Reclaim & reform probability learning

Anne Patel, *September 2019*

The language of probability

When Nate Silver (fivethirtyeight.com) predicted Donald Trump had a 30% chance of winning the 2017 US election, people assumed he wouldn't win, this underscores the problem of reasoning about probability. Change the chance situation and the language to, how surprised would you be if you tossed two coins and both came up tails? Hopefully, not that surprised. This event has a 25% chance of happening, even less than Trump winning! So, should we have been surprised that Trump won? Did Nate Silver "get it wrong"? (RNZ interview with Dr Dillon Mayhew)

We can assign probabilities based on assumptions, by looking at the sky, and thinking about the season and what the weather was like yesterday, I predict there is a 10% chance of rain today, or we can estimate probabilities from data. Language connectors describe chance events in relation to others, such as "not", "at least" and "and". By describing, interpreting and representing chance events in general situations, using proportions, students learn how to use the language of chance and about ambiguities in the use of the word "or".

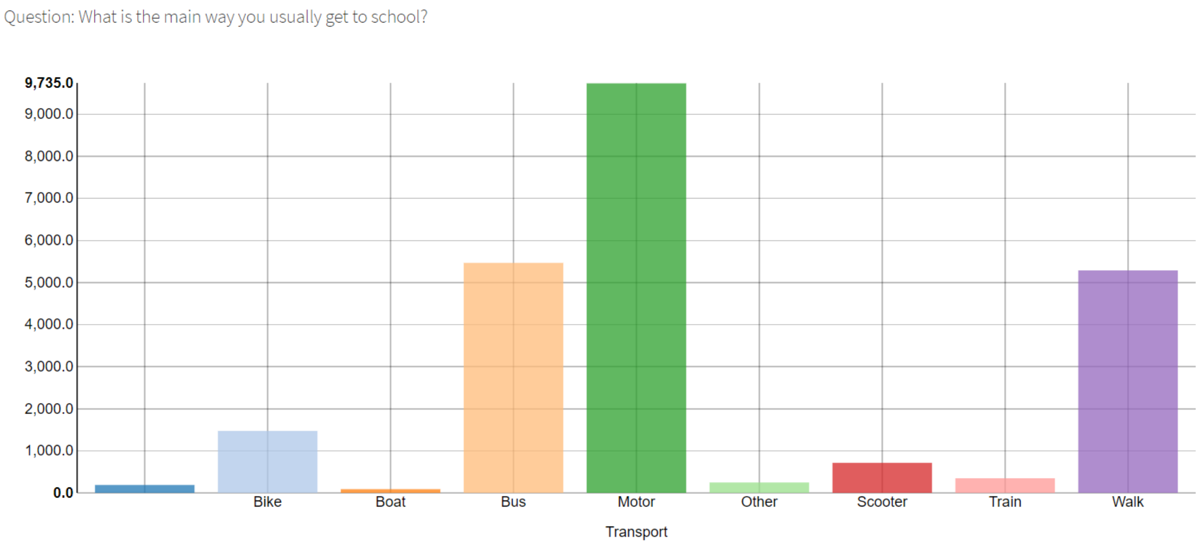
- 12 out of 25 or 48% of students do not have a pet.
- At least half of the students brought their lunch to school today.
- 64% of students use a bus or car to get to school.
- Am I safer to walk to school or go in the car?
- What is the probability a student chosen at random walks or takes a car to school?

Older student's need to consider situations involving two or three stages of chance experiments with more than one variable, using language such as "if", "given", "of", "knowing that" through estimating probabilities from data in everyday situations. Statistical words likely to have different everyday meanings for students, such as, random = weird and sample = tester, and opportunities must be given to allow the students to use statistical language correctly. This includes establishing that *risk* has the same meaning as probability or proportion. Distinguishing between relative and absolute risk is fundamental to reason with probabilities in data. Exploring these situations involves awareness of common mistakes interpreting language about chance and leads to investigating independence without any formal definitions. (See amsi.org.au).

Probability diagrams

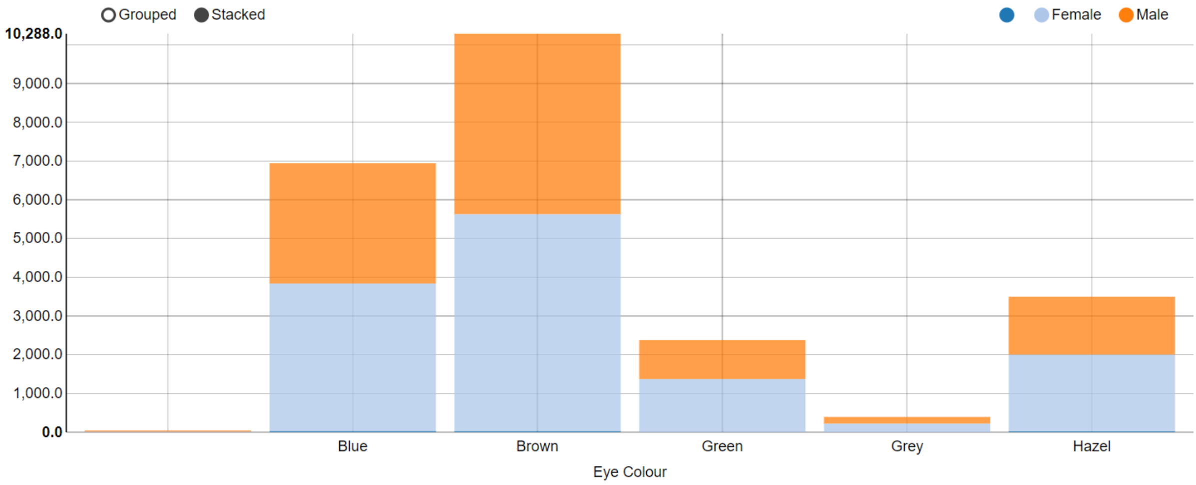
> *Probability is a* ***measure****, like length or area or weight or height, but a measure of the likeliness or chance of possibilities in some situation. Probability is a relative measure; it is a measure of chance relative to the other possibilities of the situation. Therefore, it is very important to be clear about the situation being considered. Comparisons of probabilities – which are equal, which are not, how much bigger or smaller – are therefore also of interest in modelling chance. (amsi.org.au)*

Hence, representing probabilities by areas or lengths (relative frequencies) optimises relational probabilistic reasoning, that encourages both a "sense of the problem" and correct language use when posing and answering questions arising in chance situations. The plots and tables were produced using 2019 CensusAtSchool data https://tabulator.docker.stat.auckland.ac.nz/



	Totals	Percentages
	188	0.8%
Bike	1473	6.26%
Boat	88	0.37%
Bus	5468	23.22%
Motor	9735	41.34%
Other	247	1.05%
Scooter	714	3.03%
Train	348	1.48%
Walk	5288	22.46%
Totals	23549	100%

- What is the probability a student walked or biked to school? Approximately 22% + 6% = 28% (P(W) + P(B) = P(W B)



	V1	Female	Male	Totals	Percentages
	0	25	21	46	0.2%
Blue	33	3806	3106	6945	29.49%
Brown	30	5599	4659	10288	43.69%
Green	8	1365	1006	2379	10.1%
Grey	3	221	170	394	1.67%
Hazel	24	1981	1492	3497	14.85%
Totals	98	12997	10454	23549	100%
Percentages	0.42%	55.19%	44.39%	100%	

- Given you are male, what is the probability you have green eyes? P(G|M) = 1006/10454 = 0.0962) Notice the symbolic language is reversed in the question.

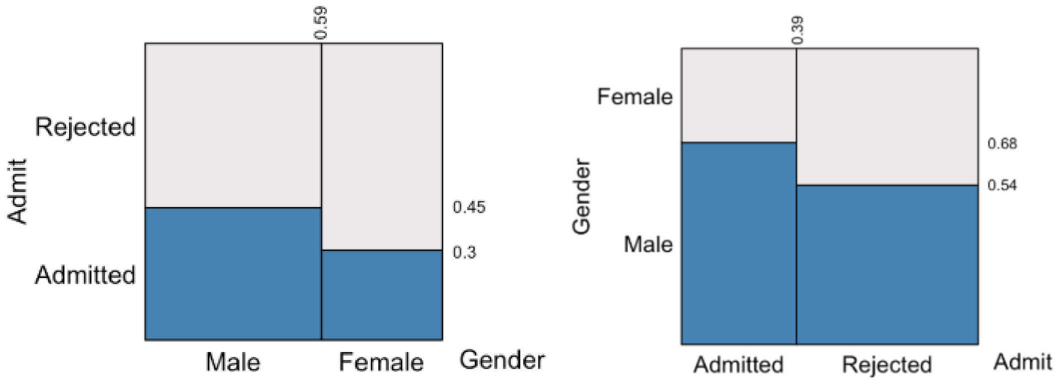
- If you have green eyes, what is the probability you are male? P(M|G) = 1006/2379 = 0.4229

Eikosograms use areas to reflect geometry in probability (see [R.W Oldford, 2018](#)).Their construction divides the unit square into:

- vertical bars with widths equalling the probabilities for each value of the conditioning variables(s),
- and then divides each vertical bar horizontally into
  - rectangles with heights equal to the conditional probabilities of each value of the response variate given the value(s) of the conditioning variable(s).

Therefore, the area of every rectangle equals the joint probability of the corresponding values of the conditioning and response variates.

Consider the example of admissions to the Berkley graduate program [R.W Oldford, (2018](#)) uses.



- Males are more likely (59%) to apply to Berkley than females (41%)
- Berkley receives 59% of it's applications from males and 49% from females
- If you are male there is a 45% chance of being accepted to Berkley
- If you are female there is a 30% chance of being accepted to Berkley
- You are 15% more likely to be accepted to Berkley if you are male?
- Being accepted to Berkley is affected by your gender.
- Being accepted to Berkley is conditional on your gender.

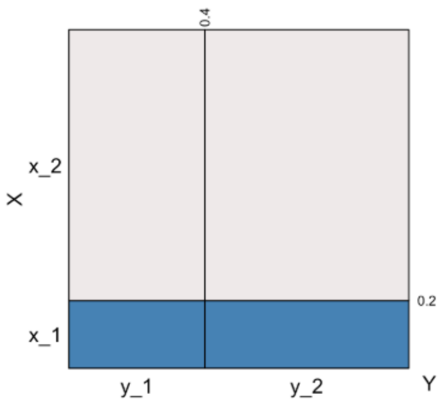
The power of the eikosogram is that the geometry reflects the probabilities of discrete random variables, when counts are superimposed on the eikosograms they also reflect data in tables, thus students can view and use multiple representations to reason about probability.

- the vertical axis shows the values of the response variable
- the horizontal axis shows the values of the conditioning variable(s)
- each axis ranges from 0 to 1
- total probability is 1, the area of the whole
- widths of rectangles are marginal probabilities
- heights of rectangles are conditional probabilities
- areas of rectangles are joint probabilities

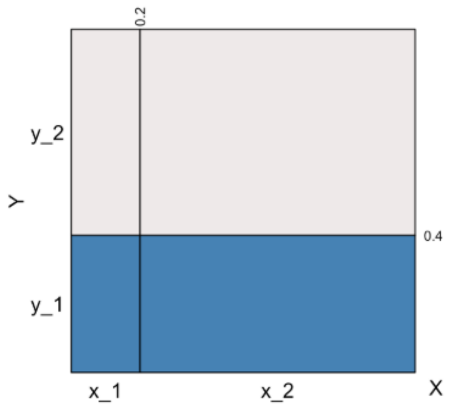
$$joint = area = height \times width = conditional \times marginal$$

- the width of an entire vertical bar is the area of that bar (since its height is 1) and is also the sum of the areas of the rectangles it contains
- the width of an entire vertical bar is the therefore the marginal probability for that value of the (horizontal) conditioning variable

Also, using eikosograms, probability comparisons and independence are easily visualised.



Independence of variates X and Y



Independence of variates X and Y

Two random variables X and Y are distributed independently of one another if their corresponding eikosgram is flat. (See Eikosograms to teach conditional and joint probability & Visualising chance.)

Visualisations using lengths and areas with extensive student experience of single event and conditioning language, may result in clearer understanding of what probabilities are used for, i.e. making decisions, calculating risk, exploring and interpreting data, and for underpinning statistical thinking and methods, including models for probability and data.

Order of teaching

Introduce conditional probability BEFORE independence, because ALL probabilities are conditional. Helen MacGilliveray (ICOTS11) advocates banning the term 'multiplication rule' instead realising P(A and B) = P(A|B)P(B) because we always multiply the probabilities of more than one event. We define independence as P(A | B) = P(A). Ross Parsonage writes:

- When second stage events are not independent of the first stage event the multiplication of probabilities is justified by P(A and B) = P(A | B) x P(B).
- When second stage events are independent of the first stage event the multiplication is justified by P(A and B) = P(A) x P(B) for independent events.

When calculating a probability of an outcome using a probability tree, there is a danger that students will multiply the probabilities on the branches without giving due consideration to the reasons why. The notation of events in the second stage should be *labelled* as conditional events when the second stage events are not independent of the first stage event.

When constructing a table from a story, which gives proportions or percentages in categories (rather than counts or frequencies), it is recommended that a table of counts is constructed rather than a table of proportions. Research has shown that information in a table of counts is more readily understood than information in a table of proportions. It is suggested that a large and easy-to-use number (such as 1000, 10 000, 100 000 or 1 000 000) is used as the sample size. This number does not need to relate to the size of any underlying population.

Multiple representations

When learning any concept, reasoning about multiple representations results in better coherence and integration. Venn diagrams for events are potentially misleading. Thus, calculate probabilities using two-way tables supplemented by Venn diagrams to allow students alternative reasoning and visualisation about chance events and how they interact. Consider the following problem with the table and diagram together providing reasoning about the formulas.

Given that $P(A) = 0.35$, $P(B) = 0.45$ and $P(A \cap B) = 0.13$

Find (a) $P(A \cup B) = P(A) + P(B) - P(A \cap B)$

$= 0.35 + 0.45 - 0.13$

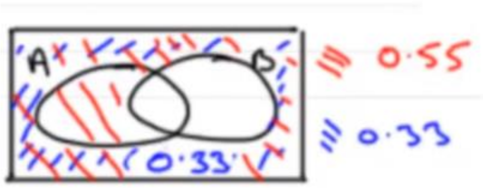
$= 0.67$

(b) $P(A' \mid B') = \frac{P(A' \cap B')}{P(B')}$

$= \frac{0.33}{0.55}$

$= 0.6$

Event	B	B′	Total
A	0.13	0.22	0.35
A′	0.32	0.33	0.65
Total	0.45	0.55	1



Visualising chance distributions

Noticing and wondering about randomness and features of distributions seen in multiple simulations is paramount to developing concepts of chance in data (see Konold & Kazak, 2008). This requires the use of simulations using technology, such as TinkerPlots. Technology allows us to appreciate the phenomenon, of increased stability in chance distributions over increasing numbers of trials, known as the Law of Large Numbers. Moreover, visualise sample variation based on random events, not available through manual tossing of coins or other physical tools of chance.

In two or more step chance experiments student need to assign probabilities to outcomes in situations involving selections either with or without replacement. There are different ways of assigning probabilities, NOT different types of probabilities.

- Estimate
- Model
- Belief

} Combination of any of these

Statistics educators are urged to incorporate probability as part of statistical investigation cycles (PPDAC) and models, see Randomness and Chance activities, Data Analysis and Modeling Activities and A models and Modelling approach. Dalrymple and Grant propose drawing students attention to common concepts in statistics and probability such as variation, distribution and sampling variation in learning experiences concerning based on chance seen in data. These experiences ensure students are aware of the chance in data and the data in chance.

Visualising and reasoning about Randomness

Humans find it difficult to identify randomness and random events, we just don't "see it" in our daily lives, and if we do, we are likely to assign a cause, other than randomness, to the event. Therefore, it falls to statistics teachers to make the invisible visible. Students need experience visualising and reasoning about randomness every year, at least three times a year.

Two activities that stimulate and expose students to randomness and human bias and follow on effects for sample selection. These indicate the need to randomly select samples as much as possible in the sampling process.

From: Rouncefield, M & Holmes, P., (1989) *Practical Statistics.* Macmillan Education Ltd: London. McIntyre, R.

Practical 6.4 Pebbles

Use a pile of about 40 or 50 pebbles, from which samples of 5 are to be taken and weighed. To take random samples, number each pebble using an adhesive label, find the weight of each pebble in the pile and calculate the population mean. It has been suggested that the largest stone will rarely be included in non-random samples so that the samples will usually underestimate the true mean.

From: G. Noether, *Teaching Statistics at its best.* (1994) pp40.
Mental random numbers: Perceived and real randomness

Ask students to write down five random selections from the digits 1,2 and 3. To analyse the data, we are only interested in the last two digits they wrote. (The first three help get into the swing of things and tapping a pencil on the desk for timing can help.)

The table gives the frequencies of nine possible outcomes of 450 students.

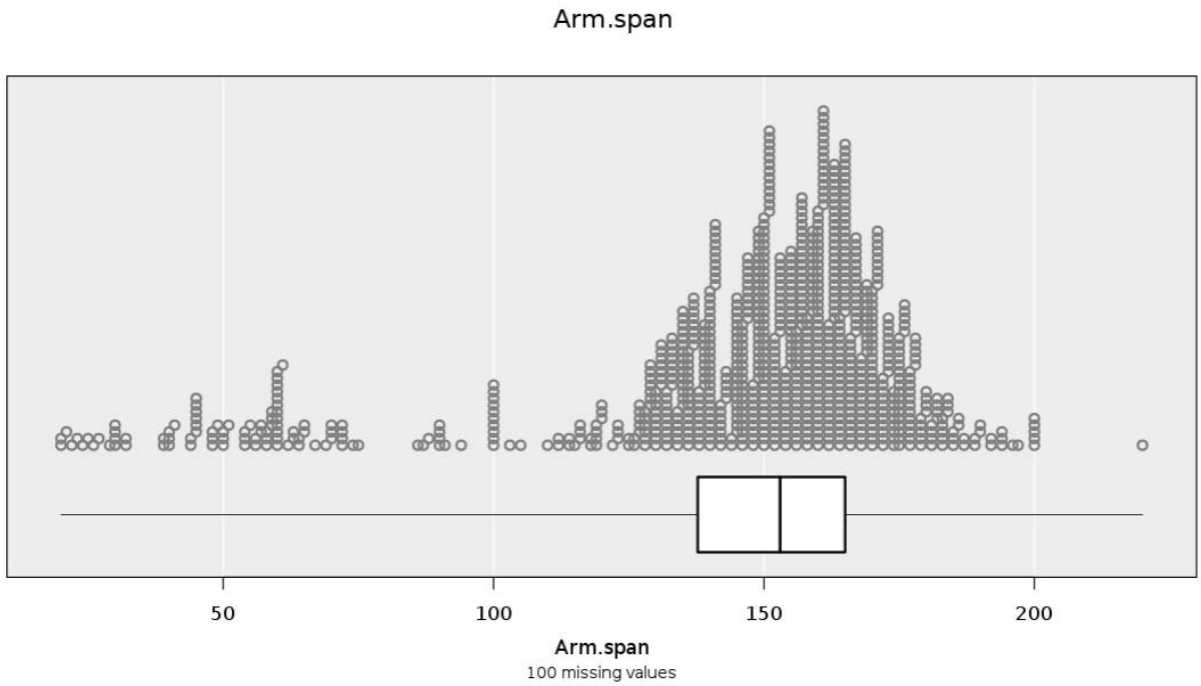
		Second digit			
		1	2	3	Totals
First digit	1	31	72	60	163
	2	57	27	63	147
	3	53	58	29	140
	Totals	141	157	152	450

Assuming the students were randomly writing down numbers, the nine combinations are equally likely, each having probability $\frac{1}{3} \times \frac{1}{3} = \frac{1}{9}$

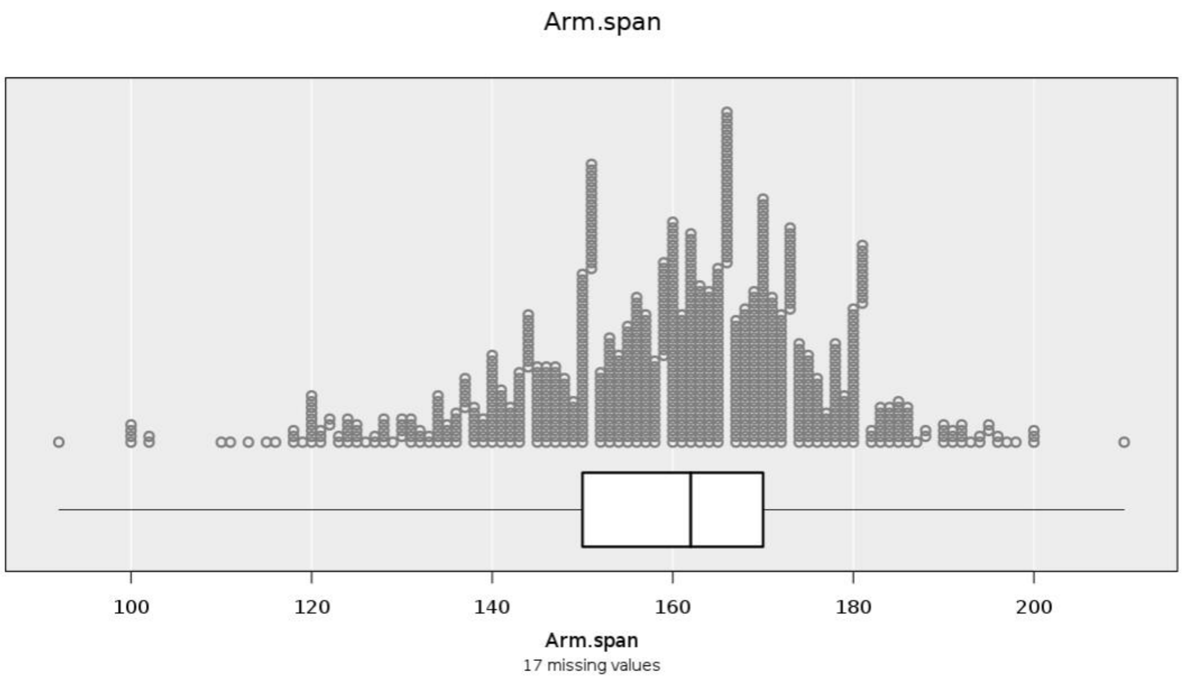
Expected frequencies for each cell then equal 50. Visual inspection tells us that the observed frequencies for 1,1 and 2,2 3,3 are smaller than expected under conditions of true randomness, while the frequencies for the non-identical digit pairs are too large. Interestingly, in spite of the lack of independence of successive digit selection, the marginal totals for both the first and second digits do not exhibit any significant deviations from theoretical frequencies. These can be confirmed by appropriate chi-square tests.

Data are everywhere

It falls to teachers of statistic to educate future citizens about the census and the use of census data. One of the easiest ways to do this is for students to take part in CensusAtSchool. By collecting measures themselves, students become aware of variation in measures. Through exploring the data, using data analysis or probabilities, students are aware of the assumptions and limitations behand data. For example, the 2005 data revealed a surprising amount of people with short arms, and 100 people out of a 1000 who did not answer the question! That’s 10% of the sample! Does it matter?



The good news is the armspan data from CensusAtSchool 2015 look a lot more normal with less missing data, so students (and their teachers) must be taking measurement data and non-response issues more seriously!

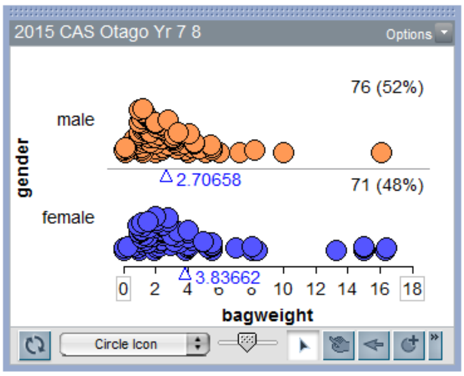


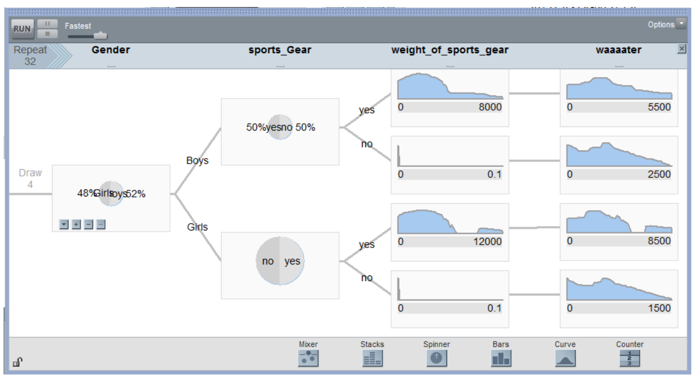
Building models

When students build models based on real data, the aim is for students to go further than *reading behind the data* to *investigating beneath the data* in order to understand that random events are ubiquitous and to develop an appreciation of what randomness looks like, how it behaves, and how its effects might begin to be quantified when making statistical inferences. (Patel & Pfannkuch, 2020). Technology such as the TinkerPlots Sampler enables students to model real world situations. See [A models and modelling approach].

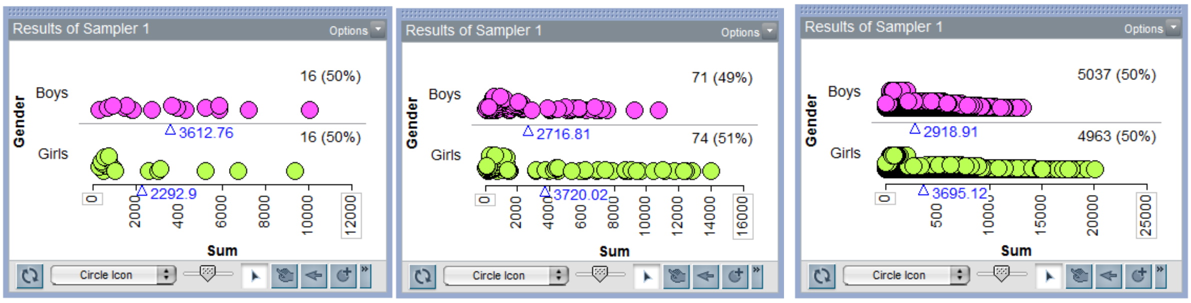
Examining a random sample of 147 Year 7 and 8 students bag weights in the Otago region (CAS 2015) lead to the construction of a model of the situation.

Real data: n=147 | Conditional Model constructed by Year 8 student





Simulated data: n=32 | Simulated data: n=145 | Simulated data: n=10000



Notice that in the small simulated sample n=32 the means are larger for males and smaller for females, by chance alone. Can we be sure there is a difference of nearly 1kg in the bag weights of males and female Year 7 and 8 students in the Otago region?

What are the factors the student considered cause bag weight? Are they all accounted for? Can we trust the model results?