Example 1: Finding the Mean, Median and Mode

9 people take a test. Their scores out of 100 are:

$$56, 79, 77, 48, 90, 68, 79, 92, 71$$

Work out the **mean**, **median**, and **mode** of their scores.

Mean: There are 9 data points. First add the numbers together and then divide the result by 9.

$$56 + 79 + 77 + 48 + 90 + 68 + 79 + 92 + 71 = 660$$

$$\text{Mean} = \frac{660}{9} = 73.3 \text{ (1 dp)}$$

Median: Firstly, put the numbers in ascending order.

$$48, 56, 68, 71, 77, 79, 79, 90, 92$$

There are 9 numbers, and $\frac{9+1}{2} = 5$, so the median must be the 5th term along.

$$\cancel{48}, \cancel{56}, \cancel{68}, \cancel{71}, 77, 79, 79, 90, 92$$

Counting along the list, we get that the median is 77.

Mode: We can see very clearly from the ordered list that there is only one repeat, 79, so the mode is 79.

Example 2: Calculating the Range

Find the **range** of $12, 8, 4, 16, 15, 15, 5, 15, 10, 8$

[1 mark]

A good way to make sure you haven't missed any numbers in determining the biggest and smallest value is to order them. Doing this, we get

$$4, 5, 8, 8, 10, 12, 15, 15, 15, 16$$

Largest − Smallest $= 16 - 4 = 12$, so the **range is** 12

Example 3: Finding the Mean – Applied Questions

There were 5 members of a basketball team who had a **mean points** score of 12 points each per game.

One of the team members left, causing the **mean point** score to reduce to 10 points each per game.

What was the **mean score** of the player that left?

[2 marks]

Step 1: Find the total for the original number of players: $5 \times 12 = 60$

Step 2: Find the total after once the mean has changed, so $4 \times 10 = 40$

Step 3: Calculate the difference between these two totals as that difference has been caused by the person who left: $60 - 40 = 20$

Therefore the **mean** score of the person who left was 20 points per game. The same method applies if a new person/amount is added, you find the old and new totals and the difference is always due to the thing which caused the change.

1) Find the mode and range of the list of number below.

$$280, 350, 320, 400, 350, 490, 590, 470, 280, 410, 350.$$

Mode: 350

Range: $590 - 280 = 310$

2) Dani recorded the heights of members of her extended family, to the nearest cm. Find the median of their heights (listed below).

163, 185, 164, 170, 188, 154, 168, 179

4 ↓ 5

Median: 154, 163, 164, 168, 170, 179, 185, 188

$\frac{n+1}{2} = \frac{8+1}{2} = 4.5^{th}$ 169 cm

3) Below is a list of recorded reaction times in seconds.

0.25, 0.34, 0.39, 0.38, 0.39, 1.67, 0.28, 0.30, 0.42, 0.46.

a) Calculate the mean of these values.

b) Identify any outliers and explain how the value of the mean would change if any outliers were to be removed from the calculation of the mean.

a) $\frac{0.25 + 0.34 + 0.39 + 0.38 + 0.39 + 1.67 + 0.28 + 0.30 + 0.42 + 0.46}{10}$

$= 0.488$ (s)

b) 1.67 Down

Question 4: The mean length of 7 planks of wood is 1.35m. When an extra plank of wood is added, the mean length of a plank of wood increased to 1.4m.
What is the length of the extra plank of wood that was added?

[2 marks]

Level 4-5

In most questions involving the mean, we are given the total and need to work out the mean from the total. In this question, we have been given the mean, so we are going to have to calculate the total from the mean.

If the mean length of 7 planks of wood is 1.35m, then the total length of all these planks of wood combined can be calculated as follows:

$$7 \times 1.35\text{m} = 9.45\text{m}$$

When the extra plank of wood is added, the mean length of a plank of wood increases to 1.4m. This means there are now 8 planks of wood, with a combined length of:

$$8 \times 1.40\text{m} = 11.2\text{m}$$

Therefore, by adding this additional plank of wood, the combined length has increased from 9.45m to 11.2m, so the length of this extra plank of wood is therefore:

$$11.2\text{m} - 9.45\text{m} = 1.75\text{m}$$

Question 5: In a rowing team, the weight of 8 women, in kilograms, is:

$$63, \quad 60, \quad 57, \quad 66, \quad 62, \quad 65, \quad 69, \quad 58$$

In order to be a more competitive team, the coach has said that each team member should try to increase their overall muscle mass which will result in a 2% gain in overall body weight.
What will be the mean weight of the team if all 8 are successful in precisely meeting this 2% weight gain?

[2 marks]

Level 4-5

In this question, we do not need to work out a 2% increase in weight for each individual team member (it would not be wrong to do so, just unnecessarily time-consuming).

The combined weight of all 8 members is:

$$63+60+57+66+62+65+69+58=500\text{kg}$$

If each team member increases their weight by 2%, then this is the same as the team increasing their combined weight by 2%. Therefore, if the team is successful in achieving this 2% weight gain, then the combined weight of the team can be calculated as follows:

$$1.02\times 500=510\text{kg}$$

Since there are 8 team members in total, then mean weight following this weight gain is:

$$510\text{kg}\div 8=63.75\text{kg}$$

PRACTICE QUESTIONS

1. Find the median of the set of numbers: 1,2,3,4,5,6,7,8,9 and 10.

a. 55
b. 10
c. 1
d. 5.5

2. Find the median of the set of numbers: 21, 3, 7, 17, 19, 31, 46, 20 and 43.

a. 19
b. 20
c. 3
d. 167

3. Find the median of the set of numbers: 100, 200, 450, 29, 1029, 300 and 2001.

a. 300
b. 29
c. 7
d. 4,080

4. The following represents age distribution of students in an elementary class. Find the mode of the values: 7, 9, 10, 13, 11, 7, 9, 19, 12, 11, 9, 7, 9, 10, 11.

a. 7
b. 9
c. 10
d. 11

5. Find the mode from these test results: 90, 80, 77, 86, 90, 91, 77, 66, 69, 65, 43, 65, 75, 43, 90.

a. 43
b. 77
c. 65
d. 90

6. Find the mode from these test results: 17, 19, 18, 17, 18, 19, 11, 17, 16, 19, 15, 15, 15, 17, 13, 11.

a. 15
b. 11
c. 17
d. 19

7. Find the mean of these set of numbers: 100, 1050, 320, 600 and 150.

a. 333
b. 444
c. 440
d. 320

8. The following numbers represent the ages of people on a bus: 3, 6, 27, 13, 6, 8, 12, 20, 5, 10. Calculate their mean of their ages.

a. 11
b. 6
c. 9
d. 110

9. These numbers are taken from the number of people that attended a particular church every Friday for 7 weeks: 62, 18, 39, 13, 16, 37, 25. Find the mean.

a. 25
b. 210
c. 62
d. 30

1. D
First arrange the numbers in a numerical sequence: 1,2,3,4,5,6,7,8,9, 10. Then find the middle number or numbers. The middle numbers are 5 and 6.
The median = 5 + 6/2 = 11/2 = 5.5

2. B
First arrange the numbers in a numerical sequence: 3,7, 17, 19, 20, 21, 31, 43, 46. Next find the middle number. The median = 20

3. A
First arrange the numbers in a numerical sequence: 29,100, 200, 300, 450, 1029, 2001. Next find the middle number. The median = 300

4. B
Simply find the most recurring number. The most occurring number in the series is 9

5. D
Simply find the most recurring number. The most occurring number in the series is 90.

6. C
Simply find the most recurring number. The most occurring number in the series is 17.

7. B
First add all the numbers 100 + 1050 + 320 + 600 + 150 = 2220. Then divide by 5 (the number of data provided) = 2220/5 = 444

8. A
First add all the numbers 3 + 6 + 27 + 13 + 6 + 8 + 12 + 20 + 5 + 10 = 110. Then divide by 10 (the number of data provided) = 110/10 = 11

9. D
First add all the numbers 62 + 18 + 39 + 13 + 16 + 37 + 25 = 210. Then divide by 7 (the number of data provided) = 210/7 = 30

1. Formula & Examples

Formula

1. Mean $\bar{x} = \dfrac{\sum x}{n}$
2. Population Variance $\sigma^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n}$
3. Population Standard deviation $\sigma = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n}}$
4. Co-efficient of Variation (Population) $= \dfrac{\sigma}{\bar{x}} \cdot 100\ \%$

Examples

1. Calculate Population Variance $\left(\sigma^2\right)$, Population Standard deviation (σ), Population Coefficient of Variation from the following data
3,13,11,15,5,4,2,3,2

Solution:

x	x^2
3	9
13	169
11	121
15	225
5	25
4	16
2	4
3	9
2	4
---	---
$\sum x = 58$	$\sum x^2 = 582$

Mean $\bar{x} = \dfrac{\sum x}{n}$

$= \dfrac{3 + 13 + 11 + 15 + 5 + 4 + 2 + 3 + 2}{9}$

$= \dfrac{58}{9}$

$= 6.4444$

Population Variance $\sigma^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n}$

$= \dfrac{582 - \dfrac{(58)^2}{9}}{9}$

$= \dfrac{582 - 373.7778}{9}$

$= \dfrac{208.2222}{9}$

$= 23.1358$

Population Standard deviation $\sigma = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n}}$

$= \sqrt{\dfrac{582 - \dfrac{(58)^2}{9}}{9}}$

$= \sqrt{\dfrac{582 - 373.7778}{9}}$

$= \sqrt{\dfrac{208.2222}{9}}$

$= \sqrt{23.1358}$

$= 4.81$

Co-efficient of Variation (Population) $= \dfrac{\sigma}{\bar{x}} \cdot 100\,\%$

$= \dfrac{4.81}{6.4444} \cdot 100\,\%$

[redacted]2. Calculate Population Variance (σ^2), Population Standard deviation (σ), Population Coefficient of Variation from the following data
85,96,76,108,85,80,100,85,70,95

Solution:

x	$x-\bar{x}=x-88$	$(x-\bar{x})^2$
85	-3	9
96	8	64
76	-12	144
108	20	400
85	-3	9
80	-8	64
100	12	144
85	-3	9
70	-18	324
95	7	49
---	---	---
$\sum x=880$	$\sum(x-\bar{x})=0$	$\sum(x-\bar{x})^2=1216$

Mean $\bar{x}=\frac{\sum x}{n}$

$=\frac{85+96+76+108+85+80+100+85+70+95}{10}$

$=\frac{880}{10}$

$=88$

Population Variance $\sigma^2=\frac{\sum(x-\bar{x})^2}{n}$

$=\frac{1216}{10}$

$=121.6$

Population Standard deviation $\sigma=\sqrt{\frac{\sum(x-\bar{x})^2}{n}}$

$=\sqrt{\frac{1216}{10}}$

$=\sqrt{121.6}$

$=11.0272$

Co-efficient of Variation (Population) $=\frac{\sigma}{\bar{x}}\cdot 100\ \%$

$=\frac{11.0272}{88}\cdot 100\ \%$

$=12.53\ \%$

Examples

1. Calculate Sample Variance $\left(S^2\right)$, Sample Standard deviation (S), Sample Coefficient of Variation from the following data 3,13,11,15,5,4,2,3,2

Solution:

x	x^2
3	9
13	169
11	121
15	225
5	25
4	16
2	4
3	9
2	4
---	---
$\sum x = 58$	$\sum x^2 = 582$

Mean $\bar{x} = \frac{\sum x}{n}$

$= \frac{3 + 13 + 11 + 15 + 5 + 4 + 2 + 3 + 2}{9}$

$= \frac{58}{9}$

$= 6.4444$

Sample Variance $S^2 = \dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n - 1}$

$= \dfrac{582 - \dfrac{(58)^2}{9}}{8}$

$= \dfrac{582 - 373.7778}{8}$

$= \dfrac{208.2222}{8}$

$= 26.0278$

Sample Standard deviation $S = \sqrt{\dfrac{\sum x^2 - \dfrac{(\sum x)^2}{n}}{n - 1}}$

$= \sqrt{\dfrac{582 - \dfrac{(58)^2}{9}}{8}}$

$= \sqrt{\dfrac{582 - 373.7778}{8}}$

$= \sqrt{\dfrac{208.2222}{8}}$

$= \sqrt{26.0278}$

$= 5.1017$

Co-efficient of Variation (Sample) $= \dfrac{S}{\bar{x}} \cdot 100\ \%$

$= \dfrac{5.1017}{6.4444} \cdot 100\ \%$

2. Calculate Sample Variance $\left(S^2\right)$, Sample Standard deviation (S), Sample Coefficient of Variation from the following data
85,96,76,108,85,80,100,85,70,95

Solution:

x	$x - \bar{x} = x - 88$	$\left(x - \bar{x}\right)^2$
85	-3	9
96	8	64
76	-12	144
108	20	400
85	-3	9
80	-8	64
100	12	144
85	-3	9
70	-18	324
95	7	49
---	---	---
$\sum x = 880$	$\sum \left(x - \bar{x}\right) = 0$	$\sum \left(x - \bar{x}\right)^2 = 1216$

Mean $\bar{x} = \dfrac{\sum x}{n}$

$= \dfrac{85 + 96 + 76 + 108 + 85 + 80 + 100 + 85 + 70 + 95}{10}$

$= \dfrac{880}{10}$

$= 88$

Sample Variance $S^2 = \dfrac{\sum \left(x - \bar{x}\right)^2}{n - 1}$

$= \dfrac{1216}{9}$

$= 135.1111$

Sample Variance $S^2 = \dfrac{\sum \left(x - \bar{x}\right)^2}{n - 1}$

$= \dfrac{1216}{9}$

$= 135.1111$

Sample Standard deviation $S = \sqrt{\dfrac{\sum \left(x - \bar{x}\right)^2}{n - 1}}$

$= \sqrt{\dfrac{1216}{9}}$

$= \sqrt{135.1111}$

$= 11.6237$

Co-efficient of Variation (Sample) $= \dfrac{S}{\bar{x}} \cdot 100 \%$

$= \dfrac{11.6237}{88} \cdot 100 \%$

$= 13.21 \%$

Example:

find the quartiles Q1, Q2, and Q3 of the following data 20, 30, 25, 23, 22, 32, 36

Solution:

Arrange data in ascending form, and n = 7 odd number

Ascending arrangement

20
22
23
25
30
32
36

q1 = (1/4) x n
= (1/4) x 7 = 1.75

q1 = 2

Q1 = 22

q2 = (2/4) x n
= (2/4) x 7 = 3.5

q2 = 4

Q2 = 25

q3 = (3/4) x n
= (3/4) x 7 = 5.25

q3 = 6

Q3 = 32

Example:

find the quartiles Q1, Q2, and Q3 of the following data 20, 30, 25, 23, 22, 32, 36, 18

Solution:

Arrange data in ascending form, and n = 8 even number

Ascending arrangement

18
20
22
23
25
30
32
36

q1 = (1/4) x n
= (1/4) x 8 = 2

q1 = mean of (2), and (3)

Q1 = (20+22)x(1/2) = 21

q2 = (2/4) x n
= (2/4) x 8 = 4

q2 = mean of (4), and (5)

Q2 = (23+25)x(1/2) = 24

q3 = (3/4) x n
= (3/4) x 8 = 6

q3 = mean of (6), and (7)

Q3 = (30+32)x(1/2) = 31

Example:

find the desiles D1, D5, and D8 of the following data 20, 30, 25, 23, 22, 32, 36

Solution:

Arrange data in ascending form, and n = 7 odd number

Ascending arrangement

20
22
23
25
30
32
36

d1 = (1/10) x n
= (1/10) x 7 = 0.7

d1 = 1

D1 = 20

d5 = (5/10) x n
= (5/10) x 7 = 3.5

d5 = 4

D5 = 25

d8 = (8/10) x n
= (8/10) x 7 = 5.6

d8 = 6

D8 = 32

Example:

find the desiles D1, D5, and D8 of the following data 20, 30, 25, 23, 22, 32, 36, 18

Solution:

Arrange data in ascending form, and n = 8 even number

Ascending arrangement
18
20
22
23
25
30
32
36

d1 = (1/10) x n
= (1/10) x 8 = 0.8

d1 = 1

D1 = 18

d5 = (5/10) x n
= (5/10) x 8 = 4

d5 = mean of (4), and (5)

D5 = (23+25)x(1/2) = 24

d8 = (8/10) x n
= (8/10) x 8 = 6.4

d8 = 7

D8 = 32

Example:

find the percentiles P8, P50, and P85 of the following data 20, 30, 25, 23, 22, 32, 36

Solution:

Arrange data in ascending form, and n = 7 odd number

Ascending arrangement
20
22
23
25
30
32
36

p8 = (8/100) x n
= (8/100) x 7 = 0.56

p8 = 1

P8 = 20

p50 = (50/100) x n
= (50/100) x 7 = 3.5

p50 = 4

P50 = 25

p85 = (85/100) x n
= (85/100) x 7 = 5.95

p85 = 6

P85 = 32

Example:

find the percentiles P8, P50, and P85 of the following data 20, 30, 25, 23, 22, 32, 36, 18

Solution:

Arrange data in ascending form, and n = 8 even number

Ascending arrangement
18
20
22
23
25
30
32
36

p8 = (8/100) x n
= (8/100) x 8 = 0.64

p8 = 1

P8 = 18

p50 = (50/100) x n
= (50/100) x 8 = 4

p50 = mean of (4), and (5)

P50 = (23+25)x(1/2) = 24

p85 = (85/100) x n
= (85/100) x 8 = 6.8

p85 = 7

P85 = 32

Question

A mutual fund achieved the following rates of growth over an 11-month period:

{3% 2% 7% 8% 2% 4% 3% 7.5% 7.2% 2.7% 2.09%}

Determine the 5^{th} decile from the data.

A. 4%

B. 3%

C. 2%

Solution

The correct answer is B.

First, you should re-arrange the data in ascending order:

{2% 2% 2.7% 2.09% 3% 3% 4% 7% 7.2% 7.5% 8%}

Secondly, you should establish the 5^{th} decile. This is simply the 50^{th} percentile and is actually the median:

$$\begin{aligned} P_{50} &= \frac{(1+11)50}{100} \\ &= 12 * 0.5 \\ &= 6 \text{ i.e. the } 6^{\text{th}} \text{ data point.} \end{aligned}$$

Therefore,

$$\text{the } 5^{\text{th}} \text{ decile} = 50^{\text{th}} \text{ percentile} = \text{median} = 3\%$$

Percentile: the value below which a percentage of data falls.

Example: You are the fourth tallest person in a group of 20

80% of people are shorter than you:

That means you are at the **80th percentile**.

If your height is 1.85m then "1.85m" is the 80th percentile height in that group.

In Order

Have the data **in order**, so you know which values are above and below.

- To calculate percentiles of height: have the data in height order (sorted by height).
- To calculate percentiles of age: have the data in age order.
- And so on.

Grouped Data

When the data is grouped:

Add up all percentages **below** the score,
plus **half** the percentage **at** the score.

Example: You Score a B!

In the test 12% got D, 50% got C, 30% got B and 8% got A

You got a B, so add up

- all the 12% that got D,
- all the 50% that got C,
- half of the 30% that got B,

for a total percentile of 12% + 50% + 15% = **77%**

In other words you did "as well or better than 77% of the class"

(Why take half of B? Because you shouldn't imagine you got the "Best B", or the "Worst B", just an average B.)

Deciles

Deciles are similar to Percentiles (sounds like decimal and percentile together), as they split the data into **10% groups**:

- The **1st decile** is the **10th percentile** (the value that divides the data so **10%** is below it)
- The **2nd decile** is the **20th percentile** (the value that divides the data so **20%** is below it)
- etc!

Example: (continued)

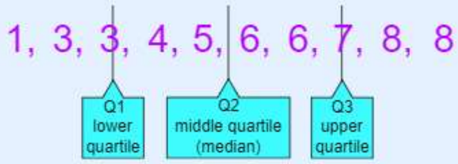
You are at the **8th decile** (the 80th percentile).

Quartiles

Another related idea is Quartiles, which splits the data into quarters:

Example: 1, 3, 3, 4, 5, 6, 6, 7, 8, 8

The numbers are in order. Cut the list into quarters:



In this case Quartile 2 is half way between 5 and 6:

Q2 = (5+6)/2 = **5.5**

And the result is:

- Quartile 1 (Q1) = **3**
- Quartile 2 (Q2) = **5.5**
- Quartile 3 (Q3) = **7**

The Quartiles also divide the data into divisions of 25%, so:

- Quartile 1 (Q1) can be called the **25th percentile**
- Quartile 2 (Q2) can be called the **50th percentile**
- Quartile 3 (Q3) can be called the **75th percentile**

What are the quartiles for the following set of numbers?

8, 11, 20, 10, 2, 17, 15, 5, 16, 15, 25, 6

A $Q_1 = 15, Q_2 = 16, Q_3 = 15.5$

B $Q_1 = 6, Q_2 = 13, Q_3 = 17$

C $Q_1 = 7, Q_2 = 13, Q_3 = 16.5$

D $Q_1 = 8, Q_2 = 13, Q_3 = 16$

What are the quartiles for the following set of numbers?

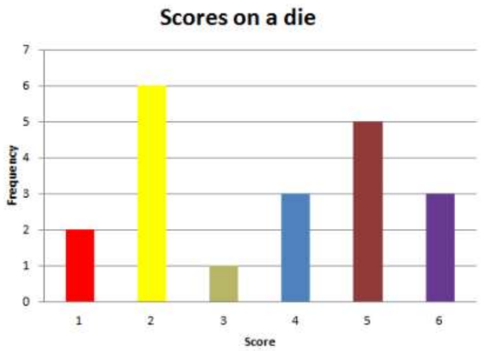
13, 18, 6, 20, 25, 11, 9, 18, 3, 30, 16, 9, 8, 23, 26, 17

A $Q_1 = 22.5, Q_2 = 10.5, Q_3 = 8.5$

B $Q_1 = 9, Q_2 = 16.5, Q_3 = 19$

C $Q_1 = 10, Q_2 = 16.5, Q_3 = 21.5$

D $Q_1 = 9, Q_2 = 16.5, Q_3 = 21.5$



The bar graph shows the scores obtained from 20 throws of a die.

What is the lower quartile?

A 2

B 2.5

C 3

D 5

Jake did a survey of the numbers of brothers and sisters of the children in his class.

He recorded the results as follows:

Number of brothers and sisters	Frequency
0	2
1	4
2	5
3	4
4	3
5	1
8	1

What is the upper quartile?

A 1

B 3

C 3.5

D 4

- Data set:
- 2,2,3,4,5,5,5,6,7,8,8,8,8,8,9,9,10,11,11,12
- What Value Exist at the **percentile** ranking of 25%?

Value $\#= \frac{percentile}{100}(n+1)$

Value $\# = \frac{25}{100}(20+1) = 5.25$

There is no "5.25th", so I take the average of the 5th & 6th values to find what value exist at the 25th percentile.

$$\frac{5+5}{2} = 5$$

Which of the following statements on interquartile range is INCORRECT?

- a. The interquartile range represents the central portion of the distribution
- b. The interquartile range is calculated as the difference between the third quartile and the first quartile
- c. The interquartile range includes about one half of the observations in the set, leaving one-quarter of the observations on each side.
- d. The interquartile range is a quarter of the range (i.e. Maximum - Minimum)

Check Answer

Right! Good job!

Q1 – Lower Quartile

At most, 25% of data is smaller than Q1.

It divides the **lower** half of a data set in half.

Q2 - Median

- The median divides the data set in half.
- 50% of the data values fall below the median and 50% fall above.

Q3 – Upper Quartile

- At most, 25% of data is larger than Q3.
- It divides the **upper** half of the data set in half.

Example…..Use the following. 2,3,5,6,8,10,12,15,18,20

- Notice that the numbers are in ascending order.
- Find P25 (the number in the 25th percentile position.

2,3,5,6,8,10,12,15,18,20

- Find P25……..
- C=np/100
- C = 10(25)/100 = 2.5 ↑
- Round up to the 3rd position.
- 2,3,**5**,6,8,10,12,15,18,20
- The answer is **5**.

Example 2 .....

2,3,5,6,8,10,12,15,18,20

- Find P60:
- C = np/100 = (10)(60)/100 = 6
- Average the numbers in the **6th and 7th position** together.
- 2,3,5,6,8,**10,12**,15,18,20
- (10+12)/2 = 11
- The answer is **11**.

Example 3.....

2,3,5,6,8,10,12,15,18,20

- Note: P75 is the same as Q3.
- C = (10)(75)/100= 7.5 ↑
- Round up to the 8th position.
- 2,3,5,6,8,10,12,**15**,18,20
- The answer is **15**.

Now you try......

- Use the following set of data to answer the next 3 questions.
- 3,4,4,6,8,10,10,12,12,12,13,15,15,15,16, 17,20,22,25,27

- A. Find P75.
- B. Find P30
- C. Find P23

Answer A....

- Find P75.
- C = (20)(75)/100 = 15
- The number will be the average of the 15th and 16th position.
- 3,4,4,6,8,10,10,12,12,12,13,15,15,15,**16**, **17**,20,22,25,27

The answer is (16+17)/2 = **16.5**

Answer B......

- Find P30.
- C = (20)(30)/100 = 6
- Average the **6th and 7th** numbers.
- 3,4,4,6,8,**10,10**,12,12,12,13,15,15,15,16, 17,20,22,25,27
- Answer: (10+10)/2 = **10**

Answer C……

- Find P23.
- C = (20)(23)/100 = 4.6
- Round up to the 5th position.
- 3,4,4,6,**8**,10,10,12,12,12,13,15,15,15,16, 17,20,22,25,27
- The answer is **8**.

Grouping of Data

1. The marks obtained by 40 students of class VIII in an examination are given below:

16, 17, 18, 3, 7, 23, 18, 13, 10, 21, 7, 1, 13, 21, 13, 15, 19, 24, 16, 2, 23, 5, 12, 18, 8, 12, 6, 8, 16, 5, 3, 5, 0, 7, 9, 12, 20, 10, 2, 23

Divide the data into five groups, namely, 0-5, 5-10, 10-15, 15-20 and 20-25, where 0-5 means marks greater than or equal to 0 but less than 5 and similarly 5-10 means marks greater than or equal to 5 but less than 10, and so on. Prepare a frequency table for the grouped data.

Solution:

Arranging the given observations in ascending order, we get them as

0, 1, 2, 2, 3, 3, 5, 5, 5, 6, 7, 7, 7, 8, 8, 9, 10, 10, 12, 12, 12, 13, 13, 13, 15, 16, 16,
16, 17, 18, 18, 18, 19, 20, 21, 21, 23, 23, 23, 24

Thus, the frequency distribution may be given as under:

Marks	Tally Marks	Frequency
0 - 5	𝍸 I	6
5 - 10	𝍸 𝍸	10
10 - 15	𝍸 III	8
15 - 20	𝍸 IIII	9
20 - 25	𝍸 II	7
	Total	40

Note: Here, each of the groups 0-5, 5-10, 10-15, 15-20 and 20-25 is called a class interval.

In class interval 10-15, the number 10 is called the lower limit and 15 is called the upper limit of the class interval.

The difference between the upper limit and the lower limit of any class interval is called the **class size**.

Thus, the class size in the above frequency distribution is 5.

The mid value of a class is called its class mark and is obtained by adding its upper and lower class limits and dividing the sum by 2.

Thus, the class mark of 0-5 is (0 + 5)/2 = 2.5

the class mark of 5-10 is (5 + 10)/2 = 7.5, etc.

Example:
The data below shows the mass of 40 students in a class. The measurement is to the nearest kg.

55	70	57	73	55	59	64	72
60	48	58	54	69	51	63	78
75	64	65	57	71	78	76	62
49	66	62	76	61	63	63	76
52	76	71	61	53	56	67	71

Construct a frequency table for the data using an appropriate scale.

Solution:
Step 1: Find the range.
The **range** of a set of numbers is the difference between the least number and the greatest number in the set.

In this example, the greatest mass is 78 and the smallest mass is 48. The range of the masses is then 78 - 48 = 30. The scale of the frequency table must contain the range of masses.

Step 2: Find the intervals
The **intervals** separate the scale into equal parts.
We could choose intervals of 5. We then begin the scale with 45 and end with 79

Step 3: Draw the frequency table using the selected scale and intervals.

Mass (kg)	Frequency
45 - 49	2
50 - 54	4
55 - 59	7
60 - 64	10
65 - 69	4
70 - 74	6
75 - 79	7

Exercise 2.1

For each of the variables listed below from the line listing in Table 2.1, identify what type of variable it is.

A. Nominal
B. Ordinal
C. Interval
D. Ratio

1. ____ Date of diagnosis
2. ____ Town of residence
3. ____ Age (years)
4. ____ Sex
5. ____ Highest alanine aminotransferase (ALT)

Exercise 2.1

1. C
2. A
3. D
4. A
5. D

Exercise 2.2

At an influenza immunization clinic at a retirement community, residents were asked in how many previous years they had received influenza vaccine. The answers from the first 19 residents are listed below. Organize these data into a frequency distribution.

2, 0, 3, 1, 0, 1, 2, 2, 4, 8, 1, 3, 3, 12, 1, 6, 2, 5, 1

Exercise 2.2

Previous Years	Frequency
0	2
1	5
2	4
3	3
4	1
5	1
6	1
7	0
8	1
9	0
10	0
11	0
12	1
Total	**19**

Another example:

In a certain game, players toss a coin and roll a dice. A player wins if the coin comes up heads, or the dice with a number greater than 4. In 20 games, how many times will a player win?

a. 13
b. 8
c. 11
d. 15

Correct Answer: A

First determine the possible number of outcomes, the sample space of this event will be:

S = { (H,1),(H,2),(H,3),(H,4),(H,5),(H,6) (T,1),(T,2),(T,3),(T,4),(T,5),(T,6) }

So there are a total of 12 outcomes and 8 winning outcomes. The probability of a win in a single event is P (W)

= 8/12 = 2/3. In 20 games the probability of a win = 2/3 × 20 = 13

1. There are 3 blue, 1 white and 4 red identical balls inside a bag. If it is aimed to take two balls out of the bag consecutively, what is the probability to have 1 blue and 1 white ball?

a. 3/28
b. 1/12
c. 1/7
d. 3/7

2. A boy has 4 red, 5 green and 2 yellow balls. He chooses two balls randomly for play. What is the probability that one is red and other is green?

a. 2/11
b. 19/22
c. 20/121
d. 9/11

3. There are 5 blue, 5 green and 5 red books on a shelf. Two books are selected randomly. What is the probability of choosing two books of different colors?

a. 1/3
b. 2/5
c. 4/7
d. 5/7

4. How many different ways can a reader choose 3 books out of 4, ignoring the order of selection?

a. 3
b. 4
c. 9
d. 12

5. There is a die and a coin. The dice is rolled and the coin is flipped according to the number the die is rolled. If the die is rolled only once, what is the probability of 4 successive heads?

a. 3/64
b. 1/16
c. 3/16
d. 1/4

6. Smith and Simon are playing a card game. Smith will win if the drawn card form the deck of 52 is either 7 or a diamond, and Simon will win if the drawn card is an even number. Which statement is more likely to be correct?

a. Smith will win more games.
b. Simon will win more games.
c. They have same winning probability.
d. Decision could not be made from the provided data.

7. A box contains 30 red, green and blue balls. The probability of drawing a red ball is twice the other colors due to its size. The number of green balls are 3 more than twice the number of blue balls, and blue are 5 less than the twice the red. What is the probability that 1st two balls drawn from the box randomly will be red?

a. 10/102
b. 11/102
c. 1/29
d. 1/30

8. Sarah has two children and we know that she has a daughter. What is the probability that the other child is a girl as well?

a. 1/4
b. 1/3
c. 1/2
d. 1

ANSWER KEY

1. A

There are 8 balls in the bag in total. It is important that two balls are taken out of the bag one by one. We can first take the blue then the white, or first white, then the blue. So, we will have two possibilities to be summed up. Since the balls are taken consecutively, we should be careful with the total number of balls for each case:

First blue, then white ball:

There are 3 blue balls; so, having a blue ball is 3/8 possible.
Then, we have 7 balls left in the bag. The possibility to have a white ball is 1/7.

P = (3/8) * (1/7) = 3/56

First white, then blue ball:

There is only 1 white ball; so, having a white ball is 1/8 possible. Then, we have 7 balls left in the bag. The possibility to have a blue ball is 3/7.

P = (1/8) * (3/7) = 3/56

Overall probability is:

3/56 + 3/56 = 3/28

2. A

Probability that the 1st ball is red: 4/11
Probability the 2nd ball is green: 5/10
Combined probability is 4/11 * 5/10 = 20/110 = 2/11

3. D

Assume that the first book chosen is red. Since we need to choose the second book in green or blue, there are 10 possible books to be chosen out of 15 − 1(that is the red book chosen first) = 14 books. There are equal number of books in each color, so the results will be the same if we think that blue or green book is the first book.

So, the probability will be 10/14 = 5/7.

4. B

Ignoring the order means this is a combination problem, not permutation. The reader will choose 3 books out of 4. So,
C(4, 3) = 4! / (3! * (4 − 3)!) = 4! / (3! * 1!) = 4

There are 4 different ways.

Ignoring the order means this is a combination problem, not permutation. The reader will choose 3 books out of 4. So,

C(4, 3) = 4! / (3! * (4 − 3)!) = 4! / (3! * 1!) = 4

There are 4 different ways.

5. A

If the die is rolled for once, it can be 4, 5 or 6 since we are searching for 4 successive heads. We need to think each case separately. There are two possibilities for a coin; heads (H) or tails (T), each possibility of 1/2; we are searching for H. The possibility for a number to appear on the top of the die is 1/6. Die and coin cases are disjoint events. Also, each flip of coin is independent from the other:

Die: 4

coin: HHHH : 1 permutation
P = (1/6) * (1/2) * (1/2) * (1/2) * (1/2) = (1/6) * (1/16)

Die: 5

coin: HHHHT, THHHH, HHHHH : 3 permutations
P = (1/6) * 3 * (1/2) * (1/2) * (1/2) * (1/2) * (1/2) = (1/6) *
(3/32)

Die: 6

coin: HHHHTT, TTHHHH, THHHHT, HHHHHT, THHHHH, HTHHHH, HHHHTH, HHHHHH : 8 permutations
P = (1/6) * 8 * (1/2) * (1/2) * (1/2) * (1/2) * (1/2) * (1/2) = (1/6) * (8/64)

The overall probability is:

Pall = (1/6) * (1/16) + (1/6) * (3/32) + (1/6) * (8/64)
= (1/6) * (1/16 + 3/32 + 8/64)
= (1/6) * (4 + 6 + 8) / 64 = (1/6) * (18/64) = 3/64

7. A

Let the number of red balls be x
Then number of blue balls = 2x − 5
Then number of green balls= 2(2x − 5) + 3 = 4x − 10 + 3

= 4x − 7

As there are total 30 balls so the equation becomes
x + 2x − 5 + 4x − 7 = 30
x = 6
Red balls are 6, blue are 7 and green are 17.
As the probability of drawing a red ball is twice than the others, let's take them as 12. So the total number of balls will be 36.

Probability of drawing the 1st red: 12/36
Probability of drawing the 2nd red: 10/34
Combined probability = 12/36 X 10/34 = 10/102

8. B

At first glance; we can think that a child can be either a girl or a boy, so the probability for the other child to be a girl is 1/2. However, we need to think deeper. The combinations of two children can be as follows:

boy + girl
boy + boy
girl + boy
girl + girl
So, the sample space is S = {BG, BB, GB, GG} where the sequence is important.
Sarah has a girl; this is the fact. So, calling this as event A, here are the possibilities:

boy + girl
girl + boy
girl + girl

Example 1

What is probability of drawing an ace from an ordinary deck of 52 playing cards?

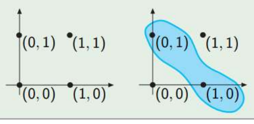
Solution. Since there are $n = 4$ aces among the $N = 52$ cards, the probability of drawing an ace is $\frac{4}{52} = \frac{1}{13}$.

Example 2

If we toss a coin 1000 times and find that it comes up heads 532 times, we estimate the probability of a head coming up to be $\frac{532}{1000} = 0.532$.

Example 4

If we toss a coin twice, the event that only one head comes up is the subset of the sample space $S = \{HH, HT, TH, TT\}$ that consists of points HT and TH. Let we use 0 to represent tails and 1 to represent heads, the sample space can be portrayed by points as in the figure where, for example, $(0, 1)$ represents tails on first toss and heads on second toss, i.e., TH.



Example 5

If someone takes three shots at a target and we care only whether each shot is a hit or a miss, describe a suitable sample space, the elements of the sample space that constitute event M that the person will miss the target three times in a row, and the elements of event N that the person will hit the target once and miss it twice.

Solution. If we let 0 and 1 represent a miss and a hit, respectively, the eight possibilities $(0, 0, 0), (1, 0, 0), (0, 1, 0), (0, 0, 1), (1, 1, 0), (1, 0, 1), (0, 1, 1)$ and $(1, 1, 1)$ may be displayed as in the figure.

Example 7

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

Solution. The sample space is $\mathcal{S} = \{HH, HT, TH, TT\}$. Since we assume that the coin is balanced, these outcomes are equally likely and we assign to each sample point the probability $\frac{1}{4}$. Letting A denote the event that we will get at least one head, we get $A = \{HH, HT, TH\}$ and

$$\begin{aligned} P(A) &= P(HH) + P(HT) + P(TH) \\ &= \frac{1}{4} + \frac{1}{4} + \frac{1}{4} = \frac{3}{4}. \end{aligned}$$

Example 8

A die is loaded 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 on a single roll of the die.

Solution. The sample space is $\mathcal{S} = \{1, 2, 3, 4, 5, 6\}$. Hence, if we assign probability w to each even number and probability $2w$ to each odd number, we find that

$$2w + w + 2w + w + 2w + w = 9w = 1$$

in accordance with Postulate 2. It follows that $w = \frac{1}{9}$ and

$$P(G) = \frac{1}{9} + \frac{2}{9} + \frac{1}{9} = \frac{4}{9}.$$

Example 11

A five-card poker hand dealt from a deck of 52 playing cards is said to be a full house if it consists of three of a hand and a pair. If all the five-card hands are equally likely, what is the probability of being dealt a full house?

Solution.

$$P(A) = \frac{n}{N} = \frac{\binom{4}{3}\binom{4}{2} \cdot 13 \cdot 12}{\binom{52}{5}} = 0.0014$$

Example 17

In a large metropolitan area, the probabilities are 0.86, 0.35, and 0.29 that a family (randomly chosen for a sample survey) owns a LCDTV set, a HDTV set or both kinds of sets. What is the probability that a family owns either or both kinds of sets?

Solution. Let $A = \{\text{a family owns LCDTV}\}$, $B = \{\text{a family owns HDTV}\}$. Since $P(A) = 0.86$, $P(B) = 0.35$, and $P(A \cap B) = 0.29$, thus

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= 0.86 + 0.35 - 0.29 \\ &= 0.92. \end{aligned}$$

Example 18

A card is drawn at random from a pack of well-shuffled playing cards. Find the probability that it is a Spade or a King.

Solution. Let $A = \{\text{a Spade is drawn}\}$, $B = \{\text{a King is drawn}\}$. Since

$$P(A) = \frac{13}{52}, \; P(B) = \frac{4}{52} \text{ and } P(A \cap B) = \frac{1}{52}$$

we have

$$\begin{aligned} P(A \cup B) &= P(A) + P(B) - P(A \cap B) \\ &= \frac{13}{52} + \frac{4}{52} - \frac{1}{52} \\ &= \frac{16}{52} = \frac{4}{13}. \end{aligned}$$

Example 21

Find the probability that a single toss of a die will result in a number less than 4 if **(a)** no other information is given and **(b)** it is given that the toss resulted in an odd number.

Solution. **(a)** Let B denote the event $\{\text{less than } 4\} = \{1, 2, 3\}$. Since B is the union of the events 1, 2, or 3 turning up, we see that

$$P(B) = P(1) + P(2) + P(3) = \frac{1}{6} + \frac{1}{6} + \frac{1}{6} = \frac{3}{6} = \frac{1}{2}$$

assuming equal probabilities for the sample points.
(b) Letting A be the event $\{\text{odd number}\} = \{1, 3, 5\}$, we see that $P(A) = \frac{3}{6}$. Also $P(A \cap B) = \frac{2}{6}$. Then

$$P(B|A) = \frac{P(A \cap B)}{P(A)} = \frac{\frac{2}{6}}{\frac{3}{6}} = \frac{2}{3}.$$

Hence, the added knowledge that the toss results in an odd number raises the probability from 1/2 to 2/3.

Example 22

The probability of a flight departing on time is $P(D) = 0.83$. The probability of a flight arriving on time is $P(A) = 0.82$. We also know that the probability that a flight both departs and arrives on time is $P(D \cap A) = 0.78$. **(a)** What is the probability of a flight arriving on time if we know it departed on time? **(b)** What is the probability of a flight departed on time if we know it arrived on time?

Solution. **(a)** $P(A|D) = \frac{P(A \cap D)}{P(D)} = \frac{0.78}{0.83} = 0.94$.
(b) $P(D|A) = \frac{P(A \cap D)}{P(A)} = \frac{0.78}{0.82} = 0.95$.

Example 24

If we randomly pick 2 television tubes in succession from a shipment of 240 television tubes of which 15 are defective, what is the probability that they will both be defective?

Solution. Let $A = \{\text{the first tube is defective}\}$ and $B = \{\text{the second tube is defective}\}$. Then

$$P(A) = \frac{15}{240} \text{ and } P(B|A) = \frac{14}{239}.$$

Thus, the probability that both tubes will be defective is

$$P(A) \cdot P(B|A) = \frac{15}{240} \cdot \frac{14}{239} = \frac{7}{1,912}.$$

Example 25

Find the probabilities of randomly drawing two aces in succession from an ordinary deck of 52 playing cards is we sample **(a)** without replacement **(b)** with replacement. **(c)** It is given that the first card was an ace. What is the probability that the second card (without replacement) is also an ace?

Solution. **(a)** $\frac{4}{52} \cdot \frac{3}{51} = \frac{1}{221}$.
(b) $\frac{4}{52} \cdot \frac{4}{52} = \frac{1}{169}$.
(c) $\frac{\frac{4}{52} \cdot \frac{3}{51}}{\frac{4}{52}} = \frac{3}{51}$.

Example 27

A box of fuses contains 20 fuses, of which 5 are defective. If 3 of the fuses are selected at random and removed from the box in succession without replacement, what is the probability that all 3 fuses are defective?

Solution. If A is the event that the first fuse is defective, B is the event that the second fuse is defective, and C is the event that the third fuse is defective, then

$$P(A) = \frac{5}{20}, \ P(B|A) = \frac{4}{19}, \text{ and } P(C|A \cap B) = \frac{3}{18},$$

and substitution into the formula yields

$$\begin{aligned} P(A \cap B \cap C) &= P(A) \cdot P(B|A) \cdot P(C|A \cap B) \\ &= \frac{5}{20} \cdot \frac{4}{19} \cdot \frac{3}{18} \\ &= \frac{1}{114}. \end{aligned}$$

Example 29

A coin is tosses three times and the eight possible outcomes, $HHH, HHT, HTH, THH, HTT, THT, TTH$ and TTT, are assumed to be equally likely. If A is the event that a head occurs on each of the first two tosses, B is the event that a tail occurs on the third toss, and C is the event that exactly two tails occur in the three tosses, show that **(a)** events A and B are independent **(b)** events B and C are dependent.

Solution. Since

$$A = \{HHH, HHT\}, \ B = \{HHT, HTT, THT, TTT\},$$

$$C = \{HTT, THT, TTH\}, \ A \cap B = \{HHT\}, B \cap C = \{HTT, THT\}$$

the assumption that the eight possible outcomes are all equiprobable yields

$$P(A) = \frac{2}{8}, \ P(B) = \frac{4}{8}, P(C) = \frac{3}{8}, \ P(A \cap B) = \frac{1}{8}, P(B \cap C) = \frac{2}{8}.$$

$$P(A) = \frac{2}{8}, \ P(B) = \frac{4}{8}, P(C) = \frac{3}{8}, \ P(A \cap B) = \frac{1}{8}, P(B \cap C) = \frac{2}{8}.$$

(a) Since

$$P(A) \cdot P(B) = \frac{2}{8} \cdot \frac{4}{8} = \frac{1}{8} = P(A \cap B),$$

the events A and B are independent.
(b) Since

$$P(B) \cdot P(C) = \frac{4}{8} \cdot \frac{3}{8} = \frac{3}{16} \neq \frac{2}{8} = P(B \cap C),$$

the events B and C are dependent.

Example 30

In the Yahtzee dice game 5 dice are rolled and a winning roll has three of a king and two of a kind (but not five of a kind). **(a)** Compute the probability of winning. **(b)** What is the probability of not getting a single winning roll in 5 tries? **(c)** What is the probability of getting 3 winners in 44 tries?

Solution. (a) $P(win) = \frac{\binom{5}{3}\cdot\binom{2}{2}\cdot 6\cdot 5}{6\cdot 6\cdot 6\cdot 6\cdot 6} = \frac{10\cdot 1\cdot 6\cdot 5}{6^5} = 0.0386.$
(b) Each roll of the 5 dice together is one try. Each try is independent of the other tries. Therefore

$$\begin{aligned} P(lose \cap lose \cap lose \cap lose \cap lose) &= P(lose)\cdot P(lose)\cdots P(lose) \\ &= (P(lose))^5 = (1 - P(win))^5 \\ &= (1 - 0.0386)^5 = 0.82. \end{aligned}$$

(c) Let consider the sequence *win*, *win*, *win*, followed by $44 - 3 = 41$ *loses*. Then, with this order we have

$$\begin{aligned} P(3win, 41lose \text{ with order}) &= (P(win))^3 \cdot (P(lose))^{41} \\ &= 0.0386^3 \cdot (1 - 0.0386)^{41} = 1.14 \times 10^{-5}. \end{aligned}$$

But there are $\binom{44}{3}$ ways to choose three winning rolls in 44 tries.
So,

$$\begin{aligned} P(3win, 41lose \text{ without order}) &= \binom{44}{3} \cdot 1.14 \times 10^{-5} \\ &= 0.15. \end{aligned}$$

Example 32

Find the probabilities of getting **(a)** three heads in three random tosses of a balanced coin; **(b)** four sixes and then another number in five random rolls of a balanced die.

Solution. (a) Multiplying the respective probabilities, we get

$$\frac{1}{2}\cdot\frac{1}{2}\cdot\frac{1}{2} = \frac{1}{8}.$$

(b) Multiplying the respective probabilities, we get

$$\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{1}{6}\cdot\frac{5}{6} = \frac{5}{7,776}.$$

Example 35

The members of a consulting firm rent cars from three rental agencies: 60 percent from agency 1, 30 percent from agency 2, and 10 percent from agency 3. If 9 percent of the cars from agency 1 need a tune-up, 20 percent of the cars from agency 2 need a tune-up, and 6 percent of the cars from agency 3 need a tune-up, what is the probability that a rental car delivered to the firm will need a tune-up?

Solution. If A is the event that the car needs a tune-up, and B_1, B_2, and B_3 are the events that the car comes from rental agencies 1, 2, or 3, we have $P(B_1) = 0.60$, $P(B_2) = 0.30$, $P(B_3) = 0.10$, $P(A|B_1) = 0.09$, $P(A|B_2) = 0.20$, and $P(A|B_3) = 0.06$. Thus,

$$\begin{aligned} P(A) &= P(B_1)\cdot P(A|B_1) + P(B_2)\cdot P(A|B_2) + P(B_3)\cdot P(A|B_3) \\ &= (0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06) \\ &= 0.12. \end{aligned}$$

Example 37

With reference to Example 35, if a rental car delivered to the consulting firm needs a tune-up, what is the probability that it came from rental agency 2?

Solution.

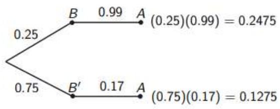
$$\begin{aligned}
P(B_2|A) &= \frac{P(B_2)\cdot P(A|B_2)}{P(B_1)\cdot P(A|B_1) + P(B_2)\cdot P(A|B_2) + P(B_3)\cdot P(A|B_3)} \\
&= \frac{(0.30)(0.20)}{(0.60)(0.09) + (0.30)(0.20) + (0.10)(0.06)} \\
&= \frac{0.060}{0.120} = 0.5
\end{aligned}$$

Observe that although only 30 percent of the cars delivered to the firm come from agency 2, 50 percent of those requiring a tune-up come from that agency.

Example 38

In a certain state, 25 percent of all cars emit excessive amounts of pollutants. If the probability is 0.99 that a car emitting excessive amounts of pollutants will fail the state's vehicular emission test, and the probability is 0.17 that a car not emitting excessive amounts of pollutants will nevertheless fail the test, what is the probability that a car that fails the test actually emits excessive amounts of pollutants?

Solution. Picturing this situation as in the figure



we find the probabilities associated with the two branches of the tree diagram are $(0.25)(0.99) = 0.2475$ and $(1 - 0.25)(0.17) = 0.1275$.

Thus, the probability that a car that fails the test actually emits excessive amounts of pollutants is

$$\frac{0.2475}{0.2475 + 0.1275} = 0.66.$$

Of course, this result could also have been obtained without the diagram by substituting directly into the formula of Bayes' theorem.

Example 12

Find the distribution function of the total of heads obtained in four tosses of a balanced coin.

Solution. Given $f(0) = \frac{1}{16}$, $f(1) = \frac{4}{16}$, $f(2) = \frac{6}{16}$, $f(3) = \frac{4}{16}$, and $f(4) = \frac{1}{16}$ from Example 3, it follows that

$$\begin{aligned}
F(0) &= f(0) = \frac{1}{16}, \\
F(1) &= f(0) + f(1) = \frac{5}{16}, \\
F(2) &= f(0) + f(1) + f(2) = \frac{11}{16}, \\
F(3) &= f(0) + f(1) + f(2) + f(3) = \frac{15}{16}, \\
F(4) &= f(0) + f(1) + f(2) + f(3) + f(4) = 1.
\end{aligned}$$

Hence, the distribution function is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{1}{16} & \text{for } 0 \le x < 1, \\ \frac{5}{16} & \text{for } 1 \le x < 2, \\ \frac{11}{16} & \text{for } 2 \le x < 3, \\ \frac{15}{16} & \text{for } 3 \le x < 4, \\ 1 & \text{for } x \ge 4. \end{cases}$$

Observe that this distribution function is defined not only for the values taken on by the given random variable, but for all real numbers. For instance, we can write $F(1.7) = \frac{5}{16}$ and $F(100) = 1$, although the probabilities of getting at most 1.7 heads or at most 100 heads in four tosses of a balanced coin may not be of any real significance.

Example 13

Find the distribution function of the random variable X of Example 4 and plot its graph.

Solution. Based on the probabilities given in the following table

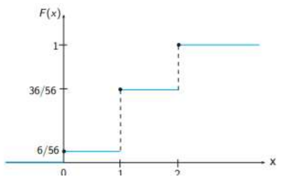
Elements of Sample Space	BB	BG	GB	GG
Probability	20/56	15/56	15/56	6/56
x	2	1	1	0

we can write $f(0) = \frac{6}{56}$, $f(1) = \frac{15}{56} + \frac{15}{56} = \frac{30}{56}$, and $f(2) = \frac{20}{56}$, so that

$$F(0) = f(0) = \frac{6}{56},$$
$$F(1) = f(0) + f(1) = \frac{36}{56},$$
$$F(2) = f(0) + f(1) + f(2) = 1.$$

Hence, the distribution function of X is given by

$$F(x) = \begin{cases} 0 & \text{for } x < 0, \\ \frac{6}{56} & \text{for } 0 \le x < 1, \\ \frac{36}{56} & \text{for } 1 \le x < 2, \\ 1 & \text{for } x \ge 2. \end{cases}$$



Example 14

Find the distribution function of the random variable that has the probability distribution

$$f(x) = \frac{x}{15} \text{ for } x = 1, 2, 3, 4, 5.$$

Solution. Since $f(1) = \frac{1}{15}$, $f(2) = \frac{2}{15}$, $f(3) = \frac{3}{15}$, $f(4) = \frac{4}{15}$, and $f(5) = \frac{5}{15}$, then

$$F(x) = \begin{cases} 0 & \text{for } x < 1, \\ \frac{1}{15} & \text{for } 1 \le x < 2, \\ \frac{3}{15} & \text{for } 2 \le x < 3, \\ \frac{6}{15} & \text{for } 3 \le x < 4, \\ \frac{10}{15} & \text{for } 4 \le x < 5, \\ 1 & \text{for } x \ge 5 \end{cases}$$

Example 16

If X has the distribution function $F(1) = 0.25$, $F(2) = 0.61$, $F(3) = 0.83$, and $F(4) = 1$ for $x = 1, 2, 3, 4$, find the probability distribution of X.

Solution. We have

$$\begin{aligned} f(1) &= F(1) = 0.25, \\ f(2) &= F(2) - F(1) = 0.61 - 0.25 = 0.36, \\ f(3) &= F(3) - F(2) = 0.83 - 0.61 = 0.22, \\ f(4) &= F(4) - F(3) = 1 - 0.83 = 0.17. \end{aligned}$$

Example 17

If X has the distribution function

$$F(x) = \begin{cases} 0 & \text{for } x < -1, \\ \frac{1}{4} & \text{for } -1 \le x < 1, \\ \frac{1}{2} & \text{for } 1 \le x < 3, \\ \frac{3}{4} & \text{for } 3 \le x < 5, \\ 1 & \text{for } x \ge 5. \end{cases}$$

find

- $P(X \le 3)$, $P(X = 3)$, $P(X < 3)$;
- $P(X \ge 1)$;
- $P(-0.4 < X < 4)$;
- $P(X = 5)$;
- the probability distribution of X.

$$F(x) = \begin{cases} 0 & \text{for } x < -1, \\ \frac{1}{4} & \text{for } -1 \le x < 1, \\ \frac{1}{2} & \text{for } 1 \le x < 3, \\ \frac{3}{4} & \text{for } 3 \le x < 5, \\ 1 & \text{for } x \ge 5. \end{cases}$$

Solution.

- $$\begin{aligned} P(X \le 3) &= \frac{3}{4} \\ P(X = 3) &= \frac{3}{4} - \frac{1}{2} = \frac{1}{4} \\ P(X < 3) &= \frac{1}{2} \end{aligned}$$

$$F(x) = \begin{cases} 0 & \text{for } x < -1, \\ \frac{1}{4} & \text{for } -1 \le x < 1, \\ \frac{1}{2} & \text{for } 1 \le x < 3, \\ \frac{3}{4} & \text{for } 3 \le x < 5, \\ 1 & \text{for } x \ge 5. \end{cases}$$

- $P(X \ge 1) = 1 - P(X < 1) = 1 - \frac{1}{4} = \frac{3}{4}$.
- $P(-0.4 < X < 4) = \frac{3}{4} - \frac{1}{4} = \frac{1}{2}$.
- $P(X = 5) = 1 - \frac{3}{4} = \frac{1}{4}$.
- $f(-1) = \frac{1}{4}$, $f(1) = \frac{1}{2} - \frac{1}{4} = \frac{1}{4}$, $f(3) = \frac{3}{4} - \frac{1}{2} = \frac{1}{4}$, $f(5) = 1 - \frac{3}{4} = \frac{1}{4}$, and 0 elsewhere.

Example of Bayes' Theorem

Imagine you are a financial analyst at an investment bank. According to your research of publicly-traded companies, 60% of the companies that increased their share price by more than 5% in the last three years replaced their CEOs during the period.

At the same time, only 35% of the companies that did not increase their share price by more than 5% in the same period replaced their CEOs. Knowing that the probability that the stock prices grow by more than 5% is 4%, find the probability that the shares of a company that fires its CEO will increase by more than 5%.

Before finding the probabilities, you must first define the notation of the probabilities.

- P(A) – the probability that the stock price increases by 5%
- P(B) – the probability that the CEO is replaced
- P(A|B) – the probability of the stock price increases by 5% given that the CEO has been replaced
- P(B|A) – the probability of the CEO replacement given the stock price has increased by 5%.

Using the Bayes' theorem, we can find the required probability:

$$P(A|B) = \frac{0.60 \times 0.04}{0.60 \times 0.04 + 0.35 \times (1 - 0.04)} = 0.067 \text{ or } 6.67\%$$

Thus, the probability that the shares of a company that replaces its CEO will grow by more than 5% is 6.67%.

Problem 1:

Let's work on a simple NLP problem with Bayes Theorem. By using NLP, I can detect spam e-mails in my inbox. Assume that the word 'offer' occurs in 80% of the spam messages in my account. Also, let's assume 'offer' occurs in 10% of my desired e-mails. If 30% of the received e-mails are considered as a scam, and I will receive a new message which contains 'offer', what is the probability that it is spam?

Now, I assume that I received 100 e-mails. The percentage of spam in the whole e-mail is 30%. So, I have 30 spam e-mails and 70 desired e-mails in 100 e-mails. The percentage of the word 'offer' that occurs in spam e-mails is 80%. It means 80% of 30 e-mail and it makes 24. Now, I know that 30 e-mails of 100 are spam and 24 of them contain 'offer' where 6 of them not contains 'offer'.

The percentage of the word 'offer' that occurs in the desired e-mails is 10%. It means 7 of them (10% of 70 desired e-mails) contain the word 'offer' and 63 of them not.

Now, we can see this logic in a simple chart.

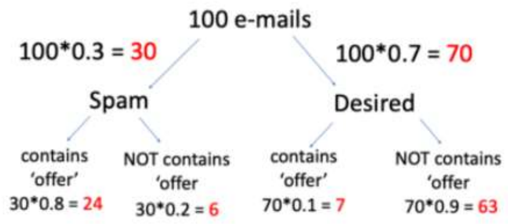


Image by author

The question was what is the probability of spam where the mail contains the word 'offer':

1. We need to find the total number of mails which contains 'offer' ;

24 +7 = 31 mail contain the word 'offer'

2. Find the probability of spam if the mail contains 'offer' ;

In 31 mails 24 contains 'offer' means 77.4% = 0.774 (probability)

NOTE: In this example, I choose the percentages which give integers after calculation. As a general approach, you can think that we have 100 units at the beginning so if the results are not an integer, it will not create a problem. Such that, we cannot say 15.3 e-mails but we can say 15.3 units.

Solution with Bayes' Equation:

A = Spam

B = Contains the word 'offer'

$$P(spam|contains\ offer) = \frac{P(contains\ offer|spam) * P(spam)}{P(contains\ offer)}$$

Image by author

P(contains offer | spam) = 0.8 (given in the question)

P(spam) = 0.3 (given in the question)

Now we will find the probability of e-mail with the word 'offer'. We can compute that by adding 'offer' in spam and desired e-mails. Such that;

P(contains offer) = 0.3*0.8 + 0.7*0.1 = 0.31

$$P(spam|contains\ offer) = \frac{0.8 * 0.3}{0.31} = 0.774$$

Image by author

As it is seen in both ways the results are the same. In the first part, I solved the same question with a simple chart and for the second part, I solved the same question with Bayes' theorem.

Problem 2:

I want to solve one more example from a popular topic as Covid-19. As you know, Covid-19 tests are common nowadays, but some results of tests are not true. Let's assume; a diagnostic test has 99% accuracy and 60% of all people have Covid-19. If a patient tests positive, what is the probability that they actually have the disease?

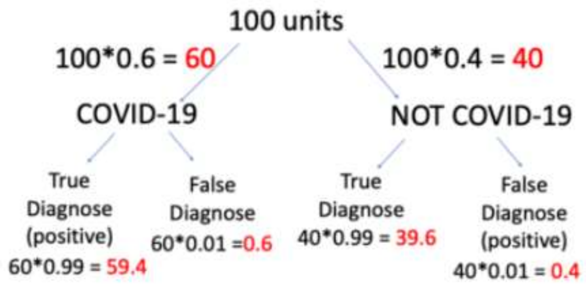


image by author

The total units which have positive results= 59.4 + 0.4 = 59.8

59.4 units (true positive) is 59.8 units means 99.3% = 0.993 probability

With Bayes';

With Bayes';

$$P(covid19|positive) = \frac{P(positive|covid19) * P(covid19)}{P(positive)}$$

image by author

P(positive|covid19) = 0.99

P(covid19) = 0.6

P(positive) = 0.6*0.99+0.4*0.01=0.598

$$P(covid19|positive) = \frac{0.99 * 0.6}{0.598} = 0.993$$

Example 1

One of two boxes contains 4 red balls and 2 green balls and the second box contains 4 green and two red balls. By design, the probabilities of selecting box 1 or box 2 at random are 1/3 for box 1 and 2/3 for box 2.

A box is selected at random and a ball is selected at random from it.

a) Given that the ball selected is red, what is the probability it was selected from the first box?

b) Given that the ball selected is red, what is the probability it was selected from the second box?

c) Compare the results in parts a) and b) and explain the answer.

Solution to

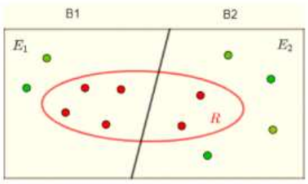
Example 1

Let us call the first box B1 and the second box B2

Let event E1 be "select box 1" and event E2 "select box 2".

Let event R be "select a red ball".

All the above information is included in the diagram below.



The probabilities of selecting one of the two boxes would are given (above) by

and $P(E_2) = 2/3$

The conditional probability that a selected ball is red given that it is selected from

box 2 is given by

$P(R|E_2) = 2/6 = 1/3$, 2 balls out of 6 are red in box 2

a)

The question is to find the conditional probability that the ball is selected from box 1

given that it is red, is given by Bayes' theorem.

$$P(E_1|R) = \frac{P(R|E1)P(E1)}{P(R|E1)P(E1) + P(R|E2)P(E2)}$$

$$= \frac{2/3 * 1/3}{2/3 * 1/3 + 1/3 * 2/3} = 1/2$$

b)

The question is to find the conditional probability that the ball is selected from box 2

given that it is red, is given by Bayes' theorem.

$$P(E_2|R) = \frac{P(R|E2)P(E2)}{P(R|E1)P(E1) + P(R|E2)P(E2)}$$

$$= \frac{1/3 * 2/3}{2/3 * 1/3 + 1/3 * 2/3} = 1/2$$

c)

The two probabilities calculated in parts a) and b) are equal.

Although there are more red balls in box 1 than in box 2 (twice as much), the probabilities calculated above are equal because the probabilities of selecting box 2 is higher (twice as much) than the probability of selecting box 1. Bayes' theorem takes all the information into consideration.

Example: The Art Competition has entries from three painters: Pam, Pia and Pablo

- Pam put in 15 paintings, 4% of her works have won First Prize.
- Pia put in 5 paintings, 6% of her works have won First Prize.
- Pablo put in 10 paintings, 3% of his works have won First Prize.

What is the chance that Pam will win First Prize?

$$P(\text{Pam}|\text{First}) = \frac{P(\text{Pam})P(\text{First}|\text{Pam})}{P(\text{Pam})P(\text{First}|\text{Pam}) + P(\text{Pia})P(\text{First}|\text{Pia}) + P(\text{Pablo})P(\text{First}|\text{Pablo})}$$

Put in the values:

$$P(\text{Pam}|\text{First}) = \frac{(15/30) \times 4\%}{(15/30) \times 4\% + (5/30) \times 6\% + (10/30) \times 3\%}$$

Multiply all by 30 (makes calculation easier):

$$P(\text{Pam}|\text{First}) = \frac{15 \times 4\%}{15 \times 4\% + 5 \times 6\% + 10 \times 3\%}$$

$$= \frac{0.6}{0.6 + 0.3 + 0.3}$$

$$= 50\%$$

A good chance!

Pam isn't the most successful artist, but she did put in lots of entries.

In Exton School, 40% of the girls like music and 24% of the girls like dance.
Given that 30% of those that like music also like dance, what percent of those that like dance also like music?

A 12%	B 50%
C 70%	D 76%

Let A = Like music and B = Like dance

Use Bayes' Theorem:

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

P(A) = 40% = 0.4
P(B) = 24% = 0.24
P(B|A) = 30% = 0.3

Therefore $P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.4 \times 0.3}{0.24} = \frac{0.12}{0.24} = 50\%$

Therefore 50% of those that like dance also like music.

75% of the children in Exton school have a dog, and 30% have a cat.
Given that 60% of those that have cat also have a dog, what percent of those that have a dog also have a cat?

A 24%	B 30%
C $33\frac{1}{3}\%$	D 40%

Let A = Have a cat and B = Have a dog

Use Bayes' Theorem:

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

P(A) = 30% = 0.3
P(B) = 75% = 0.75
P(B|A) = 60% = 0.6

Therefore $P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.3 \times 0.6}{0.75} = \frac{0.18}{0.75} = 24\%$

Therefore 24% of those that have a dog also have a cat.

35% of the children in Exton school have a tablet, and 24% have a smart phone.
Given that 42% of those that have smart phone also have a tablet, what percent of those that have a tablet also have a smart phone?

A 10.08%

B 25%

C 28.8%

D 58%

Let A = Have a smart phone and B = Have a tablet

Use Bayes' Theorem:

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

P(A) = 24% = 0.24
P(B) = 35% = 0.35
P(B|A) = 42% = 0.42

$$\text{Therefore } P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.24 \times 0.42}{0.35} = \frac{0.1008}{0.35} = 28.8\%$$

Therefore 28.8% of those that have a tablet also have a smart phone.

In a factory, machine X produces 60% of the daily output and machine Y produces 40% of the daily output.

2% of machine X's output is defective, and 1.5% of machine Y's output is defective.

One day, an item was inspected at random and found to be defective. What is the probability that it was produced by machine X?

A $\frac{2}{3}$

B $\frac{1}{2}$

C $\frac{1}{3}$

D $\frac{1}{50}$

Let A = The item was produced by machine X and B = An item chosen at random is defective.

Use Bayes' Theorem:

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

P(A) = 60% = 0.6
P(B) = 2% × 60% + 1.5% × 40% = 0.012 + 0.006 = 0.018
P(B|A) = 2% = 0.02

$$\text{Therefore } P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.6 \times 0.02}{0.018} = \frac{0.012}{0.018} = \frac{2}{3}$$

Therefore, the probability the defective item was produced by X $= \frac{2}{3}$

A test for a disease gives a correct positive result with a probability of 0.95 when the disease is present, but gives an incorrect positive result (false positive) with a probability of 0.15 when the disease is not present.

If 5% of the population has the disease, and Jean tests positive to the test, what is the probability Jean really has the disease?

A 0.05

B 0.25

C 0.33

D 0.5

Correct (you have answered this before).

Let A = A patient really has the disease and B = A patient tests positive

Use Bayes' Theorem:

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

P(A) = 5% = 0.05
P(B) = 5% × 0.95 + 95% × 0.15 = 0.0475 + 0.1425 = 0.19
P(B|A) = 0.95

$$\text{Therefore } P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.05 \times 0.95}{0.19} = \frac{0.0475}{0.19} = 0.25$$

Therefore, the probability Jean really has the disease = 0.25

Wire manufactured by a company is tested for strength.
The test gives a correct positive result with a probability of 0.85 when the wire is strong, but gives an incorrect positive result (false positive) with a probability of 0.04 when in fact the wire is not strong.

If 98% of the wires are strong, and a wire chosen at random fails the test, what is the probability it really is not strong enough?

A 0.02

B 0.04

C 0.12

D 0.15

Yes! That is Right!

Let A = A wire really is not strong enough and B = A wire fails the test

Use Bayes' Theorem:
$$P(A|B) = \frac{P(A)P(B|A)}{P(B)}$$

P(A) = 2% = 0.02
P(B) = 98% × 0.15 + 2% × 0.96 = 0.147 + 0.0192 = 0.1662
P(B|A) = 1 − 0.04 = 0.96

$$\text{Therefore } P(A|B) = \frac{P(A)P(B|A)}{P(B)} = \frac{0.02 \times 0.96}{0.1662} = \frac{0.0192}{0.1662} = 0.1155$$

Therefore, the probability a wire that fails the test really is not strong enough = 0.12 correct to two decimal places

A supermarket buys light globes (light bulbs) from three different manufacturers - Brightlight (35%), Glowglobe (20%) and Shinewell (45%).

In the past, the supermarket has found that 1% of Brightlight's globes are faulty, and that 1.5% of each of Glowglobe's and Shinewell's globes are faulty.

A customer buys a globe without looking at the manufacturer's name - in other words, it's a random choice. When she gets home, she finds the globe is faulty.

What is the probability she chose a Shinewell's globe?

Let
A_1 = The globe was a Brightlight's.
A_2 = The globe was a Glowglobe's.
A_3 = The globe was a Shinewell's.

and B = A globe chosen at random is faulty.

Use Bayes' Theorem:
$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)}$$

$P(A_1)$ = 35% = 0.35
$P(A_2)$ = 20% = 0.2
$P(A_3)$ = 45% = 0.45
$P(B|A_1)$ = 1% = 0.01
$P(B|A_2)$ = 1.5% = 0.015
$P(B|A_3)$ = 1.5% = 0.015

$$\text{Therefore } P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3)}$$
$$= \frac{0.45 \times 0.015}{0.35 \times 0.01 + 0.2 \times 0.015 + 0.45 \times 0.015}$$
$$= \frac{0.00675}{0.0035 + 0.003 + 0.00675}$$
$$= \frac{0.00675}{0.01325}$$
$$= 0.509...$$

Therefore, the probability she chose a Shinewell's globe = 0.51 correct to two decimal places

A glazier buys his glass from four different manufacturers - Clearglass (10%), Strongpane (25%), Mirrorglass (30%) and Reflection (35%).

In the past, the glazier has found that 1% of Clearglass' product is cracked, 1.5% of Strongpane's product is cracked, and 2% of Mirrorglass' and Reflection's products are cracked.

The glazier removes the protective covering from a sheet of glass without looking at the manufacturer's name - in other words, it's a random choice. He finds the glass is cracked. What is the probability it was made by Mirrorglass?

Let
A_1 = The glass was from Clearglass.
A_2 = The glass was from Strongpane.
A_3 = The glass was from Mirrorglass
A_4 = The glass was from Reflection.

and B = A glass chosen at random is cracked.

Use Bayes' Theorem:

$$P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + P(A_4)P(B|A_4)}$$

$P(A_1) = 10\% = 0.1$
$P(A_2) = 25\% = 0.25$
$P(A_3) = 30\% = 0.3$
$P(A_4) = 35\% = 0.35$
$P(B| A_1) = 1\% = 0.01$
$P(B| A_2) = 1.5\% = 0.015$
$P(B| A_3) = 2\% = 0.02$
$P(B| A_4) = 2\% = 0.02$

Therefore $P(A_3|B) = \frac{P(A_3)P(B|A_3)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2) + P(A_3)P(B|A_3) + P(A_4)P(B|A_4)}$

$$= \frac{0.3 \times 0.02}{0.1 \times 0.01 + 0.25 \times 0.015 + 0.3 \times 0.02 + 0.35 \times 0.02}$$

$$= \frac{0.006}{0.001 + 0.00375 + 0.006 + 0.007}$$

$$= \frac{0.006}{0.01775}$$

$= 0.3380...$

Therefore, the probability the glass was made by Mirrorglass = 0.34 correct to two decimal places

What is the population standard deviation for the numbers: 75, 83, 96, 100, 121 and 125?

(Try to do this yourself, without using the Standard Deviation calculator.)

A 16.9

B 17.1

C 17.6

D 18.2

Correct (you have answered this before).

1. Firstly find the mean:
Mean = (75 + 83 + 96 + 100 + 121 + 125) ÷ 6 = 600 ÷ 6 = 100

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(75 - 100)^2 + (83 - 100)^2 + (96 - 100)^2 + (100 - 100)^2 + (121 - 100)^2 + (125 - 100)^2$
$= (-25)^2 + (-17)^2 + (-4)^2 + (0)^2 + (21)^2 + (25)^2$
= 625 + 289 + 16 + 0 + 441 + 625
= 1,996

So the Variance = 1,996 ÷ 6 = 332.66...

3. The Standard Deviation is just the square root of the Variance
= √(332.66...)
= 18.2 correct to 1 decimal places

Ten friends scored the following marks in their end-of-year math exam:
23%, 37%, 45%, 49%, 56%, 63%, 63%, 70%, 72% and 82%

What was the standard deviation of their marks?

(Try to do this yourself, without using the Standard Deviation calculator.)

A 15.1%
B 15.5%
C 16.9%
D 18.6%

Correct (you have answered this before).

1. Firstly find the mean:
Mean = (23 + 37 + 45 + 49 + 56 + 63 + 63 + 70 + 72 + 82) ÷ 10 = 560 ÷ 10 = 56

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(23 - 56)^2 + (37 - 56)^2 + (45 - 56)^2 + (49 - 56)^2 + (56 - 56)^2 + (63 - 56)^2 + (63 - 56)^2 + (70 - 56)^2 + (72 - 56)^2 + (82 - 56)^2$
$= (-33)^2 + (-19)^2 + (-11)^2 + (-7)^2 + (0)^2 + (7)^2 + (7)^2 + (14)^2 + (16)^2 + (26)^2$
= 1,089 + 361 + 121 + 49 + 0 + 49 + 49 + 196 + 256 + 676
= 2,846

So the Variance = 2,846 ÷ 10 = 284.6

3. The Standard Deviation is just the square root of the Variance
= √(284.6)
= 16.9 correct to 1 decimal place

A booklet has 12 pages with the following numbers of words:
271, 354, 296, 301, 333, 326, 285, 298, 327, 316, 287 and 314

What is the standard deviation number of words per page?

A 22.6
B 22.0
C 21.9
D 21.4

You got it Right!

1. Firstly find the mean number of words per page:
Mean = (271 + 354 + 296 + 301 + 333 + 326 + 285 + 298 + 327 + 316 + 287 + 314) ÷ 12
= 3,708 ÷ 12
= 309

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(271 - 309)^2 + (354 - 309)^2 + (296 - 309)^2 + (301 - 309)^2 + (333 - 309)^2 + (326 - 309)^2 + (285 - 309)^2 + (298 - 309)^2 + (327 - 309)^2 + (316 - 309)^2 + (287 - 309)^2 + (314 - 309)^2$
$= (-38)^2 + (45)^2 + (-13)^2 + (-8)^2 + (24)^2 + (17)^2 + (-24)^2 + (-11)^2 + (18)^2 + (7)^2 + (-22)^2 + (5)^2$
= 1,444 + 2,025 + 169 + 64 + 576 + 289 + 576 + 121 + 324 + 49 + 484 + 25
= 6,146

So the Variance = 6,146 ÷ 12 = 512.166...

3. The Standard Deviation is just the square root of the Variance
= √(512.166...)
= 22.6 correct to 1 decimal place

Standard Deviation and...

When the answer was revealed they found they had guessed well (and one was the winner!)

Here is how close they each got:

-9, -7, -4, -1, 0, 2, 7, 9, 12

(A negative number shows an underestimate, a positive number shows an overestimate.)

What was the standard deviation of their errors?

(Try to do this yourself, without using the Standard Deviation calculator.)

A 3.9

B 5.5

C 6.2

D 6.8

Correct (you have answered this before).

1. Firstly find the mean:
Mean = $(-9 + -7 + -4 + -1 + 0 + 2 + 7 + 9 + 12) \div 9 = 9 \div 9 = 1$

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(-9 - 1)^2 + (-7 - 1)^2 + (-4 - 1)^2 + (-1 - 1)^2 + (0 - 1)^2 + (2 - 1)^2 + (7 - 1)^2 + (9 - 1)^2 + (12 - 1)^2$
$= (-10)^2 + (-8)^2 + (-5)^2 + (-2)^2 + (-1)^2 + (1)^2 + (6)^2 + (8)^2 + (11)^2$
$= 100 + 64 + 25 + 4 + 1 + 1 + 36 + 64 + 121$
$= 416$

So the Variance = $416 \div 9 = 46.22$

3. The Standard Deviation is just the square root of the Variance
$= \sqrt{(46.22)}$
= 6.8 (to 1 decimal place)

What is the standard deviation of the first 10 natural numbers (1 to 10)?

Use a calculator to help (but don't use the Standard Deviation Calculator).

A 2.45

B 2.87

C 3.16

D 8.25

Correct (you have answered this before).

1. Firstly find the mean:
Mean = $(1 + 2 + 3 + 4 + 5 + 6 + 7 + 8 + 9 + 10) \div 10 = 55 \div 10 = 5.5$

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(1 - 5.5)^2 + (2 - 5.5)^2 + (3 - 5.5)^2 + (4 - 5.5)^2 + (5 - 5.5)^2 + (6 - 5.5)^2 + (7 - 5.5)^2 + (8 - 5.5)^2 + (9 - 5.5)^2 + (10 - 5.5)^2$
$= (-4.5)^2 + (-3.5)^2 + (-2.5)^2 + (-1.5)^2 + (-0.5)^2 + (0.5)^2 + (1.5)^2 + (2.5)^2 + (3.5)^2 + (4.5)^2$
$= 20.25 + 12.25 + 6.25 + 2.25 + 0.25 + 0.25 + 2.25 + 6.25 + 12.25 + 20.25$
$= 82.5$

So the Variance = $82.5 \div 10 = 8.25$

3. The Standard Deviation is just the square root of the Variance
$= \sqrt{(8.25)}$
= 2.87 correct to 2 decimal places

What is the variance of the first 10 numbers of the Fibonacci sequence {0, 1, 1, 2, 3, 5, 8, 13, 21, 34} ?

You can use a calculator to do the sums, but don't use the Standard Deviation calculator.

A 10.47	B 16.88
C 109.56	D 285.01

Correct (you have answered this before).

1. Firstly find the mean:
Mean = $(0 + 1 + 1 + 2 + 3 + 5 + 8 + 13 + 21 + 34) \div 10 = 88 \div 10 = 8.8$

2. Next find the variance. To calculate the Variance, take each difference, square it, and then average the result:
$(0 - 8.8)^2 + (1 - 8.8)^2 + (1 - 8.8)^2 + (2 - 8.8)^2 + (3 - 8.8)^2 + (5 - 8.8)^2 + (8 - 8.8)^2 + (13 - 8.8)^2 + (21 - 8.8)^2 + (34 - 8.8)^2$
$= (-8.8)^2 + (-7.8)^2 + (-7.8)^2 + (-6.8)^2 + (-5.8)^2 + (-3.8)^2 + (-0.8)^2 + (4.2)^2 + (12.2)^2 + (25.2)^2$
$= 77.44 + 60.84 + 60.84 + 46.24 + 33.64 + 14.44 + 0.64 + 17.64 + 148.84 + 635.04$
$= 1{,}095.6$

So the Variance $= 1{,}095.6 \div 10 = 109.56$

The population standard deviation of the numbers 3, 8, 12, 17, and 25 is 7.563 correct to 3 decimal places.

What happens if each of the five numbers is multiplied by 3?

(You may use the standard deviation calculator help link below.)

A The standard deviation remains the same	B The standard deviation is increased by 3
C The standard deviation is multiplied by 3	D The standard deviation is multiplied by 9

You got it Right!

If each number is multiplied by 3, then the mean is also multiplied by 3.

The values of the differences, therefore, are also multiplied by 3
⇒ The values of the squares of the differences are multiplied by 9 (3^2)
⇒ The value of the variance is multiplied by 9
⇒ The value of the standard deviation is multiplied by $\sqrt{9} = 3$

Which one of the following is quantitative data?

A She is black and white.	B She has two ears.
C She has long hair.	D She has a long tail.

Correct (you have answered this before).

Quantitative data is numerical information (numbers).
The only one that is quantitative is B - **two** ears.

Which one of the following is continuous data?

A She has two eyes.

B She has five kittens.

C She weighs 5.4 kg.

D She has four paws.

Excellent ... you are right.

Discrete data can only take certain values (like whole numbers).
Continuous data can take any value (within a range).
The weight of a cat is continuous because it can take any value witin certain limits.

Which one of the following is discrete data?

A She is 45.2 cm long.

B She is 22.3 cm high.

C She weighs 5.4 kg.

D She has 30 teeth.

Yes! That is Right!

Discrete data can only take certain values (like whole numbers).
Continuous data can take any value (within a range).
The number of teeth has to be a whole number, so is discrete.

A census collects information about:

A All members of the population.

B All adult members of the population.

C A large sample of the population.

D A small sample of the population.

Correct (you have answered this before).

A Census is when you collect data for every member of the group (the whole "population").

Which one of the following is NOT quantitative data?

A The snake is 7 feet long

B The snake has two eyes

C The snake is green and yellow

D The snake has no legs

Yes! That is Right!

The snake is green and yellow is qualitative because it is descriptive.

The other three are all quantitative because they tell us about quantity. Even D tells us that the number of legs is zero, so is quantitative.

Which one of the following is discrete data?

A Sam is 160 cm tall

B Sam has two brothers and one sister

C Sam weighs 60 kg

D Sam ran 100 meters in 10.2 seconds

Excellent ... you are right.

Discrete data can only take certain values (like whole numbers)
Continuous data can take any value (within a range)

B is discrete because the numbers of brothers and sisters can only be values like 0, 1, 2 etc...

The other three are all continuous because they can take any value within a range, such as 160.3 cm or 75.35 kg.

A sample collects information about:

A All members of the population.

B All adult members of the population.

C None of the population.

D Some, but not all, of the population.

Only use a calculator if it is not cheating!

Excellent ... you are right.

A Sample is when you collect data just for selected members of the group. This can be some, but not all, of the population.

Rachel rolled a die forty five times with the following results:

Score	Frequency
1	8
2	11
3	4
4	8
5	5
6	9

What was her mean score?

A 3.4

B 3.5

C 3.62

D 7.5

Correct (you have answered this before).

Represent the scores by x and the frequencies by f, and complete the table as follows:

x	f	fx
1	8	8
2	11	22
3	4	12
4	8	32
5	5	25
6	9	54
	$\Sigma f = 45$	$\Sigma fx = 153$

$\sum fx/\sum f = 153/45 = 3.4$

Therefore Rachel's mean score was 3.4

The students in a class of 30 were given a spelling test consisting of 10 words. The numbers of words they each got correct were as follows:

3, 5, 7, 0, 2, 6, 2, 10, 5, 5, 9, 4, 1, 5, 2, 4, 5, 2, 2, 1, 0, 4, 7, 0, 2, 1, 2, 0, 9, 3

Construct a frequency table.

Use your table to calculate the mean score.

A 3.27

B 3.6

C 3.93

D 5.5

Correct (you have answered this before).

The frequency table is constructed using a tally:

Mark	Tally	Frequency
0	IIII	4
1	III	3
2	~~IIII~~ II	7
3	II	2
4	III	3
5	~~IIII~~	5
6	I	1
7	II	2
8		0
9	II	2
10	I	1

Represent the marks by x and the frequencies by f. Complete the frequency table as follows:

x	f	fx
0	4	0
1	3	3
2	7	14
3	2	6
4	3	12
5	5	25
6	1	6
7	2	14
8	0	0
9	2	18
10	1	10
	$\Sigma f = 30$	$\Sigma fx = 108$

$\sum fx/\sum f = 108/30 = 3.6$

Therefore the mean mark was 3.6

Ramiro did a survey of the number of pets owned by his classmates, with the following results:

Number of pets	Frequency
0	4
1	12
2	8
3	2
4	1
5	2
6	1

What was the mean number of pets?

A 1.8

B 2.13

C 2.57

D 30

Correct (you have answered this before).

Represent the number of pets by x and the frequencies by f, and complete the table as follows:

x	f	fx
0	4	0
1	12	12
2	8	16
3	2	6
4	1	4
5	2	10
6	1	6
	$\Sigma f = 30$	$\Sigma fx = 54$

$\sum fx/\sum f = 54/30 = 1.8$

So the mean number of pets was 1.8

The Lakers scored the following numbers of goals in their last twenty matches:
3, 0, 1, 5, 4, 3, 2, 6, 4, 2, 3, 3, 0, 7, 1, 1, 2, 3, 4, 3

Construct a frequency table.

Use your table to calculate the mean score.

A 2.5

B 2.85

C 3.5

D 4

Congratulations, that is the right answer.

The frequency table is constructed using a tally:

No of goals	Tally	Frequency
0	II	2
1	III	3
2	III	3
3	~~IIII~~ I	6
4	III	3
5	I	1
6	I	1
7	I	1

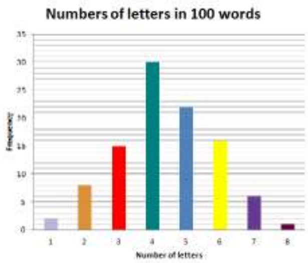
Represent the marks by x and the frequencies by f. Complete the frequency table as follows:

x	f	fx
0	2	0
1	3	3
2	3	6
3	6	18
4	3	12
5	1	5
6	1	6
7	1	7
	$\Sigma f = 20$	$\Sigma fx = 57$

$\sum fx / \sum f = 57/20 = 2.85$

Therefore the mean number of goals scored by the Lakers was 2.85

Olivia chose a 100 word passage and recorded the number of letters in each word. Her results are shown in the following bar graph:



Calculate the mean number of letters per word for this passage.

A 4.29

B 4.39

C 4.49

D 4.5

Correct (you have answered this before).

Complete a Frequency Distribution table for the data shown in the bar graph and calculate the mean from the frequency table:

Number of letters x	Frequency f	fx
1	2	2
2	8	16
3	15	45
4	30	120
5	22	110
6	16	96
7	6	42
8	1	8
	$\Sigma f = 100$	$\Sigma fx = 439$

$\sum fx / \sum f = 439/100 = 4.39$

Therefore the mean number of letters per word = 4.39

A coin is tossed three times. Find the probability of getting at least two heads.

A $\frac{3}{8}$

B $\frac{1}{4}$

C $\frac{3}{4}$

D $\frac{1}{2}$

Well Done!

The set of possible outcomes
= {(H, H, H), (H, T, H), (H, H, T), (T, H, H), (T, H, T), (T, T, H), (H, T, T), (T, T, T)}
There are 8 possible outcomes.

The set of favorable outcomes
= {(H, H, T), (H, T, H), (T, H, H), (H, H, H)}
There are 4 favorable outcomes.

Therefore,
$P(\text{At least two heads}) = \frac{4}{8} = \frac{1}{2}$

A code consists of a two digit number chosen from 00 to 99, followed by two different letters of the alphabet.
What is the probability the code is 12KY?

A $\frac{1}{67,600}$

B $\frac{1}{65,000}$

C $\frac{1}{62,500}$

D $\frac{3}{200}$

Correct (you have answered this before).

The choice of digits and the choice of letters are **independent**

$$\mathrm{P}(12) = \frac{1}{100}$$

and

$$\mathrm{P}(\mathrm{KY}) = \frac{1}{26} \times \frac{1}{25} = \frac{1}{650}$$

$$\text{So } \mathrm{P}(12\mathrm{KY}) = \frac{1}{100} \times \frac{1}{650} = \frac{1}{65,000}$$

Two cards are drawn from the top of a well-shuffled deck. What is the probability that they are both Diamonds?

A $\frac{1}{17}$

B $\frac{1}{16}$

C $\frac{1}{13}$

D $\frac{2}{17}$

Correct (you have answered this before).

For the first card picked you have 13 Diamonds to choose from the 52 cards, so the probability is:

$$\frac{13}{52} = \frac{1}{4}$$

For the next card you have 12 Diamonds to choose from 51 cards since you already picked the other one, so the probability is:

$$\frac{12}{51} = \frac{4}{17}$$

Now just multiply:

$$\frac{1}{4} \times \frac{4}{17} = \frac{1}{17}$$

A bag contains 5 red marbles, 4 green marbles and 1 blue marble.
A marble is chosen at random from the bag and not replaced; then a second marble is chosen. What is the probability that neither marble is blue?

A $\frac{3}{5}$

B $\frac{4}{5}$

C $\frac{8}{9}$

D $\frac{9}{10}$

Correct (you have answered this before).

For the first choice, there are 9 non-blue marbles in the bag out of 10 marbles altogether

$$\text{So P(1st marble is not blue)} = \frac{9}{10}$$

For the second choice, there are 8 non-blue marbles left in the bag out of 9 marbles altogether

$$\text{So P(2nd marble is not blue)} = \frac{8}{9}$$

$$\text{And so P(Both marbles are not blue)} = \frac{9}{10} \times \frac{8}{9} = \frac{8}{10} = \frac{4}{5}$$