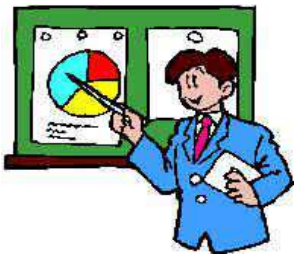
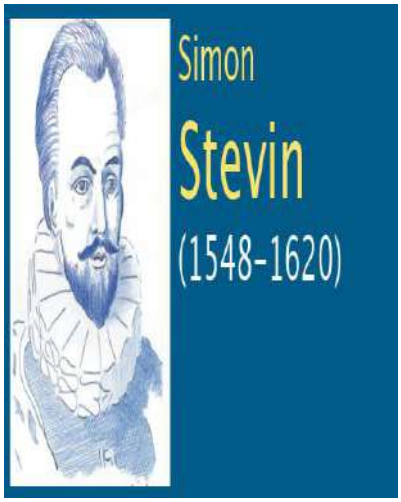
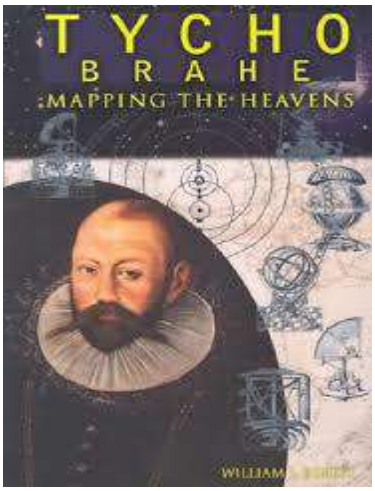
HISTORY OF MEAN, MEDIAN

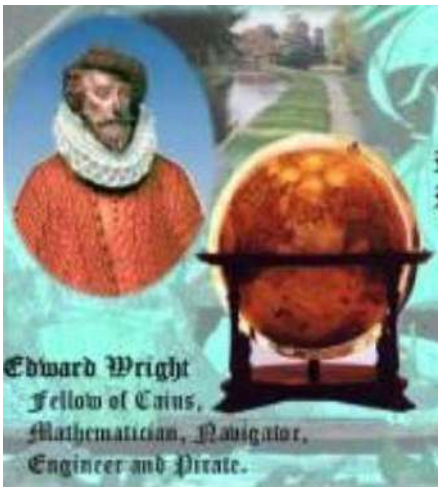
*The arithmetic **mean**, although a concept known to the Greeks, was not generalised to more than two values until the 16th century. The invention of the decimal system by Simon Stevin in 1585 seems likely to have facilitated these calculations. This method was first adopted in astronomy by Tycho Brahe who was attempting to reduce the errors in his estimates of the locations of various celestial bodies.*



*The idea of the **median** originated in Edward Wright's book on navigation (Certaine Errors in Navigation) in 1599 in a section concerning the determination of location with a compass. Wright felt that this value was the most likely to be the correct value in a series of observations.*







Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+mean+and+median

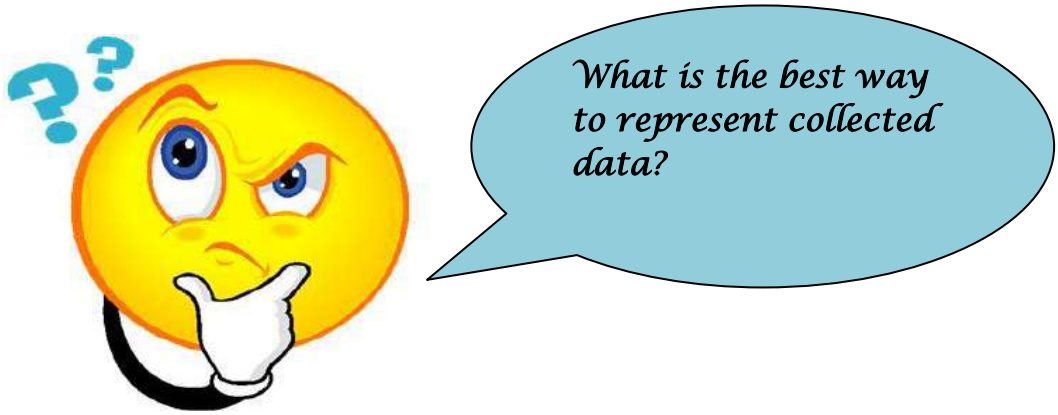
Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images

6.1 Data Representation

LEARNING OUTCOMES

Students should be able to:

- Organise numerical data into a frequency table
- Recognise different types of graphs and how they are sketched
- Identify suitable graphs for various data



We will study various graphical forms of data representation which are often used by statisticians. Some forms of representation are used more commonly than others.

We will now have a look at five different forms of graphs used by statisticians. These are:

- frequency tables
- bar graphs
- line graphs
- pie charts
- Pictograms

A worked example will be given for each one followed by an example for you to try on your own.

a) Frequency tables:

- Data that is collected is generally tabulated in frequency tables. These tables might contain data that is either grouped or ungrouped.
- The type of representation that is used depends on:
    - the nature of the data, i.e., discrete or continuous data
    - the format in which the data is given ungrouped or grouped

Definition

> ***A Frequency Table is a table that lists items and uses tally marks to record and show the number of times they occur.***

Example 6.1

- The following frequency table shows the number of Grade 2 students who chose cat, dog or hamster as their favorite pet.

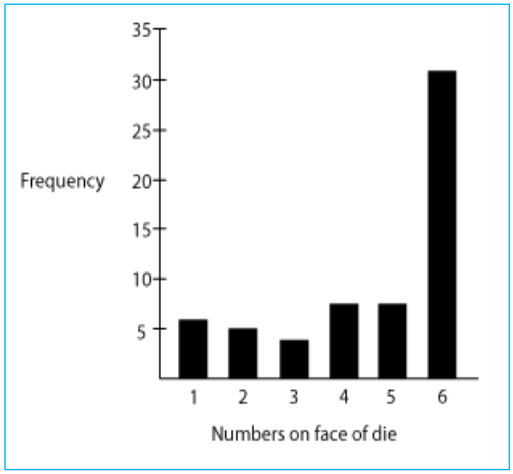
Favorite Pets		
Pet	**Tally Marks**	**Number**
	𝍸 𝍸	10
	||||	4
	𝍸 |	6

b) Bar graphs

A **bar graph** is a graph made of vertical columns or rectangles with equal widths. The height/length represents the frequency of the category.
Here is an example of a bar graph to represent the data from the following table.

Example 6.2

Occurrence of the numbers thrown with a die	
Number	**Frequency**
1	6
2	5
3	4
4	7
5	7
6	32



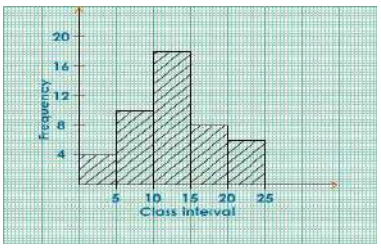
When the bars are joined to each other, the graph is called a **histogram**.

Example 6.3

Draw a histogram for the following data:

Class Interval	Frequency
0 - 5	4
5 - 10	10
10 - 15	18
15 - 20	8
20 - 25	6

Solution:



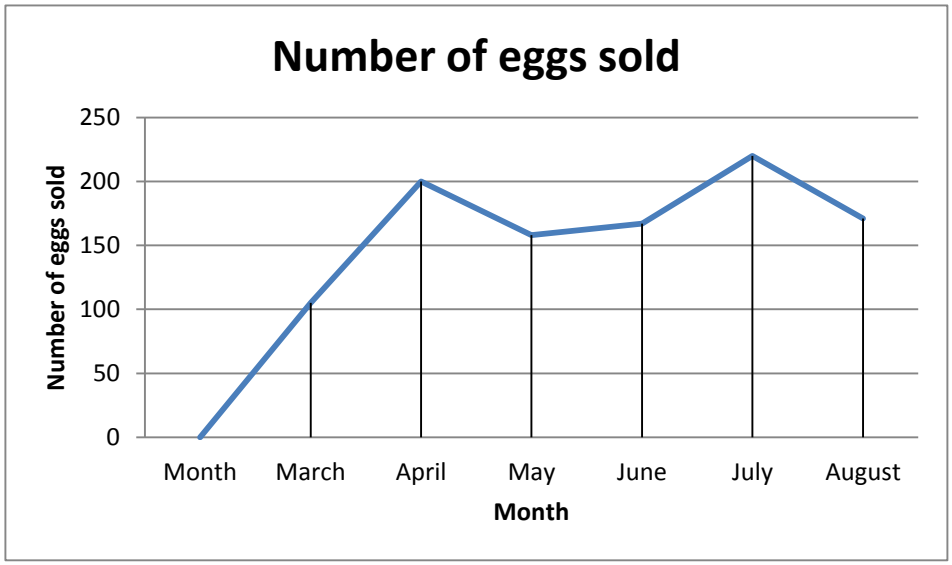
Source: http://math.tutorvista.com/statistics/histogram.html

c) **Line graphs**

A **line graph** is a graph made up of straight line segments joined together between points which are marked according to frequencies.

Month	Number of Eggs sold
March	105
April	200
May	158
June	167
July	220
August	171

Eggs sold at Corner Shop March to August



d) Pie charts

A **pie chart** is a circular representation, a bit like a pie which has been cut up into various sizes of slices. The different segments (the slices) represent the various **relative frequencies**.

The **relative frequency** of a class is the fraction of the total number of data of items belonging to the class. For example, a data set with a total number of *n* items, the relative frequency of each class is given by the following formula.

$$\textit{Relative Frequency} = \frac{\text{number of successful trials}}{\text{total number of trials}}$$

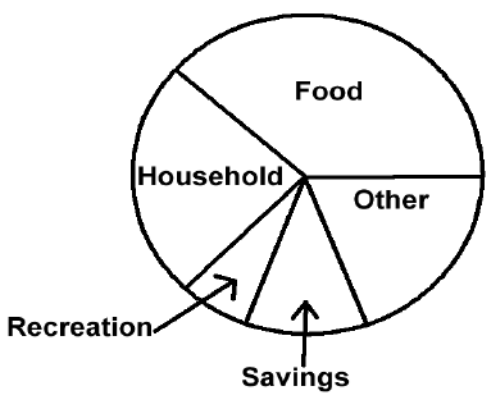
The segments are usually labeled in percentages, and the total of all the percentages should be 100%.

Example 6.4

The pie chart represents amounts spent by a family on various items in the month of September in a certain year. Take note of how the angle sizes of the segments are calculated.

Item	Amount Spent ($)	Amount Spent (%)	Angle Size of Segment
Food	2000	40	$\frac{40}{100} \times 360 = 144$
Household	1250	25	$\frac{25}{100} \times 360 = 90$
Recreation	250	5	$\frac{5}{100} \times 360 = 18$
Savings	500	10	$\frac{10}{100} \times 360 = 36$
Other	1000	20	$\frac{20}{100} \times 360 = 72$

Amounts spent from family budget on various items



e) Pictogram

A **pictogram** looks a like a bar graph. The difference is that the bars are made up of little pictures which represent certain numbers of things as is indicated in the key, which must accompany the pictogram. The picture usually relates in some way to the data being represented.

Below is a frequency table for the number of cars passing through various police check points on 22nd September in a certain year.

Check points	Number of Cars Passed through
A	250
B	520
C	135
D	330
E	405

Below is a pictogram representing the number of cars passing through the Check points A – E:

Checkpoint	No of Cars						
A							
B							
C							
D							
E							

KEY	
	Represents 100 cars
	Represents 10 cars
	Represents 5 cars

Exercise 6.1

1. i. Using the data in the table below, draw a bar graph to illustrate the information. Use graph or grid paper.

Favorite cakes in Class 4B	
Type	**Frequency**
chocolate	15
vanilla	11
carrot	6
coconut	7
apple pie	6
coffee	3

ii. What is an advantage of using a bar graph here than a pie chart?

2. Consider the following information relating to the number of buses going and arriving at various bus stations in Fiji.

City/Town	Number of buses per day
Suva	120
Lautoka	67
Nadi	45
Nausori	90
Ba	55

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+buses

Which would be a better way to represent this in a line graph or a pictogram? Why? (**Show your answer**)

6.2 Measures Of Central Tendency

LEARNING OUTCOMES

Students should be able to:

- Define measures of central tendency
- Calculate measures of central tendency
- Work out measures of central tendency from frequency table
- Identify median from a histogram
- Relate measures of central tendency to real life situations

One of the most simple ways of describing and analyzing data is to use measures of central tendency.

To a statistician, the arithmetic mean is only one of several measures of central tendency to get an idea about the center of a distribution. The idea of the center of a distribution and what it reflects depends on a number of some factors. That is why statisticians believe that there should be other measures of central tendency.

i. Mean or average

A single number, obtained through a mathematical calculation which is very often used to describe a set of data.

To calculate the mean, **we add all of the data items together and then divide that sum by the number of terms which we added.**

Example 6.5

Find the mean of 56, 34, 25, 38, 49, 80 and 73.

To do this you should add 56, 34, 25, 38, 49, 80 and 73 and then divide the sum of these numbers by 7.

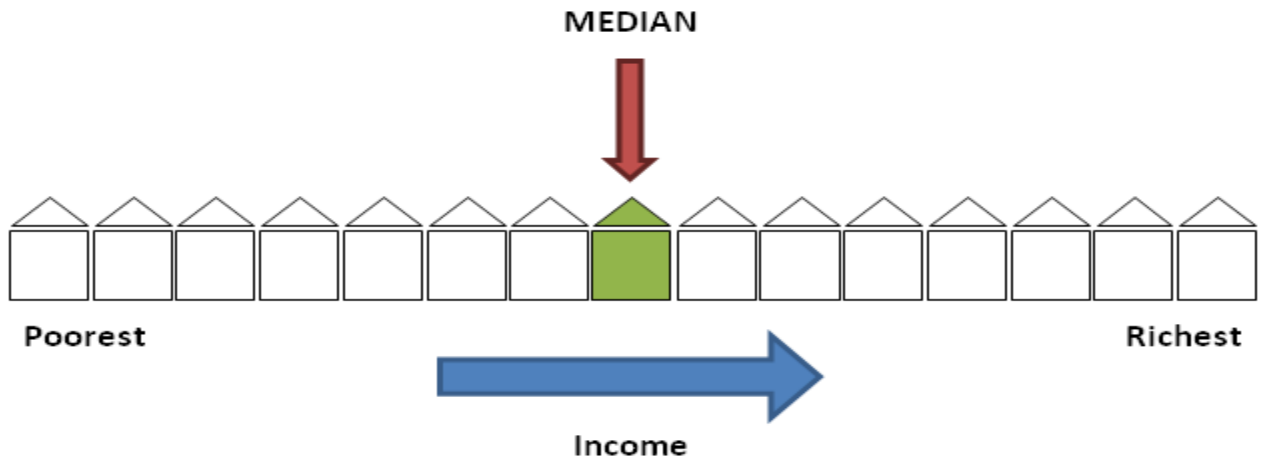
So $56 + 34 + 25 + 38 + 49 + 80 + 73 = \frac{355}{7} = 50.71$

The mean is 50.71.

ii. Median

The median of a set of data is the value which divides the set into two equal numbered parts when the data has been ranked according to size. To rank data, we put it into ascending (or

descending) numerical order. If the number of scores is odd, then the central score will be the median. If the number of scores is even, then the median will be the average of the two central scores.



Source:
https://www.google.com/search?q=images+of+median&biw=1760&bih=843&tbm=isch&tbo=u&source=univ&sa=X&ei=GIfFVNGRNIz-8QXNhoKwCw&ved=0CDMQ7Ak

Example 6.6

Calculate the median of the following data:

2, 4, 9, 10, 13, 15, 19

The data has been ranked (it is in ascending numeric order), there are seven data items (scores) and so 10, which is the central most score (the one in the middle), is the median.

Example 6.7

Calculate the median of the following data:

12, 16, 80, 0, 0, 0, 0, 0, 0, 0, 0, 0

Rank the data: 0, 0, 0, 0, 0, 0, 0, 0, 0, 12, 16, 80.

There is an even number of data items. The two central data items are both zero. The average of these two numbers is zero and so the median is zero.

iii. Mode

The mode is the score that appears the most often. It is the ***most common score***.

Why do you think educators might find the mode a useful measure of central tendency?

Source: https://www.google.com/search?q=images+of+mode+of+scores&biw=1760&bih=843&tbm=isch&tbo=u&source=univ&sa=X&ei=t4jFVPzNM6LdmAXQ9YGwBw&ved=0CDMQ7Ak

Example 6.8

What is the mode of 1, 2, 2, 3, 3, 3, 3, 4 and 15?

The 3 appears more often than any other score and so it is the mode.

What is the mode of 2, 4, 9, 9, 10, 10, 13, 13, 13, 15, 19 and 19?

The 13 appears more often than any other score and so it is the mode.

Data may have two modes, in which case it is called ***bi-modal***.

Exercise 6.2

1. A marathon race was completed by 5 participants in the times given below.

   3.2 hr, 7.1 hr, 5.4 hr, 5.8 hr, 4.7 hr

   What is the mean race time for this marathon?

2. i. Find the mean of the following numbers: 32, 24, 14, 18, 11, 10, 31. Show your working.

   ii. Is the order of numbers important while calculating the mean? Why?

3. If the rainfall in Koro Island in a certain year was 12mm in January, 16 mm in February, 80 mm in March and in the remaining months of the year no rainfall was recorded, calculate the mean rainfall for the Koro Island for that year. Show your working.

4. Calculate the median of: (Show all working)

a) 32, 24, 14, 18, 11, 10 and 31.
b) 50, 65, 35, 60, 40, 90, 50, 48, 63, 27, 68 and 53.
c) Compare your working in a) and b) above. What is different about the two workings? Why is this so?

5. What is the mode of:

1,1,2,2,3,3,3,3,4,4,4,5 and 6?

6. Below is a table showing the Maths marks of Year 9 students in a test

Marks	Frequency
0	2
2	2
3	8
4	**A**
5	2

a. If the mean test mark is 3, find the value of **A.**

b. What is the:

i. Median mark?

ii. Modal mark?

6.2.1 Class Intervals And Histograms

When data does not fall into evenly spaced categories, we can make our own categories. These categories are called class intervals. Class intervals should be equal in width and non-overlapping. A histogram is a frequency graph of class intervals.

To make a frequency distribution table, first divide the numbers over which the data ranges into intervals of equal length. Then count how many data points fall into each interval.

A *histogram* is "a representation of a frequency distribution by means of rectangles whose widths represent class intervals and whose areas are proportional to the corresponding frequencies." Online Webster's Dictionary

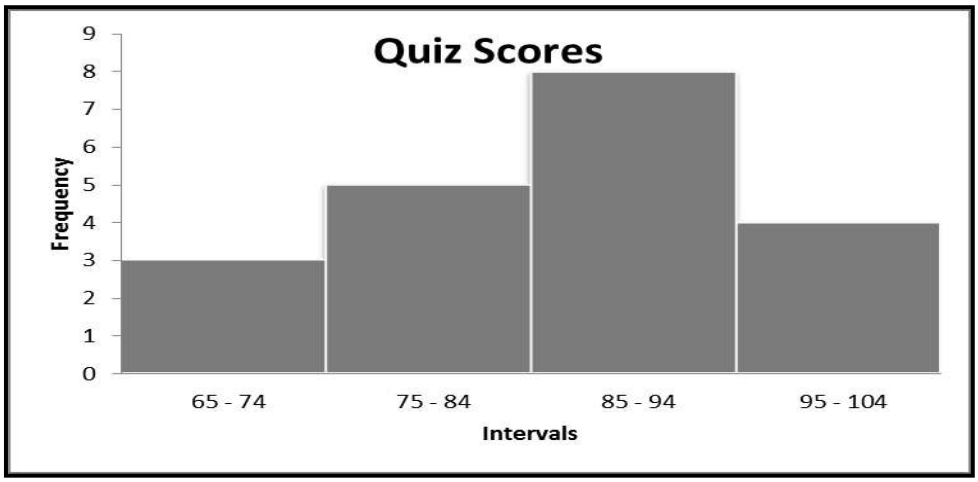
Finding The Mean Range, Mode And Median From Class Intervals.

Example 6.9

Create a frequency table and histogram of the data below on quiz grades.
67, 99, 88, 87, 95, 92, 100, 89, 73, 84, 85, 84, 93, 91, 72, 75, 98, 93, 83, 79

Solution

Intervals	Frequency
65 - 74	3
75 - 84	5
85 - 94	8
95 - 104	4
Total	20



Calculate the range and mean, median and mode

Answers

i. **Range**= Highest score – Lowest score

100 – 65 = 35

ii. **Mean**= $\frac{\text{Sum of Centre of Interval x Frequency}}{\text{Sum of frequency}}$

Intervals	Centre of Interval	Frequency	Total [Centre x Freq]
65 - 74	69.5	3	208.5
75 - 84	79.5	5	397.5
85 - 94	89.5	8	716
95 - 104	99.5	4	398
Total		**20**	**1720**

Mean= $\frac{1720}{20}$ = 86

iii. **Median**= $\frac{\text{Sum of Cumulative Frequency + 1}}{2} = \frac{21}{2} = 10.5$

Intervals	Frequency	Cumulative Frequency
65 - 74	3	3
75 - 84	5	8
85 - 94	8	16 } 10th and 11th scores
95 - 104	4	20

The median is between the 10th and the 11th scores which is in the 85-94 score range so the median interval is 85-94.

iv. **Mode**=The most common score would be shown by the highest bar so the modal interval is 85-94

Exercise 6.3

1. Below are the marks obtained by learners in Year 10, Term 4.

51 64 56 53 72 45 42 46 57 41

51 63 50 45 64 53 48 47 34 48

39 52 36 55 58 46 50 39 48 44

57 47 54 41 36 54 46 44 57 52

49 77 70 49 41 60 54 69 53 65

i. Draw a frequency table for the data.

ii. Show the data on a histogram.

iii. Find the:

- Mean
- Range
- Median
- Mode

2. Apenisa grows two different types of cucumber plants (Type A and Type B) in his greenhouse. One week he keeps a record of the number of cucumbers he picks from each type of plant.

Day	**Mon**	**Tue**	**Wed**	**Thu**	**Fri**	**Sat**	**Sun**
Type A	5	5	4	1	0	1	5
Type B	3	4	3	3	7	9	6

i. Calculate the mean, median and mode for the Type A cucumber plants

ii. Calculate the mean, median and mode of the Type B cucumber plants

iii. Which measure would you argue that there is no difference between type A and Type B plants?

iv. Which measure would you argue that Type A is the best plant?

v. Which measure would you argue that Type B is the best plant?

6.3 Measures of Dispersion

LEARNING OUTCOMES

Students should be able to:

- Identify and describe different measures of dispersion
- Calculate the measures of dispersion from an ungrouped data
- Calculate the measures of dispersion from frequency tables
- Interpret the meaning of numerical values representing measures of dispersion
- Relate the measures of dispersion to real life situations

For a more informative numerical summary, we need some measure of dispersion, sometimes called the *spread*, of observations. This gives us the information of how spread out the values of a data set is.

i. The Range

The ***range*** is defined as the difference between the largest score in the set of data and the smallest score in the set of data.

Formula: Range = Highest Score – Lowest Score

Source:
https://www.google.com/search?q=images+of+range+of+scores&biw=1760&bih=843&tbm=isch&tbo=u&source=univ&sa=X&ei=eIvFVNmnJ4GsmAWvwIHQDQ&ved=0CDEQ7Ak

Example 6.10

What is the range of the following data:
4 8 1 6 6 2 9 3 6 9

The largest score is 9; the smallest score is 1; the range is 9 - 1 = 8

ii. Quartiles and the interquartile range

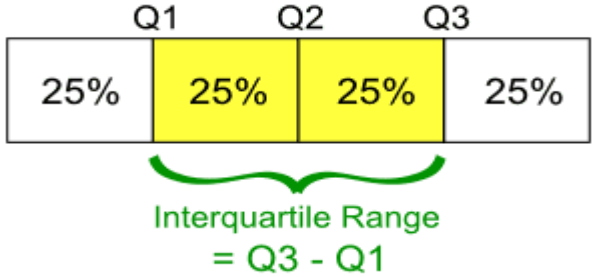
The **median** divides the data into two equal sets, the first (usually to the left of the median) and the second.

- The lower quartile is the value of the middle of the first set, where 25% of the values are smaller than Q_1 and 75% are larger. This first quartile takes the notation Q_1.
- The upper quartile is the value of the middle of the second set, where 75% of the values are smaller than Q_3 and 25% are larger. This third quartile takes the notation Q_3.
- It should be noted that the median takes the notation Q_2, the second quartile

The **interquartile range** is the difference between the lower and upper quartiles. The quartiles split the data into quarters. In other words:

- **lower quartile(Q_1)** = splits lowest 25% of data
- **second quartile** (Q_2) = cuts data set in half = also known as the median
- **third quartile** (Q_3) = **upper quartile** = splits highest 25% of data, or lowest 75%

The difference between the upper and lower quartiles is called the ***interquartile range***

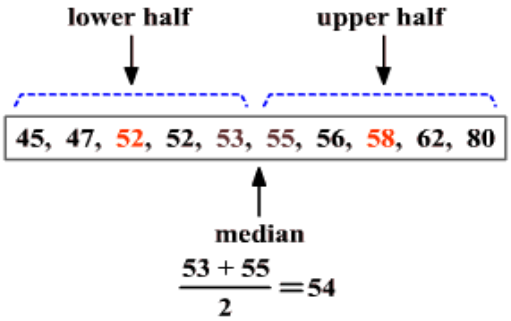


Example 6.11

For the set given below, identify the median, the middle value of the 1st set and the middle value of the 2nd set

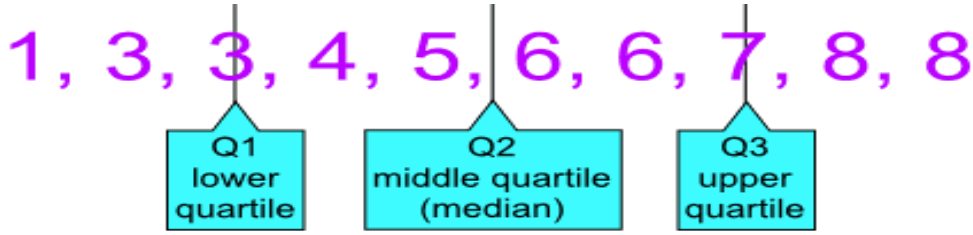
45, 47, 52, 52, 53, 55, 56, 58, 62, 80

Answer



Example 6.12

Identify the median, the upper and lower quartiles



Example 6.13

Find the upper and lower quartile of the following set of data

1, 11, 15, 19, 20, 24, 28, 34, 37, 47, 50, 57

ANSWER

First calculate the median of the data.

Since there is an even number of data values, the median is the mean of the tenth and eleventh values. In other words, the median is:

$$\boxed{\text{Median} \;=} \quad \frac{24 + 28}{2}$$

$$= \quad \mathbf{26}$$

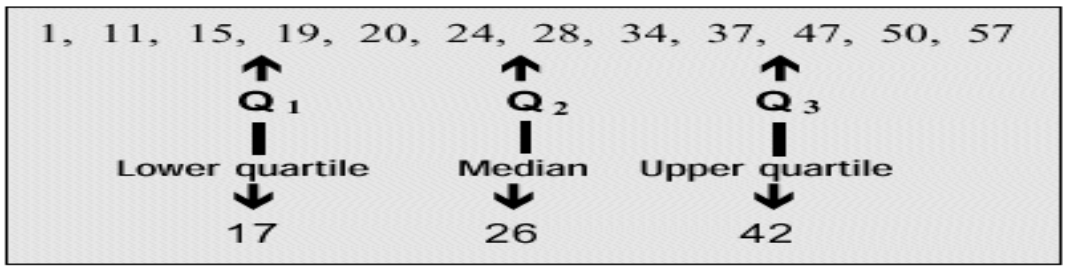
The Lower and Upper Quartiles are as follows

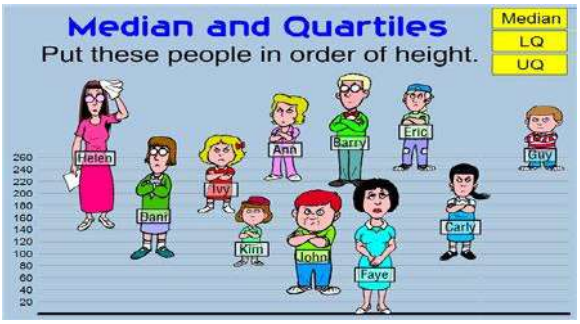
$$\boxed{\text{Lower quartile} \;=} \quad \frac{15 + 19}{2}$$

$$= \quad \mathbf{17}$$

$$\boxed{\text{Upper Quartile} \;=} \quad \frac{37 + 47}{2}$$

$$= \quad \mathbf{42}$$





Example 6.14

7, 1, 3, 6, 3, 7 may be ordered as 1, 3, 3, 6, 7, 7

i. The sample median is : 1, 3, **3, 6**, 7, 7

$= 3 + 6$

$= 9 \div 2 =$ **4 .5**

ii. The upper quartile and lower quartile is:

1**,** **3**, 3, 6, **7**, 7

$\mathbf{Q_1} = 3$ and $\mathbf{Q_3} = 7$

iii. The interquartile range is:

1**,** **3**, 3, 6, **7**, 7

$= 7 - 3 =$ **4**

The standard deviation: A measure that does not require sorting of the data and has good statistical properties is the ***standard deviation*.**

The **Standard Deviation** is a measure of how spread out numbers is.

Its symbol is **σ** (the Greek letter sigma)

Exercise 6.4

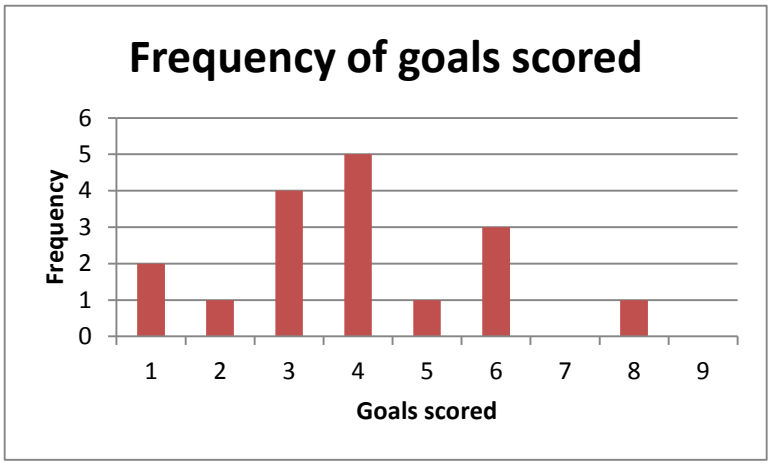
1. The list below shows the prices of soft drinks at Sunshine Supermarket:

   $1.50, $2.00, $3.20,$3.20, $4.00, $4.80, $2.40

   a. Find the median price of a soft drink
   b. Find the range

c. Find the upper and lower quartiles
d. Find the interquartile range.

2. The diagram below shows the frequency of goals scored by a soccer team in a 2014 tournament.

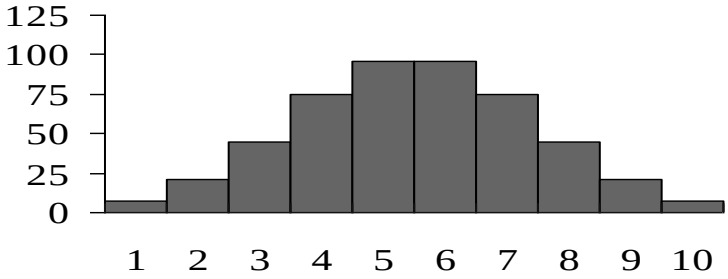


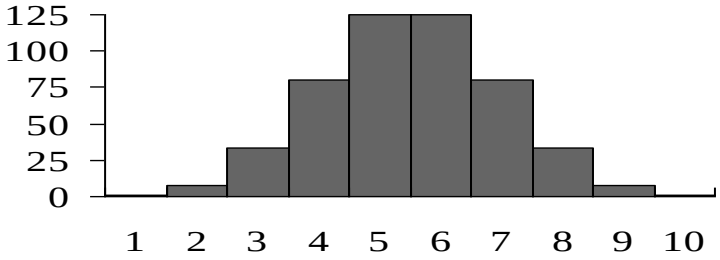
a. How many games did the team play?
b. What was the modal number of goals scored?
c. What was the median?
d. What was the mean?
e. What was the range and the interquartile range?

3. Which of the distributions of scores has the larger dispersion? In Graph I or Graph II? Explain your choice.

Graph I

Graph II





Statistics in Real life Situations

In today's information over-loaded age, statistics is a very useful subject to learn. Newspapers are filled with statistical data and anyone who is ignorant is at a risk of being misled about important real- life decisions as to what to eat, who is leading the polls, the dangers of smoking etc. Knowing a little about statistics will help you to make informed decisions. When used correctly, statistics can also tell us any trends in what happened in the past and can be useful in predicting what may happen in the future.

Exercise 6.5

Find an article from a newspaper or magazine which includes a graphical form of data representation. Read the whole article and study the graph carefully before answering the following questions.

1. What was the investigation about?
2. How is the data represented?
3. Is the presentation used to compare sets of data? If it is, then what is the comparison about?
4. Is the data represented clearly? Give reasons for your answer. Explain any abuses of statistics which you feel are present, if there are any.
5. Discuss the usefulness of the data.
6. What technology (such as computers) do you think was used to collect the data?
7. Will it be a useful exercise for the learners? Why?

PROBABILITY

Introduction

Probability is simply how likely something is to happen. Whenever we're unsure about the outcome of an event, we can talk about the probabilities of certain outcomes (how likely they are). The analysis of events governed by probability is called statistics.

Probability can be expressed as a ratio, a percentage or as a fraction

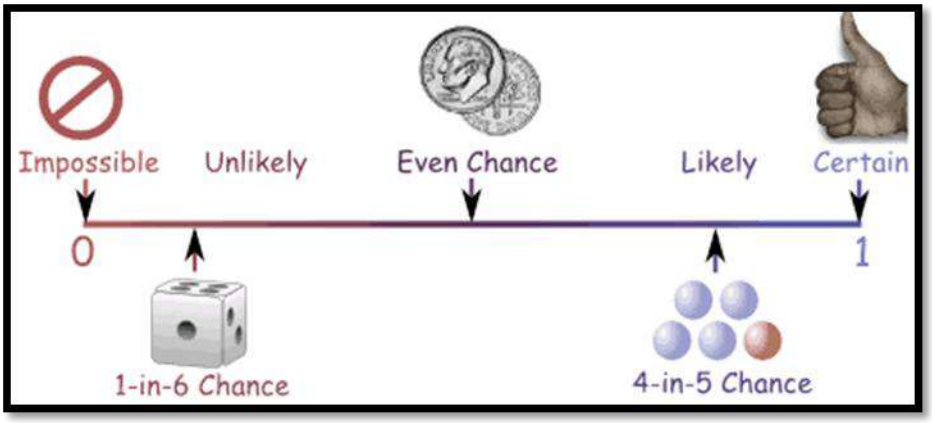
6.4 Probability Experiments

LEARNING OUTCOMES

Students should be able to:

- Define simple probability terms
- Perform simple probability experiments and determine relative frequency outcomes

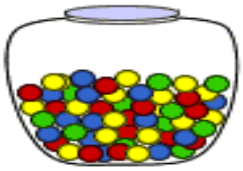
Probability we had said earlier is the likelihood or chance that something will happen. The highest probability an event can have is 1 or 100% and it means that the event will certainly happen. The lowest probability is 0% meaning that the event is certainly not going to happen.



Probability Key Terms

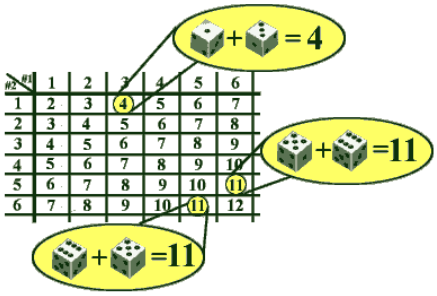
- **Experiment**

  An experiment or a trial in probability is a test to see what will happen in case you do something. A simple example is flipping a coin. When you flip a coin, you are performing an experiment to see what side of the coin you'll end up with.



- **Outcome**

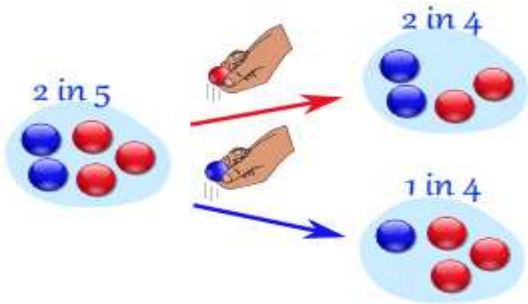
  An outcome in probability refers to a single (one) result of an experiment. In the example of an experiment above, one outcome would be heads and the other would be tails.



Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+an+outcome+in+a+probability

- **Event**

  An event in probability is the set of a group of different outcomes of an experiment. Suppose you flip a coin multiple times, an example of an event would the getting a certain number of heads.



Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+an+event+in+a+probability

- **Sample Space**

A sample space in probability is the total number of all the different possible outcomes of a given experiment. If you flipped a coin once, the sample space **S** would be given by:

$$S = (H, T)$$

If you flipped the coin multiple times, all the different combinations of heads and tails would make up the sample space. A sample space is also defined as a Universal Set for the outcomes of a given experiment.

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+a+sample+space

Exercise 6.6

1. Which of these can you be certain of?Why?
   i. Your height
   ii. Cost of your fare to school
   iii. Exact time of arrival of your school bus
2. Write the following probabilities as percentages:
   i. 0.35
   ii. 2/3
   iii. 1 out of 5
3. List the sample space for:
   i. Throwing a single die
   ii. Throwing two dice
   iii Tossing a coin

Basic Concepts Of Probability

i. ***We can use probability to make decisions.***

For example:

The weather report might say that the probability of rain is 40%. This means that there is a 40% chance that it will rain and a 60% chance that it won't (40% + 60% = 100%). This probability will help us to make decisions regarding activities like picnics or games to organize for the day.

ii. ***We can use probability to make predictions.***

For example:

The probability of tossing a single coin and getting a tail is 0.5. If we toss the coin 30 times how many times will we get tails?

Probability of tails x Number of tosses = Number of tails

0.5 x 30 = 15 tails

iii. ***We can estimate probability from observations.***

Example 6.15

In April 2014, the Levuka Islanders counted 17 sunny days out of a total of 30. They assumed then that the probability of a sunny day in April 2015 will be the same as the percentage of sunny days last April.

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+sunny+day

$$\frac{Sunny\ day}{Total\ days} = \frac{17}{30} = 0\ 57 = 56.7\%$$

Do you think the islanders are correct in saying that the weather in the future will be like in the past?

They are correct in assuming the weather will be like the previous years but they are not correct in counting the number of sunny days from just one year. It would be better to take the number of sunny days over several years.

$$\frac{\text{Sunny days in April over 10 years}}{\text{Total days in April over 10 years}} = \frac{120}{300} = 0.4 = 40\%$$

The more data we use for our estimation the better it will be.

iv. ***We can find the probability of events by calculating the relative frequency.***

Relative Frequency is the number of successful trials divided by the total number of trials.

Relative frequency = no. of successful trials / Total count

The above equation expresses relative frequency as a *proportion.* It is also often expressed as a percentage. Thus, a relative frequency of 0.40 is equivalent to a percentage of 40%.

Example 6.16

Tossing a coin

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+tossing+a+coin+10+times

Raju tosses a coin 20 times and gets a total of 14 heads. The relative frequency of the head is:

Relative frequency = no. of successful trials / Total count

= 14/ 20 = 0.7

Example 6.17

Rolling a Die

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+rolling+dice

a. Sera did an experiment in which she threw a single die 50 times. She found that the relative frequency of a “five” was 0.2. How many “fives” did she get:

Relative frequency = no. of successful trials / Total count

$$0.2 = x / 50$$

$$x = 10 \text{ “fives”}$$

b. If Sera throw the same die 150 times, how many “fives” should she expect?

Relative frequency = no. of successful trials / Total count

$$0.2 = x / 150$$

$$x = 30 \text{ “fives”}$$

Example 6.18

Drawing a card

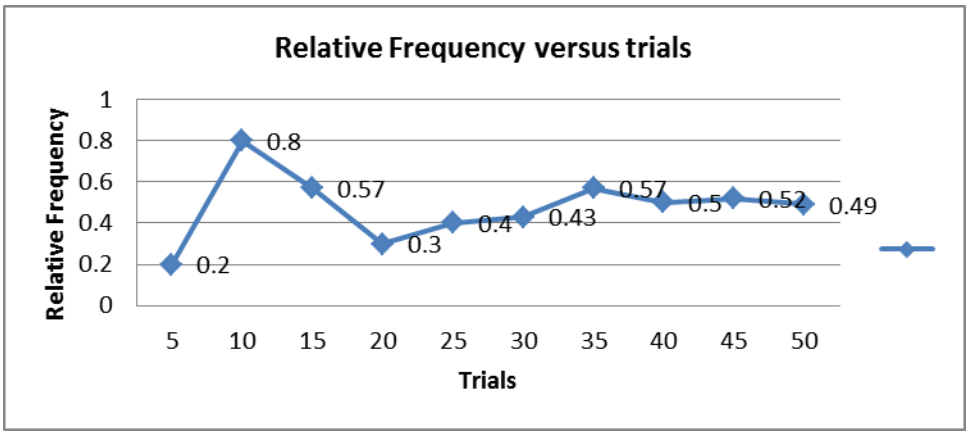
Mani did an experiment to find the probability of drawing a black card from a standard deck of 52 cards.

The table below shows the data of the relative frequencies in various trials:

Trials	Relative Frequencies
5	0.2
10	0.8
15	0.57
20	0.3
25	0.4
30	0.43
35	0.57
40	0.50
45	0.52
50	0.49

a. What is the limit of relative frequency?

We plot the data given above on a graph.



The graph above shows that the relative frequency starts very low then goes very high and then comes closer and closer to 0.5. If we keep increasing the number of trials, it will keep getting closer to 0.5. So, the limit of relative frequency is 0.5.

b. What is the probability of drawing a black card?
The probability of an event is equal to the limit of its relative frequency.
So, the probability of drawing a black card is 0.5.

Exercise 6.7

1.a) If you toss a fair coin, the probability of getting "heads" is 0.5

i. Write this probability as a percentage.

ii. If John tosses the coin 28 times how many times would you expect to get heads?

iii. Will you always get tails this many times when anyone does this experiment? Explain.

2. When a child is born the probability that it will be a girl is equal to that it will be a boy: both probabilities are 0.5

i. If 300 babies were born in CWMH how many would you expect to be girls?

ii. How many would you expect to be boys?

3. The probability of getting a surprise Accounting test in Year 10 is 0.25.

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+students+sitting+tests

If there are 180 school days in the year, how many surprise tests can students expect?

Is it certain that students will have this many surprise tests? Explain.

4. The table below shows the percentage of Rosi High School Year 10 students who passed their mathematics test over the last 5 years.

Year	1	2	3	4	5
Pass rate	90%	99%	100%	87%	91%

i. Find the average pass rate of students.

ii. What is the probability of a randomly selected student passing this year?

iii. If there are 260 students in Year 10, how many are likely to pass?

5. Some students in Year 10 investigated the probability of a box landing on its side as shown below.

i. Use the data below to calculate the relative frequency for each set of toss.

Trials	Side	Trials	Side
5	3	30	8
10	2	35	18
15	0	40	6
20	9	45	10
25	14	50	19

ii. On graph paper draw the graph of the relative frequency.

iii. What seems to be the limit of relative frequency?

iv. What is the probability of the matchbox landing on its side?

v. Is this reasonable? Explain in terms of the shape of the matchbox.

6.5 Events Of Probability

LEARNING OUTCOMES

Students should be able to:

- Identify and define the different types of events
- Describe the probability

Life is full of random events! We need to get a "feel" for them to be smart and successful people.

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+life

The throw of a dice, toss of a coin and lottery draws are all examples of random events.

When we say "Event" we mean one (or more) outcomes.

An event therefore is defined as any outcome that can occur.

Types Of Probability Events

An event can include several outcomes:

- Choosing a "King" from a deck of cards (any of the 4 Kings) is an event
- Rolling an "even number" (2, 4 or 6) is **also** an event
- Getting a Tail when tossing a coin is an event
- Rolling a "5" is an event.

Let's look at each of those types of events.

- **Random event**

In a random event all the outcomes have the same probability.

Example 6.19

Tossing a coin is a random event because “heads” and “tails” both have the same chance of occurring.

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=IMAGES+OF+FIJI+COINS

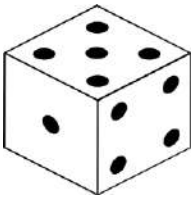
- **Equally likely outcome**

Equally likely outcomes are outcomes that have the same probability of occurring.

Source: http://amsi.org.au/teacher_modules/Chance_years_1-3.html

Example 6.20

A single die is rolled*:*



What is the probability of each outcome?

What is the probability of rolling an even number? of rolling an odd number?

Outcomes: The possible Outcomes of this experiment are 1, 2, 3, 4, 5 and 6

Probabilities:

$$P(1) = \frac{\text{\# of ways to roll a 1}}{\text{total \# of sides}} = \frac{1}{6}$$

$$P(2) = \frac{\text{\# of ways to roll a 2}}{\text{total \# of sides}} = \frac{1}{6}$$

$$P(3) = \frac{\text{\# of ways to roll a 3}}{\text{total \# of sides}} = \frac{1}{6}$$

$$P(4) = \frac{\text{\# of ways to roll a 4}}{\text{total \# of sides}} = \frac{1}{6}$$

$$P(5) = \frac{\text{\# of ways to roll a 5}}{\text{total \# of sides}} = \frac{1}{6}$$

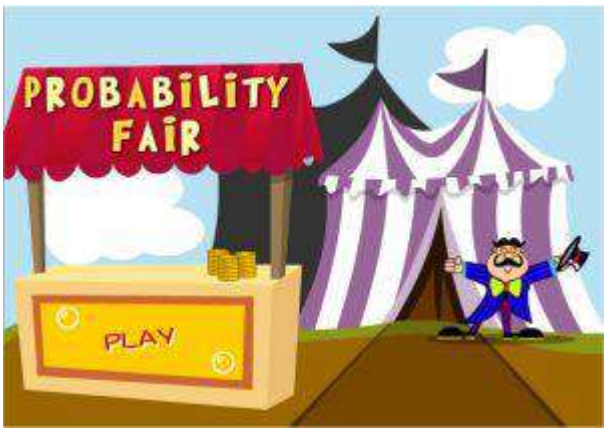
$$P(6) = \frac{\text{\# of ways to roll a 6}}{\text{total \# of sides}} = \frac{1}{6}$$

In this experiment the probability of rolling each number on the die is always one sixth. We say that the outcomes are equally likely to occur.
Knowing that the above are equally likely outcomes we can calculate their probabilities without doing the experiment i.e the probability of anyone of them is equal to 1 divided by the number of outcomes:

$$1 \div 6 = \frac{1}{6}$$

- **Fair event**

Tossing a fair coin is a random event. This means that all the outcomes are equally likely.



Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+fair+events

- **Biased event**

This means that the probabilities of the outcomes are not equal.

.

"With careful and prolonged planning, we may reduce or eliminate many potential sources of bias, but seldom will we be able to eliminate all of them. Accept bias as inevitable and then endeavor to recognize and report all exceptions that do slip thought the cracks."

Good and Hardin (2006) *Common Errors in Statistics (and How to Avoid Them)*, p. 113

Biased samples can occur two ways:

1. If the sample selected is not representative of the population to be surveyed.

Example:

The city council is going to vote on whether to keep its leash laws.

Sample 1 ~ survey dog owners
Sample 2 ~ survey people who don't have pets
Sample 3 ~ surveying some people who own dogs and some who don't.

Which sample shows the least bias?

Answer

Sample 3 ~ equal chances of yes or no answers

2. If the survey questions are worded in ways meant to lead the people surveyed to

Example:

Wording 1 ~ Should the mayor approve a bill requiring dogs that are outdoors to be on leashes at all times, instead of running over people's property?

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+dogs

Wording 2 ~ Should the mayor approve a bill requiring dogs that are outdoors to be on leashes at all times, instead of running freely to exercise?

Wording 3 ~ Should the mayor approve a bill requiring dogs that are outdoors to be on leashes at all times?

Which wording shows the least bias?

Answer

Wording 3 ~ the other two wordings make it sound like either a bad idea or good idea to let the dogs run freely.

Exercise 6.8

1. Identify more likely, less likely, equally likely, sure and impossible events:

i. Selection of a white ball from a box with 5 white balls, 8 red balls and 10 yellow balls.
ii. Selection of a black card from a deck of cards
iii. Occurrence of even number when a die is rolled.
iv. Selection of red marble from a box with 12 red marbles.
v. Selection of red marble from a box with 12 white balls.
vi. Selecting a boy for a field trip from a group of 35 students with 12 girls.

2. A fair dice is tossed:

i. Write down the set of possible outcomes.

ii. Are the outcomes equally likely?

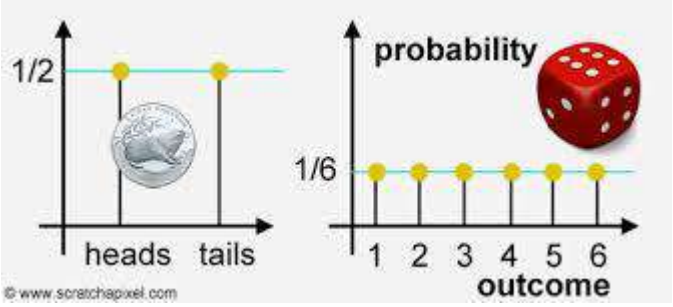
iii. What must the sum of the probabilities of all the possible outcomes

iv. Find the probability of each of the outcomes listed in iii.

3. State whether or not a sample is biased. If so, identify the source of the bias in the following scenarios.

(a). A movie theater surveyed patrons 45 years of age and older and asked what type of movie they prefer.
(b). The student body wants to survey the 8th grade class to choose a theme for the winter dance. They surveyed Mr. Walker's 3rd period art class.
(c). The patrons of a pet store were asked what their favorite type of pet is.
(d). Students are asked, "Are you willing to waste weekend time volunteering at an animal shelter?"
(e). Parents are asked, "Should we develop hand-eye coordination in our children by teaching them how to play computer games?"

Determining Probability From An Experiment



LEARNING OUTCOME

Students should be able to:

- Determine the probability from an experiment consisting of equally likely outcomes for example; drawing a colored ball from a box.

Experimental probability refers to the probability of an event occurring when an experiment was conducted.

For example, if a dice is rolled 6000 times and the number '5' occurs 990 times, then the experimental probability that '5' shows up on the dice is 990/6000 = 0.165.

On the other hand, theoretical probability is determined by noting all the possible outcomes theoretically, and determining how likely the given outcome is. Mathematically,

Exercise 6.9

1. A bag contains 7 red marbles, 3 blue marbles, and 1 green marble. What is the probability of choosing a marble that is not blue?

   A. 11/8 B. 3/11

   C. 3/8 D. 8/11

2. A bag contains 6 red, 3 blue, 2 green and 1 white balls. A ball is picked at random. What is the probablity that it is:
   1. red
   2. blue or white
   3. not green
   4. not green or red?

3. A card is selected randomly from a pack of 52. What is the probability that it is:
   1. the 2 of hearts
   2. a red card
   3. a picture card
   4. an ace
   5. a number less than 4?

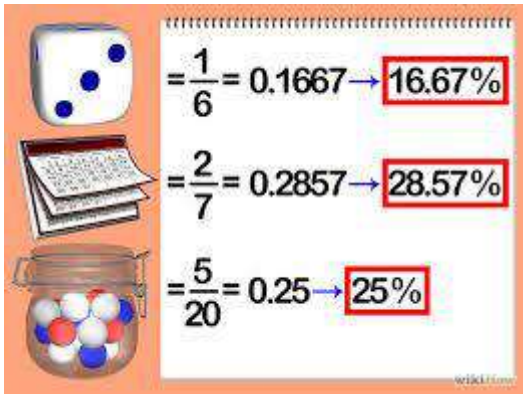
4. Even numbers from 2 -100 are written on cards. What is the probability of selecting a multiple of 5, if a card is drawn at random?

6.6 Probability Formula

LEARNING OUTCOMES

Students should be able to:

- Solve the probability of an event P (E)
- Determine the expected number of occurrences in a sequence of independent trials



Probability formula is the ratio of number of favorable outcomes to the total number of possible outcome.

Probability of an event X is symbolized by P(X). Probability of an event X is lies between $0 \leq P(X) \leq 1$.

Probability of an event = (number of ways it can happen) / (total number of outcomes)

P(X) = (no. of ways X can happen) / (Total number of outcomes)

How to find the probability:

Step 1: List the outcomes of the experiment.

Step 2: Count the number of possible outcomes of the experiment.

Step 3: Count the number of favorable outcomes.

Step 4: Use the probability formula.

Example 1: Rolling dice

When rolling a single die, there are six different outcomes: {1,2,3,4,5,6}

i. What's the probability of rolling a '1'?

ii. What's the probability of rolling a '1' or a '6'?

iii. What's the probability of rolling an even number? (i.e. rolling a 2, 4 or 6)

Using the formula from above:

P(A) = (no. of ways A can happen) / (Total number of outcomes)

- The probability of rolling a '1' is $\frac{1}{6}$
- The probability of rolling a '1' or a '6' is $\frac{2}{6}$ or $\frac{1}{3}$
- The probability of rolling an even number is $\frac{3}{6}$

Example 6.21

Two dice are rolled once. Calculate the probability that the sum of the numbers on the two dice is 6.

Solution:
Possible outcomes (Sample Space) = {(1, 1), (1, 2),.....,(1, 6), (2, 1), (2, 2),.....,

(2, 6), (3, 1), (3, 2),.....,(3, 6), ...,(4, 1), (4, 2),.....,(4, 6), (5, 1), (5,2),.....,(5, 6), (6, 1), (6, 2),.....,(6, 6)}

Total possible outcomes = **36**

No. of outcomes of the experiment that are favorable to the event that a sum of two events is 6

=> Favorable outcomes are: (1, 5), (2, 4), (3, 3), (4, 2) and (5, 1)

Number of favorable outcomes = **5**

Use, probability formula = *Number of favorable outcomes/Total number of possible outcomes*

$$= \frac{5}{36}$$

The probability of a sum of 6 is $\frac{5}{36}$

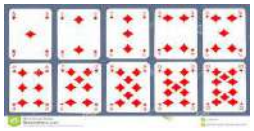
Note:

- *The probability of event A is often written as P(A)*
- *If P(A) > P(B) then event A has a higher chance of occurring than event B*
- *If P(A) = P(B) then events A and B are equally likely to occur*

Exercise 6.10

Using the above formula find:

i. A card is drawn at random from a deck of cards. Find the probability of getting the 5 of diamond.



ii. Two dice are rolled, find the probability that the sum is

a) equal to 1

b) equal to 5

c) less than 13

iii. A jar contains 4 red marbles, 6 green marbles and 10 white marbles. If a marble is drawn from the jar at random, what is the probability that this marble is white?

6.7 Properties Of Probability

LEARNING OUTCOMES

Students should be able to:

- Find out why the sum of all probabilities for an experiment is equal to one
- Examine and explain why the probability of an event is between zero and one
- Apply the above properties to real life situations

i. The probability of an event can only be between 0 and 1.

$$0 \leq P(A) \leq 1$$

In other words, the highest probability an event can have is 1 or 100% (certain to happen) and the lowest probability is 0(certain not to happen). Probabilities between 0 and 1 mean that the event might happen; the higher the probability, the greater the chance that it will happen.

ii. The sum of the probabilities of an event and its complementary is 1, so the probability of the complementary event is:

$$p(\overline{A}) = 1 - p(A)$$

iii. The probability of an impossible event is zero.

$$p(\varnothing) = 0$$

Exercise 6.11

1. Which of the following events has the highest probability of occurring?

   a. The probability of rain on Christmas day is 65%.

   b. The probability of tossing a coin and getting a "tail" is 0.5%

   c. The probability of the Fiji Rugby team winning the rugby game in Dubai is 2 out of 15

   d. The probability of rolling two dice and getting a total of 8 is 5/36

2. A lollie was picked from a bag that contains 1 blue lollie, 1 red lollie and 1 orange lollie.

   a. List the sample space

   b. What is the probability of picking a red lollie?

   c. What is the probability of picking a lollie that is not red?

   d. What is the probability of picking a white lollie?

Application Of Probability

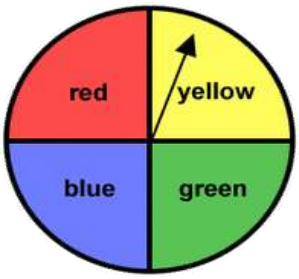
"The probable is what usually happens." - Aristotle.

It is true that the future cannot be predicted but mathematics can help determine its probability and how likely something might occur or not. In our daily lives we often use probability to make better decisions.

Here are some of the common categories to discuss:

- **Sports** – be it basketball or football and coin is tossed and both teams have 50/50 chances of winning it, a basketball player takes a free throw judging his past performance it can be determined if he will make it or not.

- **Board Games** – a game spinner with four sections, there is a 25% chance of it landing blue, since one of the 4 sections is blue. Similarly the odds of rolling one die and getting and odd number there is a 50% chance since three of the six numbers on a die are odd.

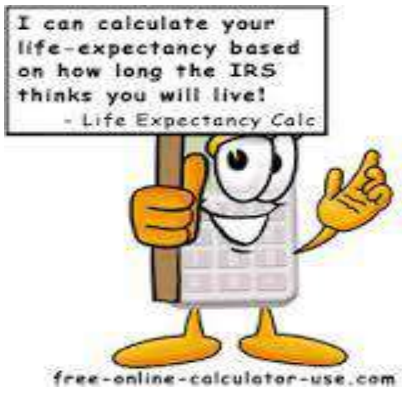


Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+SPINNERS

- **Medical Decisions** – When a patient is advised to undergo surgery, they often want to know the success rate of the operation which is nothing but a probability rate. Based on the same the patient takes a decision whether or not to go ahead with the same.

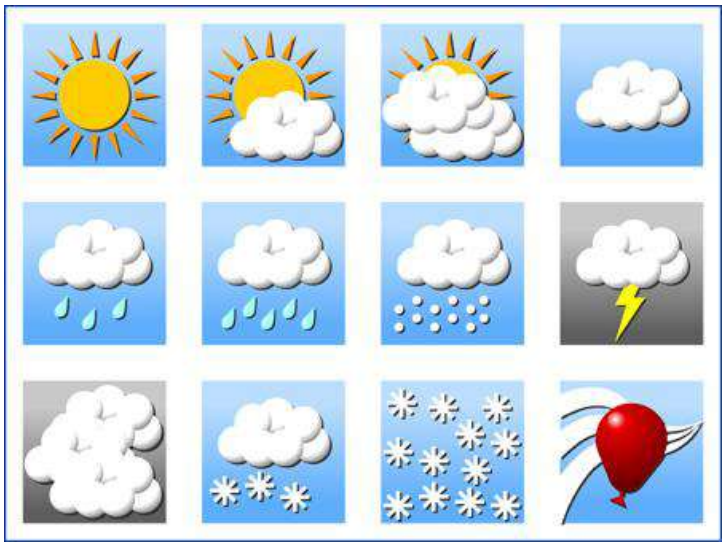
Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf 8#q=cartoon+images+of+hospital+patients

- **Life Expectancy** – this is based on the number of years the same groups of people have lived in the past. "These ages are used as guidelines by entities such as financial advisers to help clients prepare for their retirement years." – ehow.com



Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+life+expectancy

- **Weather** – when planning an outdoor activity, people generally check the probability of rain. Meteorologists also predict the weather based on the patterns of the previous year, temperatures and natural disasters are also predicted on probability and nothing is ever stated as a surety but a possibility and an approximation.



Therefore, even though we do not realize the use of mathematical probabilities in everyday life, subconsciously we use it in every step that we take.

Exercise 6.12

1. Calculate the probability of obtaining the following outcomes when rolling a die:

i. An even number.
ii. A multiple of three.
iii. A five or a six.

2. Find the probability of obtaining the following outcomes when flipping two coins:

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+boy+tossing+a+coin

i. Two heads.
ii. Two tails.
iii. One head and one tail.

4. There are 20 raffle tickets in an envelope. Eight are for a car, and the rest are blank. Find the probability of drawing a ticket for the car if:

i. Only one raffle ticket is extracted.
ii. Two raffle tickets are extracted.
iii. Three raffle tickets are extracted.

4. Students A and B have probabilities of failing an exam of 1/2 and 1/5 respectively. The probability of them both failing the examination is 1/10. Determine the probability that at least one of the two students fail.

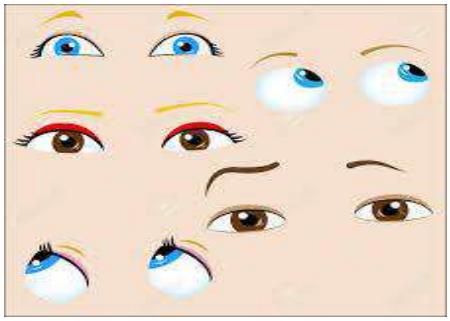
Form Four Exams Results

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+students+failing+exams

5. A father and a son go hunting. The father kills an average of 2 animals every 5 shots and the son kills one animal every 2 shots. If the two fired at the same animal at the same time, what is the probability that they will obtain a kill?

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+people+go+hunting

6. A class consists of 10 boys and 20 girls. Exactly half of the boys and half of the girls have brown eyes. Determine the probability that a randomly selected student will be a boy or a student with brown eyes.



Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+cartoon+students+who+have+brown+eyes

7. The probability that a husband and wife will still be living in 20 years is ¼ for the husband and is 1/3 for the wife. Calculate the probability that:

Source: https://www.google.com/search?q=google&ie=utf-8&oe=utf-8#q=images+of+husband+and+wife+

- Of them both being alive in 20 years.
- The husband living in 20 years but not the wife.
- The husband and the wife dying before 20 years.

GLOSSARY		
1	Bar graph	A chart that uses bars to show comparisons between categories of data. The bars can be either horizontal or vertical
2	Central tendency	A measure of central tendency is a measure that tells us where the middle of a bunch of data lies.
3	Complementary	Two events are complementary if they are the only two possible outcomes
4	Dispersion	Denotes how stretched or squeezed a distribution is
5	Event	It is any collection of outcomes of an experiment.
6	Experiment	It is any procedure that can be infinitely repeated and has a well-defined set of possible outcomes, known as the sample space
7	Frequency	The number of times the event occurred in an experiment
8	Histogram	It represents a frequency distribution by means of rectangles whose widths represent class intervals and whose areas are proportional to the corresponding frequencies: the height of each is the average frequency density for the interval
9	Interquartile range	Equal to the difference between the upper and lower quartile
10	Lower quartile	The median of the lower half of the data
11	Mean	Sum total of all the observations divided by the number of observations.
12	Median	If the data is arranged in the increasing or decreasing order then the middlemost value of the data
13	Mode	Observation has appeared the most number of times
14	Outcome	A single (one) result of an experiment
15	Pictogram	Visual presentation of data using icons, pictures, symbols
16	Pie chart	A circular statistical graph, which is divided into sectors to illustrate numerical proportion
17	Probability	The relative possibility that an event will occur, as expressed by the ratio of the number of actual occurrences to the total number of possible occurrences
18	Quartile	A set of data values are the three points that divide the data set into four equal groups
19	Random sample	It is a subset of individuals (a sample) chosen from a larger set (a population). Each individual is chosen randomly and entirely by chance, such that each individual has the same probability of being chosen at any stage during the sampling process, and each subset of *k* individuals has the same probability of being chosen for the sample as any other subset of *k* individuals
20	Range	Difference between the highest value and the lowest value
21	Relative frequency	The ratio of the actual number of favourable events to the total possible number of events
22	Sample space	The total number of all the different possible outcomes of a given experiment
23	Standard deviation	It is a measure of how numbers are spread out
24	Statistics	A branch of mathematics dealing with the collection, analysis, interpretation, and presentation of masses of numerical data
25	Upper quartile	The median of the **upper** half of the data.