$\frac{1}{9}$	$\frac{5}{36}$	$\frac{1}{6}$	$\frac{5}{36}$	$\frac{1}{9}$	$\frac{1}{12}$	$\frac{1}{18}$	$\frac{1}{36}$

Definition 2 (Poisson distribution)

From:

https://en.wikipedia.org/wiki/Poisson_distribution A **Poisson distribution** is a discrete probability distribution that expresses the probability of a given number of events occurring in a fixed interval of time if these events occur with a known constant mean rate and independently of the time since the last event.

Examples 1

- ▶ soldiers killed by horse-kicks each year in each corps in the Prussian cavalry. This example was used in a book by Ladislaus Bortkiewicz (1868–1931).
- ▶ yeast cells used when brewing Guinness beer. This example was used by William Sealy Gosset (1876–1937).
- ▶ phone calls arriving at a call center within a minute. This example was described by A.K. Erlang (1878–1929).
- ▶ goals in sports involving two competing teams.
- ▶ deaths per year in a given age group,
- ▶ jumps in a stock price in a given time interval,

Examples 2

- ▶ times a web server is accessed per minute (under an assumption of homogeneity,
- ▶ mutations in a given stretch of DNA after a certain amount of radiation,
- ▶ cells infected at a given multiplicity of infection,
- ▶ bacteria in a certain amount of liquid.
- ▶ photons arriving on a pixel circuit at a given illumination over a given time period,
- ▶ landing of V-1 flying bombs on London during World War II, investigated by R. D. Clarke in 1946.

Determine whether or not the distribution is a discrete probability distribution and select the reason why or why not.

x	-4	-3	-2
$P(X = x)$	0.55	0.39	0.06

Decide

► Yes

► No

Reason

- Since the probabilities lie inclusively between 0 and 1 and the sum of the probabilities is equal to 1.
- Since at least one of the probability values is greater than 1 or less than 0.
- Since the sum of the probabilities is not equal to 1.
- Since the sum of the probabilities is equal to 1.
- Since the probabilities lie inclusively between 0 and 1.

Consider the following data:

x	-4	-3	-2	-1	0
$P(X = x)$	0.2	0.2	0.2	0.2	0.2

Step 1 of 5: Find the expected value $E(X)$. Round your answer to one decimal place.

Step 2 of 5: Find the variance. Round your answer to one decimal place.

Step 3 of 5: Find the standard deviation. Round your answer to one decimal place.

Consider the following data:

x	-4	-3	-2	-1	0
$P(X = x)$	0.2	0.2	0.2	0.2	0.2

Step 4 of 5: Find the value of $P(X \leq -2)$. Round your answer to one decimal place.

Step 5 of 5: Find the value of $P(X < -1)$. Round your answer to one decimal place.

If you draw a card with a value of two or less from a standard deck of cards, I will pay you \$30. If not, you pay me \$6. (Aces are considered the highest card in the deck.)

Step 1 of 2: Find the expected value of the proposition. Round your answer to two decimal places. Losses must be expressed as negative values.

Step 2 of 2: If you played this game 794 times how much would you expect to win or lose? Round your answer to two decimal places. Losses must be expressed as negative values.

In the long run, which plan has the least amount of risk?

<i>Plan A</i>	
Payout	$P(\text{Payout})$
-\$50000	0.17
-\$10000	0.31
\$75000	0.52

<i>Plan B</i>	
Payout	$P(\text{Payout})$
-\$20000	0.68
\$20000	0.12
\$80000	0.2

Chapter 5 Section 2

You can skip to frame number **???add frame number later???**.
Suppose we flip a fair coin three times and we are interested in the number of heads we flip (so our random variable X counts the number of heads). Recall that the sample space is:

$$S = \{TTT, TTH, THT, HTT, HHT, HTH, THH, HHH\}$$

Let us determine the probability of each outcome using the fact that our flips are independent. Let us start with TTT (no heads) and find $P(TTT)$. By independence and the fact that its a fair coin we get

$$P(TTT) = P(T)P(T)P(T) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

What about an event with one head: TTH ?

$$P(TTH) = P(T)P(T)P(H) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

and we get the same probability for the other two permutations:

$$P(THT) = \frac{1}{8} \text{ and } P(HTT) = \frac{1}{8}.$$

And two heads: HHT ?

$$P(HHT) = P(H)P(H)P(T) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

and again the other two permutations gives the same probabilities:

$$P(HTH) = \frac{1}{8} \text{ and } P(HHT) = \frac{1}{8}.$$

And finally three heads: HHH :

$$P(HHH) = P(H)P(H)P(H) = \frac{1}{2} \cdot \frac{1}{2} \cdot \frac{1}{2} = \frac{1}{8}$$

Now let's look at $P(X = 0)$. Well $(X = 0) \equiv \{TTT\}$. So

$$P(X = 0) = P(TTT) = \frac{1}{8}$$

Now $(X = 1) \equiv \{TTH, THT, HTT\}$ and we get

$$\begin{aligned} P(X = 1) &= P(TTH, THT, HTT) \\ &= P(TTH) + P(THT) + P(HTT) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= 3 \cdot \frac{1}{8} \end{aligned}$$

If we are interested in two heads:

$$(X = 2) \equiv \{HHT, HTH, THH\}$$

$$\begin{aligned} P(X = 2) &= P(HHT, HTH, THH) \\ &= P(HHT) + P(HTH) + P(THH) \\ &= \frac{1}{8} + \frac{1}{8} + \frac{1}{8} \\ &= 3 \cdot \frac{1}{8} \end{aligned}$$

And finally three heads: $(X = 3) \equiv \{HHH\}$ gives

$$P(X = 3) = P(HHH) = \frac{1}{8}$$

What if its an unfair coin? Let's let $P(T) = .3$ and $P(H) = .7$ and run through the same steps as above. No heads:

$$P(TTT) = P(T)P(T)P(T) = 0.3 \cdot 0.3 \cdot 0.3 = (0.3)^3$$

One head:

$$P(HTT) = P(H)P(T)P(T) = 0.7 \cdot 0.3 \cdot 0.3 = 0.7(0.3)^2$$

$$P(THT) = P(T)P(H)P(T) = 0.3 \cdot 0.7 \cdot 0.3 = 0.7(0.3)^2$$

$$P(TTH) = P(T)P(T)P(H) = 0.3 \cdot 0.3 \cdot 0.7 = 0.7(0.3)^2$$

Two heads:

$$P(HHT) = P(H)P(H)P(T) = 0.7 \cdot 0.7 \cdot 0.3 = (0.7)^2 \cdot 0.3$$

$$P(HTH) = P(H)P(T)P(H) = 0.7 \cdot 0.3 \cdot 0.7 = (0.7)^2 \cdot 0.3$$

$$P(THH) = P(T)P(H)P(H) = 0.3 \cdot 0.7 \cdot 0.7 = (0.7)^2 \cdot 0.3$$

Three heads

$$P(HHH) = P(H)P(H)P(H) = 0.7 \cdot 0.7 \cdot 0.7 = (0.7)^3$$

So now calculating the probability of getting 0, 1, 2, or 3 heads is:

$$P(X = 0) = P(TTT) = (0.3)^3$$

$$\begin{aligned}P(X = 1) &= P(TTH, THT, HTT) \\&= P(TTH) + P(THT) + P(HTT) \\&= 0.7(0.3)^2 + 0.7(0.3)^2 + 0.7(0.3)^2 \\&= 3(0.7)(0.3)^2\end{aligned}$$

$$\begin{aligned}P(X = 2) &= P(HHT, HTH, THH) \\&= P(HHT) + P(HTH) + P(THH) \\&= (0.7)^2 \cdot 0.3 + (0.7)^2 \cdot 0.3 + (0.7)^2 \cdot 0.3 \\&= 3(0.7)^2(0.3)\end{aligned}$$

$$P(X = 3) = P(HHH) = (0.7)^3$$

So now we can make out probability distribution:

x	0	1	2	3
$P(X = x)$	$1(0.3)^3$	$3(0.7)(0.3)^2$	$3(0.7)^2(0.3)$	$1(0.7)^3$

Doing the same steps we can find that the probability distributions for flipping a coin twice and four times are:

$$\begin{array}{c|c|c|c} x & 0 & 1 & 2 \\ \hline P(X=x) & 1(0.7)^0(0.3)^2 & 2(0.7)^1(0.3)^1 & 1(0.7)^2(0.3)^0 \end{array}$$

$$\begin{array}{c|c|c|c|c} x & 0 & 1 & 2 & 3 \\ \hline P(X=x) & 1(0.7)^0(0.3)^3 & 3(0.7)^1(0.3)^2 & 3(0.7)^2(0.3)^1 & 1(0.7)^3(0.3)^0 \end{array}$$

$$\begin{array}{c|c|c|c|c|c} x & 0 & 1 & 2 & 3 & 4 \\ \hline P(X=x) & 1(0.7)^0(0.3)^4 & 4(0.7)^1(0.3)^3 & 6(0.7)^2(0.3)^2 & 4(0.7)^3(0.3)^1 & 1(0.7)^4(0.3)^0 \end{array}$$

I've add some information that we normally don't write to help recognize the patterns.

What patterns are we seeing?

First recalling that $P(H) = 0.7$ and $P(T) = 0.3$ we see that we have

Number of heads: x	0	1	2
Heads probability	$(0.7)^0$	$(0.7)^1$	$(0.7)^2$
Tails probability	$(0.3)^2$	$(0.3)^1$	$(0.3)^0$

Number of heads: x	0	1	2	3
Heads probability	$(0.7)^0$	$(0.7)^1$	$(0.7)^2$	$(0.7)^3$
Tails probability	$(0.3)^3$	$(0.3)^2$	$(0.3)^1$	$(0.3)^0$

Number of heads: x	0	1	2	3	4
Heads probability	$(0.7)^0$	$(0.7)^1$	$(0.7)^2$	$(0.7)^3$	$(0.7)^4$
Tails probability	$(0.3)^4$	$(0.3)^3$	$(0.3)^2$	$(0.3)^1$	$(0.3)^0$

So in general if we flip a coin n times and x is the number of heads we have the following pattern:

Number of heads: x	0	1	2	3	4	...	n
Heads probability	$(0.7)^0$	$(0.7)^1$	$(0.7)^2$	$(0.7)^3$	$(0.7)^4$...	$(0.7)^n$
Tails probability	$(0.3)^{n-0}$	$(0.3)^{n-1}$	$(0.3)^{n-2}$	$(0.3)^{n-3}$	$(0.3)^{n-4}$...	$(0.3)^{n-n} = (0.3)^0 = 1$

So in general if we flip a coin n times and x is the number of heads we have the following pattern:

Num heads: x	0	1	2	3	4	...	n
Heads prob	$(0.7)^0$	$(0.7)^1$	$(0.7)^2$	$(0.7)^3$	$(0.7)^4$...	$(0.7)^n$
Tails prob	$(0.3)^{n-0}$	$(0.3)^{n-1}$	$(0.3)^{n-2}$	$(0.3)^{n-3}$	$(0.3)^{n-4}$...	$(0.3)^{n-n} = (0.3)^0$

So far the formula for the probability of getting x heads when flipping a coin n times is:

$$P(X = x) = c_x (0.7)^x (0.3)^{n-x}$$

If we can just find a pattern for the coefficients c_x we will have a formula for calculating the probability of getting x heads when flipping a coin n times. Lets go back to calculations:

$$\begin{array}{c|c|c|c} x & 0 & 1 & 2 \\ \hline P(X = x) & 1(0.7)^0(0.3)^2 & 2(0.7)^1(0.3)^1 & 1(0.7)^2(0.3)^0 \end{array}$$

$$\begin{array}{c|c|c|c|c} x & 0 & 1 & 2 & 3 \\ \hline P(X = x) & 1(0.7)(0.3)^3 & 3(0.7)(0.3)^2 & 3(0.7)^2(0.3) & 1(0.7)^3(0.3)^0 \end{array}$$

$$\begin{array}{c|c|c|c|c|c} x & 0 & 1 & 2 & 3 & 4 \\ \hline P(X = x) & 1(0.7)^0(0.3)^4 & 4(0.7)^1(0.3)^3 & 6(0.7)^2(0.3)^2 & 4(0.7)^3(0.3)^1 & 1(0.7)^4(0.3)^0 \end{array}$$

and we see that the coefficients are:

n						
2		1	2	1		
3		1	3	3	1	
4	1	4	6	4	1	

and we see that the coefficients are:

n						
2		1	2	1		
3		1	3	3	1	
4	1	4	6	4	1	

Well that looks familiar:

				1					
				1		1			
			1		2		1		
		1		3		3		1	
	1		4		6		4		1
1		5		10		10		5	
	1		5		10		10		1

And we have a formula for each of these values:

n							
2			1	2	1		
3		1	3	3	1		
4	1	4	6	4	1		
2			2C_0	2C_1	2C_2		
3		3C_0	3C_1	3C_2	3C_3		
4	4C_0	4C_1	4C_2	4C_3	4C_4		

And there are have it, the full formula for the probability of getting x heads when flipping a coin n times with $P(H) = 0.7$ and $P(T) = 0.3$ is

$$P(X = x) = {}_nC_x(0.7)^x(0.3)^{n-x}$$

So the fully general formula for the probability of getting x heads when flipping a coin n times with $P(H) = p$ and $P(T) = q = (1 - p)$ is

$$P(X = x) = {}_nC_x p^x q^{n-x}$$

Formula 1 (Probability for a Binomial Distribution)

For a binomial random variable X , the probability of obtaining x successes in n independent trials is given by

$$P(X = x) = {}_nC_x \cdot p^x(1 - p)^{n-x}$$

where

x is the number of successes,

n is the number of trials, and

p is the probability of a success on any trial.

Definition 3 (Binomial Distribution)

The binomial distribution is a discrete probability distribution with a fixed number of independent trials, where each trial has only two possible outcomes and one of these outcomes is counted. It has the following properties:

1. The experiment consists of a fixed number, n , of identical trials.
2. Each trial is independent of the others.
3. For each trial, there are only two possible outcomes. For counting purposes, one outcome is labeled a success, and the other a failure.
4. For every trial, the probability of getting a success is called p . The probability of getting a failure is then $1 - p$.
5. The binomial random variable, X , counts the number of successes in n trials.
6. For a binomial distribution, the mean is given by $\mu = np$ and the variance is given by $\sigma^2 = np(1 - p)$.

Why tables.

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

<i>n</i>	<i>x</i>	<i>p</i>										
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>n</i>	<i>x</i>	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

		p										
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

 p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

		p										
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

 p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

n	x	p										
		0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

 p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

		p										
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

 p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

		<i>p</i>										
<i>n</i>	<i>x</i>	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>n</i>	<i>x</i>	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = ?$$

		<i>p</i>										
<i>n</i>	<i>x</i>	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
<i>n</i>	<i>x</i>	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90

p

$$n = 4, p = 0.25, x = 3, P(X \leq 3) = 0.9961$$

		p										
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
2	0	0.8100	0.6400	0.5625	0.4900	0.3600	0.2500	0.1600	0.0900	0.0625	0.0400	0.0100
	1	0.9900	0.9600	0.9375	0.9100	0.8400	0.7500	0.6400	0.5100	0.4375	0.3600	0.1900
	2	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
3	0	0.7290	0.5120	0.4219	0.3430	0.2160	0.1250	0.0640	0.0270	0.0156	0.0080	0.0010
	1	0.9720	0.8960	0.8438	0.7840	0.6480	0.5000	0.3520	0.2160	0.1562	0.1040	0.0280
	2	0.9990	0.9920	0.9844	0.9730	0.9360	0.8750	0.7840	0.6570	0.5781	0.4880	0.2710
	3	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
4	0	0.6561	0.4096	0.3164	0.2401	0.1296	0.0625	0.0256	0.0081	0.0039	0.0016	0.0001
	1	0.9477	0.8192	0.7383	0.6517	0.4752	0.3125	0.1792	0.0837	0.0508	0.0272	0.0037
	2	0.9963	0.9728	0.9492	0.9163	0.8208	0.6875	0.5248	0.3483	0.2617	0.1808	0.0523
	3	0.9999	0.9984	0.9961	0.9919	0.9744	0.9375	0.8704	0.7599	0.6836	0.5904	0.3439
	4	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000	1.0000
n	x	0.10	0.20	0.25	0.30	0.40	0.50	0.60	0.70	0.75	0.80	0.90
p												

Assume the random variable X has a binomial distribution with the given probability of obtaining a success. Find the following probability, given the number of trials and the probability of obtaining a success. Round your answer to four decimal places.

$$P(X \leq 4), n = 7, p = 0.6$$

Assume the random variable X has a binomial distribution with the given probability of obtaining a success. Find the following probability, given the number of trials and the probability of obtaining a success. Round your answer to four decimal places.

$$P(X < 5), n = 6, p = 0.7$$

Assume the random variable X has a binomial distribution with the given probability of obtaining a success. Find the following probability, given the number of trials and the probability of obtaining a success. Round your answer to four decimal places.

$$P(X = 4), n = 6, p = 0.3$$

Assume the random variable X has a binomial distribution with the given probability of obtaining a success. Find the following probability, given the number of trials and the probability of obtaining a success. Round your answer to four decimal places.

$$P(X > 4), n = 6, p = 0.4$$

The random variable X is a binomial random variable with $n = 10$ and $p = 0.3$. What is the expected value of X ? Do not round your answer.

The random variable X is a binomial random variable with $n = 8$ and $p = 0.4$. What is the standard deviation of X ?

A researcher wishes to conduct a study of the color preferences of new car buyers. Suppose that 40% of this population prefers the color green. If 12 buyers are randomly selected, what is the probability that exactly 2 buyers would prefer green? Round your answer to four decimal places.

A real estate agent has 18 properties that she shows. She feels that there is a 30% chance of selling any one property during a week. The chance of selling any one property is independent of selling another property. Compute the probability of selling at least 5 properties in one week. Round your answer to four decimal places.

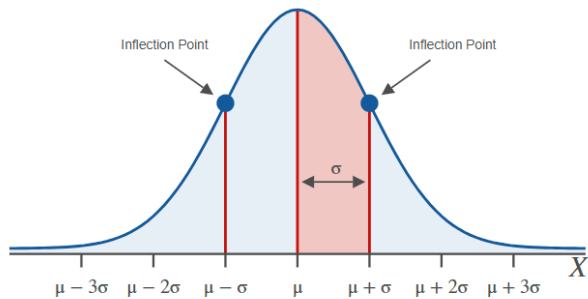
A quality control inspector has drawn a sample of 13 light bulbs from a recent production lot. If the number of defective bulbs is 1 or more, the lot fails inspection. Suppose 20% of the bulbs in the lot are defective. What is the probability that the lot will fail inspection? Round your answer to four decimal places.

A certain insecticide kills 70% of all insects in laboratory experiments. A sample of 9 insects is exposed to the insecticide in a particular experiment. What is the probability that exactly 4 insects will survive? Round your answer to four decimal places.

Chapter 6 Section 1

Definition 4 (Continuous random variable)

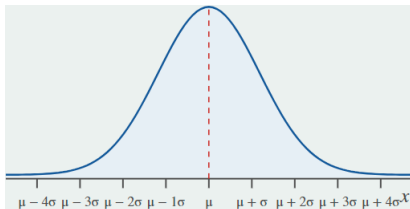
A continuous random variable is a continuous variable whose numeric value is determined by the outcome of a probability experiment.

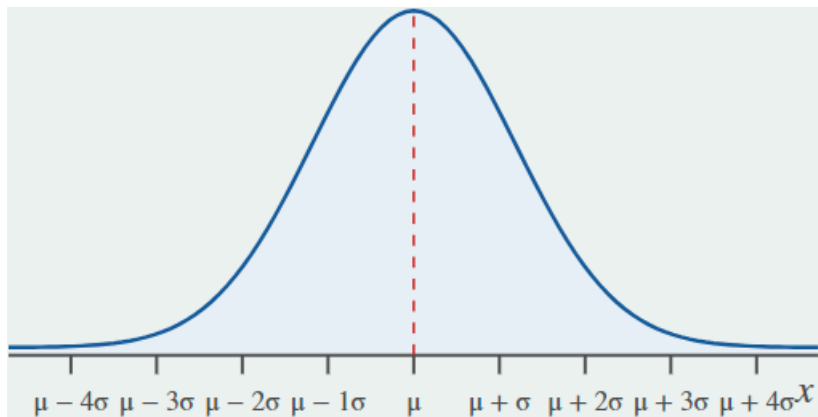


Definition 5 (Normal distribution)

A normal distribution is a probability distribution for a continuous random variable, X , defined completely by its mean and standard deviation, such that the following properties are true:

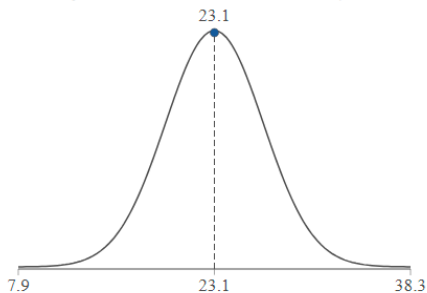
1. A normal distribution is bell-shaped and symmetric about its mean.
2. A normal distribution is completely defined by its mean, μ , and standard deviation, σ .
3. The total area under a normal distribution curve equals 1.
4. The x -axis is a horizontal asymptote for a normal distribution curve.





Given $X = 11.6$, $\mu = 23.1$, and $\sigma = 2.4$, indicate on the curve where the given X value would be.

The point can be moved by dragging or using the arrow keys. Select the Reset button to reset the point.

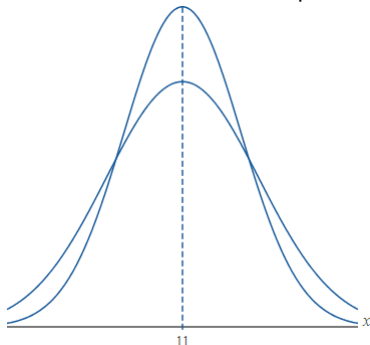
[Reset](#)

Decide which of the following statements are true.

- ☐ There are an unlimited number of normal distributions.
- ☐ For any normal distribution, only the mean and mode are equal. The median is different from the mean and mode.
- ☐ The line of symmetry for all normal distributions is $x = \mu$.
- ☐ The x -axis is a vertical asymptote for all normal distributions.

Calculate the standard score of the given X value, $x = 88.9$, where $\mu = 94.5$ and $\sigma = 95.1$. Round your answer to two decimal places.

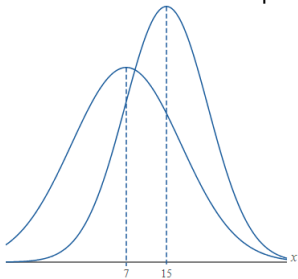
The following is a graph of two normal distributions plotted on the same x -axis.



Based on the graph above, which statement best describes the graph?

- ▶ The two distributions have equal means and different standard deviations.
- ▶ The two distributions have equal means and equal standard deviations.
- ▶ The two distributions have equal means and standard deviations that differ by 11 units.
- ▶ The two distributions have equal standard deviations and different means.

The following is a graph of two normal distributions plotted on the same x -axis.



Based on the graph above, which statement best describes the graph?

- ▶ The two distributions have equal means and standard deviations that differ by 8 units.
- ▶ The two distributions have means that differ by 8 units and equal standard deviations.
- ▶ The two distributions have means that differ by 8 units and different standard deviations.
- ▶ The two distributions have equal means and means standard deviations.

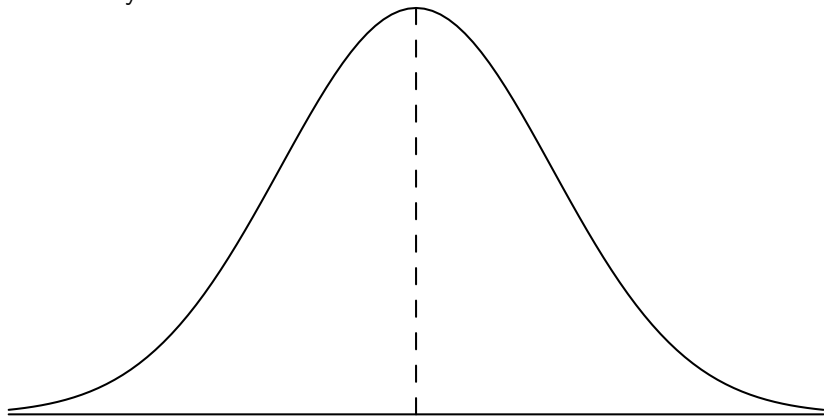
Chapter 6 Section 2

Draw Binomial with $n = 4, p = 0.5$.

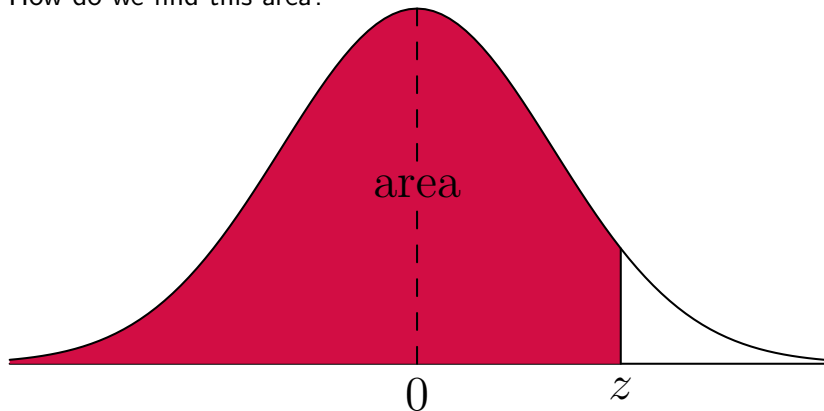
Find

1. $P(X = 0) =$
2. $P(X = 1) =$
3. $P(X = 2) =$
4. $P(X = 3) =$
5. $P(X = 4) =$
6. $P(X \leq 0) =$
7. $P(X \leq 1) =$
8. $P(X \leq 2) =$
9. $P(X \leq 3) =$
10. $P(X \leq 4) =$

Probability is area under the curve:



How do we find this area?



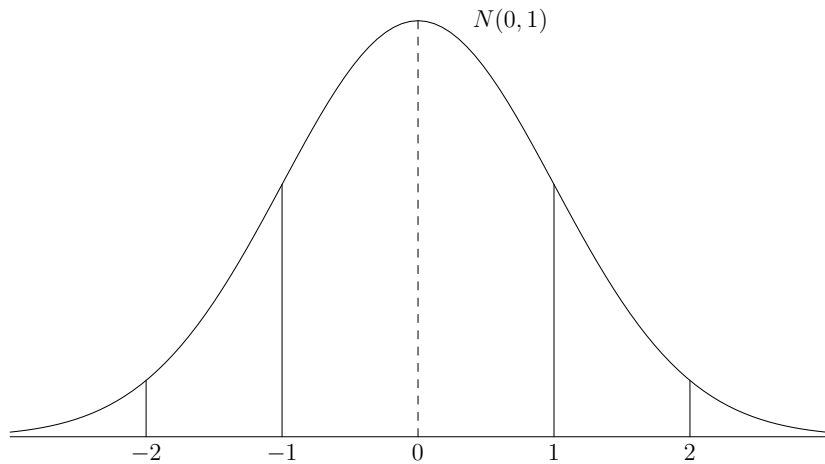
In the case of a normal distribution with mean μ and standard deviation σ its function is:

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

Definition 6 (Standard normal distribution)

The standard normal distribution is the normal distribution with $\mu = 0$ and $\sigma = 1$, such that the following properties are true:

1. The standard normal distribution is bell-shaped and symmetric about its mean.
2. The standard normal distribution is completely defined by its mean, $\mu = 0$, and standard deviation, $\sigma = 1$.
3. The total area under the standard normal distribution curve equals 1.
4. The x -axis is a horizontal asymptote for the standard normal distribution curve.



Find the area under the standard normal curve to the left of $z = -1.67$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve to the left of $z = -1.67$. Round your answer to four decimal places, if necessary.

$$z = -1.6 + 0.07^\dagger$$

[†]Technically this should say $-(1.6 + 0.07)$ or $-1.6 - 0.07$. Think of the $+$ as the word “and”.

Find the area under the standard normal curve to the left of $z = -1.67$. Round your answer to four decimal places, if necessary.

$$z = -1.6 + 0.07 = \text{row value} + \text{column value}$$

Find the area under the standard normal curve to the left of $z = -1.67$. Round your answer to four decimal places, if necessary.

$$z = -1.6 + 0.07 = \text{row value} + \text{column value}$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09

Find the area under the standard normal curve to the left of $z = -1.67$. Round your answer to four decimal places, if necessary.

$$z = -1.6 + 0.07 = \text{row value} + \text{column value}$$

z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
-1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
-1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
-1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
-1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
-1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0571	0.0559
z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09

Find the area under the standard normal curve to the right of $z = 1.38$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve to the right of $z = 2.38$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve to the right of $z = -2.37$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve between $z = 1.38$ and $z = 2.92$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve between $z = -1.44$ and $z = 1.44$. Round your answer to four decimal places, if necessary.

Find the area under the standard normal curve to the left of $z = -1.44$ and to the right of $z = 1.44$. Round your answer to four decimal places, if necessary.

Find the specified probability. Round your answer to four decimal places, if necessary.

$$P(0 < z < 2.03)$$

Use the z -score formula, $z = \frac{x - \mu}{\sigma}$, and the information below to find the mean μ . Round your answer to one decimal place, if necessary.

$$z = 1.75, \quad x = 10.9, \quad \sigma = 4.4$$

Chapter 6 Section 3

The life of light bulbs is distributed normally. The variance of the lifetime is 625 hours and the mean lifetime of a bulb is 510 hours. Find the probability of a bulb lasting for at most 550 hours. Round your answer to four decimal places.

A soft drink machine outputs a mean of 25 ounces per cup. The machine's output is normally distributed with a standard deviation of 4 ounces. What is the probability of filling a cup between 22 and 33 ounces? Round your answer to four decimal places.

The time spent waiting in the line is approximately normally distributed. The mean waiting time is 6 minutes and the standard deviation of the waiting time is 1 minutes. Find the probability that a person will wait for more than 8 minutes. Round your answer to four decimal places.

Trucks in a delivery fleet travel a mean of 120 miles per day with a standard deviation of 18 miles per day. The mileage per day is distributed normally. Find the probability that a truck drives between 150 and 156 miles in a day. Round your answer to four decimal places.

The weights of newborn baby boys born at a local hospital are believed to have a normal distribution with a mean weight of 3844 grams and a standard deviation of 612 grams. If a newborn baby boy born at the local hospital is randomly selected, find the probability that the weight will be less than 4456 grams. Round your answer to four decimal places.

Chapter 6 Section 4

ADD notes on how to do a “reverse table read”.

What value of z divides the standard normal distribution so that half the area is on one side and half is on the other? Round your answer to two decimal places.

Find the value of z such that 0.1401 of the area lies to the left of z .
Round your answer to two decimal places.

Find the value of z such that 0.14 of the area lies to the left of z .
Round your answer to two decimal places.

Find the value of z such that 0.03 of the area lies to the right of z .
Round your answer to two decimal places.

Find the value of z such that 0.05 of the area lies to the right of z .
Round your answer to two decimal places.

Find the value of z such that 0.8664 of the area lies between $-z$ and z . Round your answer to two decimal places.

Suppose SAT Writing scores are normally distributed with a mean of 497 and a standard deviation of 109. A university plans to award scholarships to students whose scores are in the top 8%. What is the minimum score required for the scholarship? Round your answer to the nearest whole number, if necessary.

Chapter 7 Section 1

Definition 7 (Sampling distribution)

The distribution of the values of a particular sample statistic for all possible samples of a given size, n .

And in particular we are interested in...

Definition 8 (Sampling distribution of sample means)

The distribution of sample means for all possible samples of a given size, n .

Add link to Pluto CLT pdf.

Formula 1 (Mean of the Sampling Distribution of Sample Means)

The mean of a sampling distribution of sample means, $\mu_{\bar{x}}$, equals the mean of the population, μ .

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

Formula 2 (Standard Deviation of the Sampling Distribution of Sample Means a.k.a. Standard Error of the Mean)

The standard deviation of a sampling distribution of sample means, $\sigma_{\bar{x}}$, equals the standard deviation of the population, σ , divided by the square root of the sample size, n .

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

Theorem 9 (The Central Limit Theorem (CLT))

For any given population with mean, μ , and standard deviation, σ , the shape of the sampling distribution of sample means will approach that of a normal distribution as the sample size increases.

For this text we will consider a sample size to be large enough if $n \geq 30$, as is common practice.

Furthermore, the larger the sample size, the better the normal distribution approximation will be.

This leads to...

Definition 10 (Distribution: Sampling Distribution of Sample Means)

A sampling distribution of sample means is the distribution of sample means for all possible samples of a given size, n , such that

1. The mean is $\mu_{\bar{x}} = \mu$.
2. The standard deviation is $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{n}}$.

A normal distribution may be used to approximate a sampling distribution of sample means if either:

- ▶ The sample size, n , is at least 30, in which case the Central Limit Theorem applies.
- ▶ The population is normally distributed.

Find the mean of the sampling distribution of sample means using the given information. Round to one decimal place, if necessary.

$$\mu = 77 \text{ and } \sigma = 14; n = 64$$

Find the standard deviation of the sampling distribution of sample means using the given information. Round to one decimal place, if necessary.

$$\mu = 78 \text{ and } \sigma = 14; n = 81$$

A study on the latest fad diet claimed that the amounts of weight lost by all people on this diet had a mean of 21.0 pounds and a standard deviation of 4.4 pounds.

Step 1 of 2: If a sampling distribution is created using samples of the amounts of weight lost by 87 people on this diet, what would be the mean of the sampling distribution of sample means? Round to two decimal places, if necessary.

Step 2 of 2: If a sampling distribution is created using samples of the amounts of weight lost by 87 people on this diet, what would be the standard deviation of the sampling distribution of sample means? Round to two decimal places, if necessary.

Chapter 7 Section 2

Formula 1 (Standard Score for a Sample Mean)

If either the sample size, n , is at least 30 or the population is normally distributed, then the standard score for a sample mean in a sampling distribution is given by

$$z = \frac{\bar{x} - \mu_{\bar{x}}}{\sigma_{\bar{x}}} = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)}$$

where \bar{x} is the given sample mean,

μ is the population mean,

σ is the population standard deviation, and

n is the sample size used to create the sampling distribution.

Suppose the horses in a large stable have a mean weight of 873 lbs, and a variance of 17,161 lbs.

What is the probability that the mean weight of the sample of horses would be less than 856 lbs if 50 horses are sampled at random from the stable? Round your answer to four decimal places.

Suppose a batch of metal shafts produced in a manufacturing company have a standard deviation of 1.5 and a mean diameter of 208 inches.

If 60 shafts are sampled at random from the batch, what is the probability that the mean diameter of the sample shafts would be greater than 208.1 inches? Round your answer to four decimal places.

The mean points obtained in an aptitude examination is 159 points with a standard deviation of 13 points.

What is the probability that the mean of the sample would be greater than 160.2 points if 60 exams are sampled? Round your answer to four decimal places.

A quality control expert at Glotech computers wants to test their new monitors. The production manager claims they have a mean life of 95 months with a standard deviation of 5 months.

If the claim is true, what is the probability that the mean monitor life would be less than 96.3 months in a sample of 84 monitors? Round your answer to four decimal places.

Chapter 8 Section 1

Definition 11 (Point Estimate)

A point estimate is a single-number estimate of a population parameter.

Definition 12 (Unbiased Estimator)

An unbiased estimator is a point estimate that does not consistently underestimate or overestimate the population parameter.

Suppose we have 1000 data points between 1 and 100:

46, 63, 17, 44, 58, \dots , 57, 16, 95, 23, 76

It has a population mean $\mu = 50.927$, population median 51.0, and population mode 97.

Suppose we take 30 random samples of size 100 and calculate the mean for each sample:

46.61	51.38	51.73	48.25	44.06	45.39
50.75	52.19	46.83	54.81	54.3	50.03
56.75	47.35	53.77	51.36	51.18	49.21
46.51	48.52	55.3	49.6	53.86	51.37
51.09	54.59	52.27	48.29	50.36	50.13

Do these sample means consistently underestimate or overestimate the population mean?

It's probably hard to see what's happening here, so let's subtract each of these sample means from the population mean.

So we have $\mu - \bar{x}$ for each sample (sorted from smallest to largest):

-5.82	-4.37	-3.88	-3.66	-3.37	-2.93
-2.84	-1.34	-1.26	-0.8	-0.45	-0.44
-0.43	-0.25	-0.16	0.18	0.57	0.8
0.9	1.33	1.72	2.41	2.64	2.68
3.58	4.1	4.32	4.42	5.54	6.87

While there are some “big” differences (such as the 6.87) the sample mean doesn't consistently underestimate or overestimate the population mean of $\mu = 50.927$.

Calculating the medians gives:

46.5	53.0	52.0	47.5	42.0	45.0
48.5	51.5	44.0	56.5	54.5	48.0
63.0	48.0	55.0	52.0	54.5	48.0
45.5	49.0	58.0	50.0	55.0	52.5
48.5	54.5	51.5	42.0	46.5	47.0

And subtracting them from the population median and sorting:

-12.0	-7.0	-5.5	-4.0	-4.0	-3.5
-3.5	-3.5	-2.0	-1.5	-1.0	-1.0
-0.5	-0.5	1.0	2.0	2.5	2.5
3.0	3.0	3.0	3.5	4.0	4.5
4.5	5.5	6.0	7.0	9.0	9.0

The sample medians do seem to consistently underestimate or overestimates the population median of 51.0 (it's a bit hard to see this with this small number of samples).

Calculating the modes gives:

11	93	92	14	7	8
72	90	97	33	67	59
23	14	99	59	8	99
10	5	58	22	71	82
23	26	71	22	14	25

And subtracting them from the population mode and sorting:

-2.0	-2.0	0.0	4.0	5.0	7.0
15.0	25.0	26.0	26.0	30.0	38.0
38.0	39.0	64.0	71.0	72.0	74.0
74.0	75.0	75.0	83.0	83.0	83.0
86.0	87.0	89.0	89.0	90.0	92.0

The sample modes definitely consistently underestimate or overestimates the population mode of 97.

Definition 13 (Interval estimate)

An interval estimate is a range of possible values for a population parameter.

Definition 14 (Level of confidence)

The level of confidence, denoted by c , is the percentage of all possible samples of a given size that will produce interval estimates that contain the actual parameter.

Definition 15 (Confidence interval)

A confidence interval is an interval estimate associated with a certain level of confidence.

Link to [Confidence Interval: Known Sigma \(Click Me\)](#)

Definition 16 (Margin of error)

The margin of error, or maximum error of estimate, E , is the largest possible distance from the point estimate that a confidence interval will cover.

Definition 17 (Critical z -values)

Critical z -values, denoted

$$-z_{\frac{\alpha}{2}} \quad \text{and} \quad z_{\frac{\alpha}{2}},$$

mark the boundaries for the area under the middle of the standard normal curve that corresponds with a particular level of confidence, c .

Formula 1 (Margin of Error of a Confidence Interval for a Population Mean (σ Known))

When the population standard deviation is known, the sample taken is a simple random sample, and either the sample size is at least 30 or the population distribution is approximately normal, the margin of error of a confidence interval for a population mean is given by

$$E = \left(z_{\frac{\alpha}{2}} \right) (\sigma_{\bar{X}}) = \left(z_{\frac{\alpha}{2}} \right) \left(\frac{\sigma}{\sqrt{n}} \right)$$

where $z_{\frac{\alpha}{2}}$ is the critical value for the level of confidence, $c = 1 - \alpha$, such that the area under the standard normal distribution to the right of $z_{\frac{\alpha}{2}}$ is equal to $\frac{\alpha}{2}$, σ is the population standard deviation, and n is the sample size.

Rounding Rule

When calculating a margin of error for a confidence interval, round to at least six decimal places to avoid additional rounding errors in the subsequent calculations of the endpoints of the confidence interval.

Formula 2 (Confidence Interval for a Population Mean)

The confidence interval for a population mean is given by

$$\bar{x} - E < \mu < \bar{x} + E$$

or

$$(\bar{x} - E, \bar{x} + E)$$

where \bar{x} is the sample mean, which is the point estimate for the population mean, and E is the margin of error.

Rounding Rule

Round the endpoints of a confidence interval for a population mean as follows:

- ▶ If sample data are given, round to one more decimal place than the largest number of decimal places in the given data.
- ▶ If statistics are given, round to the same number of decimal places as given in the standard deviation or variance.

The mean monthly water bill for 52 residents of the local apartment complex is \$73. What is the best point estimate for the mean monthly water bill for all residents of the local apartment complex?

The water works commission needs to know the mean household usage of water by the residents of a small town in gallons per day. Assume that the population standard deviation is 2.3 gallons. The mean water usage per family was found to be 18.1 gallons per day for a sample of 386 families. Construct the 80% confidence interval for the mean usage of water. Round your answers to one decimal place.

An educational psychologist wishes to know the mean number of words a third grader can read per minute. She wants to make an estimate at the 95% level of confidence. For a sample of 507 third graders, the mean words per minute read was 22.9. Assume a population standard deviation of 3.1. Construct the confidence interval for the mean number of words a third grader can read per minute. Round your answers to one decimal place.

The mean number of hours of part-time work per week for a sample of 552 college students is 22. If the margin of error for the population mean with a 99% confidence interval is 1.6, construct a 99% confidence interval for the mean number of hours of part-time work per week for all college students.

A fast food restaurant executive wishes to know how many fast food meals teenagers eat each week. They want to construct a 85% confidence interval for the mean and are assuming that the population standard deviation for the number of fast food meals consumed each week is 1.4. The study found that for a sample of 502 teenagers the mean number of fast food meals consumed per week is 3.1. Construct the desired confidence interval. Round your answers to one decimal place.

Formula 3 (Minimum Sample Size for Estimating a Population Mean)

The minimum sample size required for estimating a population mean at a given level of confidence with a particular margin of error is given by

$$n = \left(\frac{z_{\frac{\alpha}{2}} \cdot \sigma}{E} \right)^2$$

where $z_{\frac{\alpha}{2}}$ is the critical value for the level of confidence, $c = 1 - \alpha$, such that the area under the standard normal distribution to the right of $z_{\frac{\alpha}{2}}$ is equal to $\frac{\alpha}{2}$, σ is the population standard deviation, and E is the desired maximum margin of error.

Rounding Rule

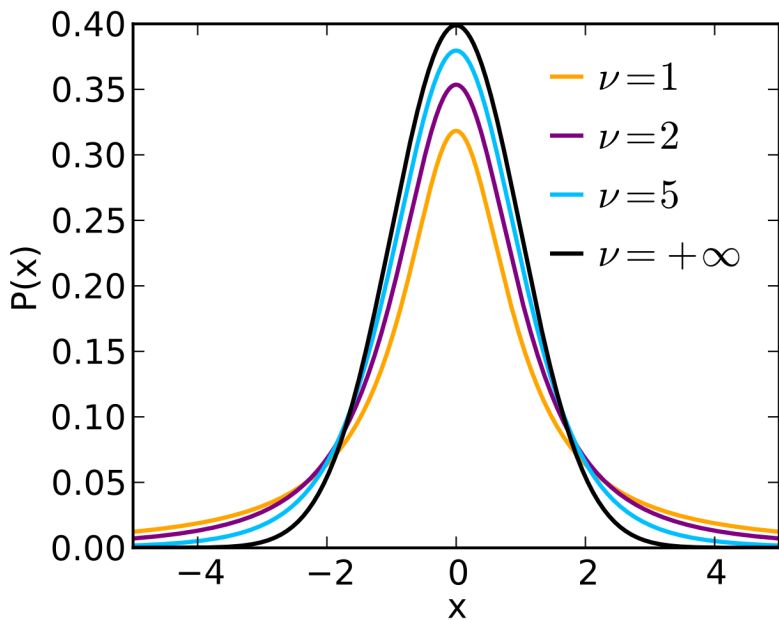
When calculating a minimum sample size, round up to the next larger whole number.

A research company desires to know the mean consumption of meat per week among males over age 23. They believe that the meat consumption has a mean of 3.8 pounds, and want to construct a 95% confidence interval with a maximum error of 0.1 pounds. Assuming a variance of 0.64 pounds, what is the minimum number of males over age 23 they must include in their sample? Round your answer up to the next integer.

Chapter 8 Section 2

Side Note Student's t -Distribution

William Gossett published his work regarding the t -distribution in 1908 while working at the Guinness Brewery in Dublin, Ireland. Because his employer required him to publish under a pseudonym, Gossett chose the name "Student." Over a century later, Gossett's work lives on, as does the pseudonym: Student's t -distribution.



Definition 1 (Distribution: Student's t -Distribution)

The Student's t -distribution is a probability distribution for a continuous random variable, X , defined completely by its number of degrees of freedom, such that the following properties are true

- 1. A t -distribution curve is symmetric and bell-shaped, centered about 0.*
- 2. A t -distribution curve is completely defined by its number of degrees of freedom, df .*
- 3. The total area under a t -distribution curve equals 1.*
- 4. The x -axis is a horizontal asymptote for a t -distribution curve.*

Rule 1 (Rounding Rule)

When calculating a t -value, round to three decimal places. This follows the convention used in the t -distribution table.

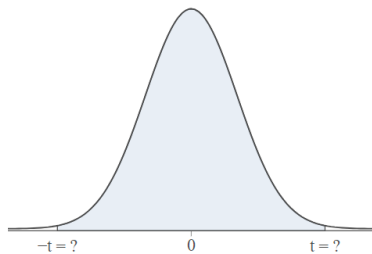
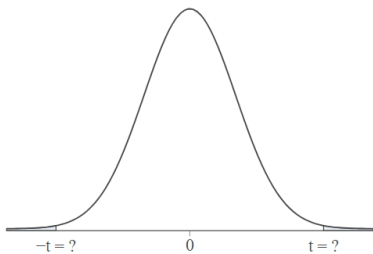
Find the value of $t_{0.10}$ for a t -distribution with 8 degrees of freedom. Round your answer to three decimal places, if necessary.

Find the value of t for a t -distribution with 32 degrees of freedom such that the area to the right of t equals 0.005. Round your answer to three decimal places, if necessary.

Find the value of t for a t -distribution with 13 degrees of freedom such that the area to the left of t equals 0.05. Round your answer to three decimal places, if necessary.

Consider the value t such that the area under the curve between $-|t|$ and $|t|$ equals 0.9.

Step 1 of 2 Select which graph best represents the given description of t .

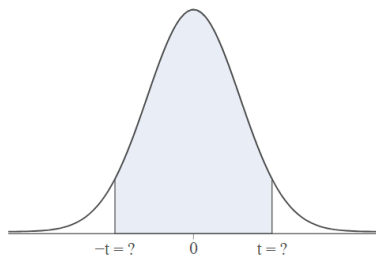
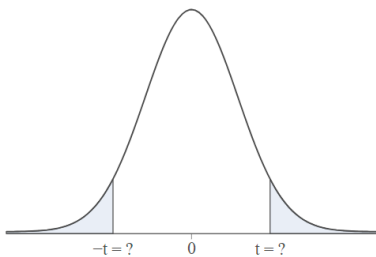
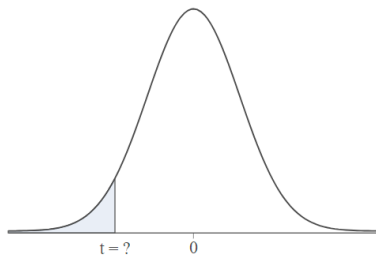
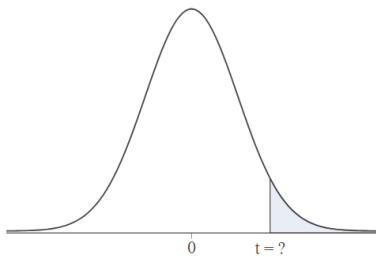


Step 2 of 2 Assuming the degrees of freedom equals 2, select the t -value from the t -distribution table.

Find the critical t -value for a 99% confidence interval using a t -distribution with 28 degrees of freedom. Round your answer to three decimal places, if necessary.

Consider the value t such that the area to the left of $-|t|$ plus the area to the right of $|t|$ equals 0.1.

Step 1 of 2 Select which graph best represents the given description of t .



Step 2 of 2 Assuming the degrees of freedom equals 30, select the t -value from the t -distribution table.

Chapter 8 Section 3

Formula 1 (Formula: Margin of Error of a Confidence Interval for a Population Mean (σ Unknown))

When the population standard deviation is unknown, the sample taken is a simple random sample, and either the sample size is at least 30 or the population distribution is approximately normal, the margin of error of a confidence interval for a population mean is given by

$$E = \left(t_{\frac{\alpha}{2}}\right) \left(\frac{s}{\sqrt{n}}\right)$$

where $t_{\frac{\alpha}{2}}$ is the critical value for the level of confidence, $c = 1 - \alpha$, such that the area under the t -distribution with $n - 1$ degrees of freedom to the right of $t_{\frac{\alpha}{2}}$ is equal to $\frac{\alpha}{2}$, s is the sample standard deviation, and n is the sample size.

Rule 1 (Rounding Rule)

When calculating a margin of error for a confidence interval, round to at least six decimal places to avoid additional rounding errors in the subsequent calculations of the endpoints of the confidence interval.

Formula 2 (Confidence Interval for a Population Mean)

The confidence interval for a population mean is given by

$$\bar{x} - E < \mu < \bar{x} + E$$

or

$$(\bar{x} - E, \bar{x} + E)$$

where \bar{x} is the sample mean, which is the point estimate for the population mean, and E is the margin of error.

Rule 2 (Rounding Rule)

Round the endpoints of a confidence interval for a population mean as follows:

- ▶ *If sample data are given, round to one more decimal place than the largest number of decimal places in the given data.*
- ▶ *If statistics are given, round to the same number of decimal places as given in the standard deviation or variance.*

A consumer affairs investigator records the repair cost for 20 randomly selected refrigerators. A sample mean of \$57.22 and standard deviation of \$25.76 are subsequently computed. Determine the 90% confidence interval for the mean repair cost for the refrigerators. Assume the population is approximately normal.

Step 1 of 2 Find the critical value that should be used in constructing the confidence interval. Round your answer to three decimal places.

Step 2 of 2 Construct the 90% confidence interval. Round your answer to two decimal places.

A physicist examines 8 seawater samples for nitrate concentration. The mean nitrate concentration for the sample data is 0.424 cc/cubic meter with a standard deviation of 0.0827. Determine the 95% confidence interval for the population mean nitrate concentration. Assume the population is approximately normal.

Step 1 of 2 Find the critical value that should be used in constructing the confidence interval. Round your answer to three decimal places.

Step 2 of 2 Construct the 95% confidence interval. Round your answer to two decimal places.

A random sample of 10 fields of rye has a mean yield of 45.4 bushels per acre and standard deviation of 2.93 bushels per acre. Determine the 80% confidence interval for the true mean yield. Assume the population is approximately normal.

Step 1 of 2 Find the critical value that should be used in constructing the confidence interval. Round your answer to three decimal places.

Step 2 of 2 Construct the 80% confidence interval. Round your answer to two decimal places.

Chapter 10 Section 1

A **hypothesis** is a theory or premise—often it is a claim that someone is making that must be investigated. Hypothesis testing is a statistical process for determining the likelihood that a given hypothesis is true.

Procedure 1 (Performing a Hypothesis Test)

1. *State the null and alternative hypothesis.*
2. *Determine which distribution to use for the test statistic, and state the level of significance.*
3. *Gather data and calculate the necessary sample statistics.*
4. *Draw a conclusion and interpret the decision.*

Definition 18 (Hypothesis Testing Terminology)

Hypothesis testing is a statistical process for determining the likelihood that a given hypothesis is true.

The **alternative hypothesis**, denoted by H_a , is a mathematical statement that describes a population parameter, and it is the hypothesis that the researcher is aiming to gather evidence to support. It is also referred to as the *research hypothesis*.

The **null hypothesis**, denoted by H_0 , is the the statement that expresses the value currently believed to be true; it will always be a statement of equality.

Shortcut	Types
Ctrl+Alt+G M	μ
Ctrl+Alt+=	\neq

The population parameter we want to test determines the distribution we use and test statistic.

Mean, μ	Normal	$z = \frac{\bar{x} - \mu}{\frac{\sigma}{\sqrt{n}}}$
Mean, μ	Student's t	$z = \frac{\bar{x} - \mu}{\frac{s}{\sqrt{n}}}$
Proportion, p	Normal	$z = \frac{\hat{p} - p}{\sqrt{\frac{p(1-p)}{n}}}$
Variance, σ^2	Chi-Squared	$\chi^2 = \frac{(n-1)s^2}{\sigma^2}$

JSU male heights: $\mu = 68$ inches, $\sigma = 3$ inches

Suppose it's believed that the average JSU male is 63 inches (someone confused the male and female average heights).

You take a sample of $n = 25$ JSU males and find the average to be $\bar{x} = 68$ inches.

Are we surprised to see an average height of 68 inches if the mean is 63 inches?

How far from the supposed mean of 63 inches is this?

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{68 - 63}{\left(\frac{3}{\sqrt{25}}\right)} = 8.33$$

That's a whole lotta standard deviations! So we should be very surprised!

A manufacturer of chocolate chips would like to know whether its bag filling machine works correctly at the 402 gram setting. Is there sufficient evidence at the 0.05 level that the bags are underfilled or overfilled? Assume the population is normally distributed.

State the null and alternative hypotheses for the above scenario.

A manufacturer of banana chips would like to know whether its bag filling machine works correctly at the 445 gram setting. Is there sufficient evidence at the 0.02 level that the bags are overfilled? Assume the population is normally distributed.

State the null and alternative hypotheses for the above scenario.

A manufacturer of banana chips would like to know whether its bag filling machine works correctly at the 426 gram setting. Is there sufficient evidence at the 0.02 level that the bags are underfilled? Assume the population is normally distributed.

State the null and alternative hypotheses for the above scenario.

A sample statistic is said to be **statistically significant** if it is far enough away from the presumed value of the population parameter to conclude that it would be unlikely for that sample statistic to occur by chance if the null hypothesis is true.

The level of confidence, c , is a measure of how certain we are that the confidence interval contains the population parameter. If the level of confidence we choose for a particular hypothesis test, c , is how certain we are that we have captured the true parameter, then $1 - c$ is the probability that we did not capture the true parameter.

The **level of significance** is $\alpha = 1 - c$. When conducting a hypothesis test, it is important to determine the level of significance you want to use before gathering the sample data, as a different level of significance might produce a different conclusion to the test.

There are two possible conclusions when performing a Hypothesis Test:

1. Reject the Null Hypothesis
2. Fail to Reject the Null Hypothesis

To report the results it is common to state whether there is or is not sufficient evidence to support the alternative hypothesis:

Reject the Null Hypothesis There is sufficient evidence ...

Fail to Reject the Null Hypothesis There is not sufficient evidence ...

Using traditional methods it takes 107 hours to receive an advanced flying license. A new training technique using Computer Aided Instruction (CAI) has been proposed. A researcher believes the new technique may reduce training time and decides to perform a hypothesis test. After performing the test on 120 students, the researcher fails to reject the null hypothesis at a 0.10 level of significance.

What is the conclusion?

- ▶ There is sufficient evidence at the 0.10 level of significance that the new technique reduces training time.
- ▶ There is not sufficient evidence at the 0.10 level of significance that the new technique reduces training time.

Using traditional methods it takes 98 hours to receive an advanced flying license. A new training technique using Computer Aided Instruction (CAI) has been proposed. A researcher believes the new technique may lengthen training time and decides to perform a hypothesis test. After performing the test on 200 students, the researcher decides to reject the null hypothesis at a 0.01 level of significance.

What is the conclusion?

- ▶ There is sufficient evidence at the 0.01 level of significance that the new technique reduces training time.
- ▶ There is not sufficient evidence at the 0.01 level of significance that the new technique reduces training time.

Chapter 10 Section 2

Hypothesis Testing: The logic

Recall the real data:

JSU male heights: $\mu = 68$ inches, $\sigma = 3$ inches

Suppose it's believed that the average JSU male is $\mu = 63$ inches (someone confused the male and female average heights). You take a sample of $n = 25$ JSU males and find the average to be $\bar{x} = 68$. Are we surprised to see an average height of 68 if the mean is 63? How far from the supposed mean is this?

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{68 - 63}{\left(\frac{3}{\sqrt{25}}\right)} = 8.33$$

That's a whole lotta standard deviations! So we should be very surprised!

Hypothesis Testing: The logic II

$$z = \frac{\bar{x} - \mu}{\left(\frac{\sigma}{\sqrt{n}}\right)} = \frac{68 - 63}{\left(\frac{3}{\sqrt{25}}\right)} = 8.33$$

In fact what is

$$P(Z \geq 8.33) = ?$$

Example 2.1

An engineer has designed a valve that will regulate water pressure on an automobile engine. The valve was tested on 120 engines and the mean pressure was 3.9 pounds/square inch (psi). Assume the population standard deviation is 0.9. If the valve was designed to produce a mean pressure of 4.1 psi, is there sufficient evidence at the 0.05 level that the valve does not perform to the specifications?

Example 2.2

A manufacturer of potato chips would like to know whether its bag filling machine works correctly at the 404 gram setting. It is believed that the machine is underfilling the bags. A 43 bag sample had a mean of 400 grams. Assume the population standard deviation is known to be 18. A level of significance of 0.1 will be used.

Example 2.3

An engineer has designed a valve that will regulate water pressure on an automobile engine. The valve was tested on 200 engines and the mean pressure was 4.2 pounds/square inch (psi). Assume the population standard deviation is 1.0. If the valve was designed to produce a mean pressure of 4.1 psi, is there sufficient evidence at the 0.02 level that the valve performs above the specifications?

- 1. State the null and alternative hypotheses.*
- 2. Find the value of the test statistic. Round your answer to two decimal places.*
- 3. Specify if the test is one-tailed or two-tailed.*
- 4. Find the p -value of the test statistic. Round your answer to four decimal places.*
- 5. Identify the level of significance for the hypothesis test.*
- 6. Make the decision to reject or fail to reject the null hypothesis.*

Example 2.4

The director of research and development is testing a new medicine. She wants to know if there is evidence at the 0.02 level that the medicine relieves pain in less than 312 seconds. For a sample of 65 patients, the mean time in which the medicine relieved pain was 310 seconds. Assume the population standard deviation is 23.

Example 2.5

An engineer has designed a valve that will regulate water pressure on an automobile engine. The valve was tested on 110 engines and the mean pressure was 4.5 pounds/square inch (psi). Assume the population variance is 0.64. The engineer designed the valve such that it would produce a mean pressure of 4.7 psi. It is believed that the valve does not perform to the specifications. A level of significance of 0.05 will be used. Find the p -value of the test statistic. Round your answer to four decimal places.

Example 2.6

An automobile manufacturer has given its car a 36.6 miles/gallon (MPG) rating. An independent testing firm has been contracted to test the actual MPG for this car since it is believed that the car performs over the manufacturer's MPG rating. After testing 170 cars, they found a mean MPG of 36.7. Assume the population standard deviation is known to be 2.3. Is there sufficient evidence at the 0.1 level to support the testing firm's claim?

Chapter 10 Section 3

Example 3.1

A manufacturer of chocolate chips would like to know whether its bag filling machine works correctly at the 431 gram setting. Based on a 26 bag sample where the mean is 441 grams and the standard deviation is 20, is there sufficient evidence at the 0.01 level that the bags are overfilled? Assume the population distribution is approximately normal.

Example 3.2

An engineer designed a valve that will regulate water pressure on an automobile engine. The engineer designed the valve such that it would produce a mean pressure of 7.9 pounds/square inch. It is believed that the valve performs above the specifications. The valve was tested on 16 engines and the mean pressure was 8.1 pounds/square inch with a standard deviation of 0.5. A level of significance of 0.01 will be used. Assume the population distribution is approximately normal. Make the decision to reject or fail to reject the null hypothesis.

Example 3.3

A manufacturer of chocolate chips would like to know whether its bag filling machine works correctly at the 436 gram setting. It is believed that the machine is underfilling the bags. A 24 bag sample had a mean of 430 grams with a variance of 225. A level of significance of 0.025 will be used. Assume the population distribution is approximately normal. Is there sufficient evidence to support the claim that the bags are underfilled?

Example 3.4

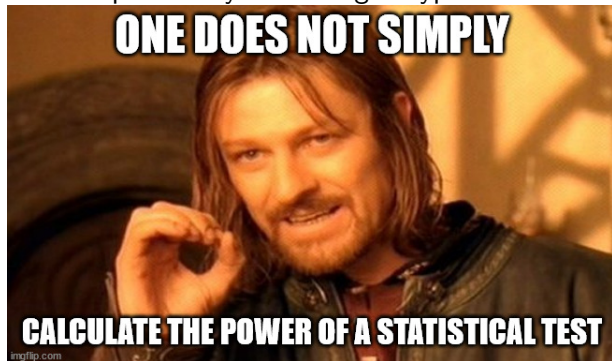
Using traditional methods, it takes 10.2 hours to receive a basic driving license. A new license training method using Computer Aided Instruction (CAI) has been proposed. A researcher used the technique with 25 students and observed that they had a mean of 10.6 hours with a standard deviation of 0.9. A level of significance of 0.05 will be used to determine if the technique performs differently than the traditional method. Assume the population distribution is approximately normal. Is there sufficient evidence to support the claim that the time taken is different from traditional methods?

Type I and Type II Errors

	H_0 True	H_0 False
Reject H_0	Type I error Probability = α False Positive	Correct inference Probability = $1 - \beta$ True Positive
Fail to Reject H_0	Correct inference Probability = $1 - \alpha$ True Negative	Type II error Probability = β False Negative

α is the significance of a statistical test.
It is the probability of making a Type I error.
We choose it.

β is the power of a statistical test.
It is the probability of making a Type II error.



	Innocent	Guilty
Guilty Verdict	Type I error	Correct
Not Guilty Verdict	Correct	Type II error

H_0 : Innocent

H_a : Guilty

	H_0 True	H_0 False
Reject H_0	Type I error	Correct inference
Fail to Reject H_0	Correct inference	Type II error

Chapter 12 Section 1

Definition 19 (Explanatory variable, Response variable)

Quantitative variables are measurements or counts. An explanatory variable, also known as the predictor variable or independent variable, is the variable whose fluctuation in value influences a change in the value of another variable. A response variable, or the dependent variable, is the variable that has its value change as a result of a change in the value of the explanatory variable. We will concentrate on two variables: an explanatory variable and a response variable. Relationship between variables will be linear relationships.

Extra graphs 2023 week 14 regular vs Georgia
See `error-type_regression_simpsons.pdf` on my Canvas page.

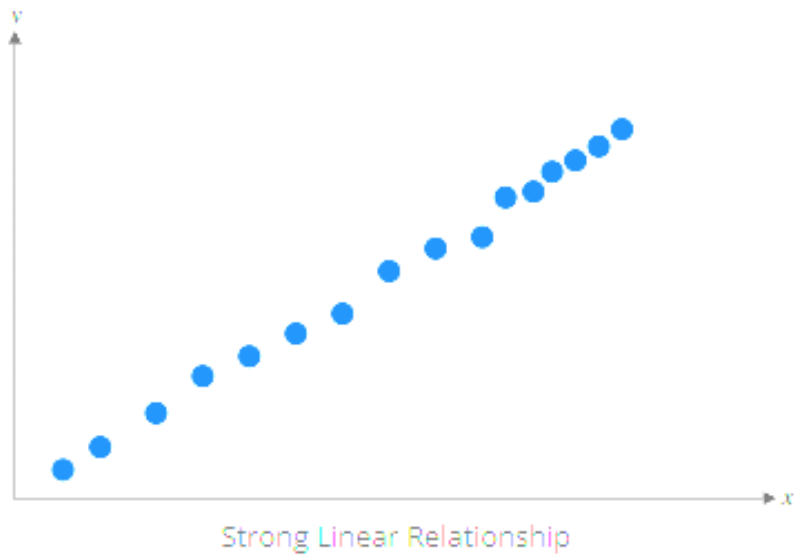
Definition 20 (Pearson Correlation Coefficient)

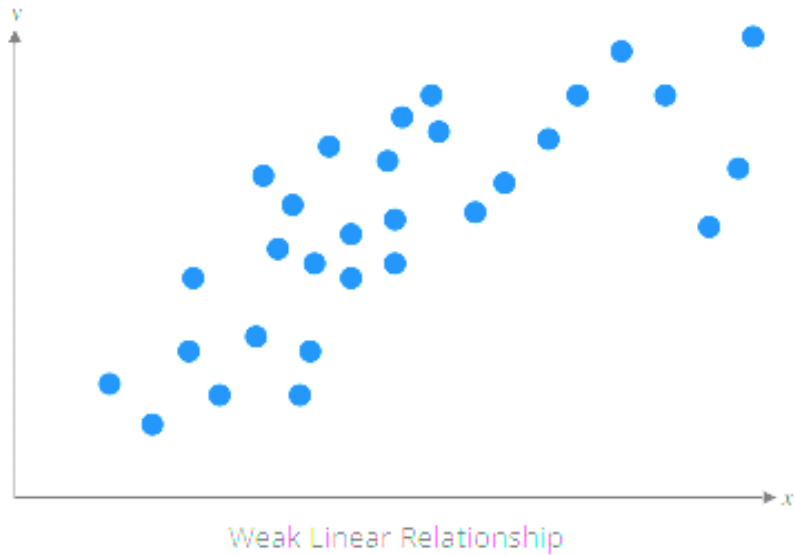
The Pearson correlation coefficient, ρ , is the parameter that measures the strength of a linear relationship between two quantitative variables in a population. The correlation coefficient for a sample is denoted by r . It always takes a value between -1 and 1 , inclusive.

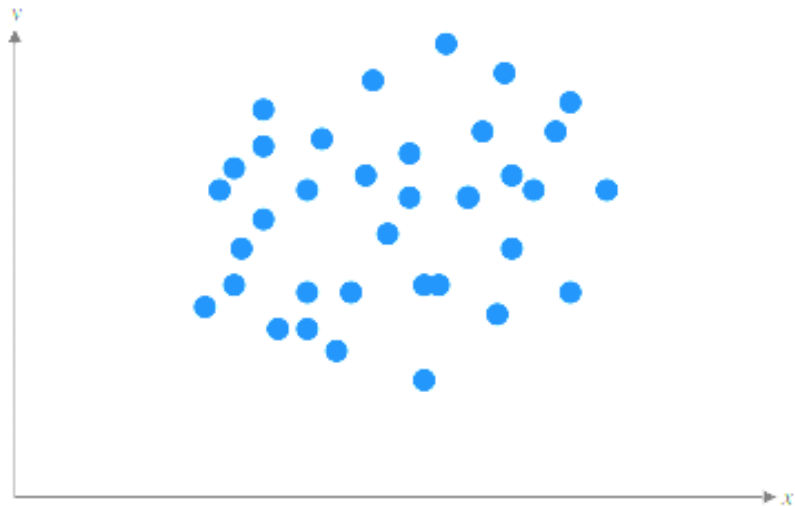
$$-1 \leq r \leq 1$$

Rule 1 (Rounding correlation coefficient)

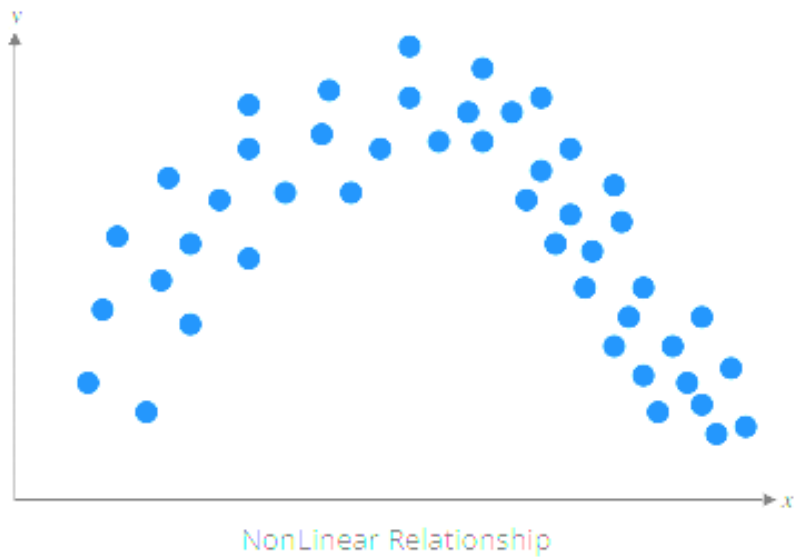
Round the correlation coefficient, r , to three decimal places.



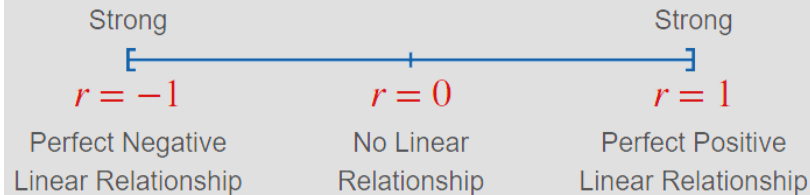




No Linear Relationship



Pearson Correlation Coefficient



Formula 1 (Pearson Correlation Coefficient)

The Pearson correlation coefficient for paired data from a sample is given by

$$r = \frac{n \sum x_i y_i - (\sum x_i)(\sum y_i)}{\sqrt{n \sum x_i^2 - (\sum x_i)^2} \sqrt{n \sum y_i^2 - (\sum y_i)^2}}$$

where n is the number of data pairs in the sample, x_i is the i^{th} value of the explanatory variable, and y_i is the i^{th} value of the response variable.

Good news everyone!



There is no reason to torture either of us with using that formula.
Instead use link: [Correlation Coefficient Calculator](#)

Test 1 (Determining Statistical Significance for the Pearson Correlation Coefficient)

A sample correlation coefficient, r , is statistically significant if

$$|r| \geq r_{\alpha}.$$

Example I

Consider the relationship between the number of bids an item on eBay received and the item's selling price. The following is a sample of 5 items sold through an auction.

Price in Dollars	110	120	150	160	190
Number of Bids	11	12	14	16	19

Calculate the correlation coefficient, r . Round your answer to three decimal places.

Example II

A study of bone density on 5 random women at a hospital produced the following results.

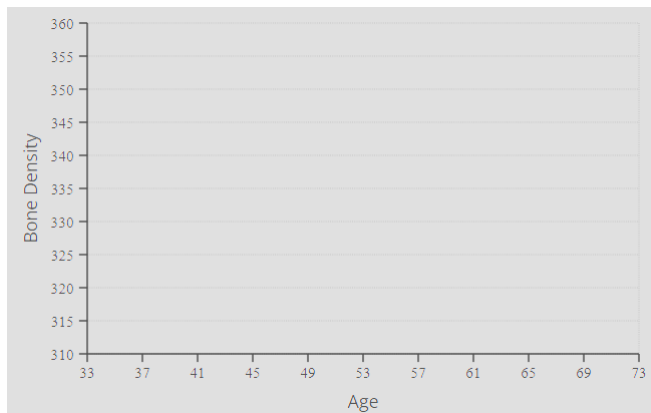
Age	41	45	57	69	73
Bone Density	350	345	330	325	315

1. Draw a scatter plot of the given data.
2. Estimate the correlation in words:

positive	negative	no correlation
----------	----------	----------------
3. Calculate the correlation coefficient, r . Round your answer to three decimal places.

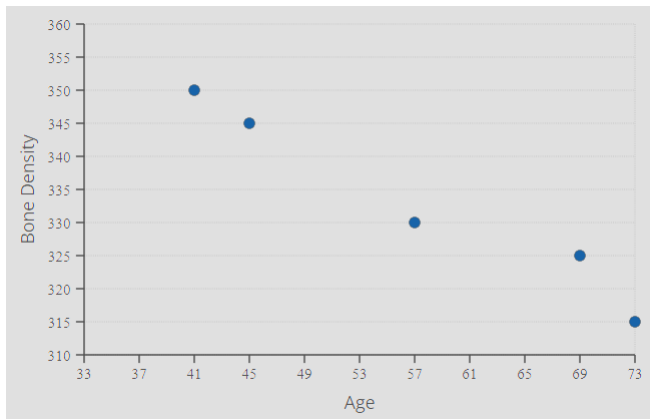
Example II (1)

Age	41	45	57	69	73
Bone Density	350	345	330	325	315



Example II (1)

Age	41	45	57	69	73
Bone Density	350	345	330	325	315



Example II (2)-(3)

Age	41	45	57	69	73
Bone Density	350	345	330	325	315

2. Estimate the correlation in words:

positive	negative	no correlation
----------	----------	----------------

3. Calculate the correlation coefficient, r . Round your answer to three decimal places.

Example III

The following data gives the number of hours 7 students spent studying and their corresponding grades on their midterm exams.

Hours Spent Studying	0	0.5	1	2	2.5	3.5	4.5
Midterm Grades	63	69	81	84	87	90	93

1. Calculate the correlation coefficient, r . Round your answer to six decimal places.
2. Determine if r is statistically significant at the 0.05 level.
3. Calculate the coefficient of determination, r^2 . Round your answer to three decimal places.

Example IIIb

2. Determine if r is statistically significant at the 0.05 level.

See the table:

Critical Values of the Pearson Correlation Coefficient (Two Tailed)

n	$\alpha = 0.05$	$\alpha = 0.01$
4	0.950	0.990
5	0.878	0.959
6	0.811	0.917
7	0.754	0.875
8	0.707	0.834
\vdots	\vdots	\vdots

Example IIIc

Definition 21 (Coefficient of Determination)

The coefficient of determination, r^2 , is a measure of the proportion of the variation in the response variable (y) that can be associated with the variation in the explanatory variable (x).

Rule 2 (Rounding coefficient of determination)

Round the coefficient of determination, r^2 , to three decimal places, just as we do for the correlation coefficient.

Example IIId

The following data gives the number of hours 7 students spent studying and their corresponding grades on their midterm exams.

Hours Spent Studying	0	0.5	1	2	2.5	3.5	4.5
Midterm Grades	63	69	81	84	87	90	93

1. Calculate the correlation coefficient, r . Round your answer to six decimal places.
2. Determine if r is statistically significant at the 0.05 level.
3. Calculate the coefficient of determination, r^2 . Round your answer to three decimal places.

Example IV

A study of bone density on 5 random women at a hospital produced the following results.

Age	33	37	45	53	61
Bone Density	345	335	325	320	315

1. Calculate the correlation coefficient, r . Round your answer to six decimal places.
2. Determine if r is statistically significant at the 0.05 level.
3. Calculate the coefficient of determination, r^2 . Round your answer to three decimal places.

Chapter 12 Section 2

More to come... this is just what I have so far.

Slope

$$m = \frac{\Delta y}{\Delta x} = \frac{y_2 - y_1}{x_2 - x_1}$$

Definition 22 (Least-Squares Regression Line)

The least-squares regression line is the line for which the average variation from the data is the smallest, also called *the line of best fit*, given by

$$\hat{y} = b_0 + b_1x$$

Linear regression should not be used to model the relationship between two variables if r is not statistically significant.

Formulas for slope:

$$b_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)(\sum x_i)}$$

and y -intercept:

$$b_0 = \frac{\sum y_i}{n} - b_1 \frac{\sum x_i}{n}$$

Bad news everyone!



There is no reason to torture either of us with using those formulas.
Instead we will use this link again: [Correlation Coefficient Calculator](#)

A prediction should **not** be made with a regression model if...

1. The data do not fall in a linear pattern when graphed on a scatter plot.
2. The correlation coefficient is not statistically significant.
3. You wish to make a prediction about a value outside the range of the sample data.
4. The population is different than that from which the sample data were drawn.

The table below gives the list price and the number of bids received for five randomly selected items sold through online auctions. Using this data, consider the equation of the regression line, $\hat{y} = b_0 + b_1x$, for predicting the number of bids an item will receive based on the list price. Keep in mind, the correlation coefficient may or may not be statistically significant for the data given. Remember, in practice, it would not be appropriate to use the regression line to make a prediction if the correlation coefficient is not statistically significant.

Price in Dollars	25	27	28	30	45
Number of Bids	1	4	6	8	9

Price in Dollars	25	27	28	30	45
Number of Bids	1	4	6	8	9

- Step 1 of 6: Find the estimated slope. Round your answer to three decimal places.
- Step 2 of 6: Find the estimated y -intercept. Round your answer to three decimal places.
- Step 3 of 6: Find the estimated value of y when $x = 30$. Round your answer to three decimal places.
- Step 4 of 6: Determine the value of the dependent variable \hat{y} at $x = 0$.
- ▶ b_0
 - ▶ b_1
 - ▶ x
 - ▶ y
- Step 5 of 6: Substitute the values you found in steps 1 and 2 into the equation for the regression line to find the estimated linear model. According to this model, if the value of the independent variable is increased by one unit, then find the change in the dependent variable \hat{y} .
- Step 6 of 6: Find the value of the coefficient of determination. Round your answer to three decimal places.

The table below gives the number of hours five randomly selected students spent studying and their corresponding midterm exam grades. Using this data, consider the equation of the regression line, $\hat{y} = b_0 + b_1x$, for predicting the midterm exam grade that a student will earn based on the number of hours spent studying. Keep in mind, the correlation coefficient may or may not be statistically significant for the data given. Remember, in practice, it would not be appropriate to use the regression line to make a prediction if the correlation coefficient is not statistically significant.

Hours Studying	1	2	4	5	6
Midterm Grades	66	68	75	80	95

Hours Studying	1	2	4	5	6
Midterm Grades	66	68	75	80	95

- Step 1 of 6:** Find the estimated slope. Round your answer to three decimal places.
- Step 2 of 6:** Find the estimated y -intercept. Round your answer to three decimal places.
- Step 3 of 6:** Determine if the statement "Not all points predicted by the linear model fall on the same line" is true or false.
- ▶ True
 - ▶ False
- Step 4 of 6:** Determine the value of the dependent variable \hat{y} at $x = 0$.
- ▶ b_0
 - ▶ b_1
 - ▶ x
 - ▶ y
- Step 5 of 6:** Find the estimated value of y when $x = 5$. Round your answer to three decimal places.
- Step 6 of 6:** Find the value of the coefficient of determination. Round your answer to three decimal places.