Statistics and probability	Descriptive statistics: collection of raw data, display of data in pictorial and diagrammatic forms, including frequency histograms, cumulative frequency graphs. Obtaining simple statistics from discrete and continuous data, including mean, median, mode, quartiles, range, interquartile range and percentiles. Calculating probabilities of simple events.

Frequency histograms vs cumulative frequency graphs

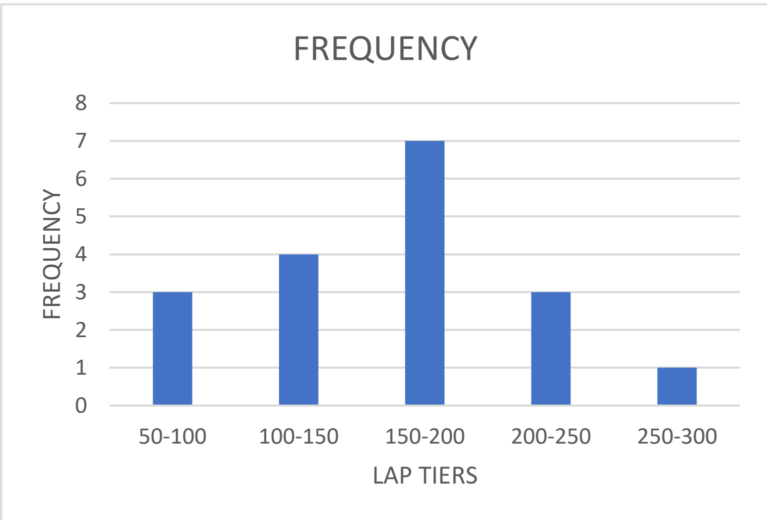
Frequency = how many times a certain event occurs

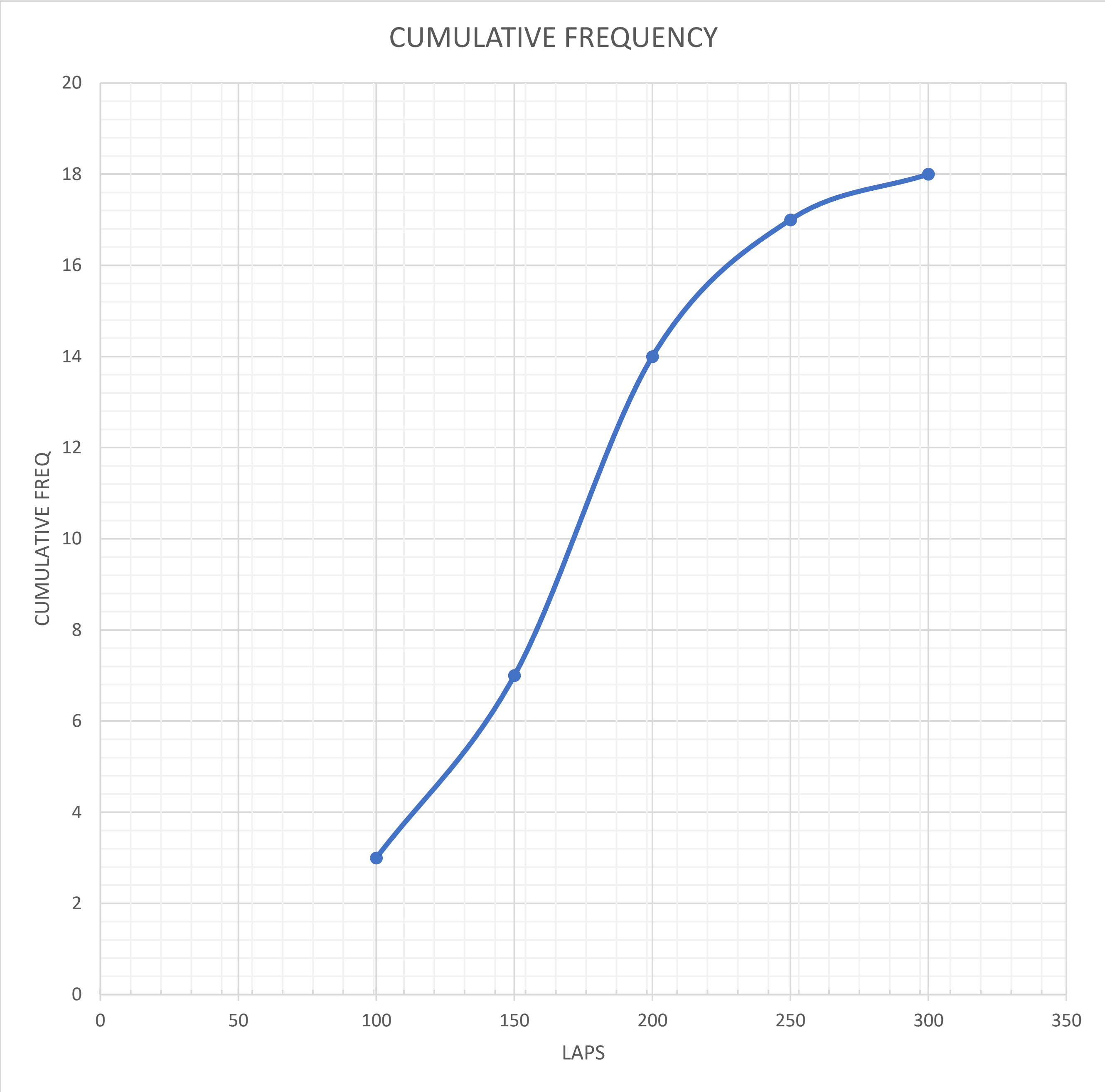
Cumulative frequency = sum of the frequency

Relative frequency = Probability of a frequency occuring

DRIVER	LAPS
HAM	153
VET	272
GRO	177
SIR	46
VER	128
RAI	148
ALO	72
HAR	246
RIC	186
ERI	137
LEC	125
MAG	166
VAN	201
GAS	222
BOT	171
STR	88
PER	178
HUL	190

LAP TIERS	*FREQUENCY*	*CUMULATIVE FREQUENCY*	*RELATIVE FREQUENCY*
50-100	3	3	1/6
100-150	4	7	4/18
150-200	7	14	7/18
200-250	3	17	1/6
250-300	1	18	1/18
TOT	18	18	1





- Always top upper tier limit is used when plotting
- S shape; ogive
- Intrapolate to find median, interquartiles and interquartile range and the 68% percentile

Ex.

PUPIL	HEIGHT (CM)
HANNAH	**166**
LEO	**178**
MARY	**156**
GEN	**175**
VINCENT	**186**
GABE	**174**
LAURA	**164**
SUSAN	**155**
TIM	**165**
CLAIRE	**169**
JIM	**180**
MICHEAL	**170**

- Plot frequency and cumulative frequency as above
- Use tiers of 10cm

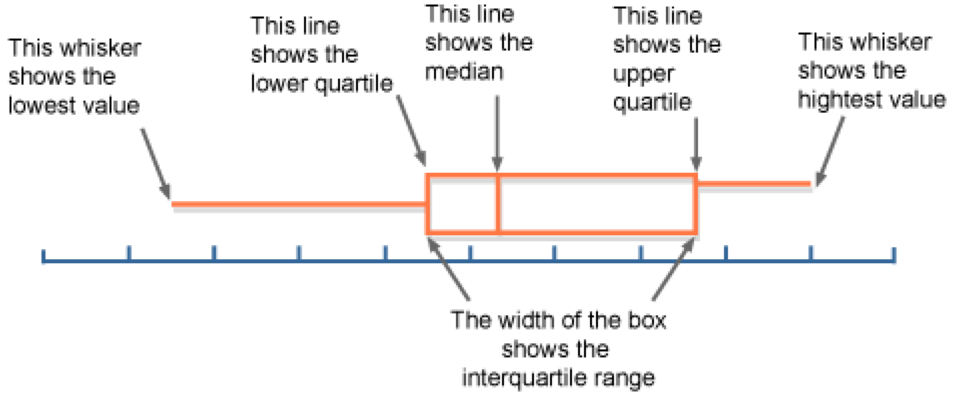
Discrete vs Continuous data

Discrete: data can only take certain values (e.g. rolling a dice)

Continuous: can take any value, such as pupils' height in a class (theoretically)

In graphs, discrete data is usually just points vs continuous data being usually lines.

Whisker plots



- Sketch a whisker plot of the exercise above.

Sample vs Population

Population: all of the elements belonging to a certain characteristic

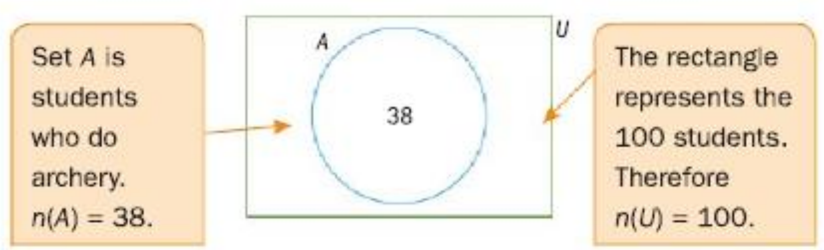
Sample: some of the elements belonging to a certain characteristic

E.g. Samples are used because it is often impossible to gather all of the population data. For instance, often voter samples are collected to give an estimation of the political outlook before a vote (aka voting polls).

Good samples are usually representative of the population composition!

Probability of events

In a school of 100 students (denoted by U) 38 students do archery (denoted by A).



The theoretical probability of an event A is $P(A) = \frac{n(A)}{n(U)}$ where $n(A)$ is the number of ways that event A can occur and $n(U)$ is the total number of possible outcomes.

Therefore P(A) = 38/100 = .38

If the probability of an event is P, in n trials you would expect the event to occur $n \times P$ times.

Expected probability. Noted as E(x)

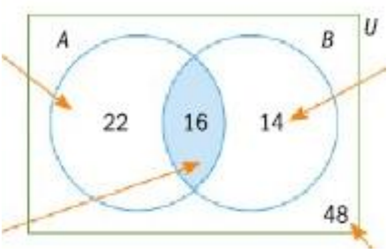
In the school there are 100 students so 100*.38=38. On the contrary, if there were 200 students, we would expect 200*.38=76 students to do archery.

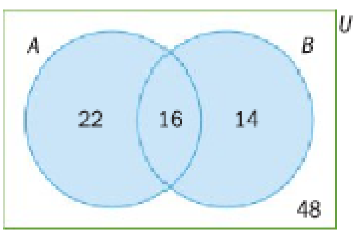
$$n(A') = n(U) - n(A)$$

denotes the *complementary set of A*. In the school case it is…

Of course, $P(A) + P(A') = 1$

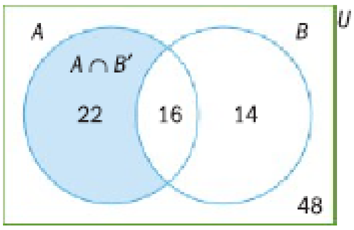
Intersection and Unions

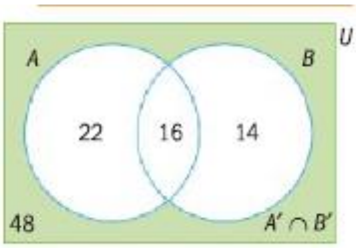




26.3.2018

You can play around with unions and intersections and complementary sets to get stuff like:

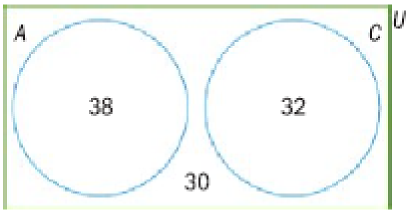




For any two events A and B
$P(A \cup B) = P(A) + P(B) - P(A \cap B)$

Mutually exclusive events

The occurrence of one affects the chance of the other happening.



→ In general, if A and B are mutually exclusive, then $P(A \cap B) = 0$ and $P(A \cup B) = P(A) + P(B)$.

Sample space diagram

They are used to outline all the possible outcomes of an experiment.

For example, if I was to toss two coins at the same time:

	Heads	**Tails**
Heads	**H, H**	**H, T**
Tails	**T, H**	**T, T**

This is really handy in many cases when calculating probabilities, e.g. rolling two dices simultaneously…

Independent events

The previous examples in sample space diagrams where independent events: the occurrence of one does *not* affect the chance of the other happening. This cannot be represented through Venn, only using sample space.

When two events A and B are independent
$P(A \cap B) = P(A) \times P(B)$

This is the **product rule for independent events**.
This is also called the multiplication rule.

Ex.

1.

In a group of 30 students, 17 play computer games, 10 play board games and 9 play neither.
Draw a Venn diagram to show this information.
Use your diagram to find the probability that:

a a student chosen at random from the group plays board games,
b a student plays both computer games and board games,
c a student plays board games but not computer games.

2.

If X and Y are two events such that $P(X) = \frac{1}{4}$ and $P(Y) = \frac{1}{8}$ and $P(X \cap Y) = \frac{1}{8}$, find

a $P(X \cup Y)$
b $P(X \cup Y)'$.

3.

The table below shows the relative frequencies of the ages of the students at a high school.

Age (in years)	Relative frequency
13	0.15
14	0.31
15	0.21
16	0.19
17	0.14
Total	**1**

a A student is randomly selected from this school. Find the probability that
  i the student is 15 years old,
  ii the student is 16 years of age or older.

There are 1200 students at this school.

b Calculate the number of 15-year-old students.

4.

A box contains board-pens of various colors. A teacher picks out a pen at random. The probability of drawing out a red pen is $\frac{1}{5}$, and the probability of drawing out a green pen is $\frac{3}{7}$.
What is the probability of drawing neither a red nor a green pen?

5.

Three unbiased coins are tossed one at a time and the results are noted. One possible outcome is that all the coins are heads. This is written HHH. Another is that the first two coins are heads and the last one is a tail. This is written HHT.
List the complete sample space for this random experiment.
Find the probability that:

a the number of heads is greater than the number of tails,

b at least two heads are tossed consecutively,

c heads and tails are tossed alternately.

6.

An air-to-air missile has probability $\frac{8}{9}$ of hitting a target. If five missiles are launched, what is the probability that the target is not destroyed?

Conditional Probability

The probability of a certain event occurring given that another different event has already occurred.

→ In general for two events A and B the probability of A occurring given that B has occurred can be found using
$$P(A \mid B) = \frac{P(A \cap B)}{P(B)}.$$

Note that,

If A and B are independent events,
$P(A \mid B) = P(A)$, $P(B \mid A) = P(B)$, $P(A \mid B') = P(A)$
and $P(B \mid A') = P(B)$.

Recall that $P(A \cap B) = P(A) \times P(B)$. in independent events.

Ex.

0.

The probability that a student takes Design Technology and Spanish is 0.1. The probability that a student takes Design Technology is 0.6. What is the probability that a student takes Spanish given that the student is taking Technology?

1.

Of the 53 staff at a school, 36 drink tea, 18 drink coffee, and 10 drink neither tea nor coffee.

a How many staff drink both tea and coffee?

One member of staff is chosen at random. Find the probability that:

b he drinks tea but not coffee,

c if he is a tea drinker he drinks coffee as well,

d if he is a tea drinker he does not drink coffee.

2.

For events A and B it is known that: $P(A' \cap B') = 0.35$; $P(A) = 0.25$; $P(B) = 0.6$. Find

a $P(A \cap B)$ **b** $P(A \mid B)$ **c** $P(B' \mid A')$.

3.

A jar contains black and white marbles. Two marbles are chosen without replacement. The probability of selecting a black marble and then a white marble is 0.34, and the probability of selecting a black marble on the first draw is 0.47. What is the probability of selecting a white marble on the second draw, given that the first marble drawn was black?

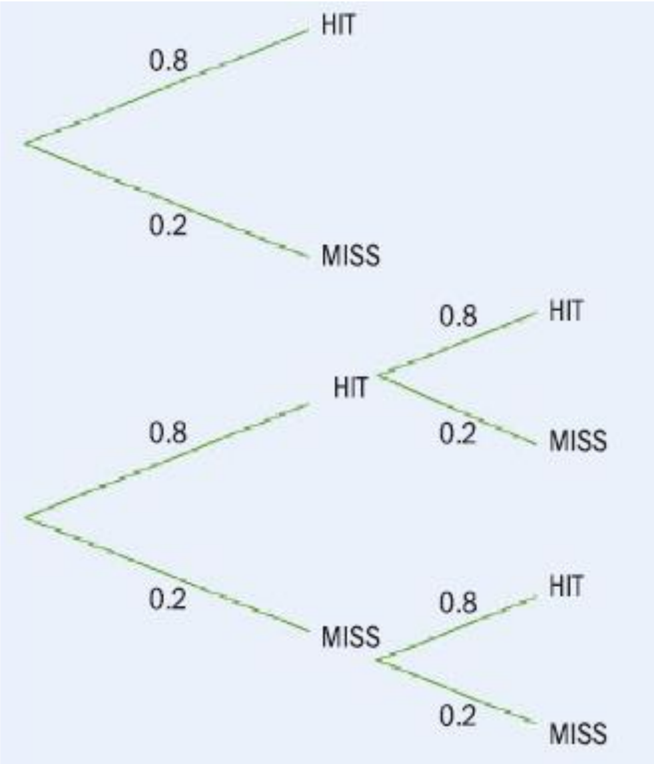
4.

Your neighbour has two children. You learn that he has a son, Sam. What is the probability that Sam's sibling is a brother?

26.3.2018

Probability tree diagrams

John is a keen archer. The probability he hits a 10 is 80%.

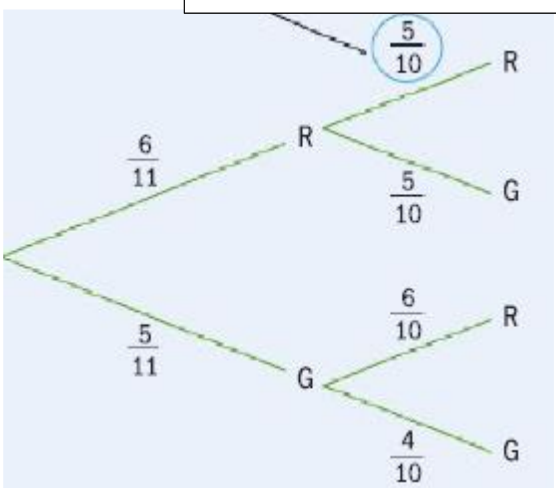


These type of diagrams are particularly useful when replacement does not occur.

Example

A bag contains respectively 6 red and 5 green marbles. With no replacement, a tree diagram would look like this:

A red marble has been drawn



Review Exs

1. In a bag, there are 20 marbles, 10 are green, 5 are red and the rest are yellow.

- Calculate the probability of drawing a green marble.
- With replacement, calculate the probability of drawing a yellow marble.
- Then, calculate the probability of drawing a second yellow marble.
- The calculate the probability of drawing two yellow marbles consecutively, with replacement. Then calculate the prob of drawing three yellow marbles again with replacement. Finally, calculate the prob of drawing a yellow and a red marble.
- Without replacement, recalculate the probabilities of the previous point.

2. In a school, there are 190 students. Among these, 55 are taking biology, 100 are taking chemistry and 19 are taking both.

- Draw a diagram representing the school.
- If a student was to be picked at random, what is the probability that he follows chemistry? What is the prob he follows neither of the subjects?
- A student gets picked at random. Given that he takes biology, what is the probability that he also takes chemistry?

3.

For events C and D it is known that:
$P(C) = 0.7 \quad P(C' \cap D') = 0.25 \quad P(D) = 0.2.$

a Find $P(C \cap D')$.

b Explain why C and D are not independent events.

4.

The two events A and B are such that $P(A) = 0.6$, $P(B) = 0.2$ and $P(A \mid B) = 0.1$.

Calculate the probabilities that:

a both of the events occur,

b at least one of the events occur,

c exactly one of the events occur,

d B occurs given that A has occurred.

5.

Two events N and M are such that $P(N) = \frac{1}{5}$ and $P(M) = \frac{1}{10}$ and $P(N \cup M) = \frac{3}{10}$.

Are N and M mutually exclusive events?

6.

Given that $P(E') = P(F) = 0.6$ and $P(E \cap F) = 0.24$

a write down $P(E)$,

b explain why E and F are independent,

c explain why E and F are not mutually exclusive,

d find $P(E \cup F')$.

7.

U and V are mutually exclusive events. $P(U) = 0.26$; $P(V) = 0.37$. Find

a $P(U \text{ and } V)$ **b** $P(U \mid V)$ **c** $P(U \text{ or } V)$.

8.

Determine the probability of getting two heads in three tosses of a biased coin for which $P(\text{head}) = \frac{2}{3}$.

9.

A pencil case contains 5 faulty and 7 working pens. A boy and then a girl each need to take a pen.

a What is the probability that two faulty pens are chosen?

b What is the probability that at least one faulty pen is chosen?

c If exactly one faulty pen is chosen, what is the probability that the girl chose it?

26.3.2018

Mean, Average, Median, Mode

Mean and average both *mean* the same thing!

Assume a sample = $\{x_1, x_2, x_3, x_4, x_5 \ldots x_n\}$ remember difference between {} and []

Average = $\frac{x1+x2+x3+\cdots+xn}{n}$

Median = $\frac{x_{greatest}+x_{smallest}}{2}$

Ex.

1. You have been fishing and caught 4 trout, of length respectively 22, 48, 52, and 36 cm. Calculate all of the above.

2. [Maximum mark: 4]

At a skiing competition the mean time of the first three skiers is 34.1 seconds. The time for the fourth skier is then recorded and the mean time of the first four skiers is 35.0 seconds. Find the time achieved by the fourth skier.