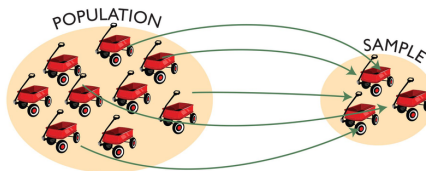


Chapter 01 : What is Statistics ?

Feras Awad

- **Data:** The information coming from observations, counts, measurements, and responses.
- **Statistics:** The science of collecting, organizing, analyzing, and interpreting data in order to make decisions.
- **Population:** the collection of all outcomes, responses, measurements, or counts that are of interest.
- **Sample:** Subset or part from the population.



- **Parameter:** Numerical description of population characteristic. For example, average age of all people in Jordan.
- **Statistic:** Numerical description of sample characteristic. For example, average age of people from three cities in Jordan.
- **Branches of Statistics:**
 1. **Descriptive Statistics:** Involves organizing, summarizing, and displaying data by tables, graphs and averages.
 2. **Inferential Statistics:** Involves using sample data to draw conclusions about population.
- **Types of Data:**
 1. **Qualitative:** Consists of all attributes, labels, or nonnumerical entries. For example: major, place of birth, eye color, ID number. Note that the mathematical computations between data are not meaningful.
 - ★ There are two types of quantitative data:
 - (a) **Nominal:** No order. For example: blood type, hair color.
 - (b) **Ordinal:** Can be ordered. For example: best five football team in the world.
 2. **Quantitative:** Numerical measurements or counts. For example: ages, temperature, weights, study-year.
 - ★ There are two types of quantitative data:
 - (a) **Discrete:** the data that takes certain values like study-year: 1, 1, 5, 4, 3, 2, 3, 1, 4, 2, 4, ...
 - (b) **Continuous:** the data that takes any value in a range or interval like weights: 63, 41, 56, 58, 49, 81, 104, 99, ...

Chapter 02 : Data and Data Organizing

Feras Awad

1 Raw Data

Raw Data is the data that recorded in the sequence in which they are collected and before they are processed or ranked.

Example 1. The following is a sample of the final grade of 20 students.

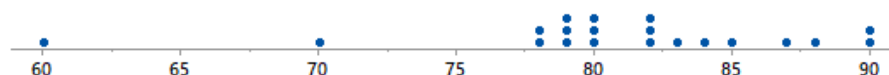
78 82 79 79 82 80 80 82 85 70
80 88 83 79 87 78 60 90 84 90

Note We can order data either in ascending order (from smallest to largest) or in descending order (from largest to smallest).

Example 2. The data in example (1) can be ordered in ascending order as follows.

60 70 78 78 79 79 79 80 80 80
82 82 82 83 84 85 87 88 90 90

Dot Diagram It is a graph of the values on the number line, where each value is represented by a dot. For example, the dot plot for the data in example (1) is as follows.



Stem-Leaf Plot Each number is separated into a leaf (the right most digit) and a stem (the remaining left most digits).

Example 3. Draw a stem-leaf plot for the data in example (1).

1. Number of leaves = number of data.

2. Since the minimum data is 60 and the maximum data is 90 then the stem values begin by 6 and end by 9.

3. The leaves are in ascending order.

4. Writing the key of the plot is so important.

Stem	Leaf
6	0
7	088999
8	00022234578
9	00

Key: 6|7=67

Example 4. Use a stem-leaf plot to display the following data.

0.8 1.5 1.6 1.8 2.1 2.3 2.4 2.5 3.0 3.9 4.0 4.0

Stem	Leaf
0	8
1	568
2	1345
3	09
4	00

Key: 2|3=2.3

2 Frequency Distributions and their Graphs

Frequency Distribution is a table that shows classes or intervals of data entries with a count of the number of entries in each class called frequencies.

Example 5.

Major (Class)	Frequency F	* Relative Frequency	** Cumulative Frequency
Math	17	$17/50 = 0.34 = 34\%$	17
Genetics	23	$23/50 = 0.46 = 46\%$	$23 + 17 = 40$
IT	10	$10/50 = 0.20 = 20\%$	$10 + 40 = 50$
Total	50	$1 = 1.00 = 100\%$	

* Relative frequency of a class = $\frac{\text{Class frequency}}{\text{Sample Size}} = \frac{F}{n}$.

** Cumulative frequency of a class = the sum of frequencies of that class and all previous classes.

- The sum of all frequencies = the sample size = the cumulative frequency of the last class.

Example 6. [Qualitative Data]

The following is the blood type of 40 students. Summarize the data in frequency distribution.

O B A O O O A O A B
 AB A B O A A A O B A
 O A A O A O A A A O
 O O A A O O O A A AB

Blood Type (Class)	Frequency F	Relative Frequency	Cumulative Frequency
O	16	0.40	16
A	18	0.45	34
B	4	0.10	38
AB	2	0.05	40
Total	40	1	

Homework 1. Fill-in the blanks for the given frequency distribution.

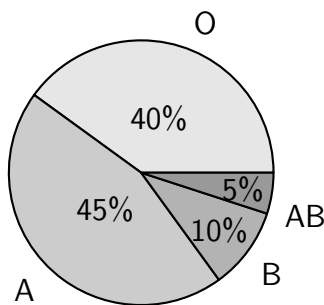
Class	F	Rel. F	Cum. F
A			32
B		0.27	
C			
D	21		100

Note Two common methods for graphically displaying qualitative or discrete data: Pie and Bar charts.

- **Pie Chart:** is a circle that is divided into sectors that represent classes.
 - The area of each sector is proportional to the frequency (or relative frequency) of each class.
 - Find the central angle θ for each sector using the corresponding frequency or relative frequency of that class, where

$$\theta = \frac{\text{Frequency}}{\text{Sample Size}} \times 360^\circ = \text{Relative Frequency} \times 360^\circ$$

Example 7. Draw a Pie chart for the data in example (6).



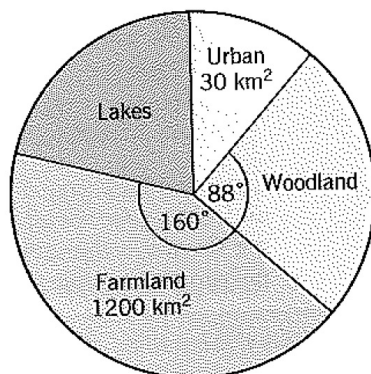
$$\theta_O = 0.40 \times 360^\circ = 144^\circ$$

$$\theta_A = 0.45 \times 360^\circ = 162^\circ$$

$$\theta_B = 0.10 \times 360^\circ = 36^\circ$$

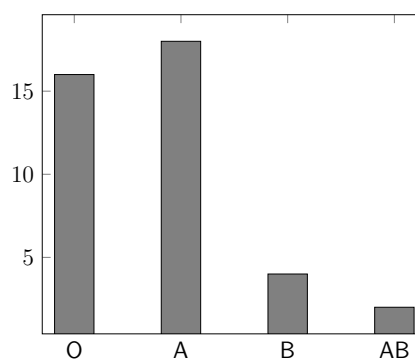
$$\theta_{AB} = 0.05 \times 360^\circ = 18^\circ$$

Homework 2. The given pie chart, shows the distribution of various land and water in a certain country. Calculate:



1. the area of woodland,
2. the angle of the urban sector,
3. the area and the angle of the lakes,
4. the total area of the country.

- **Bar Chart:** displays the distinct data classes on a horizontal axis and either the frequencies or relative frequencies of those classes on a vertical axis. For example, the bar chart of the data in example (6) is

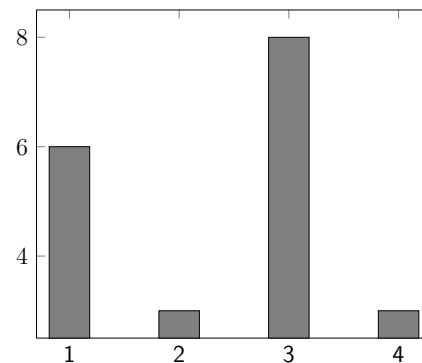


Example 8. [Discrete Data]

Summarize the following data using a frequency distribution, then draw bar chart for these data.

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

Value (Class)	Frequency F	Relative Frequency	Cumulative Frequency
1	6	0.30	6
2	3	0.15	9
3	8	0.40	17
4	3	0.15	20
Total	20	1	



3 Frequency Distribution of Continuous Data

The Goal is to group the observations according to intervals and recording the frequencies of the intervals.

Example 9. The table below shows regular frequency distribution of continuous data. There are six classes, and the frequencies for each of the six classes are 5, 8, 6, 8, 5, and 4.

Class	Frequency
1–5	5
6–10	8
11–15	6
16–20	8
21–25	5
26–30	4

Notes

- The classes do not **overlap**.
- Each class has a
 - **lower class limit**, which is the *least* number that can belong to the class,
 - an **upper class limit**, which is the *greatest* number that can belong to the class.

For example, in the frequency distribution in example (9), the lower class limits are 1, 6, 11, 16, 21, and 26, while the upper class limits are 5, 10, 15, 20, 25, and 30.

- **Class Length** = Upper Limit of a Class – Lower Limit of that Class.
- **Class Width** = distance between lower (or upper) limits of consecutive classes.
For instance, the class width in the frequency distribution shown is $6 - 1 = 5$.
- **Class Length** = **Class Width** – 1.
- **Range** = **Maximum Data** – **Minimum Data**.

To Construct a Frequency Distribution for Continuous Data:

1. Decide on the number of classes to include in the frequency distribution. The number of classes should be between 5 and 20; otherwise, it may be difficult to detect any patterns.
2. Find the class width using the formula

$$\text{Class Width} = \left\lceil \frac{\text{Range}}{\text{Number of Classes}} \right\rceil$$

where $\lceil \cdot \rceil$ is the ceiling function. For example, $\lceil 2.8 \rceil = 3$, $\lceil 1.5 \rceil = 2$, and $\lceil 6.1 \rceil = 7$.

3. Find the class limits.
 - (a) You can use the minimum data entry as the lower limit of the first class.
 - (b) To find the remaining lower limits, add the class width to the lower limit of the preceding class.
 - (c) Then find the upper limit of the first class.
 - (d) Remember that classes cannot overlap.
4. Find the frequency F for each class.

Example 10. [Quantitative Continuous Data]

For the following set of data, construct a frequency distribution that has seven classes.

128 100 180 150 200 90 340 105 85 270
200 65 230 150 150 120 130 80 230 200
110 126 170 132 140 112 90 340 170 190

1. The number of classes = 7 which is stated in the problem.
2. The minimum data entry is 65 and the maximum data entry is 340, so the range equals $340 - 65 = 275$.
3.
$$\text{Class Width} = \left\lceil \frac{\text{Range}}{\text{Number of Classes}} \right\rceil = \left\lceil \frac{275}{7} \right\rceil \approx \lceil 39.286 \rceil = 40.$$
4. The minimum data entry is a convenient lower limit for the first class. To find the lower limits of the remaining six classes, add the class width of 40 to the lower limit of each previous class. So, the lower limits of the other classes are $65 + 40 = 105$, $105 + 40 = 145$, and so on. The upper limit of the first class is 104, which is one less than the lower limit of the second class. The upper limits of the other classes are $104 + 40 = 144$, $144 + 40 = 184$, and so on.
5. Make a tally mark for each data entry in the appropriate class. For instance, the data entry 128 is in the 105 – 144 class, so make a tally mark in that class. Continue until you have made a tally mark for each of the 30 data entries.
6. The number of tally marks for a class is the frequency of that class.

Class	Frequency
65 – 104	6
105 – 144	9
145 – 184	6
185 – 224	4
225 – 264	2
265 – 304	1
305 – 344	2

The **midpoint** of a class is the sum of the lower and upper limits of the class divided by two. The midpoint is sometimes called the **class mark**.

$$\text{midpoint} = \frac{(\text{Lower class limit}) + (\text{Upper class limit})}{2}.$$

You can use the formula shown above to find the midpoint of each class, or after finding the first midpoint, you can find the remaining midpoints by adding the class width to the previous midpoint. For instance, the midpoint of the first class in Example (10) is

$$\text{midpoint} = \frac{65 + 104}{2} = 84.5.$$

Using the class width of 40, the remaining midpoints are

$$84.5 + 40 = 124.5$$

$$124.5 + 40 = 164.5$$

$$164.5 + 40 = 204.5$$

$$204.5 + 40 = 244.5$$

and so on. The midpoints, relative frequencies, and cumulative frequencies for the data in Example (10) are shown in the expanded frequency distribution below.

Class	Frequency	Relative F	Cumulative F	Midpoint
65 – 104	6	0.200	6	84.5
105 – 144	9	0.300	15	124.5
145 – 184	6	0.200	21	164.5
185 – 224	4	0.133	25	204.5
225 – 264	2	0.067	27	244.5
265 – 304	1	0.033	28	284.5
305 – 344	2	0.067	30	324.5

Homework 3. Construct a frequency distribution for the following data set using the 5 classes. In the table, include the midpoints, relative frequencies, and cumulative frequencies. Which class has the greatest frequency and which has the least frequency?

7 39 13 9 25 8 22 0 2 18 2 30 7
35 12 15 8 6 5 29 0 11 39 16 15

Homework 4. What Would You Do? You work at a bank and are asked to recommend the amount of cash to put in an ATM each day. You don't want to put in too much (security) or too little (customer irritation). Here are the daily withdrawals (in hundreds of dollars) for 30 days.

72 84 61 76 104 76 86 92 80 88 98 76 97 82 84
67 70 81 82 89 74 73 86 81 85 78 82 80 91 83

1. Construct a frequency distribution for the data. Use 8 classes.
2. If you put \$9000 in the ATM each day, what percent of the days in a month should you expect to run out of cash? Explain.
3. If you are willing to run out of cash on 10% of the days, how much cash should you put in the ATM each day? Explain.

Histogram Sometimes it is easier to discover patterns of a data set by looking at a graph of the frequency distribution. One such graph is a **frequency histogram**. So, a **frequency histogram** is a bar graph that represents the frequency distribution of a data set. A histogram has the following properties.

1. The horizontal scale is quantitative and measures the data entries.
2. The vertical scale measures the frequencies of the classes.
3. Consecutive bars must touch.

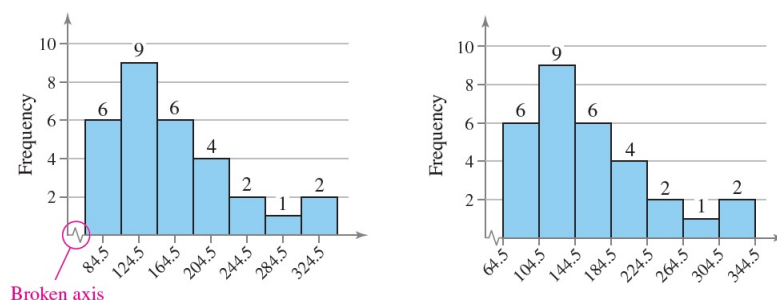
Because consecutive bars of a histogram must touch, bars must begin and end at class boundaries instead of class limits. **Class boundaries** are the numbers that separate classes without forming gaps between them. For data that are integers, subtract 0.5 from each lower limit to find the lower class boundaries. To find the upper class boundaries, add 0.5 to each upper limit. The upper boundary of a class will equal the lower boundary of the next higher class.

Example 11. Draw a frequency histogram for the frequency distribution in Example (10). First, find the class boundaries. Because the data entries are integers, subtract 0.5 from each lower limit to find the lower class boundaries and add 0.5 to each upper limit to find the upper class boundaries. So, the lower and upper boundaries of the first class are as follows.

$$\text{First class lower boundary} = 65 - 0.5 = 64.5$$

$$\text{First class upper boundary} = 104 + 0.5 = 104.5$$

The boundaries of the remaining classes are found in the same manner. To construct the histogram, choose possible frequency values for the vertical scale. You can mark the horizontal scale either at the midpoints or at the class boundaries. Both histograms are shown.



Homework 5. True or False ?

1. In a frequency distribution, the class width is the distance between the lower and upper limits of a class.
2. The midpoint of a class is the sum of its lower and upper limits divided by two.
3. Class boundaries ensure that consecutive bars of a histogram touch.

Homework 6. Use the frequency distribution shown below to find the class width, class midpoints, and class boundaries.

Class	Frequency
20–30	19
31–41	43
42–52	68
53–63	69
64–74	74
75–85	68
86–96	24

4 The Mean (Average)

The mean of raw data: Consider a sample x_1, x_2, \dots, x_n of data of size n . The sample mean is

$$\bar{x} = \frac{\text{The sum of data}}{\text{Sample Size}} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{\sum x}{n}$$

Example 12. Find the mean of the data 26, 20, 40, 24, 30.

$$\bar{x} = \frac{26 + 20 + 40 + 24 + 30}{5} = \frac{140}{5} = 28.$$

Note \bar{x} can be negative, positive, or zero. Also, \bar{x} cannot be used for qualitative data.

Homework 7. If the mean of the data 13, 8, 19, a , 1, 14 is $\bar{x} = 10$, find the value of a ?

Homework 8. Find the mean of the following data.

Stem	Leaf
1	0 5
2	3 4 6 9
3	8 8

Key: 2|4=24

Example 13. The mean of 10 numbers is 8. If an eleventh number is now added to the data, the mean becomes 9. What is the value of the new number?

Denote the new number by b . Note that $b = (\sum x)_{\text{new}} - (\sum x)_{\text{old}}$. But

$$\begin{aligned}\bar{x}_{\text{old}} &= \frac{(\sum x)_{\text{old}}}{n_{\text{old}}} \Rightarrow 8 = \frac{(\sum x)_{\text{old}}}{10} \Rightarrow (\sum x)_{\text{old}} = 8 \times 10 = 80 \\ \bar{x}_{\text{new}} &= \frac{(\sum x)_{\text{new}}}{n_{\text{new}}} \Rightarrow 9 = \frac{(\sum x)_{\text{new}}}{11} \Rightarrow (\sum x)_{\text{new}} = 9 \times 11 = 99 \\ \therefore b &= (\sum x)_{\text{new}} - (\sum x)_{\text{old}} = 99 - 80 = 19\end{aligned}$$

Homework 9. The mean of n numbers is 5. If the number 12 is now removed from the n numbers, the mean is 4. Find the value of n .

Example 14. The mean of 4 numbers is 5, and the mean of 3 other numbers is 12. What is the mean of the 7 numbers together?

$$\bar{x} = \frac{\sum x}{n} = \frac{(\sum x)_1 + (\sum x)_2}{n_1 + n_2} = \frac{n_1 \bar{x}_1 + n_2 \bar{x}_2}{n_1 + n_2} = \frac{(5)(4) + (3)(12)}{4 + 3} = \frac{56}{7} = 8.$$

The mean of discrete frequency distribution:

$$\bar{x} = \frac{\sum (\text{Class Value} \times \text{Class Freq.})}{\sum \text{Frequencies}} = \frac{\sum (x \cdot F)}{\sum F}$$

Example 15. Find the mean of the following distribution.

Value	Frequency	$x \cdot F$
1	2	2
2	5	10
4	7	28
6	6	36
Total	$\sum F = 20$	$\sum (x \cdot F) = 76$

$$\therefore \bar{x} = 76/20 = 3.8$$

Homework 10. For each of the following distributions, find the value of b for the given value of \bar{x} .

(a)

Value	Frequency
-2	5
0	2
1	b
3	9

$$\bar{x} = 1.05$$

(b)

Value	Frequency
2	2
b	5
8	2
9	1

$$\bar{x} = 5.4$$

The mean of continuous frequency distribution:

$$\bar{x} = \frac{\sum (\text{Class Midpoint} \times \text{Class Freq.})}{\sum \text{Frequencies}} = \frac{\sum (m \cdot F)}{\sum F}$$

Example 16. Find the mean of the continuous distribution of the data in example (10).

Class	Frequency	Midpoint	$m \cdot F$
65 – 104	6	84.5	507
105 – 144	9	124.5	1120.5
145 – 184	6	164.5	987
185 – 224	4	204.5	818
225 – 264	2	244.5	489
265 – 304	1	284.5	284.5
305 – 344	2	324.5	649
Total	30		4855

$$\therefore \bar{x} = 4855/30 = 971/6 \approx 161.83$$

Note that, the actual mean of the original data is $\bar{x} = \frac{4783}{30} \approx 159.43$. The difference occurs between the two means because we replace the actual data inside each class by the class midpoint.

Homework 11. Find the mean of the following continuous distribution.

Class	Frequency
1–5	9
6–10	11
11–15	16
16–20	14

5 The Mode and the Median

Mode The mode is the value that occurs with the highest frequency in a data set.

Example 17. 1. The data 77, 82, 74, 81, 79, 84, 74, 78 has mode = 74.

2. The data 18, 19, 27, 22, 29, 19, 25, 21, 22, 30 has two modes: 19 and 22.

3. The following data has no mode: 5, 9, 2, 4, 3, 8.

4. The mode of the following distribution is the category "Genetics".

Major	Frequency
Math	17
Genetics	23
IT	10

5. The mode of the following distribution is the value $x = 4$.

Value	Frequency
1	2
2	5
4	7
6	6

Median The median is the value of the middle term in a data set that has been ranked in ascending (or descending) order. To evaluate the median of a data of size n :

- Sort the data in ascending order.
- If the number of observations n is **odd**, then the median is the value of the observation numbered $\frac{n+1}{2}$.
 - If the number of observations n is **even**, then the median is the mean (average) of the observations numbered $\frac{n}{2}$ and $\frac{n}{2} + 1$.

Example 18. Find the median of each of the following data sets.

(1) 312, 257, 421, 289, 526, 374, 497.

- Sort the data in ascending order: 257, 289, 312, 374, 421, 497, 526
- Since $n = 7$ is odd, then the median is the data in position $\frac{n+1}{2} = \frac{7+1}{2} = 4$. So, median $= x_4 = 374$.

(2) 7, 8, 9, 10, 11, 12, 13, 13, 14, 17, 17, 45.

- The data is already sorted in ascending order.
- Since $n = 12$ is even, then the median is the mean of the data in positions $\frac{n}{2} = \frac{12}{2} = 6$ and $\frac{n}{2} + 1 = \frac{12}{2} + 1 = 7$. So, median $= \frac{x_6 + x_7}{2} = \frac{12 + 13}{2} = 12.5$.

(3)

Value	Frequency
1	2
2	5
4	7
6	6

- The data is already sorted in ascending order.
- Since $n = \sum F = 20$ is even, then the median is the mean of the data in positions $\frac{n}{2} = \frac{20}{2} = 10$ and $\frac{n}{2} + 1 = \frac{20}{2} + 1 = 11$. So, median $= \frac{x_{10} + x_{11}}{2} = \frac{4 + 4}{2} = 4$.

Homework 12. Find the mode and the median of the following data.

Stem	Leaf
1	0 5
2	3 4 6 9
3	8 8

Key: 2|4=24

Homework 13. If the median of the data 9, 1, x , 12 is 7, what is the value of x ?

Note Both median and mode can be positive, negative, or zero.

6 The Sample Variance and Standard Deviation

- The standard deviation (and the variance) is the most-used measure of **dispersion**.
- The value of the standard deviation tells how closely the values of a data set are **clustered around the mean**.
- Lower value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively smaller range around the mean. Larger value of the standard deviation for a data set indicates that the values of that data set are spread over a relatively larger range around the mean.
- The sample variance is **denoted by** s^2 and given by the formula

$$s^2 = \frac{\sum (x - \bar{x})^2}{n - 1} = \frac{\sum (x^2) - \frac{(\sum x)^2}{n}}{n - 1}.$$

- The quantity $x - \bar{x}$ in the above formulas is called **the deviation of the x value from the mean**. The sum of the deviations of the x values from the mean is always zero; that is, $\sum (x - \bar{x}) = 0$.
- Standard Deviation = $\sqrt{\text{Variance}} = s$.
- Both the variance and the standard deviation **cannot be negative**.

Example 19. Find the variance and standard deviation of each of the following samples.

(1) 3, 3, 3, 3, 3.

Without doing any calculations, $s^2 = s = 0$.

(2) 7, 3, 9, 15.

First of all, $\bar{x} = \frac{7 + 3 + 9 + 15}{4} = \frac{34}{4} = 8.5$.

x	$x - \bar{x}$	$(x - \bar{x})^2$
7	-1.5	2.25
3	-5.5	30.25
9	0.5	0.25
15	6.5	42.25
Total	0.0	75

$$s^2 = 75/3 = 25 \Rightarrow s = \sqrt{25} = 5.$$

(3) 7, 12, 1, 10, 4, 8

x	x^2
7	49
12	144
1	1
10	100
4	16
8	64
Total	42 374

$$\therefore s^2 = \frac{\sum (x^2) - \frac{(\sum x)^2}{n}}{n - 1} = \frac{374 - \frac{(42)^2}{6}}{5} = 16 \Rightarrow s = \sqrt{16} = 4.$$

Homework 14. Find the variance and standard deviation of the sample: 19, 13, 25, 9.

Example 20. For a set of 10 numbers, $\sum x = 270$ and $\sum (x^2) = 9054$. Find the mean and the variance.

$$\bar{x} = \frac{\sum x}{n} = \frac{270}{10} = 27$$

$$s^2 = \frac{\sum (x^2) - \frac{(\sum x)^2}{n}}{n - 1} = \frac{9054 - \frac{(270)^2}{10}}{9} = 196.$$

Homework 15. For a set of 18 numbers, $\bar{x} = 23$ and $s = 12$. Find $\sum x$ and $\sum (x^2)$.

Example 21. The numbers $a, b, 6, 4, 7$ have mean 5 and variance 4. Find the values of a, b .

$$\bar{x} = \frac{a + b + 6 + 4 + 7}{5} = 5 \Rightarrow a + b = 8 \Rightarrow \boxed{a = 8 - b}$$

$$s^2 = \frac{(a - 5)^2 + (b - 5)^2 + (6 - 5)^2 + (4 - 5)^2 + (7 - 5)^2}{4} = 4$$

$$\Rightarrow (a - 5)^2 + (b - 5)^2 = 10$$

$$\Rightarrow (3 - b)^2 + (b - 5)^2 = 10$$

$$\Rightarrow b^2 - 8b + 12 = 0$$

$$\Rightarrow (b - 2)(b - 6) = 0$$

$$\Rightarrow \text{If } b = 2 \text{ then } a = 6, \text{ and if } b = 6 \text{ then } a = 2.$$

Homework 16. For a set of 20 numbers, $\sum x = 20$ and $\sum (x^2) = 96$. For a second set of 30 numbers, $\sum x = 60$ and $\sum (x^2) = 236$. Find the mean and standard deviation for the combined set of 50 numbers.

Note The population variance is given by the formula $\sigma^2 = \frac{\sum (x - \mu)^2}{N}$ where μ is the population mean, and N is the population size.

The Variance of Discrete Frequency Distribution

$$s^2 = \frac{\sum ((x - \bar{x})^2 \cdot F)}{(\sum F) - 1} = \frac{\sum (x^2 \cdot F) - \frac{(\sum (x \cdot F))^2}{(\sum F)}}{(\sum F) - 1}$$

Example 22. Find the variance of the following distribution.

Value (x)	Frequency (F)	$x \cdot F$	$x^2 \cdot F$
1	2	2	2
2	9	18	36
3	6	18	54
4	8	32	128
Total	25	70	220

$$\text{So, } s^2 = \frac{220 - \frac{(70)^2}{25}}{25 - 1} = \frac{220 - 196}{24} = 1.$$

Homework 17. Find the standard deviation of the following distribution.

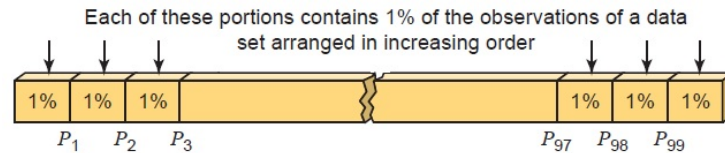
Value (x)	Frequency (F)
3	4
4	5
5	8
6	3

7 Percentiles and Quartiles

Percentiles The k th percentile P_k , is a value in a data set such that about $k\%$ of the measurements are smaller than the value of P_k and about $(100 - k)\%$ of the measurements are greater than the value of P_k .

Notes

1. The data should be sorted in increasing (ascending) order to compute percentiles.
2. Each sorted data set has 99 percentiles that divide it into 100 equal parts.



3. **median** = P_{50} .

Calculating Percentiles The (approximate) value of the k th percentile, denoted by P_k , for a ranked data in increasing order, is

$$P_k = \begin{cases} \text{The data in location } \left\lceil \frac{kn}{100} \right\rceil & \text{if } \frac{kn}{100} \text{ non-integer, where } \lceil \cdot \rceil \text{ is the ceiling function.} \\ \text{The average of the data in locations } \frac{kn}{100} \text{ and } \frac{kn}{100} + 1 & \text{if } \frac{kn}{100} \text{ is integer.} \end{cases}$$

Example 23. For the data 2, 5, 16, 21, 21, 22, 24, 26, 27, 28, 29, 32, 33, 35, 35, 37, 38, 39, 45, 50, find

1. the 36th percentile,

- $\frac{kn}{100} = \frac{36 \times 20}{100} = \frac{720}{100} = 7.2$ which is not integer.
- The location is $\lceil 7.2 \rceil = 8$.
- So, $P_{36} = 26$.

2. the 85th percentile.

- $\frac{kn}{100} = \frac{85 \times 20}{100} = \frac{1700}{100} = 17$ which is integer.
- The locations are 17 and 18.
- So, $P_{85} = \frac{38 + 39}{2} = 38.5$.

Homework 18. Find the 53rd and 70th percentiles of the following data.

Stem	Leaf
1	0 5 6
2	3 4 6 9
3	8 8 9

Key: 2|4=24

Finding Percentile Rank of a Value

$$\text{Percentile rank of } x = \frac{\text{Number of values less than } x}{\text{Data Size}} \times 100$$

Note In finding the percentile rank of a value, round your final answer to the nearest integer value.

Example 24. For the data 14, 21, 26, 27, 28, 32, 35, 35, 38, 45, 48, 50, 51, 51, 57, find

1. the percentile rank of $x = 48$.

$$\text{Percentile rank of } 48 = \frac{10}{15} \times 100 \approx 66.667 \text{ which is rounded to } 67$$

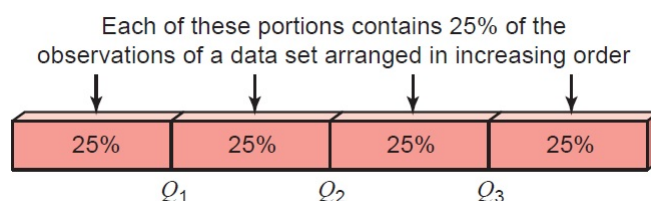
2. the percentile rank of $x = 35$.

$$\text{Percentile rank of } 35 = \frac{6}{15} \times 100 = 40.$$

Homework 19. Find the percentile rank of the number 27 for the data given in homework (18).

Quartiles are three summary measures that divide a ranked (in ascending order) data set into four equal parts. The three numbers are

1. The 1st Quartile = $Q_1 = 25\text{th Percentile}$.
2. The 2nd Quartile = $Q_2 = 50\text{th Percentile} = \text{Median}$.
3. The 3rd Quartile = $Q_3 = 75\text{th Percentile}$.



Note The difference between the third quartile and the first quartile for a data set is called the **interquartile range (IQR)**. This means 50% of the data are between Q_1 and Q_3 .

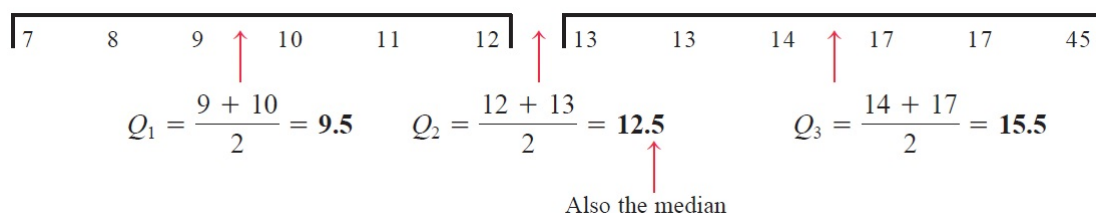
Example 25. For the data set 7, 8, 9, 10, 11, 12, 13, 13, 14, 17, 17, 45, find Q_1 , Q_2 , Q_3 and the IQR.

1. $Q_1 = P_{25} = \frac{9 + 10}{2} = 9.5$.

2. $Q_2 = \text{median} = \frac{12 + 13}{2} = 12.5$.

3. $Q_3 = P_{75} = \frac{14 + 17}{2} = 15.5$.

4. $\text{IQR} = Q_3 - Q_1 = 15.5 - 9.5 = 6$.



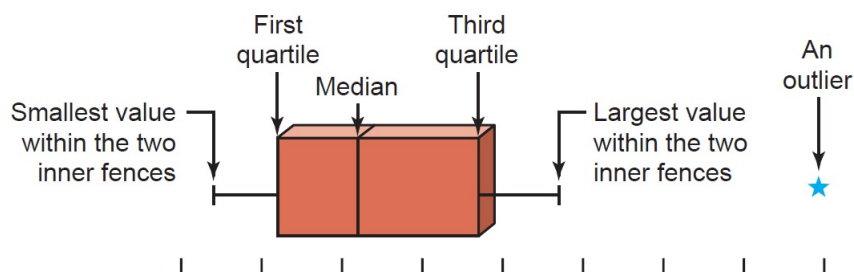
Homework 20. Find Q_1 , Q_2 , Q_3 and the IQR of the following data.

Stem	Leaf
1	0 5
2	3 4 6
3	8 9

Key: 2|4=24

Outlier is a data entry that is far removed from the other entries in the data set. One way to identify outliers is to use the interquartile range as follows:

1. Find the $IQR = Q_3 - Q_1$.
2. Find the lower and upper fences of the data using:
 - Lower Fence = $Q_1 - (1.5 \times IQR)$
 - Upper Fence = $Q_3 + (1.5 \times IQR)$
3. Any data entry x is an outlier if $x > \text{Upper Fence}$ or $x < \text{Lower Fence}$.

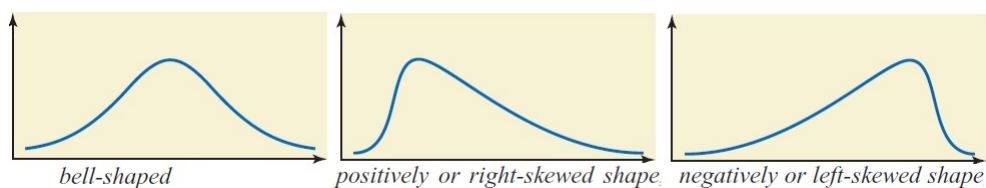


Example 26. Are there any outliers for the data set 6, 7, 8, 9, 10, 15, 16, 16, 20, 20, 23, 33, 50, 58, 104 ?

1. Since $Q_1 = P_{25} = 9$ and $Q_3 = P_{75} = 33$, then $IQR = Q_3 - Q_1 = 33 - 9 = 24$.
2. The lower fence = $Q_1 - 1.5 \times IQR = 9 - 1.5 \times 24 = -27$.
3. The upper fence = $Q_3 + 1.5 \times IQR = 33 + 1.5 \times 24 = 69$.
4. Since $x = 104 > 69$ then the data entry 104 is an outlier.

Homework 21. Does the data set 22, 24, 25, 28, 31, 32, 34, 35, 36, 41, 42, 43, 47, 49, 52, 55, 58, 59, 61, 61, 63, 65, 73, 98 contain any outliers ?

Distribution Shapes When one is describing data, it is important to be able to recognize the shapes of the distribution values. A distribution can have many shapes: Bell-Shaped (or **Symmetric** or **Normal**), positively or right-skewed shape, and negatively or left-skewed shape.



- In a normal distribution, **mean = median = mode** and $Q_3 - Q_2 = Q_2 - Q_1$.
- In a negatively-skewed distribution, **mean < median < mode** and $Q_3 - Q_2 > Q_2 - Q_1$.
- In a positively-skewed distribution, **mode < median < mean** and $Q_3 - Q_2 < Q_2 - Q_1$.
- **Coefficient of Skewness** = $\frac{\text{mean} - \text{mode}}{\text{standard deviation}} \approx \frac{3(\text{mean} - \text{median})}{\text{standard deviation}} \approx \frac{Q_3 - 2Q_2 + Q_1}{Q_3 - Q_1}$.

Example 27. For a skewed distribution, the mean is 16, the median is 20 and the standard deviation is 5. Find the coefficient of skewness.

Answer: Coefficient of Skewness = $\frac{3(16 - 20)}{5} = -2.4$.

Homework 22. For a skewed distribution, the mean is 86, the mode is 78 and the variance is 16. Find the coefficient of skewness.

8 Use of Standard Deviation

To compare variation in different data sets, you can use standard deviation when the data sets use the same units of measure and have means that are about the same. **For data sets with different units of measure or different means, use the coefficient of variation.**

The Coefficient of Variation (CV) of a sample data is $CV = \frac{s}{\bar{x}} \times 100\%$.

Example 28. The mean of the number of sales of cars over a 3-month period is 87, and the standard deviation is 5. The mean of the commissions is \$5225, and the standard deviation is \$773. Compare the variations of the two.

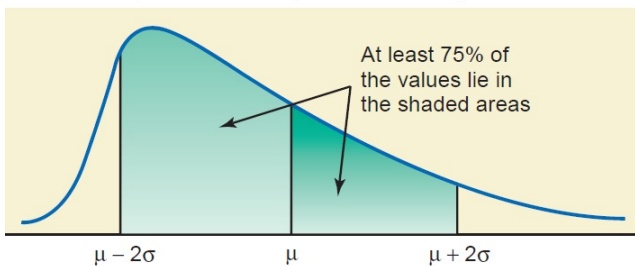
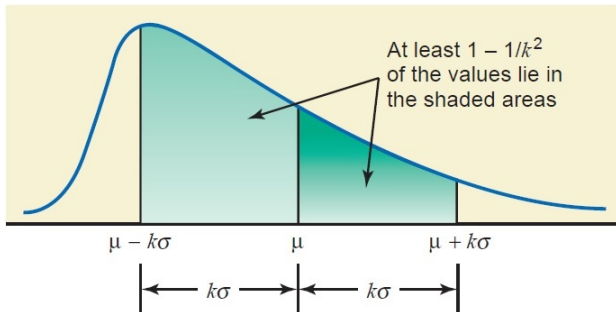
$$CV_{\text{cars}} = \frac{s}{\bar{x}} \times 100\% = \frac{5}{87} \times 100\% \approx 5.7\%$$

$$CV_{\text{commissions}} = \frac{s}{\bar{x}} \times 100\% = \frac{773}{5225} \times 100\% \approx 14.8\%$$

Since the coefficient of variation is larger for commissions, the commissions are more variable than the sales.

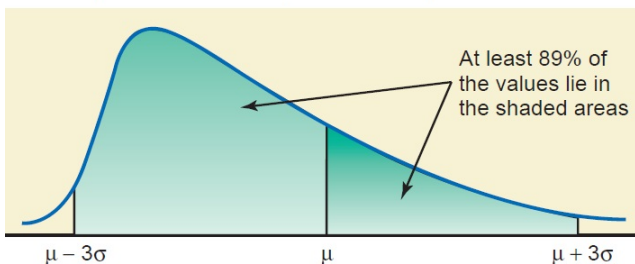
Homework 23. The yearly salaries of all employees who work for a company have a mean of \$62,350 and a standard deviation of \$6820. The years of experience for the same employees have a mean of 15 years and a standard deviation of 2 years. Is the relative variation in the salaries larger or smaller than that in years of experience for these employees?

Chebyshev's theorem The proportion of values from a data set that will fall within k standard deviations of the mean will be at least $1 - \frac{1}{k^2}$, where k is a number greater than 1 (k is not necessarily an integer).



for example, if $k = 2$, then

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(2)^2} = 1 - \frac{1}{4} \\ = 1 - .25 = .75 \text{ or } 75\%$$



If $k = 3$, then,

$$1 - \frac{1}{k^2} = 1 - \frac{1}{(3)^2} = 1 - \frac{1}{9} \\ = 1 - .11 = .89 \text{ or } 89\% \text{ approximately}$$

Example 29. The mean price of houses in a certain neighborhood is \$50000, and the standard deviation is \$10000. Find the price range for which at least 75% of the houses will sell.

Answer. Chebyshev's theorem states that 75% of the data values will fall within $k = 2$ standard deviations of the mean. Thus,

$$\bar{x} - 2s = 50000 - 2(10000) = 30000,$$

$$\bar{x} + 2s = 50000 + 2(10000) = 70000.$$

Hence, at least 75% of all homes sold in the area will have a price range from \$30000 to \$70000.

Homework 24. Jordanians spend an average of 3 hours per day on social media applications. If the standard deviation is 32 minutes, find the range in which at least 84% of the data will lie. Use Chebyshev's theorem.

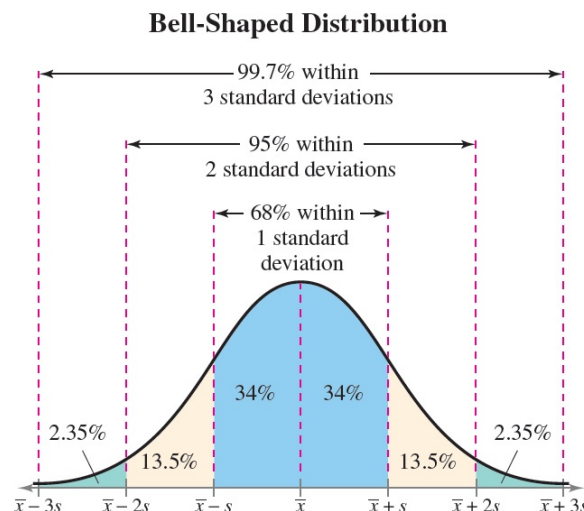
Example 30. The mean of a distribution is 80 and the standard deviation is 10. At least what percentage of the values will fall between 65 and 95 ?

Answer. Note that $65 = 80 - 10k$ and $95 = 80 + 10k$. In both cases, $k = 1.5$. Using Chebyshev's theorem, at least $1 - \frac{1}{(1.5)^2} \approx 56\%$ of values will fall between 65 and 95.

Homework 25. The mean of a distribution is 20 and the standard deviation is 2. At least what percentage of the values will fall between 12 and 28 ?

The Empirical (Normal) Rule Chebyshev's theorem applies to any distribution regardless of its shape. However, when a distribution is symmetric bell-shaped (normal), the following statements, which make up the empirical rule, are true.

- Approximately 68% of the data values will fall within 1 standard deviation of the mean.
- Approximately 95% of the data values will fall within 2 standard deviations of the mean.
- Approximately 99.7% of the data values will fall within 3 standard deviations of the mean.



Example 31. The age distribution of a sample of 5000 persons is **bell shaped** with a mean of 40 years and a standard deviation of 12 years. Determine the approximate percentage of people who are 16 to 64 years old.

Answer. We know that $\bar{x} = 40$ and $s = 12$. We find k_1 and k_2 in the following.

$$16 = \bar{x} - k_1s \Rightarrow 16 = 40 - 12k_1 \Rightarrow k_1 = 2$$

$$64 = \bar{x} + k_2s \Rightarrow 64 = 40 + 12k_2 \Rightarrow k_2 = 2.$$

Since the distance of the two points, 16 and 64, and the mean $\bar{x} = 40$ is each equal to $2s = 24$, then the interval $[16, 64]$ contains approximately $95\% \times 5000 = 4750$ observations.

Homework 26. Consider a collection of 500 observations with mean 60 and standard deviation 2. Find:

1. an interval centered at 60 which contains at least 450 observations,
2. the number of observations in the interval $[56, 62]$ assuming the distribution of the observations is bell-shaped (normal).

Standard z -Scores

- There is an old saying, “You can’t compare apples and oranges.” But with the use of statistics, it can be done to some extent.
- Suppose that a student scored 90 on a Statistics exam and 45 on an English exam. Direct comparison of raw scores is impossible, since the exams might not be equivalent in terms of number of questions, value of each question, and so on.
- A comparison of a relative standard similar to both scores can be made. This comparison uses the mean and standard deviation and is called a standard score or z -score.
- A z -score or standard score for a value x is given by the formula

$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{x - \bar{x}}{s}$$

- Note that if the z -score is positive, the score is above the mean. If the z -score is 0, the score is the same as the mean. And if the z -score is negative, the score is below the mean.

Example 32. A student scored 65 on a calculus test that had a mean of 50 and a standard deviation of 10; she scored 30 on a statistics test with a mean of 25 and a standard deviation of 5. Compare her relative positions on the two tests.

Answer. Note that $\bar{x}_{\text{calculus}} = 50$, $s_{\text{calculus}} = 10$, $\bar{x}_{\text{statistics}} = 25$ and $s_{\text{statistics}} = 5$. So,

$$z_{\text{calculus}} = \frac{65 - 50}{10} = 1.5$$
$$z_{\text{statistics}} = \frac{30 - 25}{5} = 1$$

Since the z -score for calculus is larger, her relative position in the calculus class is higher than her relative position in the statistics class.

Homework 27. Find the z -score for each test in the following, and state which is higher.

Test A	$X = 38$	$\bar{x} = 40$	$s = 5$
Test B	$X = 94$	$\bar{x} = 100$	$s = 10$

Chapter 03 : Correlation and Regression

Feras Awad

In this chapter, you will study how to describe and test the relationships between two variables when data are presented as ordered pairs (x, y) .

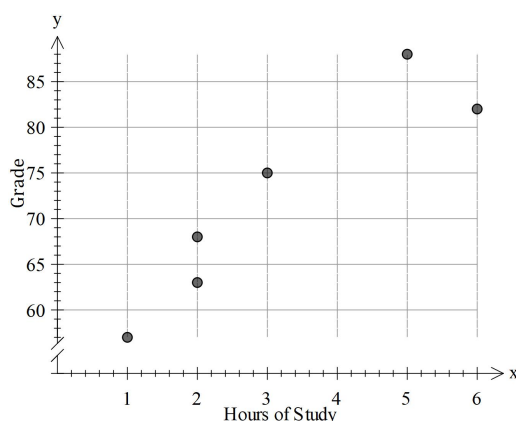
1 Correlation

If a researcher wishes to see whether there is a relationship between number of hours of study and test scores on an exam, he must select a random sample of students, determine the hours each studied, and obtain their grades on the exam. A table can be made for the data, as shown here.

Student	A	B	C	D	E	F
Studied Hours x	6	2	1	5	2	3
Grade y	82	63	57	88	68	75

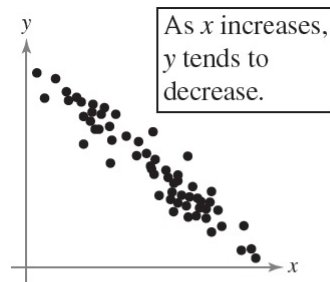
- The two variables for this study are called the independent variable and the dependent variable.
- The independent variable is the variable that can be controlled or manipulated. In this case, the number of hours of study is the independent variable and is designated as the x variable. The dependent variable is the variable that cannot be controlled or manipulated. The grade the student received on the exam is the dependent variable, designated as the y variable.
- We assume that the grade the student earns depends on the number of hours the student studied. Also, you assume that, to some extent, the student can regulate or control the number of hours he or she studies for the exam.
- The independent and dependent variables can be plotted on a graph called a **scatter plot**. The independent variable x is plotted on the horizontal axis, and the dependent variable y is plotted on the vertical axis.

Scatter Plot A scatter plot is a graph of the ordered pairs (x, y) of numbers consisting of the independent variable x and the dependent variable y .

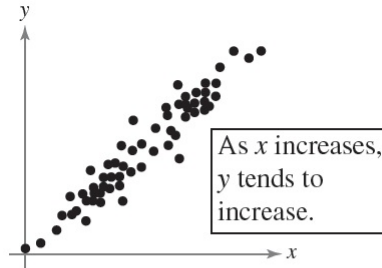


- After the plot is drawn, it should be analyzed to determine which type of relationship, if any, exists between the independent variable x and the dependent variable y .
 - Positive relationship: as x increases, y also increases.
 - Negative relationship: as x increases, y decreases.
 - No specific type of relationship.

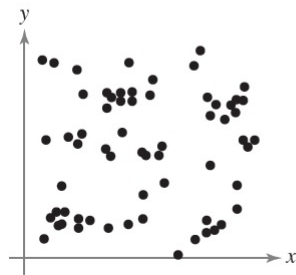
Correlation A **correlation** is a relationship between two variables (x, y) , where x is the independent variable and y is the dependent variable.



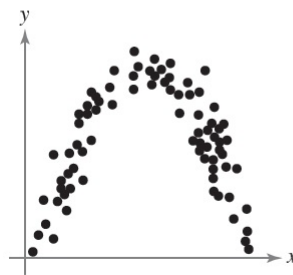
Negative Linear Correlation



Positive Linear Correlation



No Correlation

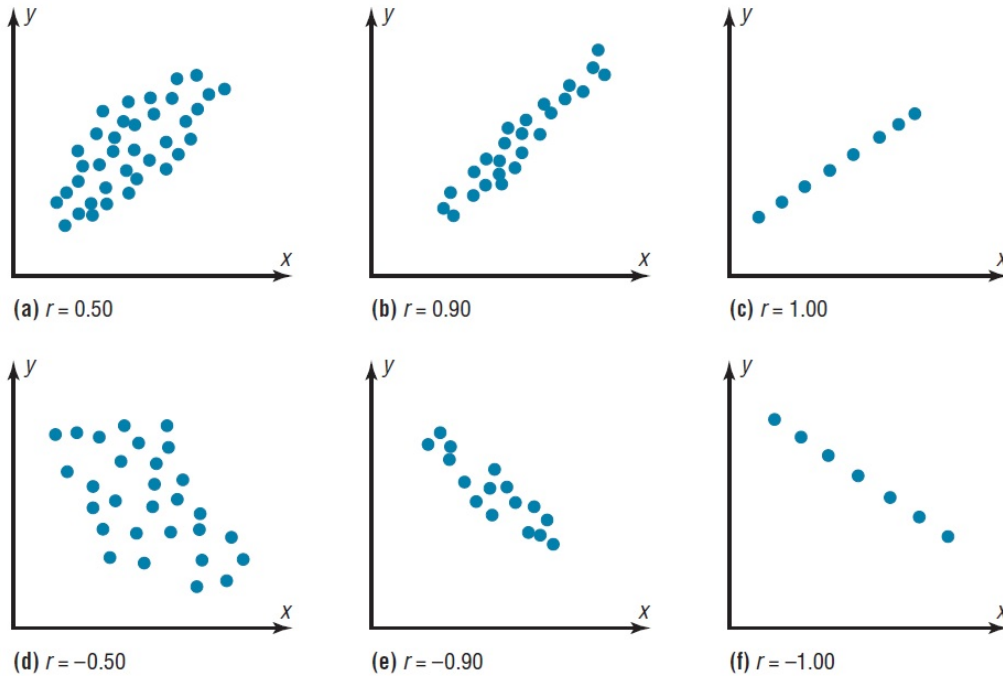


Nonlinear Correlation

Linear Correlation is a measure of association between two variables. It measures the strength of the linear relationship between two variables. Several statistics can be used to measure the correlation between two variables, but the one most commonly used is the **Pearson's correlation coefficient**, which is denoted by r .

Pearson's Correlation Coefficient (r) It is a measure of the strength between two quantitative variables; x and y that has the following properties:

1. r is unit free.
2. The value of r is between -1 and 1 . So, $r \in [-1, 1]$.
3.
 - $r = 0$ means no linear relationship exists between the two variables x and y .
 - r near zero means a little (or weak) linear relationship exists between x and y .
 - r close to -1 or 1 means a strong linear relationship exists between x and y .
4. The sign of r provides important information about the direction of association.
 - If r is positive, then as x increases, y increases linearly.
 - If r is negative, then as x increases, y decreases linearly.



Calculating the Pearson's Correlation Coefficient (r) Suppose that we have a sample of n paired observations $(x_1, y_1), (x_2, y_2), \dots, (x_n, y_n)$ then, r is computed as follows.

$$r = \frac{S_{xy}}{\sqrt{S_{xx} \times S_{yy}}}, \quad \text{where}$$

$$S_{xy} = \sum ((x - \bar{x})(y - \bar{y})) = \sum (xy) - \frac{(\sum x)(\sum y)}{n}$$

$$S_{xx} = \sum ((x - \bar{x})^2) = \sum (x^2) - \frac{(\sum x)^2}{n}$$

$$S_{yy} = \sum ((y - \bar{y})^2) = \sum (y^2) - \frac{(\sum y)^2}{n}$$

Example 1. Calculate the correlation coefficient for the studying hours and final grades given in the following table. Interpret the result in the context of the data.

Student	A	B	C	D	E	F
Studied Hours x	6	2	1	5	2	3
Grade y	82	63	57	88	68	75

Answer:

$$S_{xy} = 1476 - \frac{19 \times 433}{6} \approx 104.83$$

$$S_{xx} = 79 - \frac{19^2}{6} \approx 18.83$$

$$S_{yy} = 31935 - \frac{433^2}{6} \approx 686.83$$

$$\therefore r = \frac{104.83}{\sqrt{18.83 \times 686.83}} \approx 0.92$$

	x	y	xy	x^2	y^2
	6	82	492	36	6724
	2	63	126	4	3969
	1	57	57	1	3249
	5	88	440	25	7744
	2	68	136	4	4624
	3	75	225	9	5625
Total	19	433	1476	79	31935

The result $r = 0.92$ suggests a strong positive linear correlation. So, as the studying hours increases, the final grades tend to increase.

Homework 1. Calculate the correlation coefficient for the number of absences and final grades given in the following table. Interpret the result in the context of the data.

Student	A	B	C	D	E	F	G
Number of Absences x	6	2	15	9	12	5	8
Grade y	82	86	43	74	58	90	78

Homework 2. Calculations from a data set of pairs of $n = 36$ pairs of (x, y) values have provided the following results.

$$\sum (x - \bar{x})^2 = 530.7 \quad , \quad \sum (y - \bar{y})^2 = 235.4 \quad , \quad \sum (x - \bar{x})(y - \bar{y}) = -204.3$$

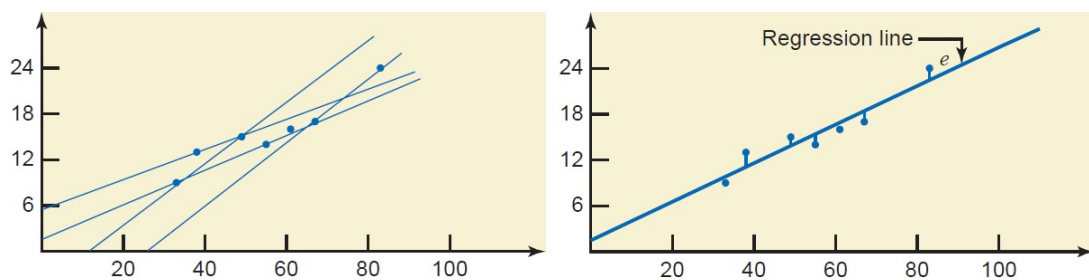
Obtain the correlation coefficient r .

Homework 3. Evaluate the value of r for the data with the following properties.

$$\begin{aligned} n = 30 \quad , \quad \sum x = 680 \quad , \quad \sum (x^2) = 20154 \\ \sum y = 996 \quad , \quad \sum (y^2) = 34670 \quad , \quad \sum (xy) = 24844 \end{aligned}$$

2 Simple Linear Regression

- After verifying that the linear correlation between two variables is significant, the next step is to determine the equation of the line that best models the data.



- This line is called a **regression line**, and its equation can be used to predict the value of y for a given value of x .

- The equation of regression line is $\hat{y} = Ax + B$ where

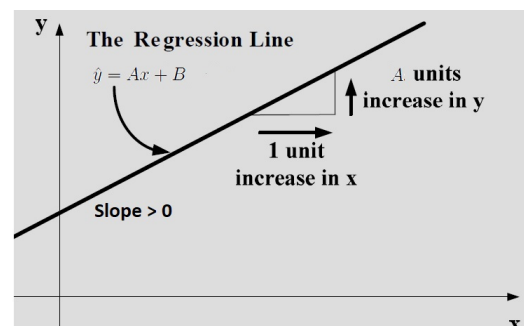
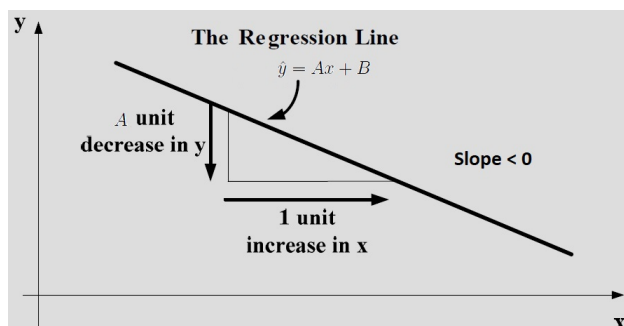
\hat{y} is the predicted value of y .

$A = \frac{S_{xy}}{S_{xx}}$ is the amount that y increases or decreases

whenever x is increased by one unit (**slope**).

$B = \bar{y} - A\bar{x}$ is the height of the line above or below the origin (**intercept**)

- If the slope A of the regression line is negative, then a one unit increase in x will result in A units decrease in y . And if the slope is positive, then a one unit increase in x will result in A units increase in y .



Example 2. Find the equation of the regression line that best fits the following data. Then predict the value of y when $x = 2.5$.

x	y	x^2	xy
0	1	0	0
1	5	1	5
2	3	4	6
3	9	9	27
4	7	16	28

Total 10 25 30 66

Since

$$S_{xx} = \sum (x^2) - \frac{(\sum x)^2}{n} = 30 - \frac{10^2}{5} = 10 \text{ and,}$$

$$S_{xy} = \sum (xy) - \frac{(\sum x)(\sum y)}{n} = 66 - \frac{10 \times 25}{5} = 16$$

then the equation of the regression line is $\hat{y} = Ax + B$ where

$$A = \frac{S_{xy}}{S_{xx}} = \frac{16}{10} = 1.6$$

$$B = \bar{y} - A\bar{x} = \frac{25}{5} - \left(1.6 \times \frac{10}{5}\right) = 1.8.$$

Hence, the equation is $\hat{y} = 1.8 + 1.6x$, and the predicated value of y when $x = 2.5$ is $\hat{y} = 1.8 + (1.6 \times 2.5) = 5.8$.

Example 3. Find the equation of the regression line that best fits the following data. Then find x if $\hat{y} = 3$.

x	y	x^2	xy
8	1	64	8
1	4	1	4
4	2	16	8
7	6	49	42
5	3	25	15

Total 25 16 155 77

Since $S_{xx} = 155 - \frac{25^2}{5} = 30$ and $S_{xy} = 77 - \frac{25 \times 16}{5} = -3$ then $A = \frac{-3}{30} = -0.1$ and $B = \frac{16}{5} - (-0.1) \times \frac{25}{5} = 3.5$. So, the regression equations is $\hat{y} = -0.1x + 3.5$. Now, if $\hat{y} = 3$ then $x = \frac{3.5 - 3}{0.1} = 5$.

Homework 4. Find the equation of the regression line for each data given in homework 1 and 3.

Homework 5. In a regression calculation for five pairs of observations, one pair of values was lost when the data were filed. The four known pairs are

x	0.1	0.2	0.4	0.3
y	0.1	0.3	0.7	0.4

If the regression line was calculated as $y = 2x - 0.1$, find the missing pair of values.

Chapter 04 : Probability and Counting

Feras Awad

In this chapter, you will learn how to determine the probability of an event.

1 Basic Concepts

- **Probability** as a general concept can be defined as the chance of an event occurring.
- A **probability (random) experiment** is a chance process that leads to well-defined results (cannot be predicted with certainty) called outcomes.
- An **outcome** is the result of a single trial of a probability experiment. A trial means for example flipping a coin once, rolling one die once, or the like.
- A **sample space** is the set of all possible outcomes of an experiment and it is usually denoted by the letter S or Ω . For example,

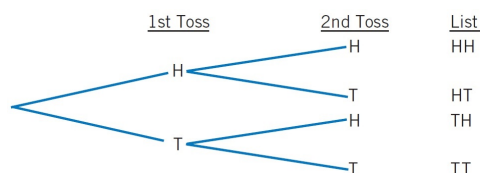
Experiment	Sample space	
Toss one coin	Head (H), Tail (T)	(finite)
Roll a die	1, 2, 3, 4, 5, 6	(finite)
Answer a question	True, False	(finite)
Baby gender	Male, Female	(finite)
Toss a coin until a head appears	H, TH, TTH, TTTH, ...	(infinite)

- Each outcome in a sample space is called an **element** of the sample space.

Example 1. Find the sample space of each of the following experiments.

1. Tossing a coin twice.

Answer. $S = \{HH, HT, TH, TT\}$



2. Tossing two coins.

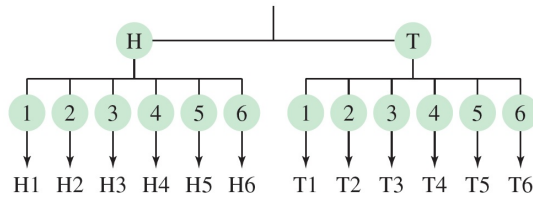
Answer. Same as tossing a coin twice.

3. Rolling a die twice. (Same as rolling two dice)

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

4. Tossing a coin and then rolling a die.

Answer. $S = \{H1, H2, H3, H4, H5, H6, T1, T2, T3, T4, T5, T6\}$



Homework 1. Find the sample space for the gender of the children if a family has three children. Use **B** for boy and **G** for girl.

- An **event** is a subset of a sample space.

Example 2. When rolling a die, the sample space is $S = \{1, 2, 3, 4, 5, 6\}$. Find the following events.

1. An even number appears: $A = \{2, 4, 6\}$. **(Compound Event)**
2. A prime number greater than 4 appears: $B = \{5\}$. **(Simple Event)**
3. A negative number appears: $C = \{\} = \Phi$. **(Impossible Event)**
4. A number less than 10 appears: $D = \{1, 2, 3, 4, 5, 6\} = S$. **(Sure Event)**

- A **Venn diagram** is a picture (a closed geometric shape such as a rectangle, a square, or a circle) that depicts all the possible outcomes for an experiment.

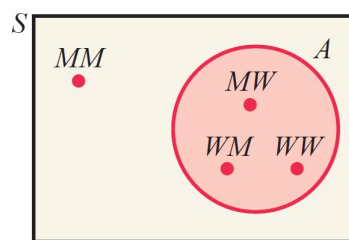
Example 3. Suppose we randomly select two workers from a company and observe whether the worker selected each time is a man or a woman.

1. Find the sample space of this experiment.

Answer: $S = \{MM, MW, WM, WW\}$.

2. Let A be the event that at most one man is selected. Find A then draw the Venn diagram of A .

Answer: $A = \{MW, WM, WW\}$.



Homework 2. Consider the experiment of choosing at random a digit from the digits $0, 1, 2, \dots, 9$.

1. What is the sample space of this experiment?
2. Find the elements of each of the following events, and draw Venn diagram of each.
 - (a) A = the number chosen is less than or equal to 3.
 - (b) B = the number chosen is between 4 and 6, inclusive.
 - (c) C = the number chosen is greater than or equal to 7.
 - (d) D = the number chosen is less than 4 or larger than 7.
 - (e) E = the number chosen is an even number.

2 Probability of an Event

Calculating Probability The probability of an event is a numerical value that represents the proportion of times the event is expected to occur when the experiment is repeated many times under identical conditions. The probability of event $A \subseteq S$ is given by

$$P(A) = \frac{\text{number of elements in } A}{\text{number of elements in } S} = \frac{n(A)}{n(S)}$$

Note The probability of any event $A \subseteq S$ must satisfy:

1. $0 \leq P(A) \leq 1$ for all events A ,
2. $P(A) = \sum_{\text{all } e \text{ in } A} P(e)$,
3. $P(S) = \sum_{\text{all } e \text{ in } S} P(e) = 1$ and $P(\Phi) = 0$.

Example 4. If we twice flip a balanced coin, what is the probability of getting at least one head?

Answer: The sample space is $S = \{HH, HT, TH, TT\}$, where H and T denote head and tail. The event A is $A = \{HH, HT, TH\}$. So, $P(A) = \frac{3}{4}$.

Example 5. The blood type of students are given by the following frequency table.

Blood Type	O	A	B	AB	Total
$n(\text{Students})$	3	5	6	1	15

If one student is randomly selected, what is the probability that this person's blood type is:

1. AB ?
Answer: $P(AB) = \frac{1}{15}$.
2. Either A or B ?
Answer: $P(A \text{ or } B) = \frac{5+6}{15} = \frac{11}{15}$.
3. Not O ?
Answer: $P(\text{Not } O) = \frac{5+6+1}{15} = \frac{12}{15} = \frac{4}{5}$.

Example 6. There are four elementary outcomes in a sample space $S = \{e_1, e_2, e_3, e_4\}$. If $P(e_1) = 0.3$, $P(e_2) = 0.4$, and $P(e_3) = 0.2$, Find $P(e_4)$.

Answer:

$$\begin{aligned} P(S) = 1 &\Rightarrow P(e_1) + P(e_2) + P(e_3) + P(e_4) = 1 \\ &\Rightarrow 0.3 + 0.4 + 0.2 + P(e_4) = 1 \\ &\Rightarrow P(e_4) = 0.1 \end{aligned}$$

Note If $S = \{e_1, e_2, \dots, e_n\}$ where $P(e_1) = P(e_2) = \dots = P(e_n)$ then the elementary events e_1, e_2, \dots, e_n are called equally likely and $P(e_i) = \frac{1}{n}$; $\forall i = 1, 2, \dots, n$.

Homework 3. A die is rolled once in such a way that each odd number is twice as likely to occur as each even number. Find $P(G)$, where G is the event that a number greater than 3 occurs.

Homework 4. A stack contains eight tickets numbered 1, 1, 2, 2, 2, 3, 3, 3. One ticket will be drawn at random and its number will be noted.

1. List the sample space and assign probabilities to the elementary outcomes.
2. What is the probability of drawing an odd-numbered ticket?

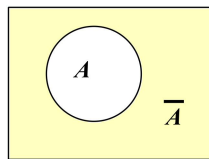
Example 7. The probabilities of males and females in a statistics class is given by the ratio 3 : 7. If we randomly select one student, find the probability that this student is male?

Answer: $P(\text{male}) = \frac{3}{3+7} = \frac{3}{10} = 0.3$. Note that $P(\text{female}) = \frac{7}{3+7} = \frac{7}{10} = 0.7$.

3 Event Relations and Laws of Probability

1. The Complement:

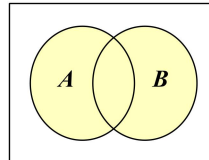
- The complement of an event A , denoted by \bar{A} is the set of all elementary outcomes that are not in A .



- The occurrence of \bar{A} means that A does not occur.
- $A \cup \bar{A} = S$ and $A \cap \bar{A} = \Phi$.
- $P(A) + P(\bar{A}) = 1$.

2. The Union:

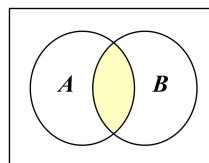
- The union of two events A and B , denoted by $A \cup B$, is the set of all elementary outcomes that are in A , B , or both.



- The occurrence of $A \cup B$ means that either A or B or both occur.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$.
- $P(A \cup B) = P(B \cup A)$.

3. The Intersection:

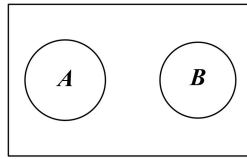
- The intersection of two events A and B , denoted by $A \cap B$, is the set of all elementary outcomes that are in A and B .



- The occurrence of $A \cap B$ means that both A and B occur.
- $P(A \cap B) = P(B \cap A)$.

4. Disjoint Events:

- Two events A and B are called disjoint or mutually exclusive if their intersection $A \cap B = \Phi$.



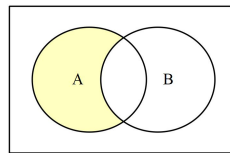
- Because disjoint events have no elementary outcomes in common, they cannot occur at the same time.
- $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

5. Exhaustive Events:

- Two events A and B are called exhaustive if $A \cup B = S$.
- $P(A \cup B) = 1$.

6. The Difference:

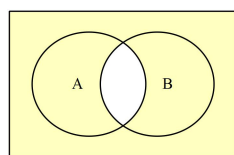
- The difference of two events A and B , denoted by $A - B$, is the set of all elementary outcomes that are in A but not in B .



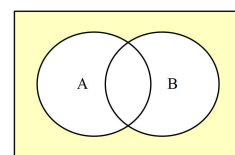
- The occurrence of $A - B$ means that A occurs but B does not occur.
- $P(A - B) = P(A \cap \overline{B}) = P(A) - P(A \cap B)$.
- In general, $P(A - B) \neq P(B - A)$.

7. De Morgan Laws:

- $\overline{A \cap B} = \overline{A} \cup \overline{B}$.



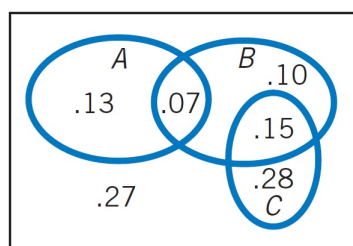
- $\overline{A \cup B} = \overline{A} \cap \overline{B}$.



8. Distributive Law:

- $A \cap (B \cup C) = (A \cap B) \cup (A \cap C)$.
- $A \cup (B \cap C) = (A \cup B) \cap (A \cup C)$.

Example 8. The accompanying Venn diagram shows three events A , B , and C and also the probabilities of the various intersections.



Determine:

1. $P(A)$.

Answer: $P(A) = 0.13 + 0.07 = 0.20$

2. $P(B - C)$.

Answer: $P(B - C) = 0.10 + 0.07 = 0.17$

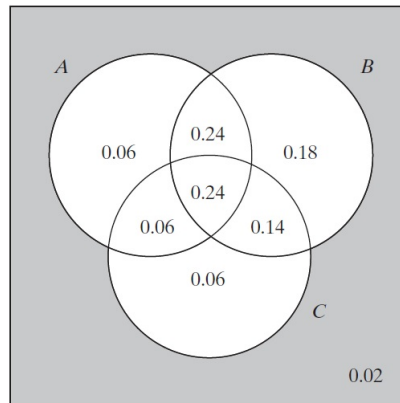
3. $P(A \cup C)$.

Answer: $P(A \cup C) = 0.20 + 0.43 = 0.63$

4. the probability that exactly one of the three events A , B , C will occur.

Answer: $P(A \cap \bar{B} \cap \bar{C}) + P(\bar{A} \cap B \cap \bar{C}) + P(\bar{A} \cap \bar{B} \cap C) = 0.13 + 0.10 + 0.28 = 0.45$

Homework 5. The accompanying Venn diagram shows three events A , B , and C and also the probabilities of the various intersections.



Determine:

1. the probability that neither B nor C occur.

2. $P(A \cap (B \cup C))$.

3. $P((A \cap B) \cup (A \cap \bar{B}))$.

Example 9. If A and B are two events such that $P(A) = \frac{19}{30}$, $P(B) = \frac{2}{5}$, and $P(A \cup B) = \frac{4}{5}$. Find $P(A \cap B)$.

Answer:

$$P(A \cup B) = P(A) + P(B) - P(A \cap B)$$

$$\frac{4}{5} = \frac{19}{30} + \frac{2}{5} - P(A \cap B)$$

$$P(A \cap B) = \frac{19}{30} + \frac{2}{5} - \frac{4}{5} = \frac{7}{30}.$$

Example 10. If A and B are two events satisfy $P(A) = 0.3$, $P(B) = 0.4$, and $P(A \cap B) = 0.1$. Find:

1. $P(A - B)$

Answer: $P(A - B) = P(A) - P(A \cap B) = 0.3 - 0.1 = 0.2$

2. the probability that at least one of the events A and B occurs.

Answer: $P(A \cup B) = P(A) + P(B) - P(A \cap B) = 0.3 + 0.4 - 0.1 = 0.6$

3. $P(\bar{A} \cap \bar{B})$

Answer: $P(\bar{A} \cap \bar{B}) = P(\overline{A \cup B}) = 1 - P(A \cup B) = 1 - 0.6 = 0.4$

Example 11. For the disjoint events A and B , if $P(A) = 0.50$ and $P(B) = 0.25$, find:

1. $P(A \cup B)$

Answer: $P(A \cup B) = P(A) + P(B) = 0.50 + 0.25 = 0.75$

2. $P(B - A)$

Answer: $P(B - A) = P(B) - P(A \cap B) = 0.25 - 0 = 0.25$

Homework 6. For the events A and B , if $P(A) = P(B)$, $P(A \cup B) = 0.7$, and $P(A \cap B) = 0.1$, find $P(\bar{A})$.

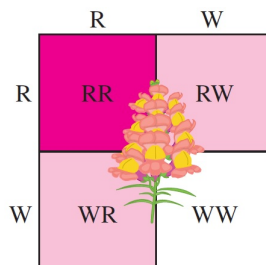
Homework 7. Given that $P(\bar{A}) = \frac{2}{3}$, $P(B) = \frac{1}{2}$, and $P(A \cap B) = \frac{1}{12}$. Find $P(A \cup B)$.

Homework 8. If A and B are two events such that $P(A) = \frac{8}{15}$, $P(B) = \frac{2}{3}$, and $P(A \cap B) = \frac{1}{5}$. Are A and B exhaustive events?

Homework 9. For two events A and B defined on a sample space S , $n(A \cap \bar{B}) = 15$, $n(\bar{A} \cap B) = 50$, and $n(A \cap B) = 2$. Given that $n(S) = 120$, how many outcomes belong to neither A nor B ?

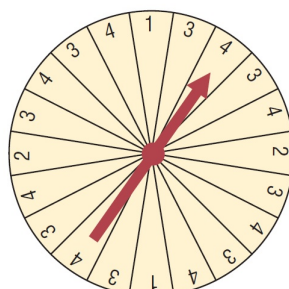
Homework 10. Events A and B are defined on a sample space S such that $P(\overline{A \cup B}) = 0.5$ and $P(A \cap B) = 0.2$. What is the probability that either A or B but not both will occur?

Homework 11. *Punnett square* is a diagram that shows all possible gene combinations in a cross of parents whose genes are known. When two pink snapdragon flowers (RW) are crossed, there are four equally likely possible outcomes for the genetic makeup of the offspring: red (RR), pink (RW), pink (WR), and white (WW), as shown in the Punnett square below. When two pink snapdragons are crossed, what is the probability that the offspring will be **(a)** pink, **(b)** red, and **(c)** white?



Homework 12. A shopping mall has set up a promotion as follows. With any mall purchase of \$50 or more, the customer gets to spin the wheel shown below. If a number 1 comes up, the customer wins \$10. If the number 2 comes up, the customer wins \$5; and if the number 3 or 4 comes up, the customer wins a discount coupon. Find the following probabilities.

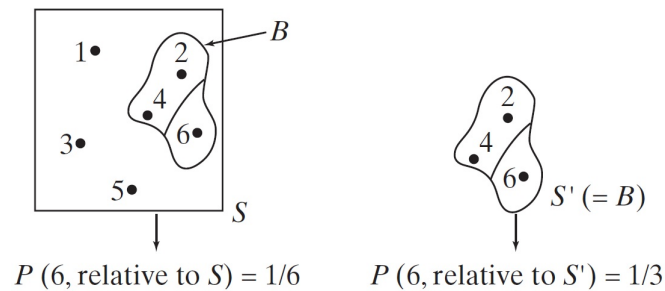
1. The customer wins \$10.
2. The customer wins money
3. The customer wins a coupon.



4 Independence and Conditional Probability

Consider a fair die being rolled, with A defined as the event “6 appears.” Clearly, $P(A) = \frac{1}{6}$. But suppose that the die has already been rolled-by someone who refuses to tell us whether or not A occurred but does enlighten us to the extent of confirming that B occurred, where B is the event “Even number appears.” What are the chances of A now? Here, common sense can help us: There are three equally likely even numbers making up the event B -one of which satisfies the event A , so the *updated* probability is $\frac{1}{3}$.

Notice that the effect of additional information, such as the knowledge that B has occurred, is to revise-indeed, to shrink-the original sample space S to a new set of outcomes S' . In this example, the original S contained six outcomes, the conditional sample space, three.



Independent Events (Multiplication Rule 1) Two events A and B are independent events if the fact that A occurs does not affect the probability of B occurring. When two events are independent, the probability of both occurring is

$$P(A \cap B) = P(A) \cdot P(B)$$

Example 12. A coin is flipped and a die is rolled. Find the probability of getting a head on the coin and a 4 on the die.

Answer: Let A be the event “a head appears on the coin”, and B be the event “a 4 appears on the die”. Then $P(A \cap B) = P(A) \cdot P(B) = \frac{1}{2} \times \frac{1}{6} = \frac{1}{12}$.

Example 13. An urn contains 3 red balls, 2 blue balls, and 5 white balls. A ball is selected and its color noted. Then it is **replaced**. A second ball is selected and its color noted. Find the probability of selecting

1. two blue balls.

Answer. $P(B \cap B) = \frac{2}{10} \times \frac{2}{10} = \frac{1}{25} = 0.04$

2. one blue ball and then *one* white ball.

Answer. $P(B \cap W) = \frac{2}{10} \times \frac{5}{10} = \frac{1}{10} = 0.10$

3. one red ball and then *one* blue ball.

Answer. $P(R \cap B) = \frac{3}{10} \times \frac{2}{10} = \frac{3}{50} = 0.06$

Homework 13. Let A and B be two mutually exclusive events. Are A and B independent events? Explain your answer.

Homework 14. Let A , B be two independent events with $P(A) = P(B)$ and $P(A \cup B) = \frac{1}{2}$. Find $P(A)$.

Result: It can be shown that if A and B are independent events, then the following events are also independent.

1. \bar{A} and \bar{B} . Then $P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B})$.
2. \bar{A} and B . Then $P(\bar{A} \cap B) = P(\bar{A}) \cdot P(B)$.
3. A and \bar{B} . Then $P(A \cap \bar{B}) = P(A) \cdot P(\bar{B})$.

Example 14. Two men, A and B are shooting a target. The probability that A hits the target is $P(A) = \frac{1}{3}$, and the probability that B shoots the target is $P(B) = \frac{1}{5}$, one independently of the other. Find the probability that

1. A misses the target.

Answer. $P(\bar{A}) = 1 - P(A) = 1 - \frac{1}{3} = \frac{2}{3}$.

2. both men hit the target.

Answer. $P(A \cap B) = P(A) \cdot P(B) = \frac{1}{3} \times \frac{1}{5} = \frac{1}{15}$.

3. at least one of them hits the target.

Answer. $P(A \cup B) = P(A) + P(B) - P(A \cap B) = \frac{1}{3} + \frac{1}{5} - \frac{1}{15} = \frac{7}{15}$.

4. none of them hits the target.

Answer. $P(\bar{A} \cap \bar{B}) = P(\bar{A}) \cdot P(\bar{B}) = \frac{2}{3} \times \frac{4}{5} = \frac{8}{15}$.

Note: The multiplication rules can be used with the complementary event rule to simplify solving probability problems involving "at least." The following examples illustrate how this is done.

Example 15. A coin is tossed 4 times. Find the probability of getting at least 1 tail.

Answer. Let A is the event of getting at least 1 tail, then \bar{A} is the event of getting no tail.

So, $P(A) = 1 - P(\bar{A}) = 1 - P(HHHH) = 1 - \left(\frac{1}{2}\right)^4 = \frac{15}{16}$.

The sample space S				$A = \text{at least 1 tail}$			
HHHH	HHHT	HHTH	HHTT	HHHT	HHTH	HHTT	
HTHH	HTHT	HTTH	HTTT	HTHH	HTHT	HTTH	HTTT
THHH	THHT	THTH	THTT	THHH	THHT	THTH	THTT
TTHH	TTHT	TTTH	TTTT	TTHH	TTHT	TTTH	TTTT

Homework 15. A couple is planning on having three children. Assuming that each child is equally likely to be of either gender and that the genders of the children are *independent*, find the probability that

1. all three children will be girls.
2. at least one child will be a girl.

Example 16. A certain type of missile hits its target 30% of the time. Determine the minimum number of missiles that must be shot at a certain target in order to obtain a chance higher than 80% of hitting the target.

Answer. The probability that all the n missiles miss the target is $(0.7)^n$. So, the probability that at least one of the n missiles hits the target is $1 - (0.7)^n$. Hence, we need

$$\begin{aligned}
 1 - (0.7)^n > 0.8 &\Rightarrow (0.7)^n < 0.2 \Rightarrow \ln((0.7)^n) < \ln(0.2) \\
 &\Rightarrow n > \frac{\ln(0.2)}{\ln(0.7)} \approx 4.51 \\
 \therefore n &= 5
 \end{aligned}$$

Homework 16.

1. Find the probability of obtaining at least one six when n dice are thrown.
2. How many dice must be thrown so that the probability of obtaining at least one six is at least 0.99?

Conditional Probability (Multiplication Rule 2)

- When the outcome or occurrence of the first event affects the outcome or occurrence of the second event in such a way that the probability is changed, the events are said to be **dependent events**.
- The probability of an event A must often be modified after information is obtained as to whether or not a related event B has taken place.
- The revised probability of A when it is known that B has occurred is called the **conditional probability** of A given B and is denoted by $P(A|B)$.
- The probability that the second event B occurs given that the first event A has occurred can be found by the formula

$$P(B|A) = \frac{P(A \cap B)}{P(A)}$$

- When two events are dependent, the probability of both occurring is

$$\begin{aligned} P(A \cap B) &= P(A) \cdot P(B|A) \\ &= P(B) \cdot P(A|B) \end{aligned}$$

Example 17. When a die was thrown the score was an odd number. What is the probability that it was a prime number?

Answer. $P(\text{Prime}|\text{Odd}) = \frac{P(\text{Prime} \cap \text{Odd})}{P(\text{Odd})} = \frac{2/6}{3/6} = \frac{2}{3}.$

Example 18. A recent survey asked 100 people if they thought women in the armed forces should be permitted to participate in combat. The results of the survey are shown.

Gender	Yes	No	Total
Male	32	18	50
Female	8	42	50
Total	40	60	100

Find the probability that

1. the respondent answered yes, **given that** the respondent was a female.

Answer. $P(\text{Yes}|\text{Female}) = \frac{P(\text{Yes} \cap \text{Female})}{P(\text{Female})} = \frac{8}{50} = 0.16.$

2. the respondent was a male, **given that** the respondent answered no.

Answer. $P(\text{Male}|\text{No}) = \frac{P(\text{No} \cap \text{Male})}{P(\text{No})} = \frac{18}{60} = 0.30.$

3. the respondent was a **female**.

Answer. $P(\text{Female}) = \frac{50}{100} = 0.50.$

4. the respondent answered **no**.

Answer. $P(\text{No}) = \frac{60}{100} = 0.60.$

Homework 17. The following table gives the two way classification of 500 students based on gender and whether or not they suffer from math anxiety.

Gender	Yes	No
Male	145	95
Female	190	70

If a student randomly selected. What is the probability that this student

1. is female given that she does not suffer from math anxiety?
2. suffers from math anxiety given that he is a male?
3. is a male?
4. does not suffer from math anxiety?

Note: The events A and B are independent if and only if $P(A|B) = P(A)$, $P(B|A) = P(B)$ or $P(A \cap B) = P(A) \cdot P(B)$.

Example 19. Let A and B be two events such that $P(A|B) = 0.40$, $P(B) = 0.25$ and $P(A) = 0.20$. Find

1. $P(B|A)$

Answer.

$$P(B|A) \cdot P(A) = P(A|B) \cdot P(B)$$

$$P(B|A) \times 0.20 = 0.40 \times 0.25$$

$$P(B|A) = \frac{0.40 \times 0.25}{0.20} = 0.5.$$

2. Determine whether A and B are independent or not.

Answer. Since $P(A|B) \neq P(A)$ then the events are not independent (**dependent**).

3. $P(A \cap B)$

Answer. $P(A \cap B) = P(A|B) \cdot P(B) = 0.40 \times 0.25 = 0.1$.

Homework 18. Let A and B be two events such that $P(A) = 0.45$, $P(B) = 0.35$ and $P(A \cup B) = 0.70$. Find $P(A|B)$.

Homework 19. Let A and B be two independent events such that $P(B) = P(A) + 0.20$ and $P(A \cap B) = 0.15$. Find $P(\overline{A}|B)$.

Homework 20. If A and B be two exhaustive events such that $P(A|B) = \frac{1}{4}$ and $P(B) = \frac{2}{3}$, find $P(A)$.

Homework 21. Two events A and B are such that $P(A) = \frac{8}{15}$, $P(B) = \frac{1}{3}$ and $P(A|B) = \frac{1}{5}$, find $P(A)$. Calculate the probabilities that

1. both events occur,
2. only one of the two events occur,
3. neither event occurs.

Note: If A and B are two events, then $A = (A \cap B) \cup (A \cap \overline{B})$. Also,

$$P(A) = P(A \cap B) + P(A \cap \overline{B})$$

Example 20. Let A and B be two events such that $P(A) = \frac{4}{7}$ and $P(A \cap \overline{B}) = \frac{1}{3}$. Find $P(A \cap B)$.

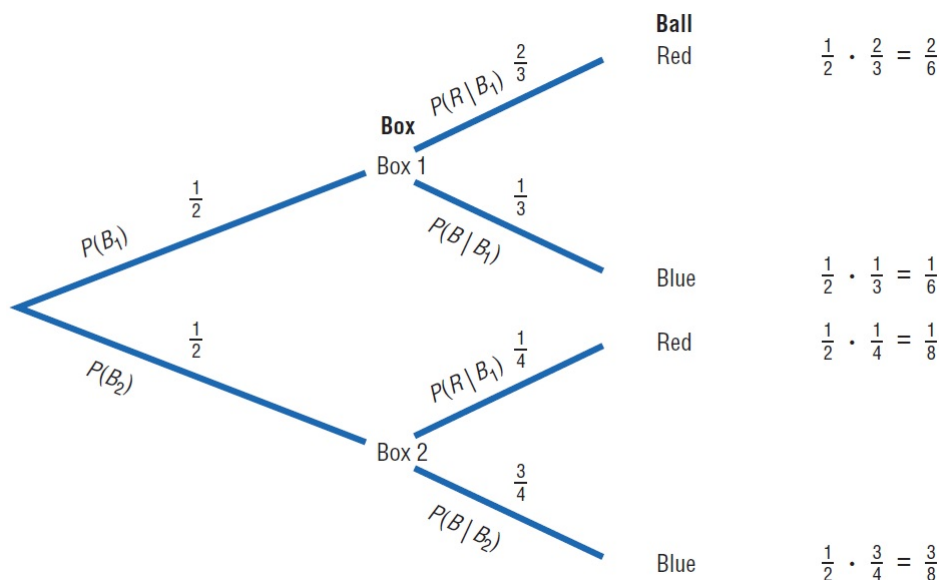
Answer.

$$\begin{aligned} P(A) &= P(A \cap B) + P(A \cap \overline{B}) \\ \frac{4}{7} &= P(A \cap B) + \frac{1}{3} \\ \therefore P(A \cap B) &= \frac{5}{21}. \end{aligned}$$

Example 21. Box (1) contains **2 red** balls and **1 blue** ball. Box (2) contains **3 blue** balls and **1 red** ball. A coin is tossed. If it falls heads up, box (1) is selected and a ball is drawn. If it falls tails up, box (2) is selected and a ball is drawn. Find the probability of selecting a red ball.

Answer.

$$\begin{aligned} P(\text{Red}) &= P(\text{Red}|\text{Box 1}) \cdot P(\text{Box 1}) + P(\text{Red}|\text{Box 2}) \cdot P(\text{Box 2}) \\ &= \left(\frac{2}{3}\right) \left(\frac{1}{2}\right) + \left(\frac{1}{4}\right) \left(\frac{1}{2}\right) \\ &= \frac{11}{24} \end{aligned}$$



Homework 22. Refer to example (21), what is the probability of selecting a blue ball.

Homework 23. Urn (1) contains 5 red balls and 3 black balls. Urn (2) contains 3 red balls and 1 black ball. Urn (3) contains 4 red balls and 2 black balls. If an urn is selected at random and a ball is drawn, find the probability it will be red.

Example 22. An urn contains 8 black balls and 3 white balls. We draw **two** balls without replacement. What is the probability that

- both balls are black?

Answer. $P(B_1 \cap B_2) = P(B_1) \cdot P(B_2|B_1) = \frac{8}{11} \times \frac{7}{10} = \frac{28}{55}$

- both balls are of the same color?

Answer. $P(\text{Same Color}) = P(B_1 \cap B_2) + P(W_1 \cap W_2) = \frac{28}{55} + \frac{3}{55} = \frac{31}{55}$

- both balls are of different colors?

Answer. $P(\text{Different Colors}) = 1 - P(\text{Same Color}) = 1 - \frac{31}{55} = \frac{24}{55}$.

5 Counting Rules

Since probability problems require knowing the total number of ways one or more events can occur, it is necessary to have a way to compute the number of outcomes in the sample spaces for a probability experiment. This is especially true when the number of outcomes is large. In order to do the computation, we use the fundamental counting rule, the permutation rules, and the combination rule.

5.1 The Fundamental Counting Rule (Product Rule)

For a sequence of n events in which the first event can occur in k_1 ways and the second event can occur in k_2 ways and the third event can occur in k_3 ways, and so on, the total number of ways the sequence can occur is $k_1 \times k_2 \times k_3 \times \cdots \times k_n$.

Example 23.

1. A die is rolled then a coin is tossed. Then there are $6 \times 2 = 12$ elements in the sample space, see part (4) from example (1).
2. Suppose there are 3 roads from city **A** to city **B** and 5 roads from **B** to **C**. Then the total number of ways (possibilities) you can get from city **A** to **C** is $3 \times 5 = 15$.
3. If you have 6 **shirts** of different colors and 4 **pants** of different colors, then there are $4 \times 6 = 24$ possibilities (ways) to wear a shirt and a pants together.

Example 24. The employees of a company are given a 4-digit identification number. How many different numbers are available if repetitions are

1. allowed ?
Answer. $10 \times 10 \times 10 \times 10 = 10^4 = 10000$.
2. NOT allowed ?
Answer. $10 \times 9 \times 8 \times 7 = 5040$.

Homework 24. A license plate consists of 2 letters and 3 digits. How many different plates can be made if repetitions are allowed? How many can be made if repetitions are not allowed?

Example 25. How many 3-digit numbers are there if repetition between digits is

1. allowed ?
Answer. $9 \times 10 \times 10 = 900$.
2. NOT allowed ?
Answer. $9 \times 9 \times 8 = 648$.

Example 26. From the set of 3-digit numbers, what is the probability of getting an even number if repetition between digits is not allowed ?

Answer.

$$\begin{aligned} P(3 - \text{digits even}) &= \frac{n(3 - \text{digits even})}{n(3 - \text{digits})} \\ &= \frac{(9 \times 8 \times 1) + (8 \times 8 \times 4)}{9 \times 9 \times 8} = \frac{328}{648} = \frac{41}{81} \end{aligned}$$

Homework 25. From the set of 3-digit numbers, what is the probability of getting a number greater than 335 if repetition between digits is allowed.

Homework 26. From the set of 4-digit numbers, what is the probability of getting a number contains two 7's if repetition between digits is allowed.

5.2 Factorial Rule

In mathematics, the factorial of a positive integer n , denoted by $n!$, is the number

$$n! = n \times (n-1) \times (n-2) \times \cdots \times 3 \times 2 \times 1.$$

For example, $1! = 1$, $3! = 3 \times 2 \times 1$, and $6! = 7 \times 6 \times 5 \times 4 \times 3 \times 2 \times 1 = 720$.

Properties

- $0! = 1$.
- $n! = n(n-1)! = n(n-1)(n-2)! = \cdots = n(n-1)(n-2) \cdots (3)(2)(1)$.

Homework 27. What is the value of $\frac{6!}{4! 2!}$?

The Rule Suppose that n distinct objects are to be drawn sequentially, or ordered in a row. Then, the number of ways to arrange n distinct objects in a row is $n!$.

Example 27. In how many ways you can arrange the letters A, B, and C ?

Answer. There are $3! = 6$ ways: ABC, ACB, BAC, BCA, CAB, and CBA.

Homework 28. In how many ways you can arrange the numbers 1, 5 and 7 ? What is the probability of getting a number greater than 400 ?

Example 28. A witness reported that a car seen speeding away from the scene of the crime had a number plate that began with V or W, the digits were 4, 7 and 8 and the end letters were A, C, E. He could not however remember the order of the digits or the end letters. How many cars would need to be checked to be sure of including the suspect car?

Answer. $2! \times 3! \times 3! = 72$.

Example 29.

1. How many ways can 3 boys and 2 girls form a line?

Answer. $5! = 120$.

2. What would the answer be if the two girls must be together?

Answer. We treat the two girls as one girl, and note that the two girls may change places. So we have $2! \times 4! = 48$ ways.

3. What would the answer be if the boys must be together and the girls must be together?

Answer. $2! \times (3! \times 2!) = 24$.

Homework 29. How many ways can we arrange 4 statistics books, 3 mathematics books, 2 biology books, and 1 computer science book on a bookshelf so that all the statistics books are together, all the mathematics books are together, all the biology books are together.

Homework 30. Ten students are placed at random in a line. What is the probability that the two youngest students are separated ?

Note: The number of different arrangements of n objects in which k_1 objects of them are the same, k_2 objects of them are the same, and so on, is $\frac{n!}{k_1! k_2! \cdots k_r!}$.

Example 30. In how many ways you can arrange the letters A, A, B, C ?

Answer. $\frac{4!}{2!} = \frac{4 \times 3 \times 2!}{2!} = 12$.

Example 31. How many ways can we arrange the letters of the word MISSISSIPPI?

Answer. The word MISSISSIPPI contains 11 letters such that there are 4 I's, 4 S's, 2 P's, and 1 M. So, the total number of ways is $\frac{11!}{4! \times 4! \times 2! \times 1!} = 34650$.

Homework 31. How many ways can we arrange the letters of the word STATISTICS?

Circular Arrangements:

- The number of ways of arranging n distinct objects in a circle (ring) is $(n - 1)!$ if clockwise and anticlockwise arrangements are different.
- If clockwise and anticlockwise arrangements are the same, then the number of ways of arranging n distinct objects in a circle (ring) is $\frac{(n - 1)!}{2}$.

Example 32. The number of different arrangements of four men A, B, C, D around a table is $3! = 6$.

5.3 Permutations Rule

The number of ways of selecting m objects from n distinct objects where order is important is

- $P_m^n = \frac{n!}{(n - m)!}$ if selection without replacement.
- n^m if selection with replacement.

Example 33. In how many ways can we choose 3 letters out of the 5 letters A, B, C, D, E if order is important and selection is

1. without replacement?

Answer. $P_3^5 = \frac{5!}{(5 - 3)!} = \frac{5!}{2!} = \frac{5 \times 4 \times 3 \times 2!}{2!} = 60$.

2. with replacement?

Answer. $n^m = 5^3 = 125$.

Example 34. A school musical director can select 2 musical plays to present next year. One will be presented in the fall, and one will be presented in the spring. If she has 9 to pick from, how many different possibilities are there?

Answer. $P_2^9 = \frac{9!}{(9 - 2)!} = \frac{9!}{7!} = \frac{9 \times 8 \times 7!}{7!} = 72$.

Homework 32. How many 4-letter code words can be made using the letters in the word pencil if repetitions are allowed? If repetitions are not allowed?

Homework 33. Six students are running for the positions of president and vice-president, and five students are running for secretary and treasurer. If the two highest vote getters in each of the two contests are elected, how many winning combinations can there be?

5.4 Combinations Rule

The number of ways of selecting m objects from n distinct objects where order is **NOT** important is

- $C_m^n = \frac{n!}{m!(n - m)!}$ if selection without replacement.
- $C_m^{n+m-1} = \frac{(n + m - 1)!}{m!(n - 1)!}$ if selection with replacement.

Example 35. In how many ways can we choose 3 letters out of the 5 letters A, B, C, D, E if order is NOT important and selection is

1. without replacement?

Answer. $C_3^5 = \frac{5!}{3! (5-3)!} = \frac{5!}{3! 2!} = \frac{5 \times 4 \times 3!}{3! 2!} = 10.$

2. with replacement?

Answer. $C_3^7 = \frac{7!}{3! (7-3)!} = \frac{7!}{3! 4!} = \frac{7 \times 6 \times 5 \times 4!}{3! 4!} = 35.$

Example 36. How many ways can an adviser choose 4 students from a class of 12 if they are all assigned the same task?

Answer. $C_4^{12} = \frac{12!}{4! 8!} = 495.$

Homework 34. In example (36), how many ways can the students be chosen if they are each given a different task?

Example 37.

1. How many ways can a committee of 6 people be selected from a group of 10 people?

Answer. $C_6^{10} = \frac{10!}{6! 4!} = 210.$

2. In a club there are 7 women and 5 men. A committee of 3 women and 2 men is to be chosen. How many different possibilities are there?

Answer. $C_3^7 \times C_2^5 = \frac{7!}{3! 4!} \times \frac{5!}{2! 3!} = 35 \times 10 = 350.$

Homework 35. There are 7 women and 5 men in a department. How many ways can a committee of 4 people be selected? How many ways can this committee be selected if there must be 2 men and 2 women on the committee? How many ways can this committee be selected if there must be at least 2 women on the committee?

Example 38. A parent-teacher committee consisting of 4 people is to be formed from 20 parents and 5 teachers. Find the probability that the committee will consist of these people.

1. All teachers.

Answer. $P(\text{All Teachers}) = \frac{C_4^5 \times C_0^{20}}{C_4^{25}} = \frac{1}{2530}.$

2. 2 teachers and 2 parents.

Answer. $P(2 \text{ teachers and } 2 \text{ parents}) = \frac{C_2^5 \times C_2^{20}}{C_4^{25}} = \frac{38}{253}.$

Homework 36. In a company there are 7 executives: 4 women and 3 men. Three are selected to attend a management seminar. Find these probabilities.

1. All 3 selected will be men.
2. 1 man and 2 women will be selected.

Homework 37. A drawer contains 11 identical red socks and 8 identical black socks. Suppose that you choose 2 socks at random in the dark. What is the probability that you get 2 matched socks?

Chapter 05 : Probability Distributions and Densities

Feras Awad

In Chapters 2 and 4, you learned how to collect and describe data and how to find the probability of an event. These skills are used in many different types of careers. In this chapter, you will learn how to create and use probability distributions and probability densities.

Knowing the shape, center, and variability of a probability distribution enables you to make decisions in inferential statistics. For example, a saleswoman can compute the probability that she will make 0, 1, 2, or 3 or more sales in a single day. Once these probabilities are assigned, statistics such as the mean, variance, and standard deviation can be computed for these events. With these statistics, various decisions can be made. The saleswoman will be able to compute the average number of sales she makes per week, and if she is working on commission, she will be able to approximate her weekly income over a period of time, say, monthly.

1 Random Variables

- The outcome of a probability experiment is often a count or a measure. When this occurs, the outcome is called a random variable.
- A **random variable** X represents a numerical value associated with each outcome of a probability experiment.
- The word random indicates that X is determined by chance.
- There are two types of random variables:
 - **Discrete:** A random variable is discrete when it has a finite or countable number of possible outcomes that can be listed. For example, the number of heads appear when tossing 3-coins is $X = \{0, 1, 2, 3\}$.
 - **Continuous:** A random variable is continuous when it has an uncountable number of possible outcomes, represented by an interval on a number line. For example, the life-time of a light bulb is $0 \leq X \leq 3000$ days.

2 Discrete Probability Distributions

A **discrete probability distribution** lists each possible value the random variable X can assume, together with its probability $P(X = x)$. A discrete probability distribution must satisfy these conditions:

1. The probability of each value of the discrete random variable is between 0 and 1, inclusive. So, $0 \leq P(X = x) \leq 1$ for all x .
2. The sum of all the probabilities is 1. So, $\sum_{\text{for all } x} P(X = x) = 1$.

Example 1. If the random variable X represents the number of heads appear when 3-coins is tossed, then the probability distribution of X is given by

$X = x$	0	1	2	3
$P(x)$	$\frac{1}{8}$	$\frac{3}{8}$	$\frac{3}{8}$	$\frac{1}{8}$

No heads	One head			Two heads			Three heads
TTT $\frac{1}{8}$	TTH $\frac{1}{8}$	THT $\frac{1}{8}$	HTT $\frac{1}{8}$	HHT $\frac{1}{8}$	HTH $\frac{1}{8}$	THH $\frac{1}{8}$	HHH $\frac{1}{8}$
$\frac{1}{8}$	$\frac{3}{8}$			$\frac{3}{8}$			$\frac{1}{8}$

Example 2. A probability distribution can be written in formula notation such as

$$P(X = x) = \frac{1}{x} \text{ where } X = 2, 3, 6.$$

The distribution is shown as follows:

$X = x$	2	3	6
$P(x)$	$\frac{1}{2}$	$\frac{1}{3}$	$\frac{1}{6}$

Homework 1. Construct a probability distribution for a family with 4 children. Let X be the number of girls.

Example 3. The discrete random variable X has probability distribution shown below.

$X = x$	-3	-2	-1	0	1
$P(x)$	0.10	0.25	0.30	0.15	c

Find:

1. the value of c .

Answer.

$$\begin{aligned} \sum P(X = x) = 1 &\Rightarrow P(-3) + P(-2) + P(-1) + P(0) + P(1) = 1 \\ &\Rightarrow 0.10 + 0.25 + 0.30 + 0.15 + c = 1 \\ &\Rightarrow c = 0.20. \end{aligned}$$

2. $P(-3 \leq X < 0)$.

Answer. $P(-3 \leq X < 0) = P(-3) + P(-2) + P(-1) = 0.10 + 0.25 + 0.30 = 0.65$.

3. $P(X > -1)$.

Answer. $P(X > -1) = P(0) + P(1) = 0.15 + 0.20 = 0.35$.

4. $P(-1 < X < 1)$.

Answer. $P(-1 < X < 1) = P(0) = 0.15$.

Homework 2. The discrete random variable X has probability distribution given by the formula

$$P(X = x) = c(3 - x) \quad X = 0, 1, 2, 3.$$

Find the value of c , then evaluate $P(X \leq 2)$.

3 Continuous Probability Densities

A function with values $f(x)$, defined over the set of all real numbers, is called a **probability density function** of the continuous random variable X if and only if

$$P(a \leq X \leq b) = \int_a^b f(x) dx \text{ where } a, b \in \mathbb{R} \text{ and } a \leq b.$$

A function can serve as a probability density of a continuous random variable X if its values, $f(x)$, satisfy the conditions

1. $f(x) \geq 0$ for all $x \in \mathbb{R}$.

2. $\int_{-\infty}^{\infty} f(x) dx = 1$.

Note: If X is a continuous random variable and a and b are real constants with $a \leq b$, then $P(X = a) = 0 = P(X = b)$ and

$$P(a \leq X \leq b) = P(a \leq X < b) = P(a < X \leq b) = P(a < X < b).$$

Example 4. A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx^2 & : 0 \leq x \leq 3 \\ 0 & : \text{otherwise} \end{cases}$$

Find:

1. the value of k .

Answer. $\int_{-\infty}^{\infty} f(x) dx = 1 \Rightarrow \int_0^3 kx^2 dx = 1 \Rightarrow \left[\frac{k}{3} x^3 \right]_0^3 = 1 \Rightarrow 9k = 1 \Rightarrow k = \frac{1}{9}$.

2. $P(0 \leq X < 1)$.

Answer. $P(0 \leq X < 1) = \int_0^1 \frac{1}{9} x^2 dx = \frac{1}{27}$.

3. $P(X \leq 2)$.

Answer. $P(X \leq 2) = \int_0^2 \frac{1}{9} x^2 dx = \frac{8}{27}$.

4. $P(X \geq 1)$.

Answer. $P(X \geq 1) = \int_1^3 \frac{1}{9} x^2 dx = \frac{26}{27}$.

Homework 3. A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx & : 0 \leq x \leq 4 \\ 0 & : \text{otherwise} \end{cases}$$

Find:

1. the value of k .

2. $P(1 \leq X < 2.5)$.

Homework 4. A continuous random variable X has probability density function

$$f(x) = \begin{cases} kx^3 & : 0 \leq x \leq c \\ 0 & : \text{otherwise} \end{cases}$$

such that $P\left(X \leq \frac{1}{2}\right) = \frac{1}{16}$ Find the values of the constants c and k .

4 The Expected Value (Mean) and the Variance

You can measure the center of a probability distribution for a discrete random variable X with its mean and measure the variability with its variance and standard deviation. The mean and the variance of a discrete random variable X are defined as follows.

- The expected value (mean) of X is given by $\mu_X = E(X) = \sum xP(x)$.
- The variance of X is $\sigma_X^2 = \text{Var}(X) = \sum (x - \mu)^2 P(x)$.

Note: The variance of a discrete random variable can be evaluated using the formula $\text{Var}(X) = E(X^2) - (E(X))^2$ where $E(X^2) = \sum x^2 P(x)$.

Example 5. In a family with two children, let X be the number of girls. Find the mean and the variance of X .

Answer. Since the possibilities of the two children are BB, BG, GB, GG then the probability distribution of X is

$X = x$	0	1	2
$P(x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$

1. The mean is $\mu_X = E(X) = \sum xP(x) = \left(0 \times \frac{1}{4}\right) + \left(1 \times \frac{1}{2}\right) + \left(2 \times \frac{1}{4}\right) = 1$.
2. Since $E(X^2) = \sum x^2 P(x) = \left(0^2 \times \frac{1}{4}\right) + \left(1^2 \times \frac{1}{2}\right) + \left(2^2 \times \frac{1}{4}\right) = \frac{3}{2}$ then the variance is $\text{Var}(X) = E(X^2) - (E(X))^2 = \frac{3}{2} - (1)^2 = \frac{1}{2}$.

Example 6. A random variable X has probability distribution given below. Find $E(X)$ and $\text{Var}(X)$.

x	$P(x)$	$xP(x)$	$x^2 P(x)$	
-2	0.30	-0.60	1.20	
-1	0.10	-0.10	0.10	$\therefore E(X) = -0.2$
0	0.15	0	0	$\therefore \text{Var}(X) = 2.10 - (-0.2)^2 = 2.06$
1	0.40	0.40	0.40	
2	0.05	0.10	0.40	
		-0.2	2.10	

Homework 5. Find the mean and the variance of each random variable in examples (1, 2, 3) and in homeworks (1, 2).

Homework 6. A random variable X can take the values 10 and 20 only. If $E(X) = 16$, find the probability distribution of X .

Homework 7. The discrete random variable X can take the values 0, 1, 2, and 3 only. Given $P(X \leq 2) = 0.9$, $P(X \leq 1) = 0.5$, and $E(X) = 1.4$, find $P(X = 0)$.

Mean and Variance for Continuous Random Variables: If X is a continuous random variable with probability density function $f(x)$, then

- $E(X) = \int_{-\infty}^{\infty} x f(x) dx$.
- $\text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2$.

Example 7. A random variable X has probability density function $f(x) = \begin{cases} \frac{3}{64}x^2 & : 0 \leq x \leq 4 \\ 0 & : \text{otherwise} \end{cases}$.

Find $E(X)$ and $\text{Var}(X)$.

Answer.

$$1. E(X) = \int_{-\infty}^{\infty} xf(x) dx = \int_0^4 \frac{3}{64}x^3 dx = 3.$$

$$2. \text{Var}(X) = \int_{-\infty}^{\infty} x^2 f(x) dx - (E(X))^2 = \int_0^4 \frac{3}{64}x^4 dx - (3)^2 = \frac{3}{5} = 0.6.$$

Homework 8. Find the mean and the variance of each random variable in example (4) and in homework (3).

Note: If a and b are constants, then

1. $E(a) = a$ and $\text{Var}(a) = 0$,
2. $E(aX + b) = aE(X) + b$ and $\text{Var}(aX + b) = a^2 \text{Var}(X)$ where X is a random variable either discrete or continuous.

Example 8. Let X be a random variable with mean $E(X) = 1$ and variance $\text{Var}(X) = 0.5$. Then

1. $E(2X - 3) = 2E(X) - 3 = -1$
2. $\text{Var}(4 - 5X) = (-5)^2 \text{Var}(X) = 12.5$

Homework 9. Find the mean and the variance of the random variable X with $E(2 - X) = 0.5$ and $\text{Var}(4X + 1) = 12$.

5 The Binomial Distribution

Many types of probability problems have only two outcomes or can be reduced to two outcomes. For example, when a coin is tossed, it can land heads or tails. When a baby is born, it will be either male or female. In a basketball game, a team either wins or loses. A true/false item can be answered in only two ways, true or false. Other situations can be reduced to two outcomes. For example, the number 5 can be classified as appears or does not appear when a die is rolled. A multiple-choice question, even though there are four or five answer choices, can be classified as correct or incorrect. Situations like these are called **binomial experiments**.

A binomial experiment is a probability experiment that satisfies the following four requirements:

1. There must be a fixed number of trials.
2. Each trial can have only two outcomes or outcomes that can be reduced to two outcomes. These outcomes can be considered as either success or failure.
3. The outcomes of each trial must be independent of one another.
4. The probability of a success must remain the same for each trial.

Binomial Probability Formula: In a binomial experiment, the probability of exactly k successes in n trials is

$$P(X = k) = C_k^n p^k (1 - p)^{n-k}$$

where p is the probability of a success, and $1 - p$ is the probability of a failure.

Example 9. A coin is tossed 3 times. Find the probability of getting exactly two heads.

Answer: $P(2 \text{ heads}) = C_2^3 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^1 = \frac{3}{8}$.

Example 10. A die is tossed 4 times. Find the probability of getting at most 3 fives.

Answer:

$$\begin{aligned} P(\text{at most 3 fives}) &= P(\text{no five}) + P(1 \text{ five}) + P(2 \text{ fives}) + P(3 \text{ fives}) \\ &= 1 - P(4 \text{ fives}) = 1 - C_4^4 \left(\frac{1}{6}\right)^4 \left(\frac{5}{6}\right)^0 = \frac{1295}{1296} \end{aligned}$$

Homework 10. Is this a binomial distribution? Explain.

X	0	1	2	3
$P(X)$	0.064	0.288	0.432	0.216

Homework 11. A student takes a 10-question, multiple-choice physics exam with four choices for each question and guesses on each question. Find the probability that the student will pass the exam. **Hint:** your answer should be approximately 0.078.

Example 11. If $X \sim B(n = 4, p)$ and $P(X = 4) = 0.0256$, find the value of p .

Answer. Since $P(X = 4) = 0.0256 = C_4^4 p^4 (1 - p)^0$ then $p^4 = 0.0256$. Hence, $p = \sqrt[4]{0.0256} = 0.4$.

Homework 12. If $X \sim B(n, p = 0.6)$ and $P(X < 1) = 0.0256$, find the value of n .

Note: If $X \sim B(n, p)$ then $E(X) = np$ and $\text{Var}(X) = np(1 - p)$.

Example 12. A die is rolled 480 times. Find the mean and variance of the number of 3s that will be rolled.

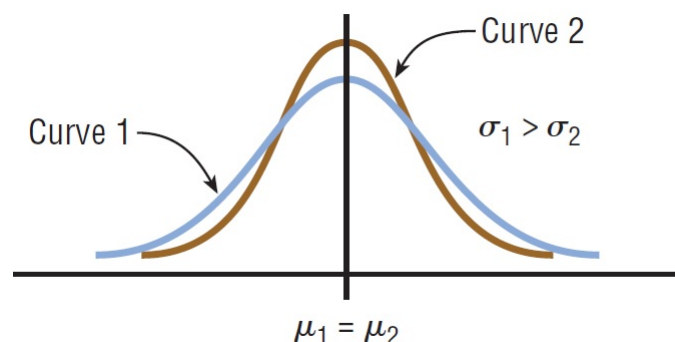
Answer. $E(X) = 480 \times \frac{1}{6} = 80$ and $\text{Var}(X) = 480 \times \frac{1}{6} \times \frac{5}{6} = \frac{20}{3}$.

Homework 13. If $X \sim B(n, p)$ with mean 5 and variance 4, find the values of n and p .

6 Normal Probability Distributions

A **normal distribution** is a continuous, symmetric, bell-shaped distribution of a variable. The mathematical equation for a normal distribution curve is

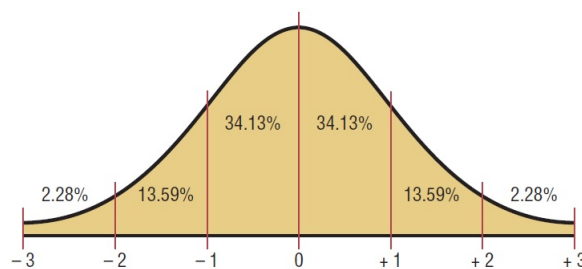
$$f(x) = \frac{1}{\sigma\sqrt{\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2}; \quad x \in (-\infty, \infty)$$



Properties of Normal Distributions:

- A normal distribution curve is bell-shaped.
- The mean, median, and mode are equal and are located at the center of the distribution.
- A normal distribution curve is unimodal (it has only one mode).
- The curve is symmetric about the mean, which is equivalent to saying that its shape is the same on both sides of a vertical line passing through the center.
- The curve is continuous; that is, there are no gaps or holes.
- The curve never touches the x axis.
- The total area under a normal distribution curve is equal to 1.

The Standard Normal Distribution is a normal distribution with mean $\mu = 0$ and variance $\sigma^2 = 1$.

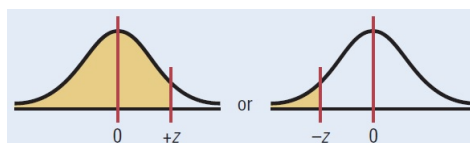


All normally distributed variables can be transformed into the standard normally distributed variable by using the formula for the standard score:

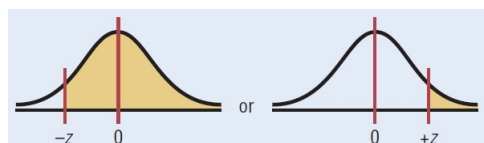
$$z = \frac{\text{value} - \text{mean}}{\text{standard deviation}} = \frac{X - \mu}{\sigma}$$

After you use the formula above to transform an x -value to a z -score, you can use the **Standard Normal Table**. The table lists the cumulative area under the standard normal curve to the left of z for z -scores from -3.49 to 3.49 . So,

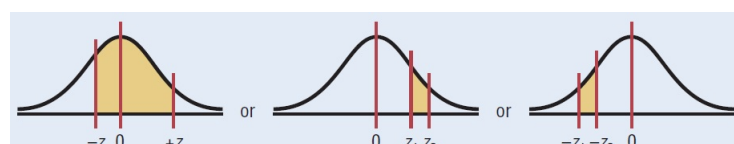
- $P(Z \leq a)$ is evaluated directly from the table.



- $P(Z \geq a) = 1 - P(Z \leq a) = P(Z \leq -a)$.



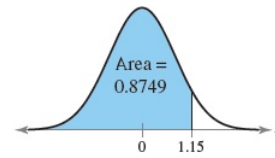
- $P(a \leq Z \leq b) = P(Z \leq b) - P(Z \leq a)$



Example 13. Let $Z \sim N(\mu = 0, \sigma = 1)$. Then:

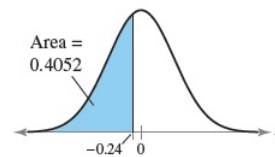
1. $P(Z \leq 1.15) = 0.8749$.

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



2. $P(Z \leq -0.24) = 0.4052$.

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



3. $P(Z > -1.84) = 1 - P(Z < -1.84) = 1 - 0.0329 = 0.9671$. Note that, since normal distributions are symmetric, then $P(Z > -1.84) = P(Z < 1.84) = 0.9671$.

4. $P(0.21 < Z \leq 1.07) = P(Z \leq 1.07) - P(Z < 0.21) = 0.8577 - 0.5832 = 0.2745$.

Homework 14. Evaluate each of the following probabilities.

1. $P(Z < -1.71)$

2. $P(Z > 0.89)$

3. $P(0 \leq Z < 2.43)$

Example 14. Find the value of c in each of the following if $Z \sim N(\mu = 0, \sigma = 1)$.

1. $P(Z \leq c) = 0.2709$.

Answer. From the table, $c = -0.61$

2. $P(Z > c) = 0.1038$.

Answer.

$$P(Z > c) = 0.1038 \Rightarrow 1 - P(Z < c) = 0.1038$$

$$\Rightarrow P(Z < c) = 0.8962 \Rightarrow c = 1.96$$

3. $P(-0.60 < Z < c) = 0.4991$.

Answer.

$$P(-0.60 < Z < c) = 0.4991 \Rightarrow P(Z < c) - P(Z < -0.60) = 0.4991$$

$$\Rightarrow P(Z < c) - 0.2743 = 0.4991$$

$$\Rightarrow P(Z < c) = 0.7734 \Rightarrow c = 0.75$$

Homework 15. Let $Z \sim N(\mu = 0, \sigma = 1)$. Find c if $P(Z > c) = 0.5987$.

Example 15. Given that $X \sim N(\mu = 60, \sigma = 4)$, find $P(55 \leq X \leq 63)$.

Answer.

$$\begin{aligned} P(55 \leq X \leq 63) &= P\left(\frac{55 - 60}{4} \leq \frac{X - 60}{4} \leq \frac{63 - 60}{4}\right) \\ &= P(-1 \leq Z \leq 0.75) = P(Z \leq 0.75) - P(Z \leq -1) \\ &= 0.7734 - 0.1587 = 0.6147 \end{aligned}$$

Homework 16. Given that $X \sim N(\mu = 50, \sigma = 4)$, find

1. $P(X \leq 46.4)$
2. $P(39 \leq X \leq 56)$

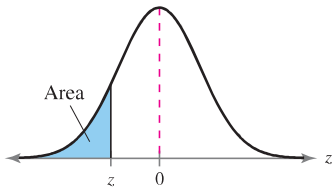
Example 16. Given that $X \sim N(\mu = 10, \sigma = 2)$, find the value of c if $P(c \leq X \leq 12) = 0.6826$.

Answer.

$$\begin{aligned} P(c \leq X \leq 12) = 0.6826 &\Rightarrow P\left(\frac{c - 10}{2} \leq \frac{X - 10}{2} \leq \frac{12 - 10}{2}\right) = 0.6826 \\ &\Rightarrow P\left(\frac{c - 10}{2} \leq Z \leq 1\right) = 0.6826 \\ &\Rightarrow P(Z \leq 1) - P\left(Z \leq \frac{c - 10}{2}\right) = 0.6826 \\ &\Rightarrow 0.8413 - P\left(Z \leq \frac{c - 10}{2}\right) = 0.6826 \\ &\Rightarrow P\left(Z \leq \frac{c - 10}{2}\right) = 0.1587 \\ &\Rightarrow \frac{c - 10}{2} = -1 \\ &\Rightarrow c = 8 \end{aligned}$$

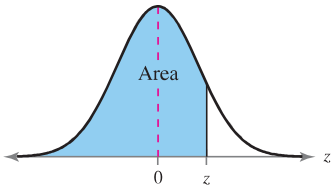
Homework 17. Given that $X \sim N(\mu = 50, \sigma = 4)$, find c if $P(X > c) = 0.2266$.

Table —Standard Normal Distribution



z	.09	.08	.07	.06	.05	.04	.03	.02	.01	.00
− 3.4	.0002	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003	.0003
− 3.3	.0003	.0004	.0004	.0004	.0004	.0004	.0004	.0005	.0005	.0005
− 3.2	.0005	.0005	.0005	.0006	.0006	.0006	.0006	.0006	.0007	.0007
− 3.1	.0007	.0007	.0008	.0008	.0008	.0008	.0009	.0009	.0009	.0010
− 3.0	.0010	.0010	.0011	.0011	.0011	.0012	.0012	.0013	.0013	.0013
− 2.9	.0014	.0014	.0015	.0015	.0016	.0016	.0017	.0018	.0018	.0019
− 2.8	.0019	.0020	.0021	.0021	.0022	.0023	.0023	.0024	.0025	.0026
− 2.7	.0026	.0027	.0028	.0029	.0030	.0031	.0032	.0033	.0034	.0035
− 2.6	.0036	.0037	.0038	.0039	.0040	.0041	.0043	.0044	.0045	.0047
− 2.5	.0048	.0049	.0051	.0052	.0054	.0055	.0057	.0059	.0060	.0062
− 2.4	.0064	.0066	.0068	.0069	.0071	.0073	.0075	.0078	.0080	.0082
− 2.3	.0084	.0087	.0089	.0091	.0094	.0096	.0099	.0102	.0104	.0107
− 2.2	.0110	.0113	.0116	.0119	.0122	.0125	.0129	.0132	.0136	.0139
− 2.1	.0143	.0146	.0150	.0154	.0158	.0162	.0166	.0170	.0174	.0179
− 2.0	.0183	.0188	.0192	.0197	.0202	.0207	.0212	.0217	.0222	.0228
− 1.9	.0233	.0239	.0244	.0250	.0256	.0262	.0268	.0274	.0281	.0287
− 1.8	.0294	.0301	.0307	.0314	.0322	.0329	.0336	.0344	.0351	.0359
− 1.7	.0367	.0375	.0384	.0392	.0401	.0409	.0418	.0427	.0436	.0446
− 1.6	.0455	.0465	.0475	.0485	.0495	.0505	.0516	.0526	.0537	.0548
− 1.5	.0559	.0571	.0582	.0594	.0606	.0618	.0630	.0643	.0655	.0668
− 1.4	.0681	.0694	.0708	.0721	.0735	.0749	.0764	.0778	.0793	.0808
− 1.3	.0823	.0838	.0853	.0869	.0885	.0901	.0918	.0934	.0951	.0968
− 1.2	.0985	.1003	.1020	.1038	.1056	.1075	.1093	.1112	.1131	.1151
− 1.1	.1170	.1190	.1210	.1230	.1251	.1271	.1292	.1314	.1335	.1357
− 1.0	.1379	.1401	.1423	.1446	.1469	.1492	.1515	.1539	.1562	.1587
− 0.9	.1611	.1635	.1660	.1685	.1711	.1736	.1762	.1788	.1814	.1841
− 0.8	.1867	.1894	.1922	.1949	.1977	.2005	.2033	.2061	.2090	.2119
− 0.7	.2148	.2177	.2206	.2236	.2266	.2296	.2327	.2358	.2389	.2420
− 0.6	.2451	.2483	.2514	.2546	.2578	.2611	.2643	.2676	.2709	.2743
− 0.5	.2776	.2810	.2843	.2877	.2912	.2946	.2981	.3015	.3050	.3085
− 0.4	.3121	.3156	.3192	.3228	.3264	.3300	.3336	.3372	.3409	.3446
− 0.3	.3483	.3520	.3557	.3594	.3632	.3669	.3707	.3745	.3783	.3821
− 0.2	.3859	.3897	.3936	.3974	.4013	.4052	.4090	.4129	.4168	.4207
− 0.1	.4247	.4286	.4325	.4364	.4404	.4443	.4483	.4522	.4562	.4602
− 0.0	.4641	.4681	.4721	.4761	.4801	.4840	.4880	.4920	.4960	.5000

Table —Standard Normal Distribution (*continued*)

[illegible]