Level 2 Simulation



A

B

C



**By Liz Sneddon**

### Name:

### What is Simulation?

Simulation is the imitation of the operation of a real-world process or system over time.

There are many things that we cannot calculate the theoretical probability for. For example, what is the probability that the photocopier in M8 is going to jam today? We can collect data to give us an idea of what the probability might be, and then we can use this data to run a simulation to come up with a more accurate prediction.

Think of a flight simulator for example:



We don’t want to put trainee pilots into a real airplane to learn to fly, so instead they put them in a simulator, where they practice again and again and again with a computer model, before they go up in a real aircraft.

This is a little like what we are going to do for this assessment. We are going to be given a situation, and then design a simulation for that situation. We will then run the simulation and evaluate how good the simulation design was.

### 

### Problem

There are two types of investigation question that you could be asked to do.

Either:

Investigate the **probability** that an event occurs

Or:

Investigate the **average number of times** that an event occurs.

## Example

If Mrs Sneddon has 4 children, what is the probability that 3 of her children have blue eyes. The probability of any one child having blue eyes is 0.7.

Or:

What is the average number of children Mrs Sneddon has to have before she has two children in a row (one after the other) that have blue eyes. The probability of any one child having blue eyes is 0.7.

The investigation question will be given to you as part of the description of the situation you need to simulate.

### Plan

# Decide what tool to use

You have several tools that you can use to simulate probabilities. They include:

* Coin (gives a 0.5 chance of success)

A

B

C

* Die (gives a 1/6 chance of success)
* Spinner (probability depends on how many sections there are)
* Calculator (can match **ANY** probability)

The key step is to match the probability of the example to the tool.

Remember that probabilities can be written as a decimal, fraction, or percentage.

For Level 2, decimals are the most common form.

## Example 1:

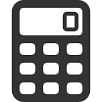
I want to simulate having a child. Each child has a 50% (or 0.5) chance of being a girl.

Tool choices:

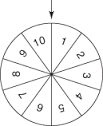
* Coin
  + where a head matches a girl, and
  + a tail matches a boy.
* Die

A

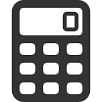
B

* + where the numbers 1, 2 and 3 match to a girl, and
  + the numbers 4, 5 and 6 match to a boy.
* Spinner - have 2 equal sized sections, then
  + the girl matches one side, and
  + the boy matches the other side.
* Calculator - generate the numbers 1 or 2,
  + the number 1 matches a girl, and
  + the number 2 matches a boy

## Example 2:

I want to simulate a 70% (or 0.7) chance of getting rain today.

Tool choices:

* Spinner – with 10 equal categories,
  + the numbers 1, 2, 3, 4, 5, 6, 7 match rain occurring, and
  + the numbers 8, 9, 10 match no rain.
* Calculator – generate the numbers 1 to 10,
  + where the numbers 1 to 7 match rain occurring, and
  + the numbers 8 – 10 match no rain.

## Exercise:

Decide which tools you can use to simulate the following situations.

1. The chance of eating a healthy lunch today is 0.2
2. The probability of missing the bus today is 0.2
3. The probability of completing homework tonight is 0.5

# Decide allocation of random numbers to probabilities and outcomes

## Step 1:

You must know what all the possible outcomes are.

Each outcome (e.g. blue eyes) has a single probability associated with it.

All probabilities must add up to 1.

## Example

Probability of blue eyes = 0.7

## Exercise:

List the outcomes for the following situations:

1. Rolling a dice:

Outcomes:



1. Tossing a coin:

Outcomes:



1. The weather today:

Outcomes:

1. Eye colour:



Outcomes:

## Step 2:

Calculators and computers can generate random numbers between 0 and 1.

If we use the formula “RANDBETWEEN” in Google Sheets, it will turn it into a whole number.

## Exercise:

1. Open a Google Sheet. Type in the formula: **=RANDBETWEEN(1,10)**

Write down the random number you generated \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. In the next cell, type in the formula: **=RANDBETWEEN(1,100)**

Write down the random number you generated \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Step 3:

Say you have 10 random numbers, between 1 and 10, you now want to match these random numbers and outcomes to the **probability** of each outcomes.

If the probability that you are simulating is to **one** **decimal place** (e.g. 0.2), then generate **10 random numbers** and allocate these to the outcomes.

If the probability that you are simulating is to **two** **decimal places** (e.g. 0.45), then generate **100 random numbers** and allocate these to the outcomes.

## Example

Probability of blue eyes = 0.7.

Because the probability is to one decimal place, I want to generate 10 random numbers between 1 and 10 and allocate them.

To match to blue eyes, I want 7 random numbers, because .

**Allocation:**

1, 2, 3, 4, 5, 6, 7 – these numbers match to having blue eyes, and

8, 9, 10 – these numbers match to **not** having blue eyes.

## Full Example

The probability of a child being born with blue eyes is 0.3, being born with brown eyes is 0.4, being born with green eyes is 0.2, and being born with grey eyes is 0.1.

Decide the allocation of random numbers to the probabilities and outcomes. Then describe the allocations.

**Step 1:** What are all the different outcomes?

The outcomes are: blue eyes, brown eyes, green eyes or grey eyes.

These probabilities all add up to 1. ()

**Step 2:** How many random numbers will you have?

The probabilities are shown to 1 decimal place, so I want 10 random numbers.

I choose the numbers 1, 2, 3, 4, 5, 6, 7, 8, 9, and 10.

**Step 3:** Match up the random numbers and outcomes to the **probability** of each outcome. I’m going to put this info into a table:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** | **Blue eyes** | **Brown eyes** | **Green eyes** | **Grey eyes** |
| **Probability** | 0.3 | 0.4 | 0.2 | 0.1 |
| **Random number matching** | 1, 2, 3 | 4, 5, 6, 7 | 8, 9 | 10 |

The numbers 1, 2 and 3 represent a child having blue eyes, because there are three outcomes, which matches the probability of blue eyes of

The numbers 4, 5, 6 and 7 represent a child having brown eyes, because there are four outcomes, which matches the probability of brown eyes of

The numbers 8 and 9 represent a child having green eyes, because there are two outcomes, which matches the probability of green eyes of

The number 10 represents a child having grey eyes, because there is one outcome, which matches the probability of grey eyes of

## Exercise:

Decide the allocation of random numbers to the probabilities and outcomes for each of the following situations.

1. A family was going to the pet store to choose 6 pets. The probability that they would chose a dog is 0.4, the probability that they would chose a cat is 0.3, the probability that they would chose a fish is 0.2, and the probability that they would chose a bird is 0.1.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** |  |  |  |  |
| **Probability** |  |  |  |  |
| **Random number matching** |  |  |  |  |

**Description:**

1. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** |  |  |  |  |
| **Probability** |  |  |  |  |
| **Random number matching** |  |  |  |  |

**Description:**

# Decide the number of trials

Simulation data is discrete or counting data.

A minimum sample size or number of trials of **50** is required.

## Example

When simulating the eye color of children, we want to generate 50 random numbers, matching the random numbers to the eye color of 50 children.

# Equation to generate random numbers

We will be using Google Sheets to do the simulation and will use the formula:

**=RANDBETWEEN(a,b)**

## Example

I want to generate random numbers between 1 and 10.

Formula: =RANDBETWEEN(1,10)

## Exercise:

Make equations to generate 10 random numbers using either a calculator or the computer from the following information.

1. Generate 10 random numbers between 1 and 5.

Random Numbers:

1. Generate 10 random numbers between 4 and 20.

Random Numbers:

1. Generate 10 random numbers between 1 and 100.

Random Numbers:



# Interpreting random numbers

For each situation, once we have generated random numbers, we need to match this back to the outcome that it represents.

## Example

Here is the random number allocation table for the colour of a child’s eyes:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** | **Blue eyes** | **Brown eyes** | **Green eyes** | **Grey eyes** |
| **Probability** | 0.3 | 0.4 | 0.2 | 0.1 |
| **Random number matching** | 1, 2, 3 | 4, 5, 6, 7 | 8, 9 | 10 |

Using the formula “=RANDBETWEEN(1,10) I generated the following random numbers:

3, 4, 4, 7, 5, 9, 1, 4, 8

What do these random numbers represent?

Each number represents the colour of another child’s eyes. So I match up the first number 3 to the outcome, and this represents a child with blue eyes. I do the same now for all the random numbers:

3, 4, 4, 7, 5, 9, 1, 4, 8

Blue, Brown, Brown, Brown, Brown, Green, Blue, Brown, Green

Therefore there would be 9 children, of which 5 had Brown eyes, 2 had Blue eyes, and 2 had Green eyes.

## Exercise:

Write the equations, generate 10 random numbers, and explain what they represent for the following situations.

1. A family was going to the pet store to choose a pet. The probability that they would chose a dog is 0.4, the probability that they would chose a cat is 0.3, the probability that they would chose a fish is 0.2, and the probability that they would chose a bird is 0.1.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade.

# Describe a trial, including the finishing point

We need to combine the previous skills together: Tool, Allocation of random number to outcomes and probabilities, Equation for generating random numbers, and Deciding the number of trials.

## Example

The probability of a child being born with blue eyes is 0.3, being born with brown eyes is 0.4, being born with green eyes is 0.2, and being born with grey eyes is 0.1.

A married couple are planning to keep having children until they have all four different colored eyes in their family.

**Tool**

The tool I am going to use is a random number generator on Google Sheets.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Outcome** | **Blue eyes** | **Brown eyes** | **Green eyes** | **Grey eyes** |
| **Probability** | 0.3 | 0.4 | 0.2 | 0.1 |
| **Random number matching** | 1, 2, 3 | 4, 5, 6, 7 | 8, 9 | 10 |

Formula: =RANDBETWEEN(1,10)

Number of trials = 50

**Description of simulation**

We want to run a simulation, with 50 trials, generating random numbers in each trial.

One trial consists of generating random numbers until we have one of each of the 4 outcomes (Blue eyes, Brown eyes, Green eyes and Grey eyes).

This means that in each trial, I will keep generating random numbers until I have one of each eye colour. Then I will move on to the next trial, and in total I will generate 50 trials.

## Exercise:

Write a full plan for each of the following situations:

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

### Data

# Create a results table

In the assessment, our results have to be recorded in sufficient detail that the examiner can check the accuracy of the solution.

This means you need to record which random numbers you generate, and how many numbers you generate in each trial.

You will be using Google Sheets.

Remember your sample size is at least 50, so you will need at least 50 rows to record the data.

## Example

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample** | **1st simulation** | **2nd simulation** | **3rd simulation** | **Etc** | **Total Frequency** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **3** |  |  |  |  |  |
| **…** |  |  |  |  |  |
| **50** |  |  |  |  |  |

## Example

The probability of a child being born with blue eyes is 0.3, being born with brown eyes is 0.4, being born with green eyes is 0.2, and being born with grey eyes is 0.1.

A married couple are planning to keep having children until they have all four different colored eyes in their family.

**Data**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample** | **Child 1** | **Child 2** | **Child 3** | **…** | **Total Frequency** |
| **1** |  |  |  |  |  |
| **2** |  |  |  |  |  |
| **…** |  |  |  |  |  |
| **50** |  |  |  |  |  |

## Exercise:

Create a results table in Google Sheets for each of the following situations. Make sure that you have each table has 50 rows because you will need to use this to do simulations in the next exercise.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

# Do the simulation and Record the results

There are several parts to doing the simulation, so rather than explaining it with screen shots, please watch the video with an example before trying the exercises for yourself.

## Exercise:

You should have a Google Sheet set up for the simulations. Generate random numbers and complete 50 trials for each situation.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

### Analysis

If you go back to the problem from the beginning, there are two types of investigation questions that you could be asked.

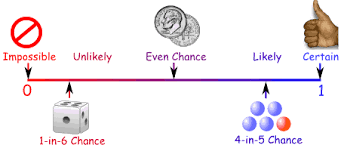
Either:

Investigate the **probability** that an event occurs

Or:

Investigate the **average number of times** that an event occurs.

# Calculating a probability

We want to know what the chance of an event occurring in the long-run is.

That is, for our simulation, what is a good estimate of the chance of an event occurring.

General formula:

## Example

A married couple are planning to keep having children until they have all four different colored eyes in their family.

What is the chance that if they have 10 children, they would have all four eye colors in their children?

**Data**

(I have shown a sample size of 10, but you will have your sample size of 50.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample** | **Blue eyes** | **Brown eyes** | **Green eyes** | **Grey eyes** | **Total Frequency** |
| **Random Number** | **1, 2, 3** | **4, 5, 6, 7** | **8, 9** | **10** |  |
| **1** | |||||| | ||| | ||| | | | 13 |
| **2** | ||||| | |||||| | |||| | | | 16 |
| **3** |  |  |  |  | 8 |
| **4** |  |  |  |  | 12 |
| **5** |  |  |  |  | 18 |
| **6** |  |  |  |  | 16 |
| **7** |  |  |  |  | 16 |
| **8** |  |  |  |  | 15 |
| **9** |  |  |  |  | 12 |
| **10** |  |  |  |  | 17 |

**Analysis**

## Exercise:

For each of the scenarios that you have been working on in the previous exercises, you have now created a results table for each.

Using each of these results tables, calculate the probability as requested below.

Then write the result in a sentence.

1. A family was going to the pet store to choose several pets. What is the probability that the family needs to buy at least 12 pets before they have at least one dog, one cat, one fish and one bird?
2. What is the chance that in the 8 assessments available this year, that Julie get at least one Excellence and one Merit this year.

# Calculate a measure of center (mean or median)

The center is trying to find where the middle of the data is. There are two measures we can use, both of which estimate where the center is.

The mean is when you add up all the frequencies and divide by your sample size.

The median is when you list all the numbers in order from smallest to largest, and the median is the number that is in the middle position.

## Example

A married couple are planning to keep having children until they have all four different colored eyes in their family. What is the average number of children they need to have before they are likely to have all four eye colors?

**Data**

(I am only showing a sample size of 10, but you will need to do this with your sample size of 50.)

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sample** | **Blue eyes** | **Brown eyes** | **Green eyes** | **Grey eyes** | **Total Frequency** |
| **Random Number** | 1, 2, 3 | 4, 5, 6, 7 | 8, 9 | 10 |  |
| **1** | |||||| | ||| | ||| | | | 13 |
| **2** | ||||| | |||||| | |||| | | | 16 |
| **3** |  |  |  |  | 8 |
| **4** |  |  |  |  | 12 |
| **5** |  |  |  |  | 18 |
| **6** |  |  |  |  | 16 |
| **7** |  |  |  |  | 16 |
| **8** |  |  |  |  | 15 |
| **9** |  |  |  |  | 12 |
| **10** |  |  |  |  | 17 |

**Mean:**

**Median:**

List the numbers in order from smallest to largest.

8, 12, 12, 13, 15, 16, 16, 16, 17, 18

There are two numbers in the middle: 15 and 16.

So, add them together and divide by 2.

# 

Luckily for us, instead of needing to calculate this by hand, we can copy and paste the data (Total Frequency) into **NZGrapher**. Plot the **dot plot** and **box and whisker plot**, along with the **summary statistics**.

Watch the video on how to do this.

## Exercise:

For each of the scenarios that you have been working on in the previous exercises, you have now created a results table for each. Using each of these results tables, calculate the mean and median for the data.

1. A family was going to the pet store to choose several pets. What is the average number of pets the family needs to buy before they have at least one of each type?

Mean =

Median =

1. What is the average number of assessments that Julie needs to sit this year for her to be reasonably certain that she gets at least one Excellence and one Merit this year.

Mean =

Median =

### Conclusion

There is one thing you need to do in your conclusion:

* Answer the investigation question

# Answer the investigation question in context (center or probability)

The investigation question could either be about where the center is (e.g. the “average”), or it could be about what the probability or chance of an event is.

You need to write a sentence describing the center or probability, answering this investigation question.

## Example

**Probability investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family.

What is the chance that if they have 10 children, they would have all four eye colors in their children?

**Analysis**

**Conclusion**

The chance that if the married couple have 10 children, that they would have all four eye colors, is 1 out of 5, or 0.2 or 20%.

## Example

**Average investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family. What is the average number of children they need to have before they are likely to have all four eye colors?

**Analysis**

**Conclusion**

The mean number of children the married couple need to have until they have all four different colored eyes in their family is 14.3 children.

The median number of children the married couple need to have until they have all four different colored eyes in their family is 15.5 children.

## Exercise:

Use your results from previous exercises to help answer the following investigation question and write a conclusion.

1. A family was going to the pet store to choose several pets. What is the average number of pets the family needs to buy before they have at least one of each type?
2. A family was going to the pet store to choose several pets. What is the probability that the family needs to buy at least 12 pets before they have at least one dog, one cat, one fish and one bird?
3. What is the average number of assessments that Julie needs to sit this year for her to be reasonably certain that she gets at least one Excellence and one Merit this year.
4. What is the chance that in the 8 assessments available this year, that Julie get at least one Excellence and one Merit this year.

****

### Plan

# Describe at least one assumption

There are many things that we have to assume in order for the conclusion of the simulation to be valid and reliable.

For Merit, the one certain, and most important assumption, is that the outcomes must be independent of one another.

## Example

For our example of the four different colored eyes in the children of a married couple, we are **assuming** that the **eye color of each child is independent** (and not connected in any way) to the eye color of another child.

However, if we think about what role genetics play in determining eye color, the eye color of the parents is critical in determining what eye color the children can possibly have.

For example, if both parents have blue eyes, then as this is a recessive characteristic, the ONLY possible eye color that their children can have is blue.

## Exercise:

Decide whether or not it is valid to assume independence for each, justifying and explaining why or why not.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

# Clear explanation of allocation of probability to outcomes

When you are explaining how to allocate the probability, you need to connect the outcome, to the probability, and then to the random numbers.

## Example 1:

The first outcome is having a **child with blue eyes** and this has a probability of **0.3**. Random numbers of **1, 2 and 3** represent getting this outcome, because 3 random numbers out of 10 matches the probability of 0.3 = 3/10.

The second outcome is having a **child with brown eyes** and this has a probability of **0.4**. Random numbers of **4, 5, 6 and 7** represent getting this outcome, because 4 random numbers out of 10 matches the probability of 0.4 = 4/10.

The third outcome is having a **child with green eyes** and this has a probability of **0.2**. Random numbers of **8 and 9** represent getting this outcome, because 2 random numbers out of 10 matches the probability of 0.2 = 2/10.

The fourth outcome is having a **child with grey eyes** and this has a probability of **0.1**. Random number of **10** represents getting this outcome, because 1 random number out of 10 matches the probability of 0.1 = 1/10.

## Exercise:

For each of the scenarios below, give a detailed explanation of how you allocate the probability to the outcome.

You may find it helpful to look back at what you wrote in the Achieved skill section.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

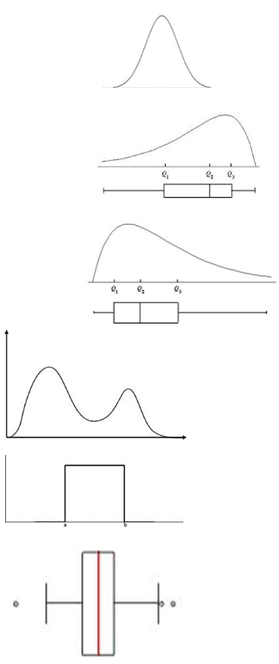
### Analysis

# Describe the shape of the distribution

Once the graph(s) has been drawn, then it is important to interpret the graph and explain what it shows.

One key feature of a graph is its shape, as this gives an indication of what distribution the simulated data may be from.

Here is a brief reminder of the different shaped distributions:



**Normal distribution**

(hill/mound shapes, symmetric, bell shaped curve)

**Left skewed**

(Tail is on the left hand side)

**Right Skewed**

(tail is on the right hand side)

**Multimodal**

(there is more than one peak)

**Uniform**

(the sides are straight and it looks like a box)

### Conclusion

# Results discussed in context, with reference to estimates

The results of the simulation are like a sample taken from a larger population.

The results are estimates, and we can’t know the exact population distribution.

Our simulation gives us a good estimate of the long-run-probability or average.

## Example

**Probability investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family.

What is the chance that if they have 10 children, they would have all four eye colors in their children?

**Analysis**

**Conclusion**

The chance that if the married couple have 10 children, that they would have all four eye colors, is 1 out of 5, or 0.2 or 20%.

This chance is an estimate of the true population probability. It is our best guess at what the true probability may be.

## Example

**Average investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family. What is the average number of children they need to have before they are likely to have all four eye colors?

**Analysis**

**Conclusion**

The mean number of children the married couple need to have until they have all four different colored eyes in their family is 14.3 children.

The median number of children the couple need to have is 15.5 children.

This mean and median are estimates of the true population center. It is our best guess at what the true center of the distribution may be.

## Exercise:

Use your results from previous exercises to help answer the following investigation question. Write a conclusion in context with a discussion on the summary statistics being estimates.

1. A family was going to the pet store to choose several pets. What is the average number of pets the family needs to buy before they have at least one of each type?
2. A family was going to the pet store to choose several pets. What is the probability that the family needs to buy at least 12 pets before they have at least one dog, one cat, one fish and one bird?
3. What is the average number of assessments that Julie needs to sit this year for her to be reasonably certain that she gets at least one Excellence and one Merit this year.
4. What is the chance that in the 8 assessments available this year, that Julie get at least one Excellence and one Merit this year.

# Sampling variability

The variation in a sample statistic from sample to sample. Or in our case, the variation in a sample statistic from simulation to simulation.

Suppose a *simulation* is run and a sample statistic, such as a *sample mean*, is calculated. If a second simulation of the same size is run from the same *population*, it is almost certain that the sample mean calculated from this simulation will be different from that calculated from the first simulation.

If further sample means are calculated, by repeatedly running simulations of the same size from the same population, then the differences in these sample means illustrate sampling variation.

If I run another simulation …

* When another simulation is run, you will produce different random numbers, therefore your data will differ from simulation to simulation.
* However, the long-run-probability or center is fixed for the population distribution, so each simulation should represent this
* This means that the analysis and conclusion are likely to remain the same.

## Example

**Probability investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family. What is the chance that if they have 10 children, they would have all four eye colors in their children?

**Analysis**

**Conclusion**

The chance that if the married couple have 10 children, that they would have all four eye colors, is 1 out of 5, or 0.2 or 20%.

If I ran another simulation, I would have different data values, as I would generate different random numbers on my calculator.

However, I would expect the estimates of the chance of having all four eye colors to be similar to what I have estimated in this simulation.

## Example

**Average investigation question:**

A married couple are planning to keep having children until they have all four different colored eyes in their family. What is the average number of children they need to have before they are likely to have all four eye colors?

**Analysis**

**Conclusion**

The mean number of children the married couple need to have until they have all four different colored eyes in their family is 14.3 children.

The median number of children the couple need to have is 15.5 children.

If I ran another simulation, I would have different data values, as I would generate different random numbers on my calculator.

However, I would expect the estimates of the average number of children they needed to have to be similar to what I have estimated in this simulation.

## Exercise:

Use your results from previous exercises to help answer the following investigation question and write a discussion on sampling variability for each situation.

1. A family was going to the pet store to choose several pets. What is the average number of pets the family needs to buy before they have at least one of each type?
2. A family was going to the pet store to choose several pets. What is the probability that the family needs to buy at least 12 pets before they have at least one dog, one cat, one fish and one bird?
3. What is the average number of assessments that Julie needs to sit this year for her to be reasonably certain that she gets at least one Excellence and one Merit this year.
4. What is the chance that in the 8 assessments available this year, that Julie get at least one Excellence and one Merit this year.

### 

### Plan

# Describe a second assumption

There are many other assumptions that you can discuss in every simulation. Many of these assumptions apply specifically to the context or situation.

Some common ideas to think about are:

* Is the probability of each outcome going to remain constant throughout the situation? For example, is the chance of rain today likely to be the same as the chance of rain tomorrow?
* The random numbers generated on the calculator are true random numbers.
* The sample you have in your simulation is representative of the true distribution.
* etc

## Example

For our example of the four different colored eyes in the children of a married couple, we are assuming that the eye color of each child is independent (and not connected in any way) to the eye color of another child.

One assumption made here is that the probability of each eye color remains constant for each child that they have. This assumption may be reasonable as each child will have a mix of genetics inherited from each parent.

Another assumption is that the simulation that we have run is a representative sample from the true population. This is a reasonable assumption as we have taken a sample size of 50 and generated random numbers without any bias.

A third assumption about the children’s eye color remaining independent may not hold true if any of the children were to be identical twins or triplets for example. In this case, these identical siblings would not be independent in their eye color.

A fourth assumptions may be that all other factors in their lifestyle, that may affect the genetic selection of eye color, remain constant. For example, we assume that the food that they eat remains similar for each pregnancy, and that the food a person eats has no effect on the eye color of a child during pregnancy.

## Exercise:

Discuss what assumptions may exist for each of the following situations.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

# Effect of the assumptions on the simulation

If any of the assumptions we discussed were not upheld (i.e. they were false), then we need to think how this would affect the validity and reliability of our simulation, the analysis and conclusion.

## Example

An analogy for you. If the house that you lived in was built on a shaky foundation, would you trust that house to hold up in every storm?

The answer of no is the same that applies to simulation. If the assumptions that our model is based upon are shaky, then the model itself will be shaky, and therefore the results and conclusions will be shaky.

If there is only a small crack in the foundation of your house, then there is still a good chance that the house will hold up. Similarly with our model. If one of the assumptions has only got a small effect, then the model will still hold up most of the time.

If however there is a large crack, then it is likely the house will fall down. Similarly with our model. If one of the assumptions is very likely to be false, then the model, the results and the conclusion are all likely to fall over and be wrong too.

## Exercise:

Think about the assumptions you discussed for each of the following situations. Write a discussion for each about how this may or may not affect your conclusion.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.

### Conclusion

# Statistical Insight shown

## Exercise:

Explain which of these graphs is more accurate and has a smaller spread.

**Sample Size**

If we generate more random samples, say 1000 rather than 50, we get **more** accurate.

* This means that the spread decreases.
* Therefore the accuracy of our conclusion is likely to increase.

If we have a larger number of trials …

* The data will be more representative of the population.
* The results will be more accurate.
* The conclusion will be more accurate, and therefore the long-run-probability and center estimates will be more accurate.

## Exercise:

Write a discussion for each of the situations below about how the sample size affects the accuracy and reliability of the conclusions.

1. A family was going to the pet store to choose several pets. The probability that they would find a dog is 0.4, the probability that they would find a cat is 0.3, the probability that they would find a fish is 0.2, and the probability that they would find a bird is 0.1. The family want to get at least one dog, one cat, one fish and one bird.
2. For each assessment this year, Julie thinks she has a 32% chance of getting Not Achieved, a 48% chance of getting Achieved, a 17% chance of getting a Merit, and the remainder is the chance of her getting an Excellence grade. Julie wants to get at least one Excellence and one Merit this year.