STUDY TIME

LEVEL 3 STATISTICS

PROBABILITY CONCEPTS

NCEA Workbook Answers

Section One
The Foundations

1. Key Terms and Equations

a. The actual values of a population, rather than the probability they happen.

b. An outcome or set of outcomes that we are interested in.

c. The chance of something an event occurring, measured between 0 and 1.

d. "The probability of", for example P(Z ∩ Y) is the same as saying "the probability of A and B".

e. The intersection of two events is every outcome where both events happen.

f. The union of two events is everything that is either or both events.

g. Probabilities calculated based on something else that has happened.

h. The probability of a particular group having a problem.

i. When we compare the probabilities of two groups having the same problem.

j. How different all the bits of data in the sample are.

k. How often an event happens.

l. A consistent frequency value that comes out of conducting many trials.

m. $\dfrac{\textit{number of favourable outcomes}}{\textit{total number of outcomes}}$

n. $\dfrac{\textit{number of favourable outcomes}}{\textit{total number of outcomes}} \times 100$

o. $\dfrac{\textit{trials of interest}}{\textit{total number of trials}}$

p. $\dfrac{\textit{Group A with L}}{\textit{Total Group A}}$

q. $\dfrac{\textit{Risk of A}}{\textit{Risk of B}}$

2. Types of events

2.1 The Union of Events:

a. The union of events is everything that fits into either or both events or categories. P(A ∩ B) represents the probability of A OR B occurring.

b. You have to subtract P(A ∩ B) because it is included in both the P(A) and the P(B), so you have to remove it so that you do not count it more than once.

2.2 Complementary Events:

a. Complementary events are events which cannot happen together. There is no possible way both outcomes can occur at the same time.

b. The probabilities of complementary events must always add to one.

c. P(A') = 1 – P(A)

2.3 Independent Events:

a. Independent events are events which do not influence one another. One event happening does not change the probability of the other event happening.

b. You multiple the probabilities together. For example, $P(A \cap B) = P(A) \times P(B)$ if you are certain the events are independent.

c. When two events are independent the probability of one event occurring does not influence the probability of the other event occurring (from (a)). Therefore, if B has occurred, the probability that A occurs is just the probability of A because it is not influenced by B, and vice versa.

2.4 Mutually Exclusive Events:

a. Mutually exclusive events are events that cannot happen at the same time.

b. Complementary events are mutually exclusive events. However, when complementary events combine, they fill the entire sample space where mutually exclusive events do not.

c. The intersection of two mutually exclusive events is zero.

d. The union of two mutually exclusive events is calculated by adding the probabilities of each event. E.g. $P(A \cup B) = P(A) + P(B)$

3. Types of Probability

3.1 True Probability:

a. True probability is the (almost always) unknown actual probability that an event will occur in a given situation.

b. We can never absolutely know the true probability of an event occurring in a specific situation. For example, when we roll a dice we may assume that the true probability of getting a 5 is 1/6, however, the dice may be slightly irregular influencing how it rolls, or the surface that we a throwing it onto may influence the final position of the dice changing the probability and not reflecting the true probability.

c. We can estimate the true probability of an event in a given situation by using the theoretical probability, which will provide a close as possible estimate of the true probability.

3.2 Theoretical Probability:

a. Theoretical probability is the calculated chance of the occurrence of events.

b. P(pottery stall and crepe stall) = P(pottery stall) × P(crepe stall)

= 0.78 × 0.97

= 0.7566 (4 d.p.)

c. *i.* P(correct first cup and first saucer) = $\frac{1}{5}$

= 0.2

= 20%

ii. P(correct second teacup and saucer) = $\frac{1}{4}$

= 0.25

= 25%

iii. P(two teacup attempts) = $\frac{1}{5} \times \frac{1}{4} = \frac{1}{20}$

= 0.05 = 5%

iv. P(more than two attempts to match) = 1 – P(two teacup attempts)

= $1 - \frac{1}{20}$

= $\frac{19}{20}$

= 0.95 = 95%

v. The probability that it will take more than two attempts to match the correct cup with the correct saucer is 0.95 or 95%.

3.3 Experimental Probability:

a. Experimental probability is the probability found by conducting numerous trials.

b. *i.* P(4 attempts) = $\frac{90}{500}$

= 0.18

= 18%

ii. P(less than four attempts) = P(2 attempts) + P(3 attempts)

$= \frac{25}{500} + \frac{50}{500}$

$= \frac{75}{500}$

= 0.15

= 15%

iii. What is the probability that a person took more than four attempts to match the teacups?

$P(x > 4) = P(x = 5) + P(x = 6) + P(x = 7) + P(x = 8) + P(x = 9)$

$= \frac{110}{500} + \frac{115}{500} + \frac{55}{500} + \frac{45}{500} + \frac{10}{500}$

$= \frac{335}{500}$

= 0.67

= 67%

OR

$P(x > 4) = 1 - (P(x < 4) + P(x = 4))$

= 1 - (0.18 + 0.15)

= 1 - 0.33

= 0.67

= 67%

iv. He matched the cups faster than most of the people he surveyed at the market. This is because 67% of people surveyed took more than four attempts, and only 15% of people surveyed took less than four attempts.

c. *i.* P(two attempts to match) = $\frac{25}{500}$

= 0.05

= 5%

ii. P(more than two attempts) = $1 - (\frac{25}{500})$

= 1 – 0.05

= 0.95

iii. The experimental probability is the same as the theoretical probability that was identified previously.

4. Venn Diagrams

a. You should use a Venn diagram when you have at least two groups or events and the events are not mutually exclusive.

b. *i.* P(Only ride the elevator) = 0.3

= 30%

P(Only climb the Memorial Tower) = 0.2

= 20%

P(Ride the elevator and climb the Memorial Tower) = 0.45

= 45%

P(Did not ride the elevator or climb the Memorial Tower) = 0.05

= 5%

ii. P(rides the elevator) = P(only rides the elevator) + P(rides the elevator and climbs the tower)

= 0.30 + 0.45

= 0.75

= 75%

iii. P(climbs the tower) = P(only climbs the tower) + P(rides the elevator and climbs the tower)

= 0.20 + 0.45

= 0.65

= 65%

iv. P(either ride elevator or climbs the memorial tower)

= P(Ride the elevator) + P(Climb Memorial tower) – P(both)

= 0.75 + 0.65 - 0.45

= 0.95

= 95%

OR

P(either ride elevator or climbs the memorial tower)

= 1 – P(neither option)

= 1 – 0.05

= 0.95

= 95%

c. *i.* P(no attractions) = $\frac{77}{1300}$

= 0.0592

= 5.92%

ii. P(all three attractions) = $\frac{100}{1300}$

= 0.0769

= 7.69%

iii. What we want to find is the following:

P(dinosaur slide or octopus swing) = P(dinosaur slide) + P(octopus slide) – P(dinosaur slide and octopus swing)

However, we need to work out each of the components first:

P(dinosaur slide) = $\frac{368 + 234 + 100 + 23}{1300}$

= $\frac{725}{1300}$

= 0.5577

P(octopus swing) = $\frac{467 + 234 + 100 + 21}{1300}$

= $\frac{822}{1300}$

= 0.6323

P(dinosaur slide and octopus swing) = $\frac{234 + 100}{1300} = \frac{334}{1300}$ = 0.2569

Then we can put it back in to the first equation:

P(dinosaur slide or octopus swing) = 0.5577 + 0.6323 – 0.2569

= 0.9331

= 93.31%

OR

P(dinosaur slide or octopus swing) = $\frac{467 + 234 + 100 + 21 + 23 + 368}{1300}$

= 0.9331

= 93.31%

iv. What is the probability that a person either only went on the train or only went on the octopus swing?

P(only went on the train) = $\frac{10}{1300}$

$= 7.692 \times 10^{-3}$

P(only went on the octopus swing) = $\frac{467}{1300}$ = 0.3592

P(only train or only octopus swing) = $7.692 \times 10^{-3} + 0.3592$

= 0.3669

= 36.69%

OR

P(only train or only octopus swing) = $\frac{10 + 467}{1300}$

= 0.3669

= 36.69%

d. *i.* Number of people who visited either = 32 – 10

= 22

ii. (museum or art gallery) = museum + art gallery – (museum and art gallery)

2 = 19 + 11 – (museum and art gallery)

P(museum and art gallery) = (19 + 11) – 22

= 8

iii. How many people just visited the museum?

Only museum = 19 – 8

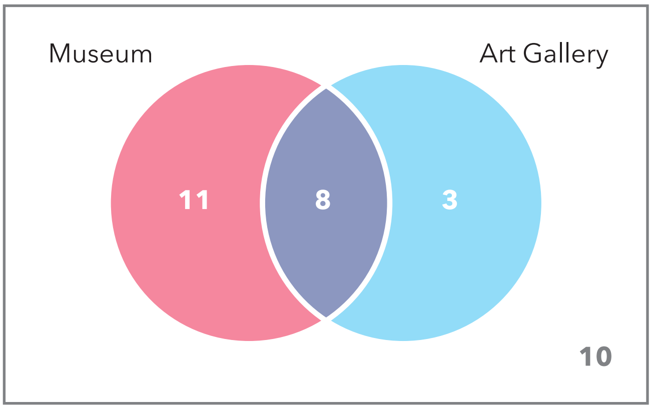
= 11

iv. How many people just visited the art gallery?

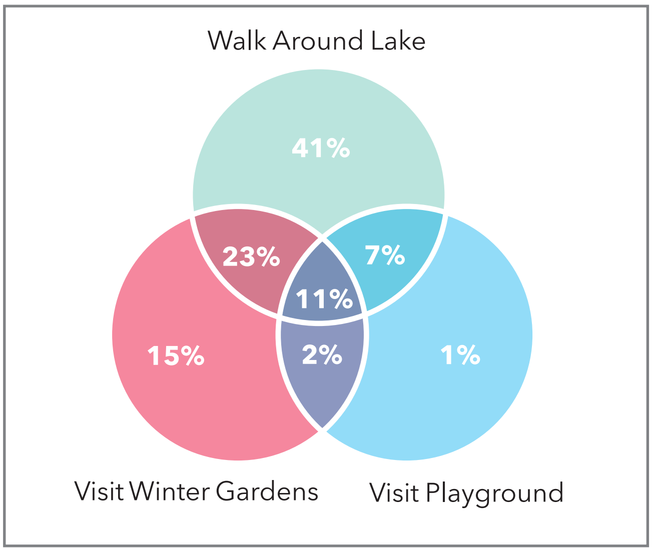
Only art gallery = 11 – 8

= 3

v.



e. *i.*



ii. P(Visited gardens) = 15 + 23 + 11 + 2

= 51%

iii. P(either walked around lake or visited playground)

= 1 – P(only visited gardens)

= 1 – 0.15

= 0.85

= 85%

iv. P(only walk around lake) = 41%

P(only visit gardens) = 15%

P(only visit playground) = 1%

P(only walk around lake or only visited gardens or only visited playground) = 41 + 15 + 1

= 57%

v. P(walk around lake and visit gardens) = 23 + 11

= 34%

P(walk around lake and visit playground) = 7 + 11

= 18%

P(walk around lake and visit gardens or walk around lake and visit playground but not all three) = 34 + 18 – (2 × 11)

= 30%

Need to remove the 11% of people who did all three twice because it was counted in both probabilities.

OR

P(walk around lake and visit gardens or walk around lake and visit playground but not all three) = 23 + 7 = 30%

Taken directly from the corresponding circles of the Venn Diagram.

5. Two-Way Tables

a. *i.* P(been on Paddle Steamer) = $\frac{1960}{4000}$

= 0.49

= 49%

ii. P(in Primary School) = $\frac{1200}{4000}$ = 0.3

P(not in Primary School) = 1 – 0.3

= 0.7

OR

P(not in Primary School) = P(in Intermediate School) + P(in High School)

= $\frac{1000}{4000} + \frac{1800}{4000}$

= 0.7

iii. P(in Intermediate School and has not been on Paddle Steamer) = $\frac{580}{4000}$

= 0.145

= 14.5%

iv. P(been on Paddle Steamer | in Primary School)

= $\frac{790}{1200}$

= 0.6583 (4 d.p.)

= 65.83%

v. P(in High School | been on Paddle Steamer)

= $\frac{750}{1960}$

= 0.3827 (4 d.p.)

= 38.27%

b. *i.*

	Good surf at Kai Iwi	Poor surf at Kai Iwi	Total
Good surf at Castlecliff	37%	30%	67%
Poor surf at Castlecliff	5%	28%	33%
Total	42%	58%	100%

ii. P(good surf at Castlecliff) = 67%

iii. P(good surf at Castlecliff and good surf at Kai Iwi) = 37%

iv. P(good surf Castlecliff | poor surf Kai Iwi) = $\frac{30}{58}$

= 0.5172

= 51.72%

v. P(good surf Kai Iwi | poor surf Castlecliff) = $\frac{5}{33}$

= 0.1515

= 15.15%

vi. P(good surf at Castlecliff and good surf at Kai Iwi) = 37%

= 0.37

Total number of days = 2 × 4

= 8 days

Expected number of days = total number of days × P(good surf at Castlecliff and good surf at Kai Iwi)

Expected number of days = 8 × 0.37

= 2.96

Tamara should expect that 2 days of the next four weekends will have good surf at both Castlecliff Beach and Kai Iwi Beach.

c. *i.*

	Know about the blockhouse	**Does not know about the blockhouse**	**Total**
Male	0.22	0.28	0.5
Female	0.23	0.27	0.5
Total	0.45	0.55	1

ii. P(female and does not know about blockhouse) = 0.27

= 27%

iii. P(knows about blockhouse | male) = $\frac{0.22}{0.5}$

= 0.44

= 44%

iv. P(know about the Cameron Blockhouse) = 0.45

Expected number of people = P(know about blockhouse) × number of people surveyed

= 0.45 × 134

= 60.3

Bob should expect that 60 people know about the Cameron Blockhouse.

v. P(female and knows about blockhouse) = 0.23

P(all four are female and know) = 0.23^4

= 2.7984×10^{-3}

= 0.0027984

vi.
- Assumption that the three people randomly selected are independent from each other to allow us to multiply the probabilities.
- Assumption that the survey size is sufficiently large so that replacement is not required to keep the probabilities relatively constant

OR

- Assumption that when one person is randomly selected, they are replaced back into the survey to so that the probability remains constant.

6. Probability Trees

a. *i.* P(walk and picnic) = 0.60 × 0.45

= 0.27

ii. P(walk and no picnic) = 0.60 × 0.55

= 0.33

P(drive and no picnic) = 0.40 × 0.20

= 0.08

P(no picnic) = 0.33 + 0.08

= 0.41

iii. P(no picnic | drives) = 0.20 (can read this directly off the tree, go to the drive branch then read the probability of no picnic of the subsequent branches)

OR

P(no picnic | drives) = P(no picnic and drives) ÷ P(drives)

= (0.40 × 0.20) ÷ 0.4

= 0.08 ÷ 0.4

= 0.2 (NOTE: this is the same as above)

iv. P(walk | picnic) = P(walk and picnic) ÷ P(Picnic)

Before we can calculate the P(walk | picnic) we need to work out the probability that a person has a picnic.

P(Picnic) = P(walk and picnic) + P(drive and picnic)

= (0.60 × 0.45) + (0.40 × 0.80)

= 0.59

Then we can go back to the formula and plug in the numbers.

P(walk | picnic) = P(walk and picnic) ÷ P(Picnic)

= (0.60 × 0.45) ÷ 0.59

= 0.4576 (4 d.p.)

v. P(picnic) = P(walk and picnic) + P(drive and picnic)

= (0.60 × 0.45) + (0.40 × 0.80)

= 0.59

Expected number to have a picnic = 170 × 0.59

= 100.3

You would expect that 100 people will have a picnic on Sunday.

b. *i.* P(two patterned socks) = ($^{26}/_{40}$) × ($^{25}/_{39}$)

= $^{650}/_{1560}$

= 0.4167 (4 d.p.)

ii. P(patterned and plain) = $(^{26}/_{40}) \times (^{14}/_{39})$

$= {}^{364}/_{1560}$

P(plain and patterned)

$= (^{14}/_{40}) \times (^{26}/_{39})$

$= {}^{364}/_{1560}$

P(plain and patterned in any order) = P(patterned and plain) + P(plain and patterned)

$= {}^{364}/_{1560} + {}^{364}/_{1560}$

$= {}^{728}/_{1560}$

= 0.4667 (4 d.p.)

iii. P(patterned sock first | plain sock second) = P(patterned sock first and plain sock second) ÷ P(plain sock second)

First we need to work out the P(plain sock second):

P(plain sock second) = P(patterned first and plain second) + P(plain first and plain second)

$= (^{26}/_{40} \times {}^{14}/_{39}) + (^{14}/_{40} \times {}^{13}/_{39})$

= 0.2333 + 0.1167

= 0.3497

P(patterned sock first | plain sock second)

= P(patterned sock first and plain sock second) ÷ P(plain sock second)

$= (^{26}/_{40} \times {}^{14}/_{39}) \div 0.3497$

$= {}^{0.2333}/_{0.3497}$ = 0.6671 (4 d.p.)

iv. P(patterned sock first | patterned sock second)

= P(patterned sock first and patterned sock second) ÷ P(patterned sock second)

First we need to work out the P(patterned sock second):

P(patterned sock second)

= P(patterned first and patterned second) + P(plain first and patterned second)

= $(^{26}/_{40} \times {}^{25}/_{39}) + (^{14}/_{40} \times {}^{26}/_{39})$

= 0.4167 + 0.2333 = 0.65

P(patterned sock first | patterned sock second)

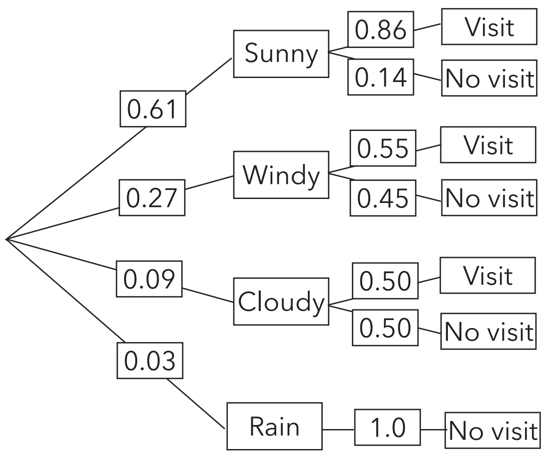
= P(patterned sock first and patterned sock second) ÷ P(patterned sock second)

= $(^{26}/_{40} \times {}^{25}/_{39}) \div 0.65$

= $^{0.416}/_{0.65}$

= 0.6411 (4 d.p.)

c. *i.*



ii. P(visit)

= P(sunny and visit) + P(windy and visit) + P(cloudy and visit) + P(rain and visit)

= (0.61 × 0.86) + (0.27 × 0.55) + (0.09 × 0.5) + (0.03 × 0)

= 0.7181 (4 d.p.)

iii. P(no visit)

= P(sunny and no visit) + P(windy and no visit) + P(cloudy and no visit) + P(rain and no visit)

= (0.61 × 0.14) + (0.27 × 0.45) + (0.09 × 0.5) + (0.03 × 1)

= 0.2819

OR

P(no visit) = 1 – P(visit)

= 1 – 0.7181

= 0.2819 (NOTE: This is the same as the working above)

iv. P(no visit | sunny) = 0.14 (read this directly from the tree, go to the sunny branch then see the probability of no visit following this)

OR

P(no visit | sunny) = P(no visit and sunny) × P(sunny)

= 0.61 × $^{0.14}/_{0.61}$

= 0.14 (NOTE: This is the same as the working above)

v. P(cloudy | no visit) = P(cloudy and no visit) ÷ P(no visit)

P(no visit) = 0.2819 (from part *iii.*)

P(cloudy and no visit) = 0.09 × 0.5

= 0.045

P(cloudy | no visit) = $^{0.045}/_{0.2819}$

= 0.1596 (4 d.p.)

vi. P(visit) = 0.7181 (From part *ii.*)

Total number of days in the two-week period = 14 days.

Expected number of days = Total number of days × P(visit)

= 14 × 0.7181
= 10.0534

Deve is expected to visit the park 10 days over a two-week period.

7. Risk and Relative Risk

a. *i.* Risk under 30 crashes = $^{150}/_{900}$

= 0.1667 (4 d.p.)

ii. Risk over 30 crashes = $^{250}/_{1100}$

= 0.2273 (4 d.p.)

iii. Relative risk = risk over 30 crashes ÷ risk under 30 crashed

= $(^{250}/_{1100}) \div (^{150}/_{900})$

= 1.3636 (4 d.p.)

iv. A person who is over 30 is 1.36 times more likely to crash while mountain biking to the Bridge to Nowhere than a person who under 30.

b. *i.* Risk of a person with a guide falls out = $^{170}/_{210}$

= 0.8095 (4 d.p.)

ii. Risk of a person without a guide falls out = $^{490}/_{550}$

= 0.8909 (4 d.p.)

iii. Relative risk

= Risk of a person without a guide falls out ÷ Risk of a person with a guide falls out

= $(^{490}/_{550}) \div (^{170}/_{210})$

= 1.1

The relative risk of a person with a without a guide falling out is 1.1 times more likely than a person with a guide falling out.

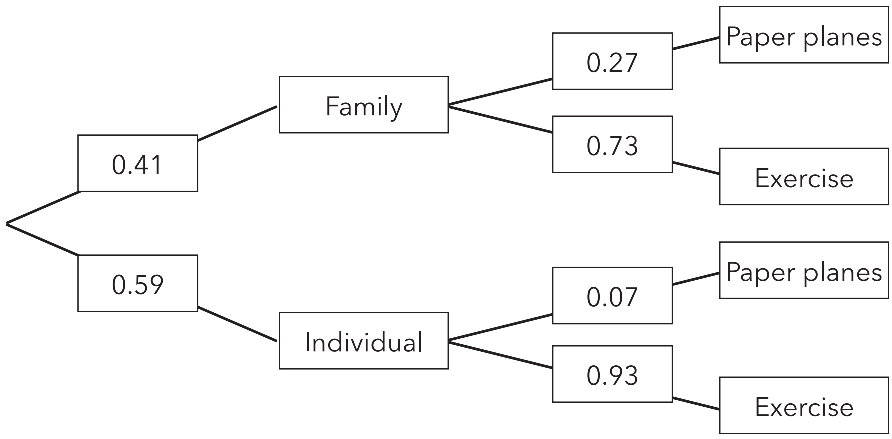
iv. Susan's thinking is correct as the relative risk of a person falling out without a guide is only 1.1 time more likely than a person who is with a guide falling out.

Section Two
Exam Skills & Mixed Practice

1. Using Probability Trees to Assess Probabilities:

1.

a. *i.*



ii. P(family group and climbs tower for exercise) = 0.41 × 0.73

= 0.2993 (4 d.p.)

b. *i.* Let Event A be a family group climbs the tower.

Let Event B be a fly's paper planes.

If events A and B are independent then P(A and B) = P(A) × P(B)

ii. P(family group and paper planes) = 0.41 × 0.27

= 0.1107 (4 d.p.)

iii. P(family group climbs tower) = 0.41

iv. P(paper planes) = P(family and paper planes) + P(individual and paper planes)

= (0.41 × 0.27) + (0.59 × 0.07)

= 0.1107 + 0.0413

= 0.152 (3 d.p.)

v. P(family group climbs tower) × P(paper planes) = 0.41 × 0.152

= 0.0623 (4 d.p.)

vi. P(family group and paper planes) = P(family group climbs tower) × P(paper planes)

$0.1107 \neq 0.0623$

Therefore, the events are not independent because the probability of a family group and fly's paper planes does not equal the probability of a family group climbs the tower multiplied by the probability that a person fly's paper planes.

c. *i.* P(family group and exercise) = 0.41 × 0.73

= 0.2993 (4 d.p.)

ii. P(individual and exercise) = 0.59 × 0.93

= 0.5487 (4 d.p.)

iii. P(family group and exercise) compared to P(individual and exercise)

$0.2993 < 0.5487$

iv. Vincent is incorrect in his thinking because the probability of a family group and exercise is smaller than the probability of and individual and exercise. Therefore, it is more likely that a person is an individual who climbs the tower for exercise than a family to climb the tower for exercise.

d. *i.* P(individual and fly's paper plane) = 0.59 × 0.07

= 0.0413 (4 d.p.)

ii. P(all three people who climb the two to fly paper planes)

= 0.0413 × 0.0413 × 0.0413

$= 7.0445 \times 10^{-5}$

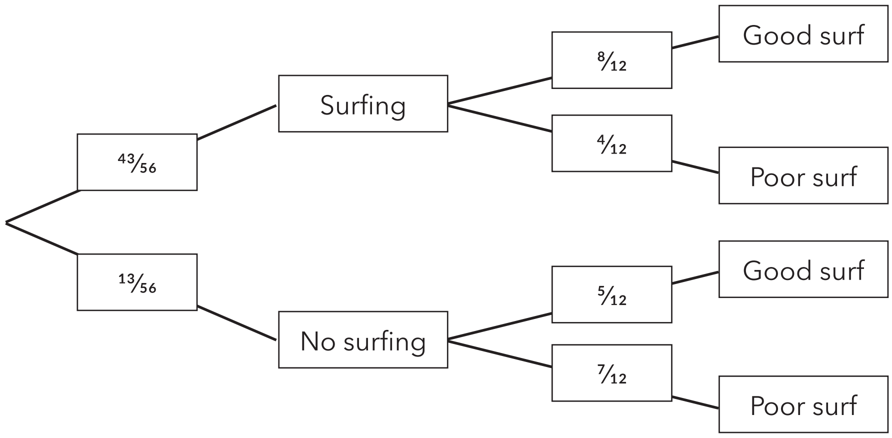
= 0.000070445

iii. Assuming the activities of the three individuals do not influence each other. Therefore, you need to assume that all three of the people are independent.

iv. Assuming the sample that Vincent took was sufficiently large that sampling without replacement is not required for the probabilities to remain relatively consistent if and individual is removed from the sample.

2.

a. *i.*



ii. P(good surf day) = P(surfing and good surf day) + P(No surfing and good surf)

$= (\frac{43}{56} \times \frac{8}{12}) + (\frac{13}{56} \times \frac{5}{12})$

$= 0.5119 + 0.0967$

$= 0.6086$ (4 d.p.)

b. *i.* For the events "goes surfing" and "good surf" to be mutually exclusive then the probability of "goes surfing and good surf" should be zero – P(goes surfing and good surf) = 0.

ii. P(goes surfing and good surf) = $(\frac{43}{56} \times \frac{8}{12})$

$= 0.5119$

iii. P(goes surfing and good surf) = 0.5119

$0.5119 \neq 0$

iv. As the probability of "goes surfing and good surf" does not equal zero then the two events of "goes surfing" and "good surf" are not mutually exclusive.

c. *i.* P(poor surf) = P(goes surfing and poor surf) + P(does not go surfing and poor surf)

$= (\frac{43}{56} \times \frac{4}{12}) + (\frac{13}{56} \times \frac{7}{12})$

$= 0.25595 + 0.13542$

$= 0.3914$ (4 d.p.)

ii. P(surfing | poor surf) = P(surfing and poor surf) ÷ P(poor surf)

$= (\frac{43}{56} \times \frac{4}{12}) \div 0.3914$

$= 0.6539$ (4 d.p.)

iii. P(no surfing | poor surf) = P(no surfing and poor surf) ÷ P(poor surf)

= ($^{13}/_{56}$ × $^{7}/_{12}$) ÷ 0.3914

= 0.346 (3 d.p.)

iv. P(surfing | poor surf) > P(no surfing | poor surf) as 0.6539 > 0.346.

v. Given that it is a poor surf day it is more likely that Tamara will go surfing than not go surfing.

d. *i.* Some beaches in New Zealand never have good surf due to the type of beach and the weather conditions. Therefore, given that Tamara goes surfing the probability that there is good surf will always be lower than 70%.

ii. The model that Tamara created is based off data that she collected during the summer. Therefore, it is likely that this cannot be applied to the other season because the surf conditions at the beach she normally surfs at will be different due to changes in the weather conditions. If it does not hold for the beach the data collected at to create the model, then the model is not going to hold for any other beach in a season other than summer.

iii. Tamara's decision to go surfing may be affected by other external factors such as the weather, because she may be less likely to go surfing if it is windy or rainy. This will affect the conditional probability that there is good surf given that Tamara goes surfing resulting in the model predicted probability being incorrect.

2. Using Two-way Tables to Assess Probabilities:

1.

a. *i.*

	Visit Cafe	Feed the ducks	Walk	Total
Not Retired	11	17	43	71
Retired	35	23	44	102
Total	46	40	87	173

ii. P(retired | café) = $^{35}/_{46}$ = 0.7609

b. *i.* P(retired) = $^{102}/_{173}$ = 0.5896 (4 d.p.) = 58.96%

ii. Based on the data that Connor collected, the Council should not consider upgrading the path because the percentage of people who visit Rotokawau/Virginia Lake is 58.96% which is less than 60%.

iii. • The data that was collected by Connor was from only one Saturday in summer. This only provides a snapshot of a singular day and not an average over the summer or over a year. If longer time frame data was collected, then the percentage of the retired people visiting Rotokawau/Virginia Lake may increase about the 60% threshold set by the council.

- Between years the number of people who are retired in Whanganui will vary, increasing or decreasing slightly between years. Based on the survey taken by Connor, slightly less than 60% of people but this may decrease next year if people move away from Whanganui or increase if more people retire or more people move to Whanganui to retire.

- The survey that Connor undertook only accounts for people that walked round the lake. Connor would be better to ask people if they wanted to walk round the lake, and whether they could or could not do so with the current state of the paths. This would give the council a better idea of whether they should upgrade the path or not, as the probabilities would better reflect the people who walk round the lake or would like to walk round the lake..

c. *i.* $P(A' \cap B') = 1 - P(A \cup B)$

$= 1 - 0.63$

$= 0.37$

ii. $P(A' \cap B) = 1 - P(A \cup B')$

$= 1 - 0.68$

$= 0.32$

iii. $P(A') = P(A' \cap B') + P(A' \cap B)$

$= 0.37 + 0.32$

$= 0.69$

iv. $P(A) = 1 - P(A')$

$= 1 - 0.69$

$= 0.31$

v. The probability that a person walks round the lake in fine weather is 0.31

2.

a. *i.*

	Old School Room	Whanganui Social History	Geological World	Total
School Group	203	211	84	498
Not School Group	54	101	90	245
Total	257	312	174	743

ii. P(not school group and Whanganui social history) = $^{101}/_{743}$

b. *i.* P(geological world | school group) = $^{84}/_{498}$

= 0.1687 (4 d.p.)

ii. P(geological world | not school group) = $^{90}/_{245}$

= 0.3673 (4 d.p.)

iii. P(geological world | school group) compared to P(geological world | not school group)

0.1687 < 0.3673

iv. It is more likely that a visitor will go to the geological world exhibition if they are not associated with a school group than if the visitor is associated with a school group.

c. *i.* P(old school room) = $^{257}/_{743}$

= 0.3459 (4 d.p.)

ii. Expected number of visitors = 0.3459 × 800

= 276.7

276 people are expected to visit the old school room next year.

iii. P(Whanganui social history) = $^{312}/_{743}$

= 0.4199 (4 d.p.)

iv. Expected number of visitors = 0.4199 × 800

= 335.9

335 people are expected to visit the Whanganui social history exhibition next year.

Section Two: **Putting it into Context** *v.* P(geological world) = $^{174}/_{743}$

= 0.2342 (4 d.p.)

vi. Expected number of visitors = 0.2342 × 800

= 187.3

187 people are expected to visit the geological world exhibition next year.

vii. The two exhibitions with the highest expected visitors are the old school room and the Whanganui social history exhibition.

viii. The museum should upgrade the old school room and the Whanganui social history exhibition.

d. *i.* **P(incorrect) =** P(Whanganui social history and old school room) + P(geological world and old school room) + P(Old school room and Whanganui social history) + P(geological world and Whanganui social history) + P(old school room and geological world) + P(Whanganui social history and geological world)

= (2 + 1 + 7 + 6 + 3 + 1) ÷ 50

= 20 ÷ 50

= 0.4

OR

P(incorrect) = 1 – P(correct)

= 1 – [P(school room and school room) + P(Whanganui social history and Whanganui social history) + P(geological world and geological world)]

= 1 – [(5 + 17 + 8) ÷ 50]

= 1 – (30 ÷ 50) = 1 – 0.6 = 0.4

ii.
- P(correct) = P(school room and school room) + P(Whanganui social history and Whanganui social history) + P(geological world and geological world)

  = (5 + 17 + 8) ÷ 50

  = 0.6

  = 60%

- P(correctly predicting old school room) = $\frac{5}{8}$

  = 0.625 or 62.5%

- P(correctly predicting the Whanganui social history) = $^{17}/_{30}$

  = 0.5667 or 56.7%

- P(correctly predicting geological world) = $^{8}/_{12}$

  = 0.6667 or 66.7%

- The model does the worst job of correctly predicting people visiting the Whanganui social histories exhibition.

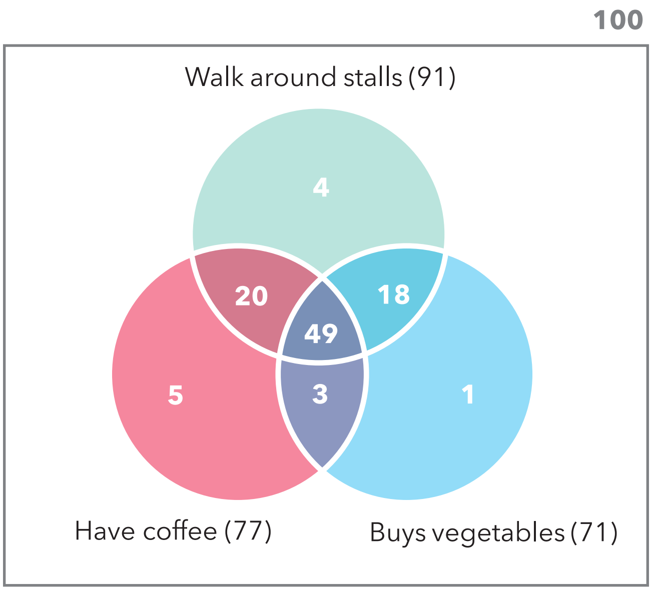
- The model has a reasonably good rate of predicting whether visitors will go to the old school room and the geological world exhibition.

- The model does a reasonable job of correctly predicting the exhibitions that visitors will go to. Overall, it correctly predicts 60% of visits. Of all three of the exhibitions, the Whanganui social histories has the lowest predictions at 56.7%.

3. Using Venn Diagrams to Assess Probabilities:

1.

a. *i.*



ii. P(buys coffee and goes for walk but not buy vegetables)

$= {}^{20}/_{100}$

= 0.2

= 20%

b. *i.* P(walk round stalls | buy coffee)

$= {}^{69}/_{77}$

= 0.8961 (4 d.p.)

ii. P(buys vegetables | buy coffee)

$= {}^{52}/_{77}$

= 0.6753 (3 d.p.)

iii. P(walk round stall | buy coffee) ÷ P(buys vegetables | buy coffee)

0.8961 ÷ 0.6753 = 1.327 (3 d.p.)

iv. A person is 1.3 times more likely to walk round the stalls given they are there to buy coffee than to buy vegetables given they are there to buy coffee.

v. The claims that have been made by both Alister and the Riverside market organisers are both incorrect. The Riverside market organisers claim of 1.5 times is closest but it is still higher than the calculated value of 1.3.

c. *i.* P(buys vegetables) = ${}^{71}/_{100}$

= 0.71

ii. Expected number of people = 0.71 × 200

= 113.6

= 113

iii. Alister should recommend that the stall owner does come next week because the expected number of people who go to buy vegetables is 113.

iv. Potential reasons:

- If the weather is bad, then less people are going to go to the market so the expected number of people will be lower than predicted because the total number of people is also lower.
- Some people do not need to buy vegetables each week so the probability that a person goes to the market to buy vegetables each week could vary slightly.
- If other people recommend the stall then there is likely to be more customers going to the stall next week compared to when Alister took the survey.
- More people are likely to go to the market during the school holidays and when other events are on so the probability that someone buys vegetables is likely to fluctuate between weeks.

a. *i.* People who used the BBQ, played on the flying fox and hired a beetle = 37

ii. People who hired a beetle and played on the flying fox = 44 – 37

= 7

iii. People who hired a beetle and used the BBQ = 48 – 37

= 11

iv. People who played on the flying fox and used the BBQ = 56 – 37

= 19

v. Only beetle = 97 – 7 – 37 – 11

= 42

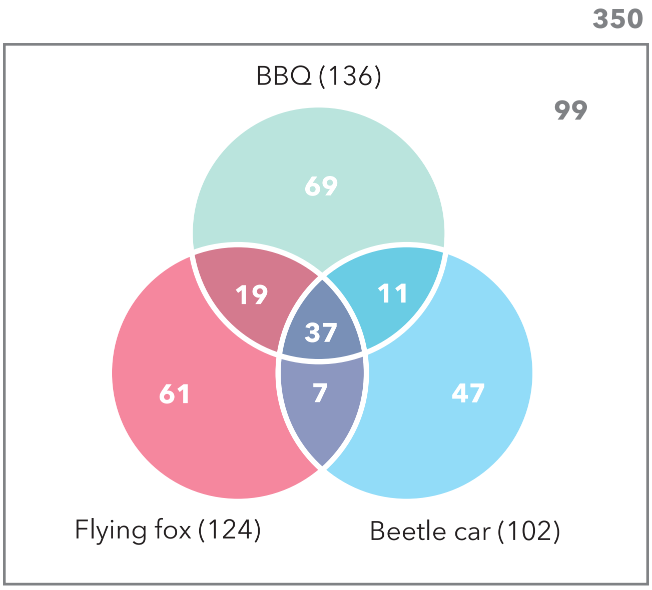
vi. Only flying fox = 112 – 19 – 37 – 7

= 45

vii. Only use BBQ = 120 – 19 – 37 – 11

= 53

viii.



ix. P(Number of people no BBQ, no beetle car, no flying fox) = 350 – 69 – 61 – 47 – 19 – 11 – 7 – 37

= 99

P(no BBQ, no beetle car, no flying fox) – $^{99}/_{350}$ = 0.2828 (4 d.p.)

b. *i.* If the events F and B are independent then the P(F | B) = P(F)

ii. P(F | B) = $^{69}/_{136}$ = 0.5073 (4 d.p.)

iii. P(F) = $^{124}/_{350}$ = 0.3542 (4 d.p.)

iv. P(F | B) = P(F)

0.5073 ≠ 0.3542, therefore the two events are not independent.

v. The events of a person uses the BBQ's and a person plays on the flying fox are not independent, so Laura is incorrect in her thinking. If the two events were independent then the occurrence of one event will not influence the other but the probability of plays on the flying fox given uses a BBQ is different from the probability of a person playing of the flying fox.

c. *i.* P(only BBQ) = $^{69}/_{350}$

= 0.1971 (4 d.p.)

ii. P(only beetle car) = $^{47}/_{349}$

= 0.1346 (4 d.p.)

iii. P(only flying fox) = $^{61}/_{348}$

= 0.1753 (4 d.p.)

iv. P(all three events) = 0.1971 × 0.1346 × 0.1753

= 0.004651

v. That the occurrence of each of the events is independent of each other.

4. Simulations and graphs to assess probabilities:

1.

a. *i.* P(North America | canoe the Whanganui River) = 23%

ii. P(No guide | North America) = 55%

iii. P(North America and without a guide) = 0.23 × 0.55

= 0.1265

= 12.65%

b. *i.* No, Susan is not correct in her thinking.

ii. Susan would have looked at the percentages of the people from Oceania which did the trip with a guide (42%) and the percentages of the people from South America which did the trip with a guide (39%).

iii. These are conditional probability based on what continent the tourists are from, not based on whether they did the trip with a guide or without a guide. These are the wrong conditional probabilities.

iv. In order to Susan to reason whether a tourist who does the trip with guide is from Oceania or South America, then she would need to work out the conditional probabilities using the information in the graph and the table combined.

c. *i.* P(Oceania and no tour guide) = 0.58 × 0.45

= 0.261 (3 d.p.)

ii. P(South America and no tour guide) = 0.61 × 0.08

= 0.0488 (4 d.p.)

iii. P(Oceania and no tour guide or South America and no tour guide)

= 0.261 + 0.0488

= 0.3098 (4 d.p.)

iv. P(Oceania and no tour guide or South America and no tour guide) ÷ P(no guide)

= 0.3098 ÷ 0.5033

= 0.6155

= 61.55%

The probability that a tourist is from Oceania or South America given that they did not take a guide on the trip is 61.55%.

2.

a. *i.* P(under 30) = $^{67}/_{532}$

= 0.1259 (4 d.p.)

ii. Expected count = 35 × 0.1259

= 4.4

= 4 people

Section Two: **Putting it into Context** *iii.* P(31 to 64) = $^{63}/_{532}$

= 0.1184 (4 d.p.)

iv. Expected count = 35 × 0.1184

= 4.1

= 4 people

v. P(over 65) = $^{402}/_{532}$

= 0.7556 (4 d.p.)

vi. Expected count = 35 × 0.7556

= 26.4

= 26 people

vii.

Age	Expected	Observed
Under 30	4	8
30 to 64	4	7
Over 65	26	20

viii. Olive does have the right to be suspicious that Bob's selection was not at random, because the observed counts are different from the expected counts calculated. For the under 30 and 30 to 64 age groups the observed counts are approximately double what the expected values are.

b. *i.* The simulation will help to show the variation (or distribution) of the expected number of people in each of the age groups that are being investigated. This will give an indication of the range of values that can be expected for each of the age groups.

ii. When looking at the simulation results you can work out the probability or likelihood of each of the outcomes. The likelihood can be calculated by dividing the number of simulations that show the outcome of interest and dividing it by the total number of outcomes. This will allow for the likelihood of each outcome can be compared to the observed number of outcomes.

c. *i.* Number of trail resulting in 7 or more = 62

ii. P(seven or more) = $^{62}/_{1000}$

= 0.062

= 6.2%

iii. It is unlikely that Bob randomly selected the sample because simulation showed that the probability that 7 or more people between 31 and 64 is 6.2%.

iv. Olive should conclude that other factors did influence the sample rather than it being completely random because the probability of getting 7 or more people between 31 and 64 is 6.2%. While this is very small, it is not impossible that this could occur.

v. Other factors include:

- Bob did not follow the instructions of randomly selecting.
- Bob may have wrongly recorded the results.
- Bob may have not have mixed the groups up before selected or between selections which could have influenced the results.

Section Three
Practice Exam

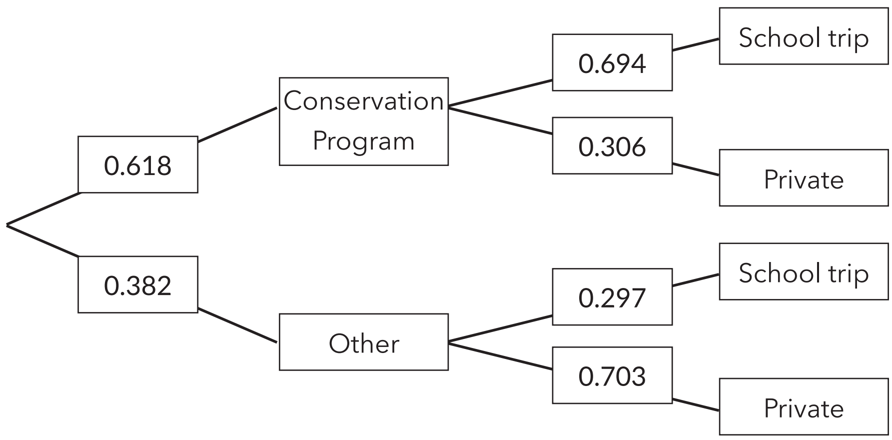
Question One:

Bushy Park Sanctuary is a 100-hectare predator-free native bird sanctuary set amongst a virgin lowland forest. It is home to a wide range of native flora and fauna which visitors can observe as they walk around the park.

Each year approximately 1500 people visit the sanctuary. Of the people who visit, Bushy Park notes whether they are part of a school trip or a personal trip and whether their visit is part of a conservation program or for another reason, such as walking.

- 61.8% of visitors are part of a conservation program.
- 69.4% of conservation program visitors are there as part of a school trip.
- 70.3% of other reason visitors are there as private visitors.

a. *i.*



P(school trip) = P(conservation and school) + P(Other and school)

= (0.618 × 0.694) + (0.382 × 0.297)

= 0.428892 + 0.113454

= 0.5423 (4 d.p.)

ii. P(conservation and school) = 0.618 × 0.694

= 0.428892

P(other and private) = 0.382 × 0.703

= 0.268546

It is more likely that a person is there for a conservation program and part of a school trip than being there for other reasons and as a private visitor. This is because the probability of conservation and school (0.4289) is greater than the probability of other and private (0.2685).

b. If events P and C are independent then the P(T | C) = P(T) as the occurrence of one of the events will not influence the occurrence of the other event.

P(T) = P(conservation and private) + P(Other and private)

= (0.618 × 0.306) + (0.382 × 0.703)

= 0.189108 + 0.268546

= 0.457654

P(T | C) = 0.306

As P(T | C) ≠ P(T) the events are not independent.

c.

	Conservation	**Walk**	**Homestay**	**Total**
Male	463	250	52	765
Female	464	219	52	735
Total	927	469	104	1500

P(female and walk) = $^{219}/_{1500}$

= 0.146

d. P(two female and conservation program, one male and conservation program)

= 3 × $^{464}/_{1500}$ × $^{463}/_{1499}$ × $^{463}/_{1498}$

= 3 × 0.3093 × 0.3089 × 0.3097

= 0.0888 (4 d.p.)

Need to assume:

- That the people were randomly selected.
- That the people are independent of each other, so the probability of one person does not influence the probability of another.

Question Two:

a. *i.* P(not Toutouwai | not Ratanui) = $^{598}/_{748}$

= 0.7995 (4 d.p.)

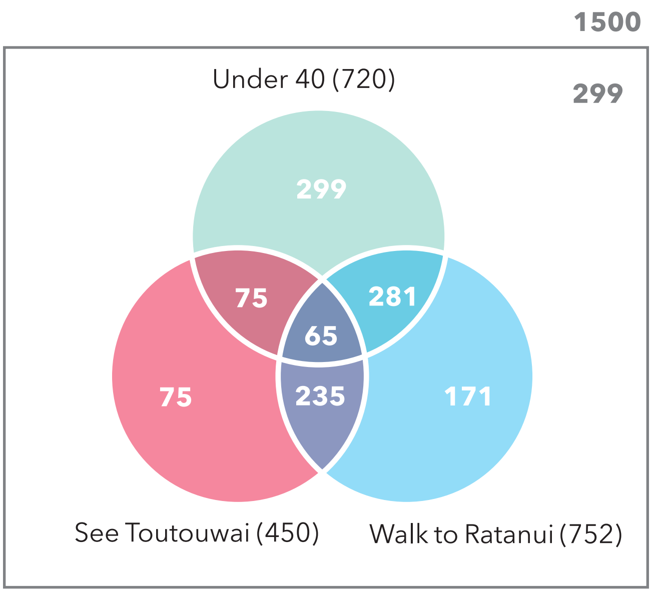
ii. For events to be mutually exclusive P(A and B) = 0.

The events of "over 40" and "walks to see Ratanui" are not mutually exclusive because the P("over 40" and "walks to see Ratanui") does not equal zero. The P("over 40" and "walks to see Ratanui") = $^{406}/_{1500}$

= 0.2707 (4 d.p.)

0.2707 ≠ 0.

b. *i.*



P(under 40 but did not walk to see Ratanui and did not see a Toutouwai) = $^{299}/_{1500}$

= 0.1993 (4 d.p.)

ii. P(over 40, walk to see Ratanui, see a Toutouwai) = $^{235}/_{1500}$

= 0.1567 (4 d.p)

Expected number of visitors = 0.1567 × 1300

= 203.71

= 203 people

Bushy Park should upgrade the path given the information that was collected and the expected number of visitors. This is because it is expected that next year 203 people will be over the age of 40, walk to see Ratanui and see a Toutouwai.

iii. Possible reasons:

- The data collected is only correct for the year that it was collected and the probabilities between years may vary.
- The expected number of visitors is only an estimate for the number of visitors, so this number may be higher or lower.
- Visitor demographic may vary due to recommendations from other people or recommendations to schools may cause more young people to go to the park.
- Toutouwai numbers may vary between years making it more or less likely to see a toutouwai.

Question Three:

i.

Time of Day	Sighting		
	Less than or equal to two	More than two	Total
Morning	336	202	538
Afternoon	339	101	440
Total	675	303	98

P(less than or equal to two birds | morning) = $^{336}/_{538}$

= 0.6245 (4 d.p.)

ii. P(two or more | morning) = $^{202}/_{538}$

= 0.3755 (4 d.p.)

P(two or more | afternoon) = $^{101}/_{440}$

= 0.2295 (4 d.p.)

Ratio = P(two or more | morning) ÷ P(two or more | afternoon)

= 0.3755 ÷ 0.2295

= 1.636 (3 d.p.)

The claim made by Bushy Park is correct, as you are 1.6 times more likely to see two or more Toutouwai in the morning than in the afternoon. This is slightly higher than the claim made by Bushy Park but it is close.

iii. P(W ∪ N) = 0.11 therefore the P(W′ ∩ N′)

= 1 – 0.79

= 0.21

P(W ∪ N′) = 0.45 therefore the P(W′ ∩ N)

= 1 – 0.45

= 0.55

P(W′) = 0.21 + 0.55

= 0.76

P(W) = 1 – 0.76

= 0.24

The probability that a sighting of a Toutouwai is affected by the weather is 0.24.

b.

	2013	2015	2017
Number of dead Toutouwai found	4	9	6
Number of Toutouwai in Bushy Park	47	54	62

i. Risk in 2013 = $^4/_{47}$

= 0.0851 (4 d.p.)

Risk in 2015 = $^9/_{54}$

= 0.1667 (4 d.p.)

Risk in 2017 = $^6/_{62}$

= 0.0968 (4 d.p.)

2015 was the year that had the greatest risk of a Toutouwai dying, as the probability of a Toutouwai dying is highest in 2015.

ii. Possible reasons:

- Not all the dead Toutouwai were found by staff at Bushy Park.
- The total number of Toutouwai in Bushy park may not be known.

iii. Possible considerations:

- The most recent data that DOC has to estimate the risk is 2017 which is not the most up to date and the risk of finding a Toutouwai dead may have drastically changed between 2017 and present.
- Bushy Park is a predator free sanctuary so the risk of a Toutouwai dying in other locations may be higher due to predators.
- Conditions in the park, such as weather or habitat may have changed since the last data was collected, altering the risk, and not making it valid to compare to other regions.
- Human intervention in other regions, may increase or decrease the risk of a Toutouwai dying.

