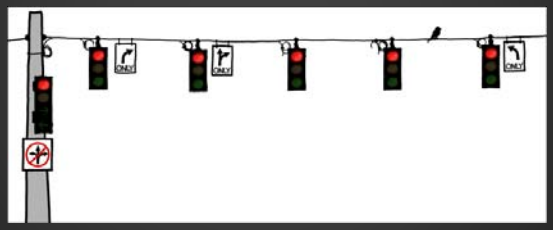
Stimulating Simulations

Teaching AS91268 with a focus on probability modelling



Anna Martin
University of Auckland
anna-marie.martin@auckland.ac.nz

Questions to consider throughout this workshop.....and when planning ☺

Why do we USE simulations?

- What are some real/genuine examples of how simulations are used to inform or make decisions?

Why do we TEACH simulations?

- What are the key concepts we want students to understand for this topic?

Why do we ENJOY teaching simulations?

- What aspect do you have fun with and wish the students enjoyed too?

Why create graphs of the simulation results?

statistics = variation

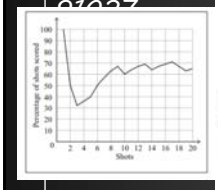
probability = uncertainty

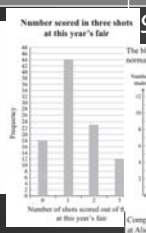
data = visualisation

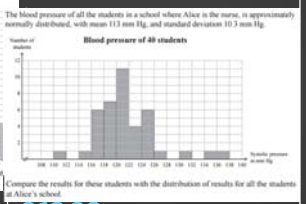
And context

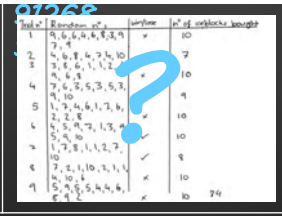
Why create graphs of the simulation results?

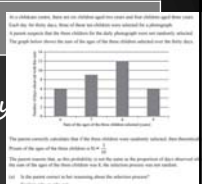
Level 1	Level 2	Level 3
91037	91267	91585 Probability
		91586 Probability dis
		91 Ex

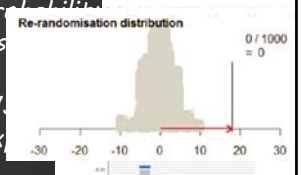


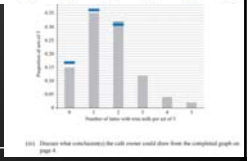










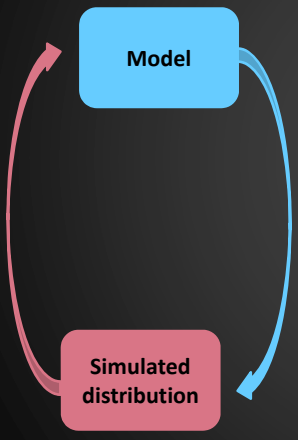


Why create graphs of the simulation results?

Yeah but.....

What should we focus on?

It is intended that students be provided with a situation to investigate. This must be one where results cannot easily be obtained through using theoretical probability models.



The model.....theoretical focus

- *Setting up a simulation to solve a tricky probability problem*
- *Setting up a simulation(s) to inform a decision making process*
- *Use simulated data to get point estimates of probability and/or mean → a substitute for the theoretical answer!*
- *Exploring variation, middle 95/n%, etc·*

The simulated distribution.....inferential focus

- *Working with something observed (a value or distribution from a sampling or experiment situation) and exploring "chance acting alone": random variation and features such as sample size and known processes, using assumed models*
- *Experimental distributions and model selection*
- *Exploring variation, middle 95/n%, etc·*

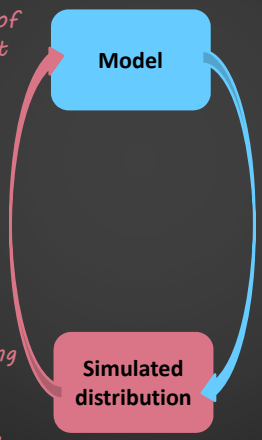
Quick examples…

Example A

For five of the 12 seasons of "The Block", the team that went first in the auction order won the competition. Does being first in the auction order increase the chance of winning?

Example B

Rolling a six sided die that is biased in some way unknown to you, and getting an estimate for the probability of rolling a six – how many times should you roll the die? [Note rolling the die is not a simulation]



Example C

Anna has a set of five keys, one which opens the deadlock and one which opens the door lock. If she selects a key at random to use to open each lock, what is the probability it will take her more than five attempts to open both locks? What is the mean number of attempts she will make? (See NZQA Level 3 sample paper Probability 2013)

Introducing simulations…

Inferential focus

My goals:

Keep the focus on simulating a random process using a model – move through different "probability tools"

Use a familiar context (binomial but without calling it that) which lends itself easily to a simulation (students can easily understand how chance could be acting)

See variation within and between small samples…… and have fun!

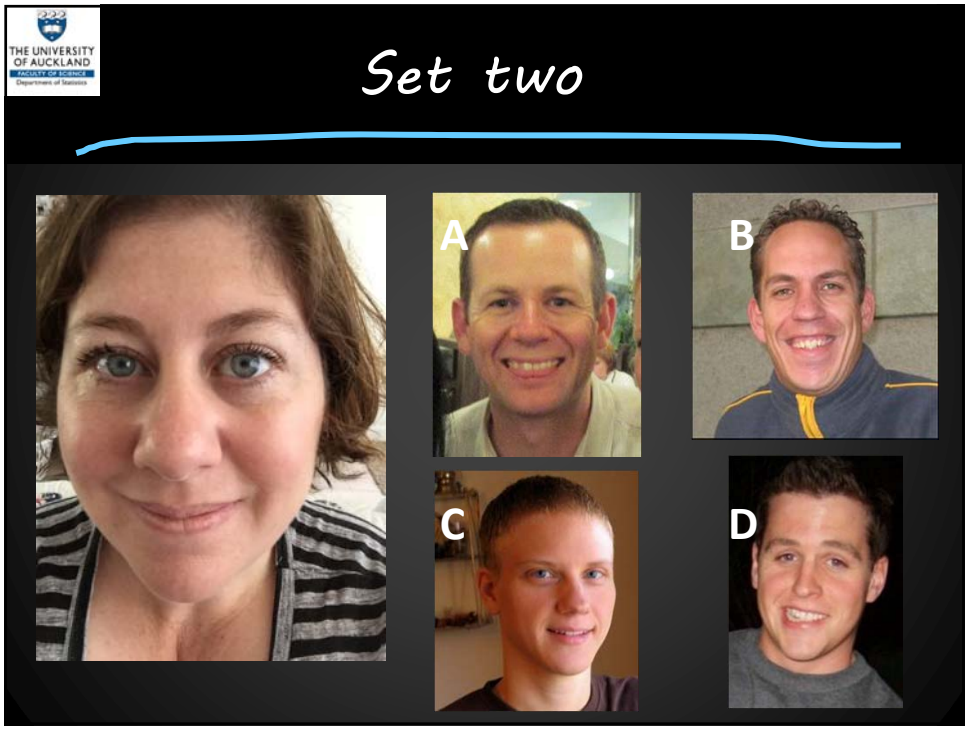
Family resemblance…

"You can tell they're related"…… or can you?

Can you tell which of the following people are related to me?

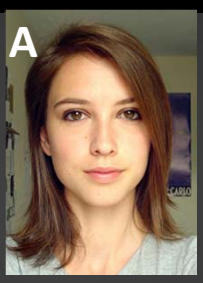
For each set of photos, write down the letter for the person who you think is related to me:-)


THE UNIVERSITY OF AUCKLAND
Set one
A
B
C
D


THE UNIVERSITY OF AUCKLAND
Set two
A
B
C
D

Set three

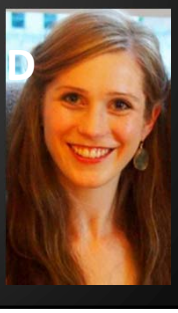
A



B

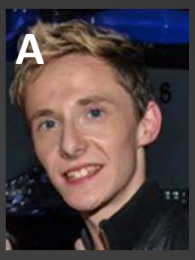
C

D



Set four

A



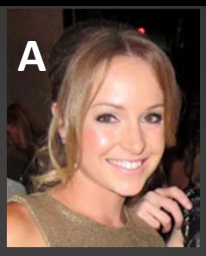
B

C

D

Set five

A



B

C

D

So how well did you do?

Set 1

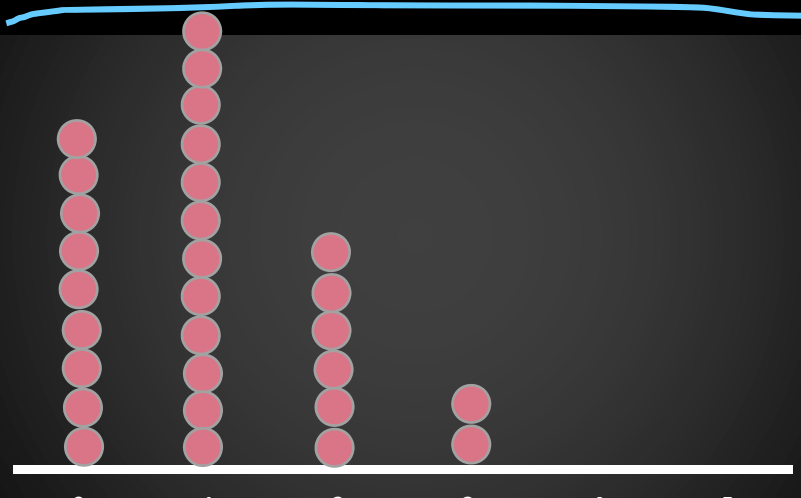
Set 2

Set 3

Set 4

Set 5

Our results



What's going on?

Discuss

- The shape of the distribution
- The mean and standard deviation (estimate these)
- Any unusual features/outliers
- What you think would be a very "unlikely" result if you were just guessing
- What do you think will happen if we conducted this with more students?

Discuss

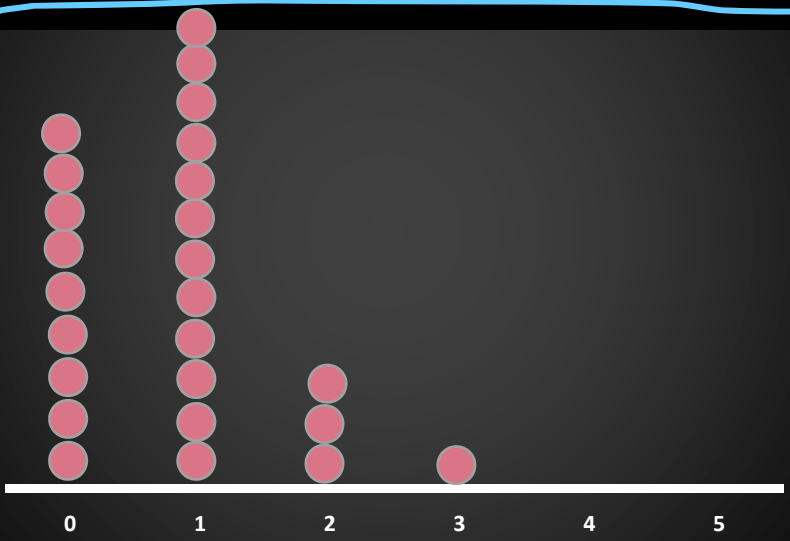
- Does this show evidence of anyone or the class as a whole being able to spot family resemblance?
- Are there any unexpected results?
- What if people were just guessing, and that's why they got it right?

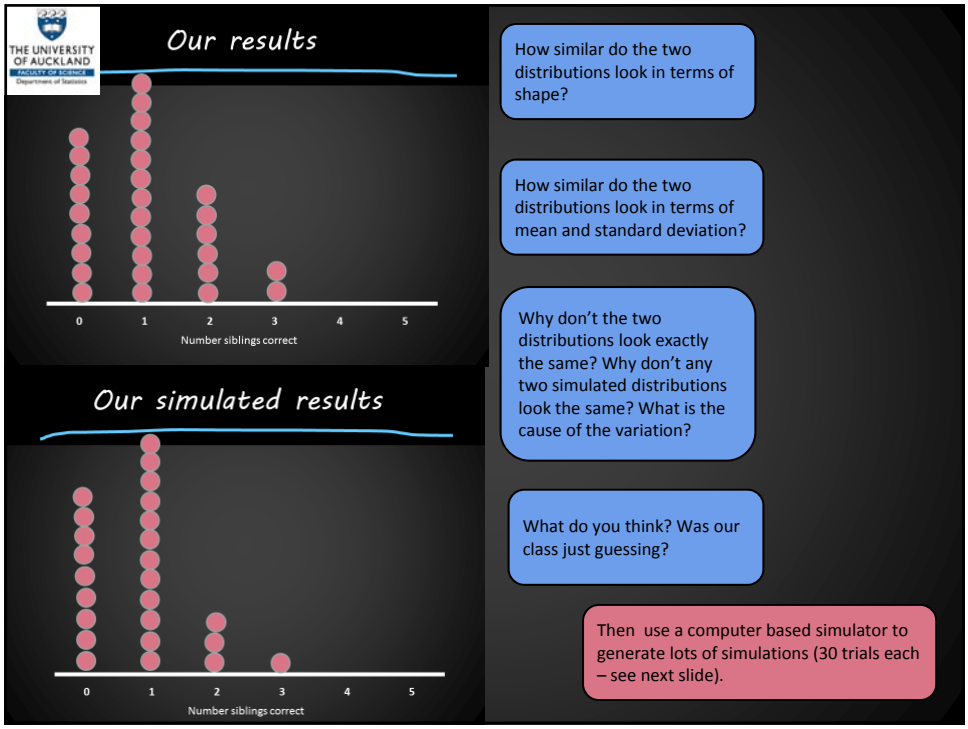
For our simulation, we want to use a **tool** that will allow us to mimic this "guessing" or random process. We want to **see** what kind of results you would get (e.g. how many correct out of 5) if you were just guessing which person was my sibling, so our **model** will be based on a probability of 25% of getting the correct sibling.

Progressing through tools…

See handout for activity for the different tools used....

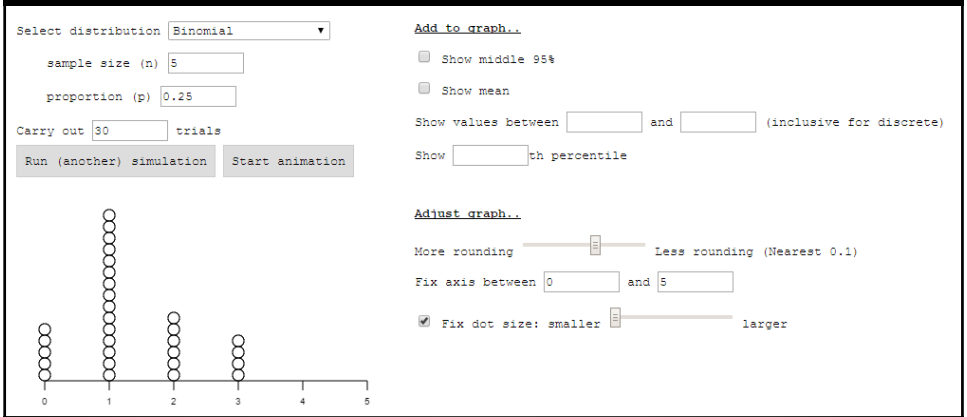
Our simulated results




THE UNIVERSITY OF AUCKLAND
Our results
0 1 2 3 4 5
Number siblings correct
Our simulated results
0 1 2 3 4 5
Number siblings correct
How similar do the two distributions look in terms of shape?
How similar do the two distributions look in terms of mean and standard deviation?
Why don't the two distributions look exactly the same? Why don't any two simulated distributions look the same? What is the cause of the variation?
What do you think? Was our class just guessing?
Then use a computer based simulator to generate lots of simulations (30 trials each – see next slide).


THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE
Department of Statistics
Information about the online simulation tool demonstrated…
The online simulation tool demonstrated during the "family resemblance" activity is available from:
https://www.stat.auckland.ac.nz/~martin/probsim/
If you have any questions about this tool, please email me ☺
anna-marie.martin@auckland.ac.nz

Using probsim



These are the settings I used
Click the button "Start animation" to generate a new simulation every second

Probability and inferential ideas that are important for level three.....

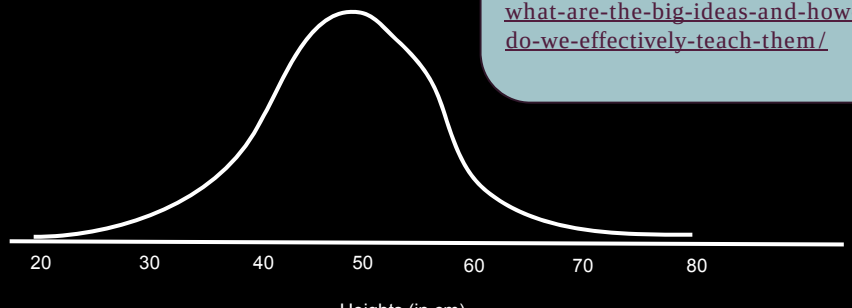
- What is the difference between true probability versus model estimates versus experimental estimates? What is randomness? What is independence?
- What does "chance acting alone" mean? What does chance variation look like? Is what I have observed unlikely if chance is acting alone?
- What would I expect to see? What is likely, what is unlikely? What is the central 95% of outcomes? Why do I see variation? Can I put a number value on this variation?
- What assumptions do you need to make when selecting probability models? Why won't the model be a perfect fit? What are the key visual features of different models? Why are we only estimating probabilities or means?

Distribution 2

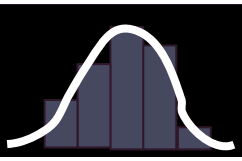
Sketching shapes of model distributions

My distribution is of heights of flowers in cm.
I am symmetrical.
I have a mean and median of 50 cm.
I am normally distributed.
My standard deviation is around 10 cm.

See my workshop from the statistics day 2014 for more slides like these sketching ones... http://new.censusatschool.org.nz/resource/probability-distributions-what-are-the-big-ideas-and-how-do-we-effectively-teach-them/

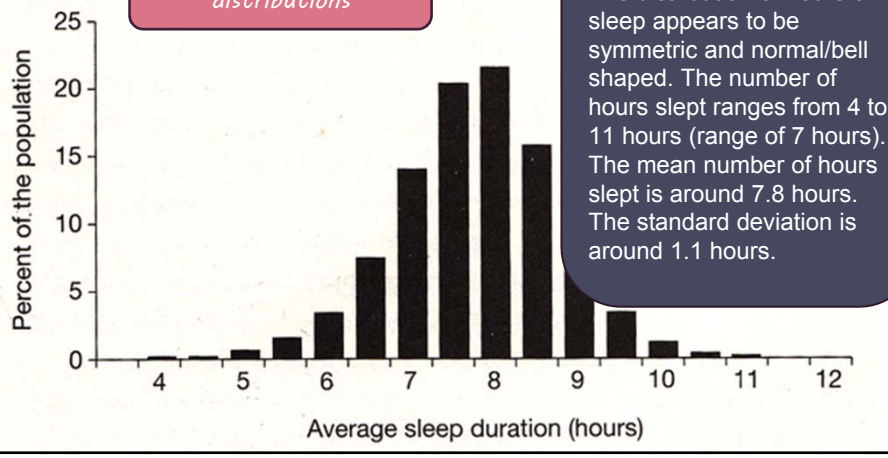


Features to describe IN CONTEXT:
Spread/range, median, mean, shape, variation/standard deviation.



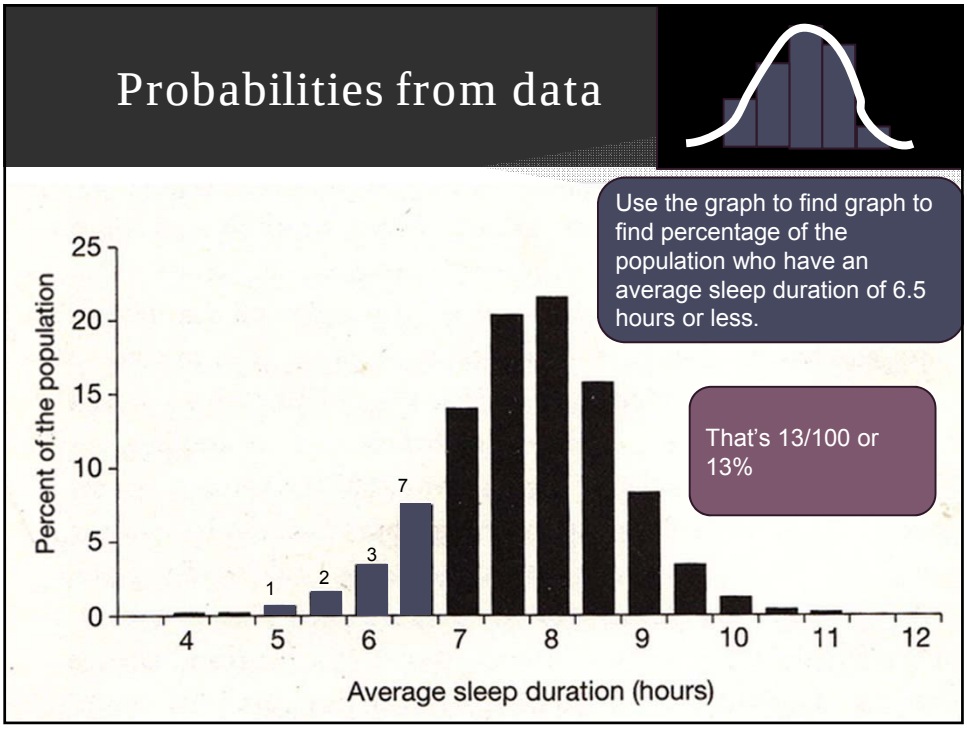
Describe as much as you can about the distribution below.

Looking at experimental distributions

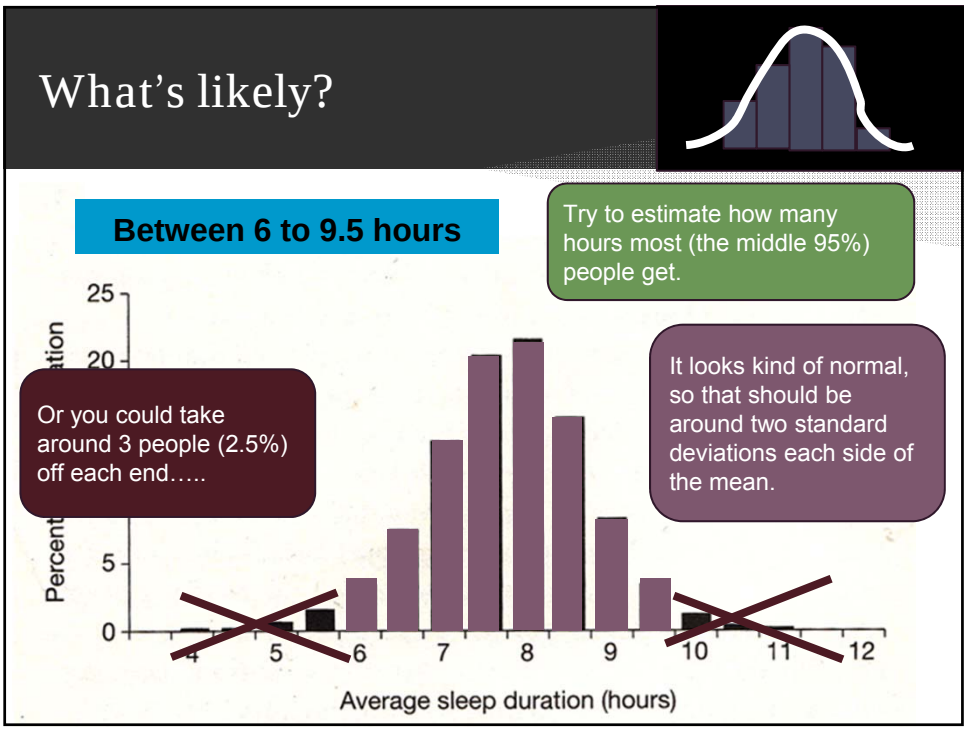


The distribution of hours of sleep appears to be symmetric and normal/bell shaped. The number of hours slept ranges from 4 to 11 hours (range of 7 hours). The mean number of hours slept is around 7.8 hours. The standard deviation is around 1.1 hours.

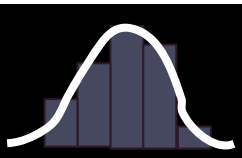
Probabilities from data



What's likely?



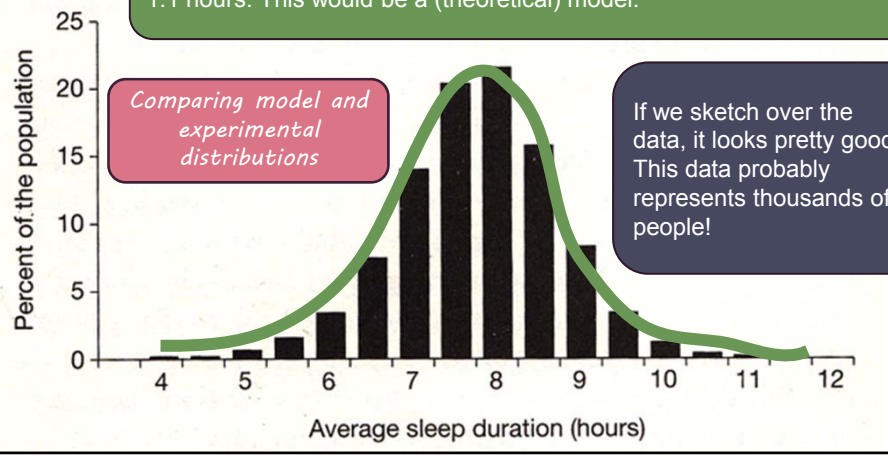
Comparing data with theoretical distributions



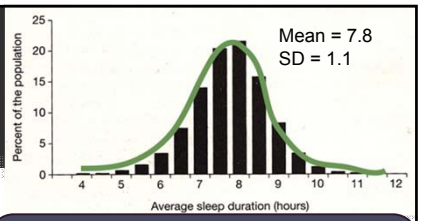
Based on this population data, we could say that the hours of sleep is normally distributed with a mean 7.8 hours and a standard deviation of 1.1 hours. This would be a (theoretical) model.

Comparing model and experimental distributions

If we sketch over the data, it looks pretty good. This data probably represents thousands of people!

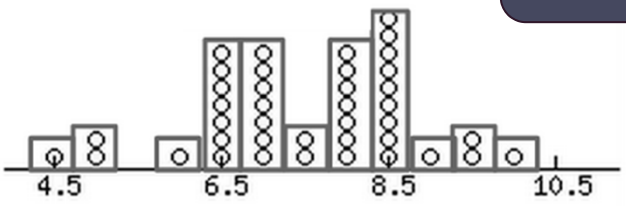


Comparing data with theoretical distributions



Suppose we asked 40 male teachers to measure their average sleep duration (in hours) per night.

Compare these results (the distribution of hours of sleep) with the model.



How similar are they? Compare features.

Hours sleep per night for Gender (Male)

Understanding impact of sample size

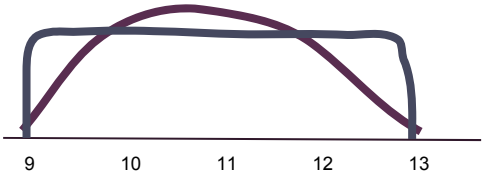
What are some reasons for the different features?

Small samples, heaps of variation

What do you think the distribution of the number of smarties in a mini box looks like?

You can only fit so many smarties in each box, and they also take into account weight, so the number of smarties should range between 9 and 13.

What kind of features??

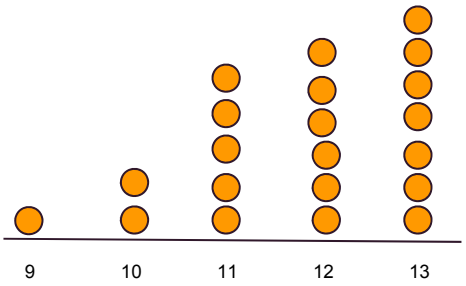


Let's investigate with a sample :-)

Considering assumptions for models

Small samples, heaps of variation

Our samples!!



Comparing real data.....

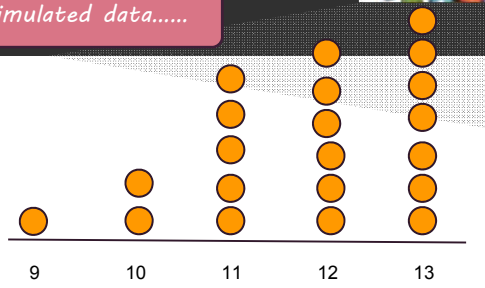
Small samples, heaps of variation

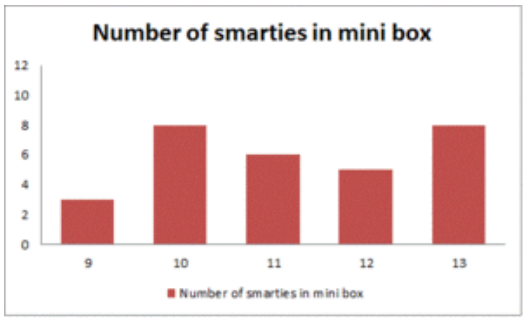
With simulated data.....

What do you think the distribution of the number of smarties in a mini box looks like?

From a sample of 30, we won't be able to get a clear picture of the shape of the distribution

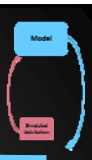
Here are a few different samples of 30 smarties boxes.





Introducing simulations...

Theoretical focus



My goals:

Focus on model construction e.g. structure, logic, algorithm to model chance situation – the model not the probability tool

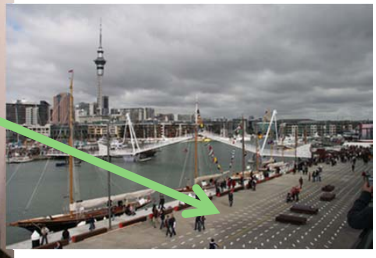
Build understanding of independence and the need for assumptions when modelling

Improve mathematical probability thinking

Promote the use of simulations as a decision making tool – changing the model, what ifs?

See you latte

This is Emma. She owns a coffee cart which she uses to sell coffee down by the waterfront.



Working ~~9 to 5~~ 7 to 4

She normally opens her coffee cart for nine hours a day - from 7am to 4pm.

That's potentially lots of coffee!!!!

Decisions, decisions...

She has to decide how much milk to buy each day for her coffee cart.

What does she need to consider to make this decision?

What to simula[illegible]?

Thinking about assumptions at the beginning of the modelling process...

Customers?
How many customers does she get per day? What coffees do they order? How much milk does each coffee need?

Got milk?

Time?
How many hours can she stay open if she only stocks up on milk once at the start of the day? What about queues? Do people wait patiently or go somewhere else?

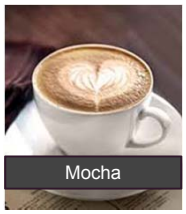
Milk?
What type? Full, skim, soy, other? How many bottles of each? How many bottles can she fit? Can she get more if she runs out? Does she keep milk over until the next day?

So many things to consider and so many assumptions to make ☺

Coffee addiction

Simplified situation

Sarah sells four types of coffee



Emma reckons she can get about 12 coffees out of each 2 litre bottle of milk

Different people like different types of coffee, so Sarah never knows exactly what the next customer will order. However, they all like full fat milk!

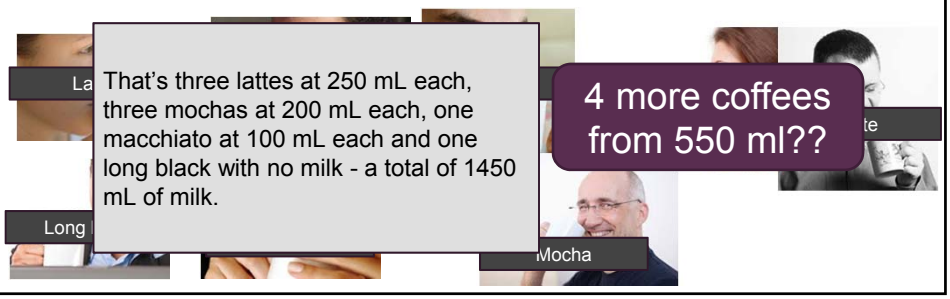
However, she reckons that 40% of customers order lattes, 35% order mochas, 15% order macchiatos and 10% order long blacks.

How much milk?

How likely is it that Emma will be able to make 12 coffees from 2 litres (2000 mL) of milk?

It depends on whether the customers order lattes, mochas, macchiatos and long blacks, as these use different amounts of milk.

For example… here are eight people. Let's randomly assign them a coffee based on those probabilities….



That's three lattes at 250 mL each, three mochas at 200 mL each, one macchiato at 100 mL each and one long black with no milk - a total of 1450 mL of milk.

4 more coffees from 550 ml??

Problem

Time for a coffee? ☺

Sarah wants to know how much milk she needs to make 12 coffees.

Design a simulation to investigate this problem.

You need to estimate the mean number amount of milk used for 12 coffees and the probability she will not have enough milk to make 12 coffees.

Problem Plan Data Analysis Conclusion

Plan:

- Need to model uncertainty about how much milk each coffee will use
- Need to assume there is no pattern to what type of coffees are ordered – it is unpredictable
- Need to simulate 12 coffees being purchased (see tool for details)
- Need to assume that the type of coffee ordered is independent from another
- Need to record how much milk was used for each coffee and then add this up (see tool for details)
- Need to assume (to keep things simple) that exactly the amounts stated are used for each type of coffee

Analysis:

- What appears to be the possible outcomes (the sample space)?
- How much milk is she likely to use?
- What is the shape of the distribution?
- Any unusual features/outliers?

For these 10000 trials:

- What was the mean amount of milk used?
- What percentage of trials used 2000 mL or more of milk?

THE UNIVERSITY OF AUCKLAND
FACULTY OF SCIENCE
Department of Statistics

Information about the online simulation tool demonstrated...

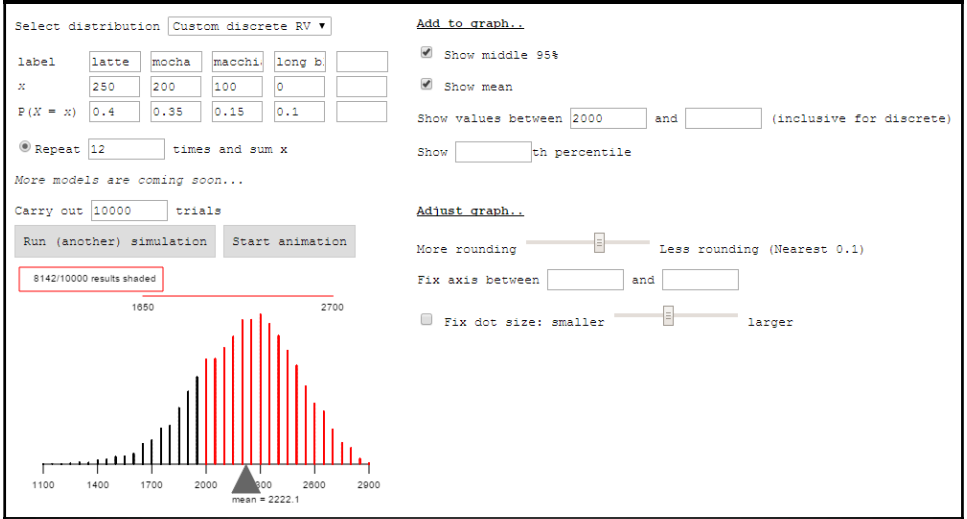
The online simulation tool demonstrated during the "see you latte" activity is available from:

https://www.stat.auckland.ac.nz/~martin/probsim/

If you have any questions about this tool, please email me ☺

anna-marie.martin@auckland.ac.nz

Using probsim



These are the settings I used for "see you latte"

Questions to stimulate higher level thinking…

Randomness - what specifically is the random process for the situation? Why can you model it?	Independence - what specific things are you assuming don't influence each other? Why does independence matter?	How will the results of the simulation help someone make their decision? Can you make any recommendations?
Number values given - why are you assuming these will stay the same? Will these always be the same? Could they be higher or lower? How would this affect your simulation?	Probabilities given - why are you assuming these will stay the same? Will these always stay the same? Will things run out or change? How would this affect your simulation?	What would real people actually do? How would the situation be like in real life? What other factors could influence the results? How could the simulation be re-designed to take these into account?
What are some issues that would affect the accuracy of your simulation? How would they make an affect?	What other factors would influence the decision making? Cost vs benefit?	How could you adjust your model/simulation? What would you expect to see if you changed an aspect of the design? Change it and see!