Level 1 – AS91037– 1.12 – 4 Credits – External

Demonstrate understanding of chance and data

Context

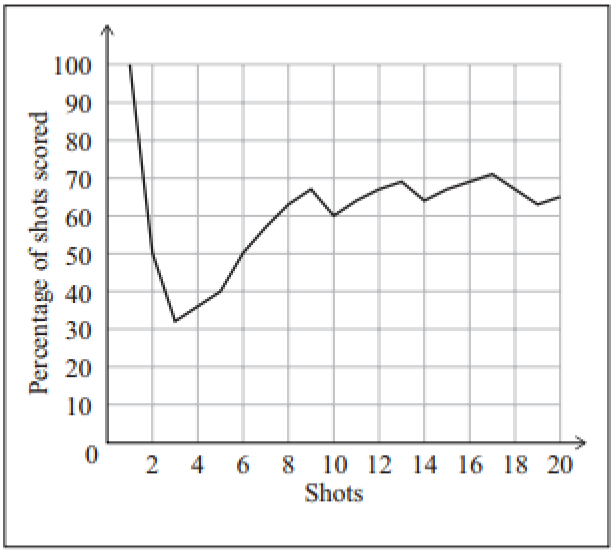
Level 1	Level 2	Level 3 statistics
1.12 Chance and data 4 credits [Ext]	2.12 Probability 4 credits [Ext]	3.13 Probability concepts 4 credits [Ext]

Sample question: Basketball

Part One: Levi has been wondering about how good he is at basketball. Levi thinks that his probability of scoring is 50% since he can either miss or score on the shot. Do you agree?

Part Two: Levi's father says that, when he was the same age, his chance of scoring was 60%. If Levi's father had taken 200 shots when he was the same age, how many shots would he have scored on?

Part Three: Levi begins an experiment. He takes some shots and records whether he scores or he misses. A graph of his results is below. Based on this experiment, what is the probability that Levi will score on his next shot [shot 21]? How confident can you be that this probability is correct?



Passport

Use this to track your progress through this workbook.

Chance

Theoretical and experimental probability	First time	Revision
Theoretical probability		
Experimental probability		

Calculations using theoretical probability

Conditional probability		
Expected value		
Combining events		

Calculations using experimental probability

Long run relative frequency		

Data

Statistical literacy [from 1.10 multivariate data]

Averages [mean, median, and mode]		
Quartiles [upper and lower]		
Spread [range, interquartile range, and standard deviation]		
Reading box and whisker plots		

Statistical literacy

Frequency tables		
Reading line graphs		
Reading pie graphs		
Reading scatter graphs		
Reading histograms and bar graphs		
Interpreting information		

Lesson One: Theoretical and experimental probability

Theoretical probability is what we expect to happen based on theory whereas experimental probability is what we expect to happen based on experiments.

Unfortunately, neither theoretical nor experimental probability can tell us what will actually happen. Stink.

Mathematicians decided that probabilities will be on a scale from 0 [impossible] to 1 [certain]. Obviously we almost never deal with absolute impossibilities or certainties, there's a small but non-zero chance you'll be bonked on the head by a meteorite in the next few minutes. Therefore we spend most of our time somewhere in between the extremes.

We denote the probability of an event happening as P(event). So the probability of event A occurring is P(A) and the probability of the meteorite-thing happening is P(bonked on the the head by a meteorite in the next few minutes). Although you could just shorten it to P(meteorite).

Fractions, decimals, and percentages
We can choose whether we write probabilities as fractions, decimals, and percentages. Each option has its own advantage:

- Fractions: Easy to calculate and to use in calculations
- Decimals: Easy to compare different probabilities
- Percentages: Easy for the general public to understand

Lesson One questions

Q1. What's the difference between theoretical and experimental probability? [Show off what you know]

Q2. Which one best predicts what will actually happen? [Again, show off what you know]

Q3. A pack of cards is shuffled. Calculate the theoretical probability that the card drawn is a:
a) Black card
b) Yellow card
c) Card with a number on it
d) Card without a number on it
e) Jack
f) Jack of hearts
g) Number card between 4 and 7 inclusive

Q4. Two different coloured dice are thrown, red and yellow. Calculate the following theoretical probabilities:
a) The total is 7
b) The total is more than 7
c) The total is at least 7
d) Getting a double
e) The number on the red die is more than the number on the yellow die
f) Only prime numbers are shown on the dice

Q5. Two identical dice are thrown. Calculate the same theoretical probabilities as for question 4.

Q6. For each part in questions 3-5, write each probability like a mathematician would, e.g.

Q3. a) $P(black) = \frac{1}{2}$ Q4. a) $P(total = 7) = \frac{1}{6}$

Lesson Two: Conditional probability and expected value

Conditional probability
Conditional probability involves calculating probabilities given that… [insert condition here]. For example, draw two cards from a deck. What's the probability that the second card is red given that the first card was black? There are only 51 cards left after the first choice and 26 red cards left so $\frac{26}{51}$, slightly more than $\frac{1}{2}$

We can say that $P(2nd\ card = red\ given\ that\ 1st\ card = black) = \frac{26}{51}$

Unsurprisingly, mathematicians have decided to shorten the 'given that' by using a symbol.
$P(2nd\ card = red \mid 1st\ card = black) = \dfrac{26}{51}$

You could even go one step further if you would like.
$P(red \mid black) = \dfrac{26}{51}$

Expected value
Expected value is the number of times you expect something to occur. For example if you have a $\frac{3}{5}$ or 60% chance of scoring in basketball and you have 10 shots, you'd expect to get 6 of them in.

Lesson Two questions

Q1. Why did you use theoretical probability in the lesson one questions rather than experimental probability? [Show off what you know]

Q2. How is conditional probability different to what you were calculating in lesson one? [Again, show off what you know]

Q3. Some cards are drawn from a deck. Calculate the probability that the card drawn is a: [show off your mathematical notation]
a) Queen given that two 10s have been drawn
b) Number card given that 8 picture cards have been drawn
c) Number card given that 8 number cards have been drawn
d) Ace given that 2 aces have been drawn
e) King given that all the red cards have been drawn
f) King given that a red card has been drawn

Q4. Fill in the totals in the table below. Calculate the probability that a randomly chosen student is: [again, show off your mathematical notation]

	Boy	Girl	Total
Year 11	11	5	
Year 12	5	6	
Year 13	5	2	
Total			

a) A boy
b) A girl
c) A year 11
d) A boy given that they're year 11
e) A girl given that they're year 12
f) A year 12 given that they're a girl
g) A boy given that they're year 13
h) A year 13 given that they're a boy

Q5. Two different coloured dice are thrown, red and yellow. Calculate the following theoretical probabilities: [guess what I'm going to ask you to show off?]
a) The total is at least 9
b) The total is a prime number

Lesson Three: Combining events

All good things come in threes. But what are the chances of that? What's the probability that you roll a 6 on a die and then roll a 5? You'll probably find a probability tree helpful at this point.

Lesson Three questions

Q1. What is the probability of the following events?
a) Flipping 5 tails in a row with a coin?
b) Flipping 5 in a row [either heads or tails]?
c) Flipping 10 heads in a row?

Q2. For every part of question 1, consider if I gave you $100 every time you did what was asked. What is the expected value for each attempt?

Q3. Let's make the challenge in question 2 fair. How much would you need to pay me each time you lost to drop the expected value to 0 for each attempt?

Lesson Four: Experimental probability

Experimental probability is just as good as theoretical probability for giving us an idea of what will happen. A big reason why you'd probably prefer to use theoretical probability is it's much easier to calculate. On the other hand, a big reason why you'd choose to use experimental probability is when there is no theory. For example, what's the chance that a meteorite will bonk you on the head in the next few minutes?

A theoretical approach would surely involve counting and tracking all of the asteroids currently in our solar system [if you were doing it properly, you'd track all of the asteroids that will be near Earth over your lifetime]. Alternatively, an experimental approach would involve waiting for a meteorite to hit you and counting how many do. Actually, it would involve counting how many people have been hit by meteorites and dividing by the total number of people.

Experimental probability is also known as the "long-run relative frequency" and it's important to know why. Let's break it down word-by-word, starting at the end [because, reasons].

Frequency: Counting how often something happens, e.g. being hit by a meteorite or getting a basketball shot in.
Relative: Counting how often something *doesn't* happen or more accurately, counting the total number of times attempted.
Long-run: Doing the counting over a long period. If you only counted the first few basketball shots you wouldn't get any useful data. Only after 'enough' attempts will you get a good picture [and even then it's not guaranteed. Why?]

Have a look back at the graph from the sample question.

Lesson Four questions

Q1. What's the difference between theoretical and experimental probability? [Do I need to remind you to show off what you know?]

Q2. Which one best predicts what will actually happen?

Q3. Fill out the totals in the table below and answer the following questions: [need I encourage you to show off your mathematical notation?]

	Left-handed	Right-handed	Ambidextrous	Total
Year 4	4	233	26	
Year 5	10	1,166	173	
Year 6	90	1,327	16	
Year 7	144	2,379	21	
Year 8	210	2,937	30	
Year 9	466	6,570	48	
Year 10	445	5,532	58	
Year 11	376	3,111	244	
Year 12	16	1,181	172	
Year 13	7	764	108	
Total				

Source: A random sample from censusatschool.org.nz

a) What is the probability that a random student picked from this sample is left-handed?
b) What about right-handed?
c) What about ambidextrous?
d) What about ambidextrous given that they're year 11?
e) What about year 11 given that they're ambidextrous?
f) Which year group has the highest relative frequency of left-handed students?
g) Which year group has the highest relative frequency of right-handed students?
h) Which year group has the highest relative frequency of ambidextrous students?

Lesson Five: Statistical literacy part 1

Statistical literacy [noun]: "the ability to understand statistics."[1]

Most, if not all, of the statistics standards involve statistical literacy. The massive upshot of this is that you can revisit the same skills to further develop them. Around half of the statistical literacy required for this standard has already been covered in 1.10 multivariate data.

There are 10 summary statistics:

Measures of centre	Quartiles	Extremes	Measures of spread
Mean Median Mode	Lower quartile Upper quartile	Minimum Maximum	Range Inter-quartile range Standard deviation

Mean
The mean is the easiest to calculate but is not the most useful as it's affected by extreme values. To calculate the mean add all the values together and divide by the number of values. If a news article or other public statistic mentions the "average" without specifying which one it is, it'll be the mean.

Median
The median is harder to calculate because it requires the data to be in numerical order. The strength of the median is that it's unaffected by extreme values. To calculate the median, order the data from smallest to largest and select the middle number. If there are two middle values, find halfway between them.

Mode
The mode is laughably easy to calculate but hardly useful as a measure of centre. To find the mode, find the value that occurs most often. There can be more than one mode.

Lower quartile
While the median gives us the middle of the data, the lower quartile gives us the lower quarter. This is useful in seeing the shape of the data. To calculate the lower quartile, keep the data ordered from smallest to largest and select the number halfway between the minimum and the median.

Upper quartile
The upper quartile gives us the upper quarter of the data. When combined with the lower quartile and median, it gives us a strong picture of the shape of the data. To calculate the upper quartile, keep the data ordered from smallest to largest and select the number halfway between the median and the maximum.

---

[1] https://en.wikipedia.org/wiki/Statistical_literacy [that most reputable source]

Minimum
The minimum helps give us a better picture of the shape of the data but it's not as useful as the lower quartile. It is the smallest value in the data.

Maximum
The maximum also helps give us a better picture of the shape of the data but again it's not as useful as the upper quartile. It is the largest value in the data.

Range
The range is easy to calculate but is easily influenced by extreme values so isn't very useful as a measure of spread. To calculate the range, take the maximum value minus the minimum value.

Inter-quartile range
The inter-quartile range is easy to calculate and isn't influenced by extreme values making it very useful. It gives the spread of the middle 50% of the data. To calculate the inter-quartile range, take the upper quartile minus the lower quartile.

Standard deviation
The standard deviation is hard to calculate and hard to interpret. But it is an excellent measure of spread so you can be sure that data with a higher standard deviation will have a larger spread. Use a calculator to calculate the standard deviation. See the appendix for how to calculate standard deviation by hand.

Standard deviation is most powerful when combined with the mean. 68% of the data is within ± 1 standard deviation from the mean. 95% of the data is within ± 2 standard deviations from the mean. And 99.7% is within ± 3 standard deviations.

Lesson Five questions

Q1. Calculate the ten summary statistics and make a comparison between the two samples. Don't make an inference about the population.

Sizes of kiwi birds (cm):

Male	42, 32, 30, 40, 46, 36, 44, 30, 46, 43, 33, 31, 36, 38, 40, 41, 38, 43, 46, 40
Female	49, 36, 47, 38, 52, 33, 47, 43, 37, 40, 33, 46, 46, 33, 37, 45, 48, 48, 39, 33

Q2. Calculate the ten summary statistics and make a comparison between the two samples. Don't make an inference about the population.

Gender	Age	Homework/day	Region
female	9	0:30:00	Waikato Region
female	10	0:15:00	Waikato Region
female	12	2:15:00	Auckland Region
female	12	0:45:00	Wellington Region
female	9	0:15:00	Auckland Region
female	13	2:15:00	Auckland Region
female	10	0:30:00	Auckland Region
female	12	1:00:00	Northland Region
female	14	0:00:00	Waikato Region
female	15	0:45:00	Canterbury Region
male	13	1:30:00	Otago Region
male	10	0:30:00	Auckland Region
male	8	0:15:00	Other North Island
male	14	0:30:00	Auckland Region
male	14	1:00:00	Auckland Region
male	9	0:30:00	Auckland Region
male	10	0:00:00	Auckland Region
male	8	0:45:00	Bay of Plenty Region
male	15	2:00:00	Waikato Region
male	13	0:00:00	Waikato Region

Q3. Calculate the ten summary statistics and make a comparison between the three samples. Don't make an inference about the population.

Year	Gender	Reaction speed	Region
10	female	0.389	Northland Region
10	male	0.436	Canterbury Region
10	female	0.386	Wellington Region
10	male	0.822	Auckland Region
10	female	0.488	Wellington Region
10	female	0.556	Auckland Region
10	female	0.468	Auckland Region
10	female	0.388	Auckland Region
10	female	0.403	Auckland Region
10	male	0.825	Auckland Region
11	male	37.9	Waikato Region
11	male	0.383	Waikato Region
11	male	1.088	Auckland Region
11	female	0.868	Other North Island
11	male	1.783	Bay of Plenty Region
11	female	0.529	Waikato Region
11	male	0.382	Northland Region
11	female	0.52	Other North Island
11	female	1.135	Canterbury Region
11	male	0.418	Auckland Region
12	male	0.415	Bay of Plenty Region
12	female	43.4	Northland Region
12	male	0.445	Otago Region
12	female	0.457	Canterbury Region
12	female	0.463	Waikato Region
12	female	0.491	Auckland Region
12	female	0.368	Auckland Region
12	female	0.382	Auckland Region
12	female	0.419	Wellington Region
12	female	0.355	Wellington Region

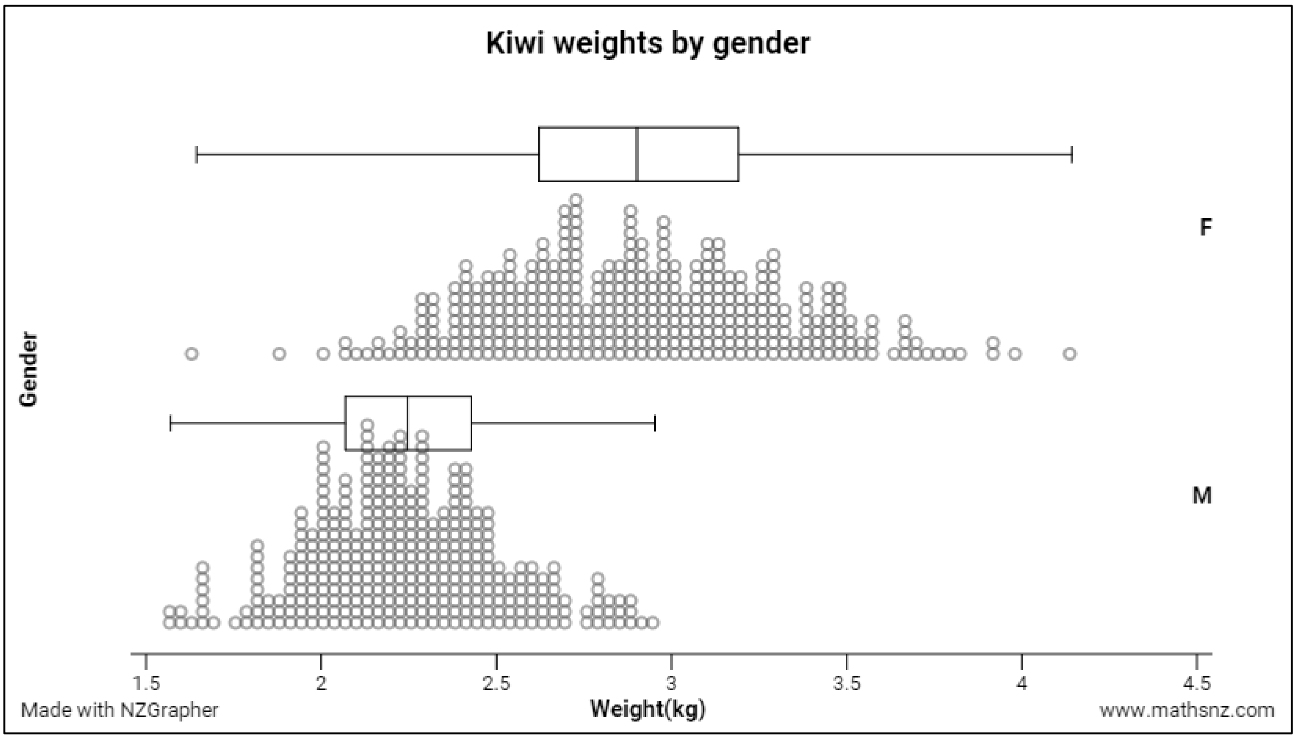
Lesson Six: Statistical literacy part 2

A lot of statistics are shown as percentages. And a lot of statistics are shown in graphs or pictures. You'll now learn about how the graphs are made so you can better interpret them.

Frequency table
A frequency table is a tally chart with a fancy name. Easy to interpret and very good at keeping counting mistakes to a minimum [which are frustratingly common].

Title: How Do We Get to School?		
Categories	Tallies	Total
Walk	卌 II	7
Bike	III	3
Car	IIII	4
Bus	卌 卌 II	12

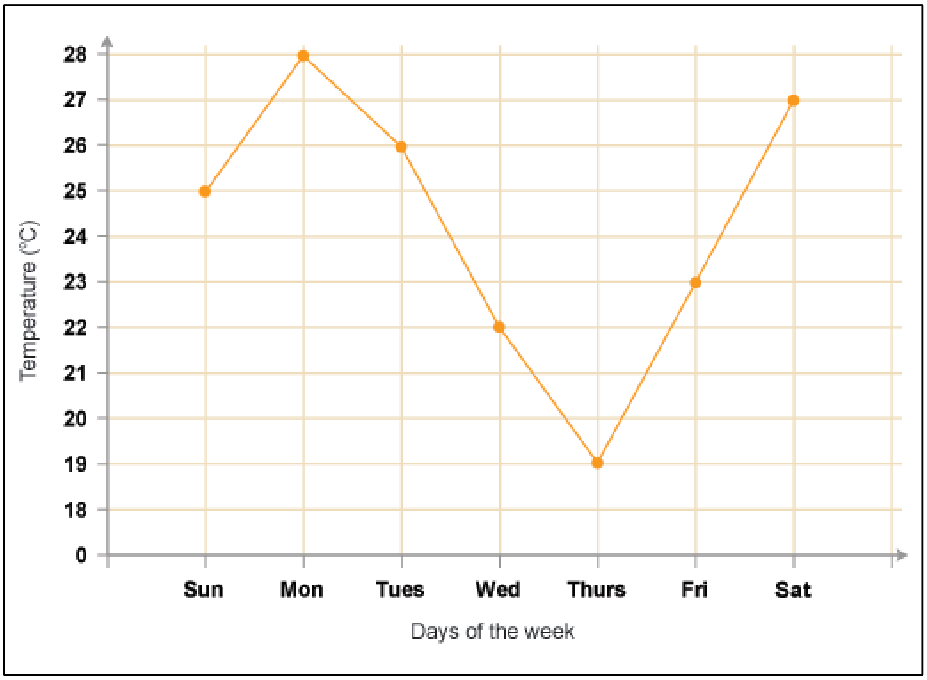
Box and whisker plot
You know how useful these are. SO useful! They require a moderate amount of work but are mind-bogglingly powerful displays. Here is an example combined with a dot plot.

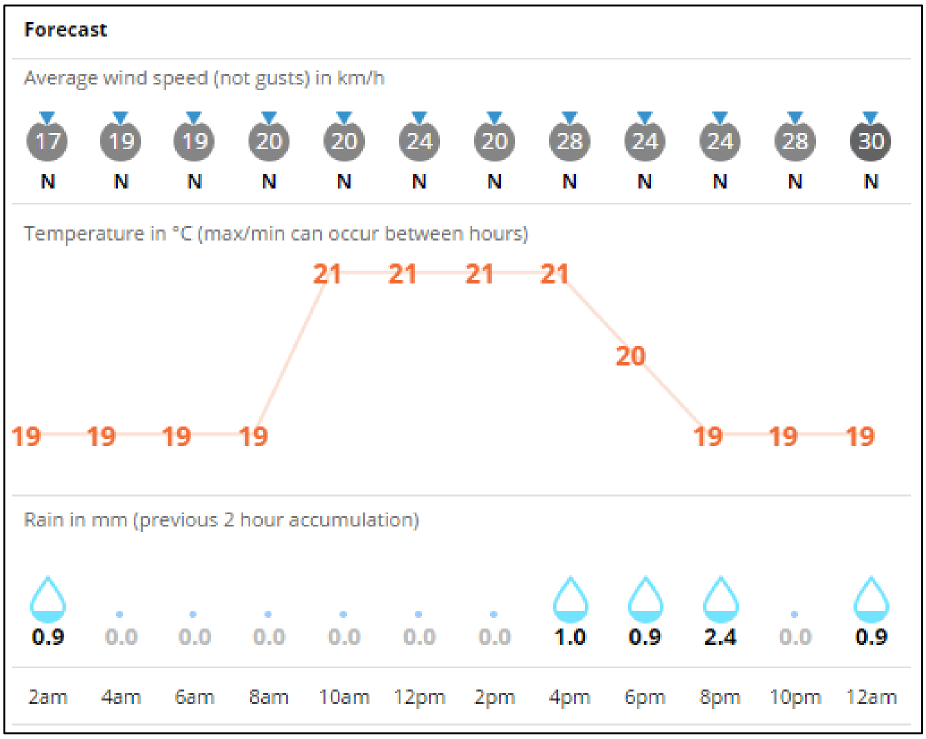


Line graph

Line graphs are useful for finding trends and patterns in data, especially over time. In most cases we can *interpolate* [estimate values in between the points listed]. For example, it is not meaningful to speak of the maximum temperature between Wednesday and Thursday in this first graph, there is no such value. But in the second graph we can interpolate that the temperature at 5pm is around 20.5^{o}C.

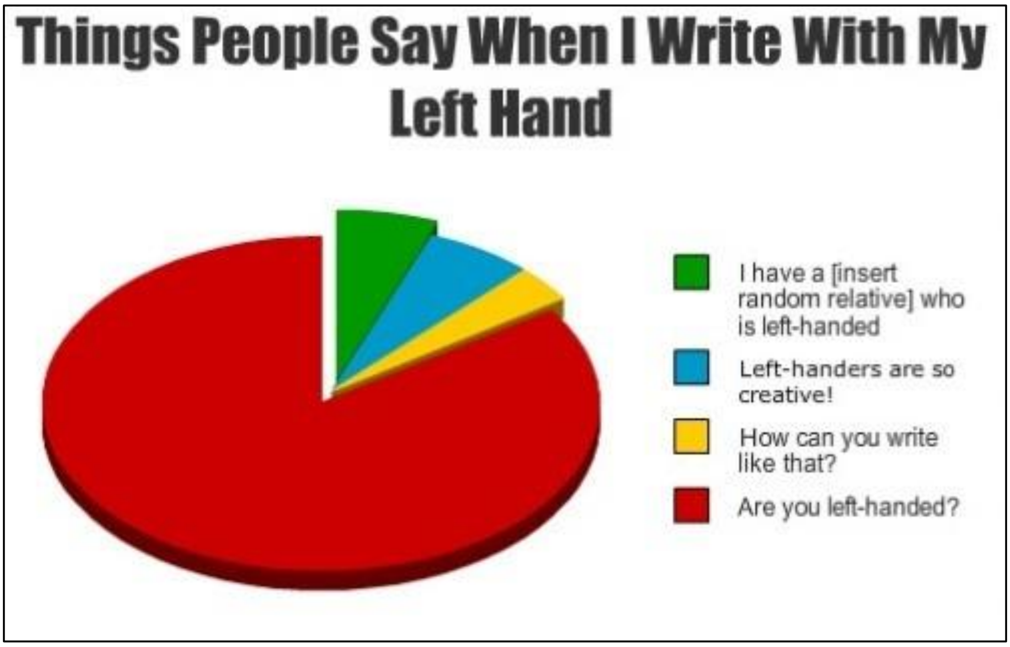
Sometimes we can *extrapolate* [estimate values outside the data range]. For example, we can estimate that the temperature at 2am the next day is around 19^{o}C in the second graph. We can't make a confident estimate about the temperature next Sunday in the first graph.





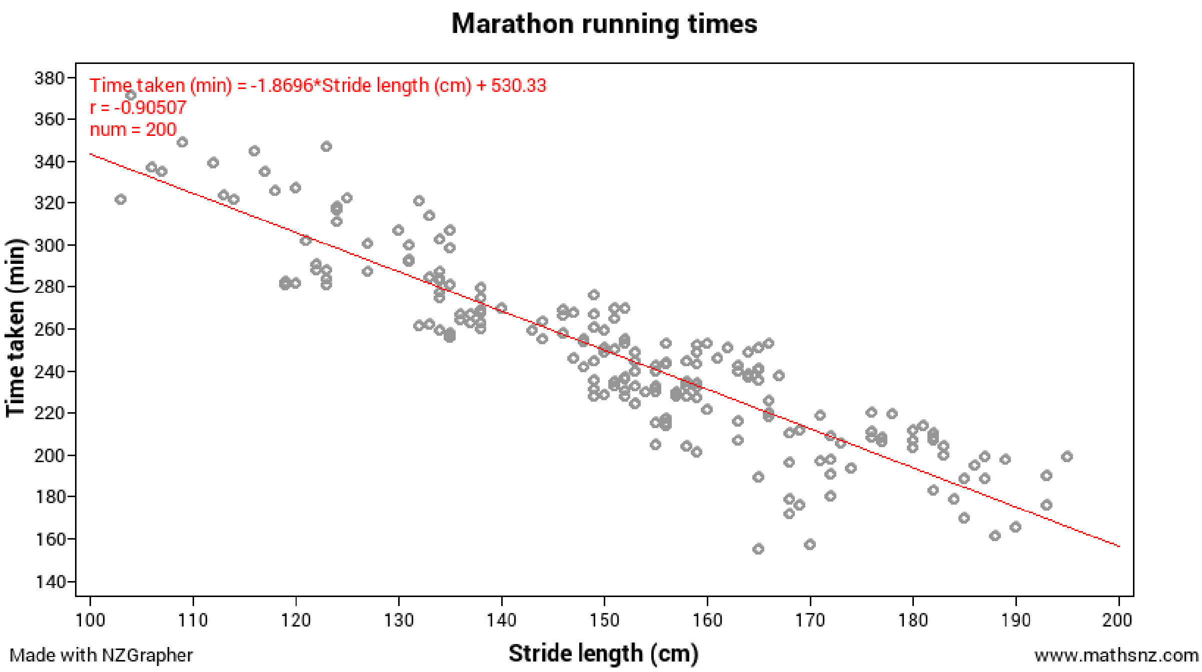
Pie graph

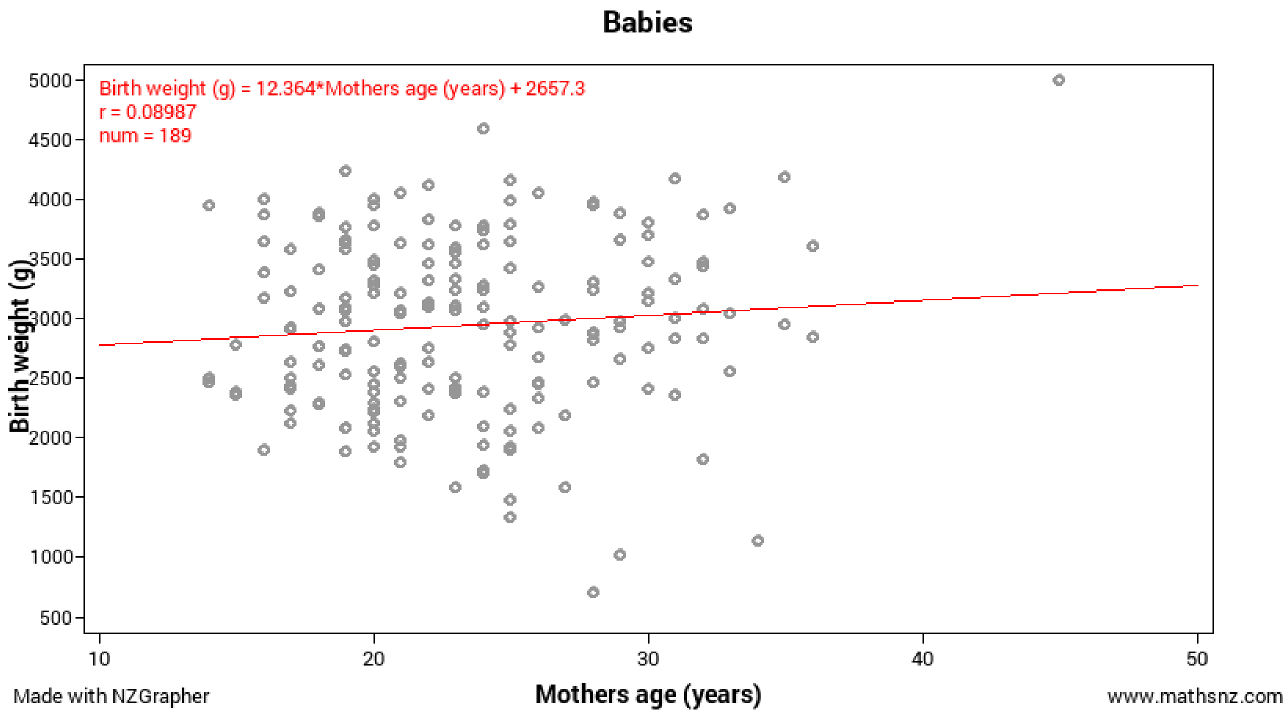
Pie graphs are amusingly-named displays of how a quantity is divided amongst several categories, e.g. time divided amongst activities or money divided amongst people. Each proportion is calculated from a 360° circle, e.g. if school takes up $\frac{1}{5}$ of your day then it takes $\frac{1}{5} \times 360° = 72°$ of the circle.



Scatter graph

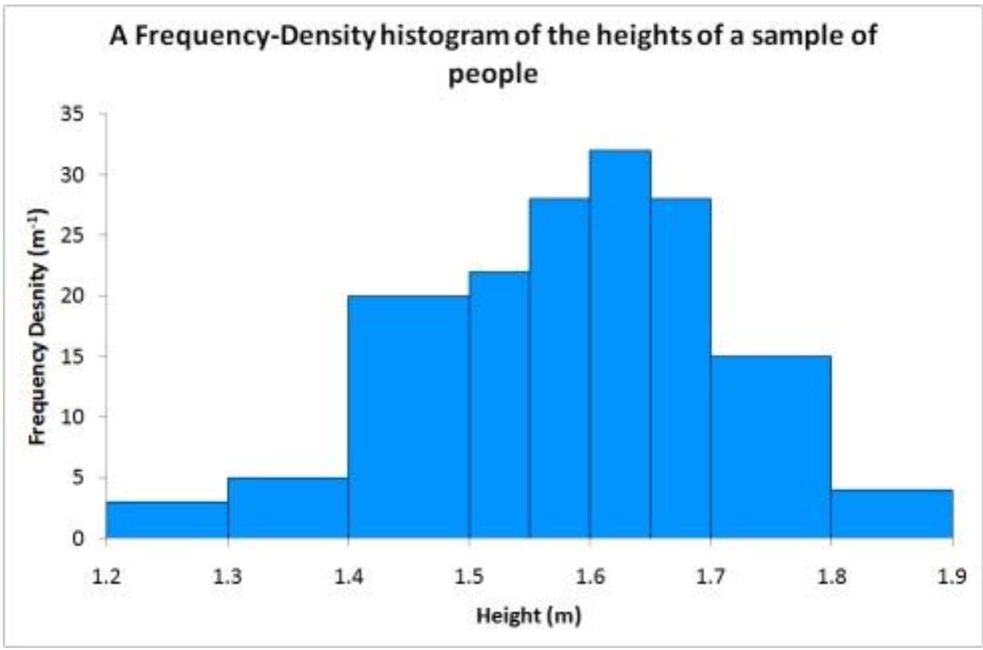
Scatter graphs are used to determine if there is a relationship between two variables. If there is a strong relationship, the points will be close to the line of best fit. If there is a poor or no relationship, they won't. From the following two graphs we can see there is a strong [negative] relationship between marathon running times and stride length while there is no relationship between a mother's age and their baby's birth weight.





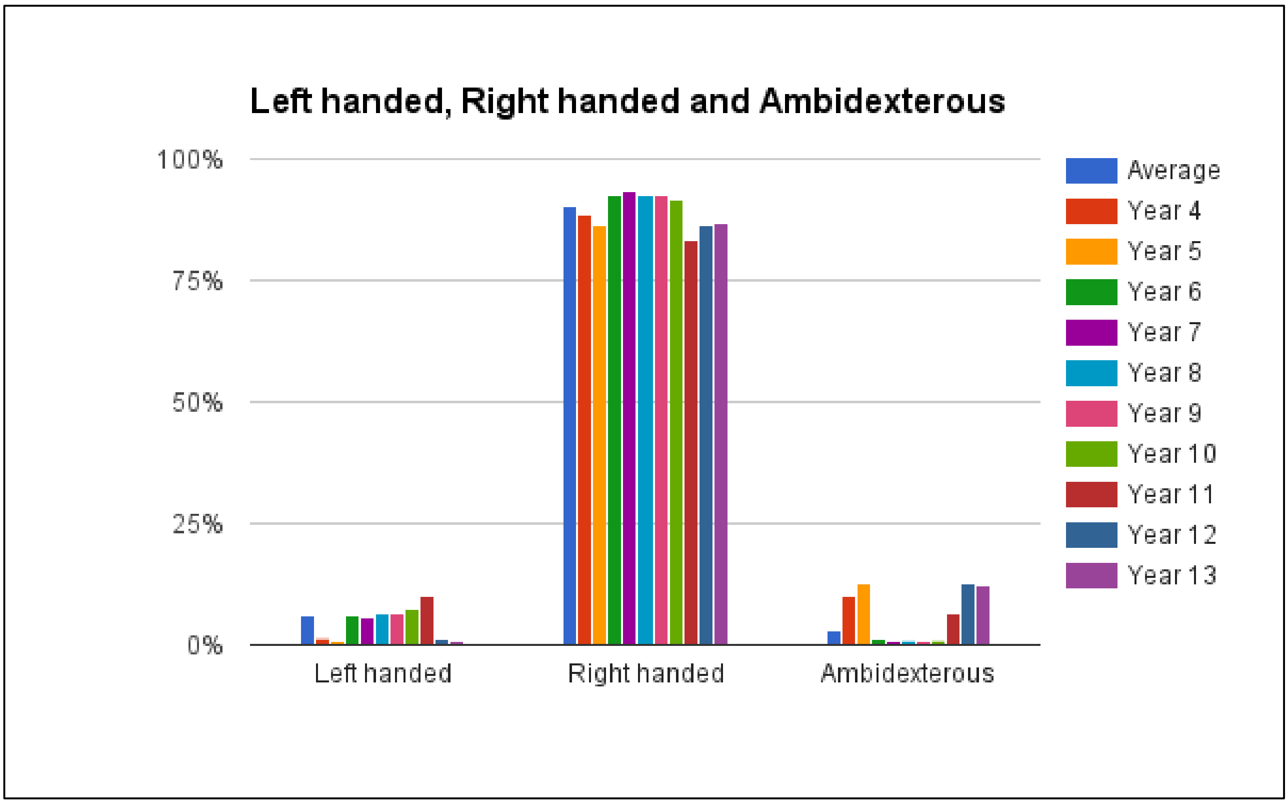
Histograms

Histograms are confusing. But they don't have to be. In one sense, they're just like bar graphs except instead of discrete categories on the x-axis, they have ranges of continuous data. Unfortunately, the frequency of those ranges aren't measured by the height of the rectangles but by the area of the rectangles. This isn't a problem as long as each range is the same width along the x-axis. Pray that it is [in most cases it is].



Bar graphs

Ah, bar graphs. Everyone's favourite graph. There's just something about it that's unassuming and simple. Bar graphs use discrete categories rather than continuous data. The height of the bars tell the frequency of each category. Bar graphs are excellent when we want to know which category has the biggest or smallest frequency.



Interpreting information

You will be asked to interpret information that is displayed in a table, graph, or chart. I encourage you to take you time with this, get to know the data, casually at first, and see how things turn out. But seriously, the data is telling you the answer, it's just hidden at first.

When you have an answer, back it up with evidence from the data.

Lesson six questions

The best way to practice interpreting data is to look at real data. Look at the past exams and select the questions on data [not chance].

Practice externals

Past exam papers can be used for practice and can be found on the Tāwari maths website: davidstarshaw.github.io. Note that in 2010 and earlier, the standard is different so these exams will be less useful than more recent ones.

Homework tracking

As you do homework for this standard, keep track of it all right here. Do around 30 minutes for each class time.

Week	Day of the week	Length of time	What did you do?
Term 2, Week 5			
Term 2, Week 6			
Term 2, Week 7			
Term 2, Week 8			