3

9 1 5 8 4 R

Level 3 Mathematics and Statistics (Statistics), 2021 v1

91584 Evaluate statistically based reports

Credits: Four

RESOURCE BOOKLET

Refer to this booklet to answer the questions for Mathematics and Statistics (Statistics) 91584.

Check that this booklet has pages 2 – 4 in the correct order and that none of these pages is blank.

YOU MAY KEEP THIS BOOKLET AT THE END OF THE EXAMINATION.

REPORT 2

More NZ school leavers with no qualifications

After recent concerns about the number of teenagers leaving school with no qualification, the Ministry of Education conducted surveys using samples of school leavers in 2018 and 2019.

12 per cent of last year's (2019) school leavers did not have an NCEA qualification, up from 9.5 per cent in 2018, which was the first year an increase in the statistic had been recorded.

The group represented 732 out of the 6,100 young people in the 2019 school leavers sample, and numbered a few hundred more than the equivalent group in the 2018 school leavers sample.

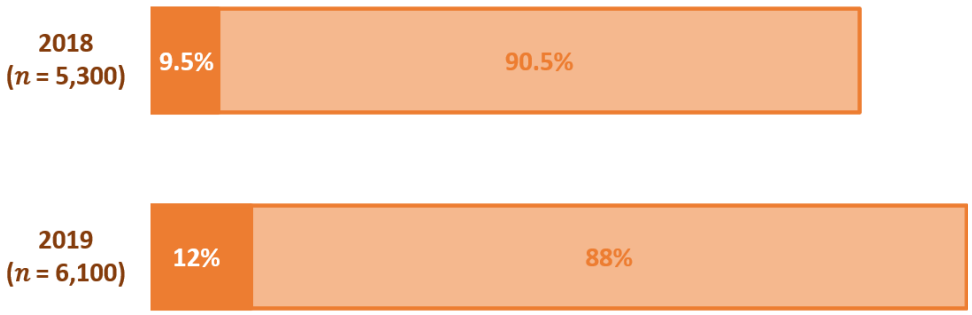
The figures showed about half, 368, were Pākehā, and 328 were Māori. Considered by gender, more than half, 423, were boys. At some schools, more than 30 per cent of school leavers last year had no qualification.

The data showed more than 30 per cent of the students leaving some low-decile schools last year had no NCEA qualification, and at some high-decile schools the figure was about 15 per cent.

Considered by region, Gisborne had the highest rate of unqualified school leavers, at 19 per cent, followed by Northland and Manawatū with 15 per cent each.

The principal of Gisborne Boys' High School, Andrew Turner, said one of the reasons for the figure was the "truck load" of job opportunities in the region. "The downside is that we're seeing more and more of our young people getting to that 15, 16 years of age - 16 is the legal age to leave school - who turn 16 during Year 11 and haven't got NCEA level 1 yet, and think 'doesn't matter, I'm going to work, I can get $19.10 an hour, that's a priority for the family, that's more important'. And off they go."

Figure 1: Percentage of school leavers without NCEA



Adapted from:
https://www.nzherald.co.nz/nz/the-trend-is-of-concern-school-leavers-with-no-qualifications-rises/MO352VZCPPGA4DD3XOHNHCBYGE/

3

9 1 5 8 4

Level 3 Mathematics and Statistics (Statistics), 2021 v1

91584 Evaluate statistically based reports

Credits: Four

Achievement	Achievement with Merit	Achievement with Excellence
Evaluate statistically based reports.	Evaluate statistically based reports, with justification.	Evaluate statistically based reports, with statistical insight.

You should attempt ALL the questions in this booklet.

Pull out Resource Booklet 91584R from the centre of this booklet.

Show ALL working.

Make sure that you have the Formulae and Tables Booklet L3–STATF.

If you need more space for any answer, use the page(s) provided at the back of this booklet and clearly number the question.

YOU MUST HAND THIS BOOKLET TO THE SUPERVISOR AT THE END OF THE EXAMINATION.

TOTAL

QUESTION ONE

ASSESSOR'S USE ONLY

*Refer to **Report 1** in the resource booklet to answer the following questions.*

(a) Explain whether this study is an observational study or an experiment.

(b) Identify and describe the explanatory and response variables in the study.

(c) The researchers stated that "a more diverse green space was more protective".
How would the researchers have made sure that some children were exposed to a more diverse green space than others?

QUESTION TWO

ASSESSOR'S USE ONLY

*Refer to **Report 2** in the resource booklet to answer the following questions.*

(a) Report 2 states that 6 100 school leavers were included in the 2019 survey.

(i) Calculate the margin of error for the 2019 survey.

(ii) Explain why the margin of error is needed to interpret any claims made from this survey.

(b) The 95% confidence interval for the percentage of school leavers with an NCEA qualification in 2019 is approximately (86.7%, 89.3%).

How can the sample size be altered so that the width of this confidence interval is reduced by 25%?

ASSESSOR'S USE ONLY

(c) Report 3 claims that "We observed a 24% jump in the chance of good perceived mental health when people increased their activity from two to three days a week".

Describe the statistical analysis that could have been used to make this claim, and discuss what other information you would need to evaluate this claim.

(d) The heading for this report is "New research confirms physical activity is tied to healthy mental wellbeing".

Discuss ONE potential issue with the study design used for the survey conducted, in respect to the heading for this report.

Q	Expected Coverage	Achievement (c)	Merit (j)	Excellence (i)
(d)	This is an observational study. The researchers were not in control of the treatment levels (children with limited green spaces and children with more green spaces), so they cannot make a causal claim. The researchers used the children who had limited green spaces as a control to compare with the children who were exposed to more green spaces. The researchers would have selected the children who are demographically similar for this study. The researchers would have compared the occurrence of asthma among the children who were exposed to more green space with the occurrence of asthma among those with limited green spaces. The researchers then would have come to their conclusion after observing that the occurrence of asthma among the children exposed to more green spaces is less than for the children with limited green spaces.	Mentions this is an observational study or a causal claim cannot be made.	Describes that the two groups of children were compared, allowing the researchers to reach a conclusion based on the observed differences between groups.	Explains the demographical similarity of the children in two groups , and how the researchers would have compared one group of children with the other (baseline) to quantify any differences.
(e)	The results obtained from the study are extended globally through the claim that 300 million asthma-sufferers would decrease by 15 per cent. This is a potential issue as this study was conducted in NZ environments using only NZ children. Moreover, what constitutes a green space (trees, grass and flowers etc) would be different overseas. Therefore, the quantified decrease in the number of asthma-suffers is likely to be not applicable outside NZ.	Describes ONE potential issue with extending the results.	Describes ONE potential issue with extending the results. AND Attempts to explain why it could limit extending the results.	Describes ONE potential issue with extending the results. AND Describes why it could limit extending the results by using specific features of the report/ study.

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Attempt at one part of the question.	1 of c	2 of c OR 1 of j	3 of c OR 1 of c and 1 of j	2 of j	3 of j	1 of i	2 of i

Q	Expected Coverage	Achievement (c)	Merit (j)	Excellence (i)
THREE (a) (i)	Snowball sampling is a recruitment technique in which research participants are asked to assist researchers in identifying other potential participants.	Described snowball sampling.		
(a) (ii)	• Participants with many acquaintances and colleagues are more likely to be recruited into the sample. • Initial participants tend to nominate people that they know well. This way, the researchers are not in control of sampling, there is no guarantee that the sample obtained is representative of the population. Perhaps the sample would be likely to end up consisting mostly of personal trainer/gym goers who are on social media.		One potential example of snowball sampling is described.	One potential example of snowball sampling is described. AND Described how snowball sampling could lead to a non-representative sample.
(b)	Explanatory variable: Duration of physical activity per week Response variable: Perceived mental health	Explanatory variable AND response variable are described.		
(c)	The variables Duration of physical activity per week and perceived mental health could have been used to construct a scatterplot, with duration of physical activity per week as the explanatory variable, and perceived mental health as the response variable. A linear model would be fitted to the data, and the gradient of this model would have been used to quantify the relationship between Duration of physical activity per week and perceived mental health. To evaluate this claim, one would need to know the strength of the linear relationship between these two variables.	Use of scatter plot to display data.	The gradient of the linear model is linked to the claim / inference made. OR The need to assess the strength of evidence is discussed, in terms of being uncertain about the amount of variation.	

3

9 1 5 8 5

Level 3 Mathematics and Statistics (Statistics), 2021 v1

91585 Apply probability concepts in solving problems

Credits: Four

Achievement	Achievement with Merit	Achievement with Excellence
Apply probability concepts in solving problems.	Apply probability concepts, using relational thinking, in solving problems.	Apply probability concepts, using extended abstract thinking, in solving problems.

You should attempt ALL the questions in this booklet.

Show ALL working.

Make sure that you have the Formulae and Tables Booklet L3–STATF.

If you need more space for any answer, use the page(s) provided at the back of this booklet and clearly number the question.

YOU MUST HAND THIS BOOKLET TO THE SUPERVISOR AT THE END OF THE EXAMINATION.

TOTAL

QUESTION ONE

ASSESSOR'S USE ONLY

(a) A general practitioner in Auckland collected his patient files. Each patient file contains the medical history of the patient including whether the patient is immunocompromised and whether the patient had a cold last year.

The general practitioner found that:

- 38% of the patients are immunocompromised
- 24% of the patients who are not immunocompromised had a cold last year
- About 41.1% of the patients had a cold last year.

(i) Out of 108 patients who saw the general practitioner last year, find the expected number of patients who were not immunocompromised and had a cold last year.

(ii) Calculate the probability that a randomly chosen patient is immunocompromised and had a cold last year.

ASSESSOR'S USE ONLY

(b) A large medical clinic has a team of dermatologists, urologists and gynaecologists working in the same building.

The following are some notes made on the 120 patients who visited the clinic last month:

- 26 saw only a dermatologist
- 2 saw a gynaecologist, a dermatologist and a urologist
- 51 saw a gynaecologist
- 14 saw a dermatologist and a gynaecologist
- 15 only came in to pick up their prescriptions (and did not see any doctor)
- 5 saw a urologist and a gynaecologist
- 9 saw a dermatologist and a urologist

(i) Calculate the proportion of patients who saw only a urologist last month.

(ii) Two patients are randomly selected from this group. Both patients saw exactly two of the three specialists last month. One of the two specialists was a dermatologist for both patients.
Calculate the probability that both of these patients saw the same two specialists last month.
Support your answer with statistical statements and reasoning.
State any assumption(s) made.

QUESTION TWO

ASSESSOR'S USE ONLY

Studies have shown that eczema and asthma are linked to inflammation that is often caused by a strong reaction to environmental allergens.
A health researcher obtained the medical records of 1550 randomly chosen individuals for her research.

Upon analysing this data, the health researcher found that:
• 152 individuals were diagnosed with asthma
• Of the 170 individuals who were diagnosed with eczema, 62 were not diagnosed with asthma.

(a) One individual was randomly chosen from the health researcher's data.

(i) Calculate the probability that the individual was diagnosed with eczema or asthma.

(ii) It is claimed that an individual who was diagnosed with asthma is three times as likely to be also diagnosed with eczema than to not be diagnosed with eczema.
Does the health researcher's data support this claim?
Support your answer with appropriate statistical statements.

ASSESSOR'S USE ONLY

(b) Over the years, medical researchers deduced that asthma and eczema are related to a group of respiratory conditions such as hay fever.

Using research results that are based on a larger number of individuals, the medical researchers determined that:

- 5% of individuals were diagnosed with asthma and eczema, but not hay fever
- 7% of individuals were diagnosed with asthma and hay fever, but not eczema
- 4% of individuals were diagnosed with eczema and hay fever, but not asthma
- 2% of individuals were diagnosed with all of asthma, eczema and hay fever
- 16% of individuals were not diagnosed with any of asthma, eczema or hay fever

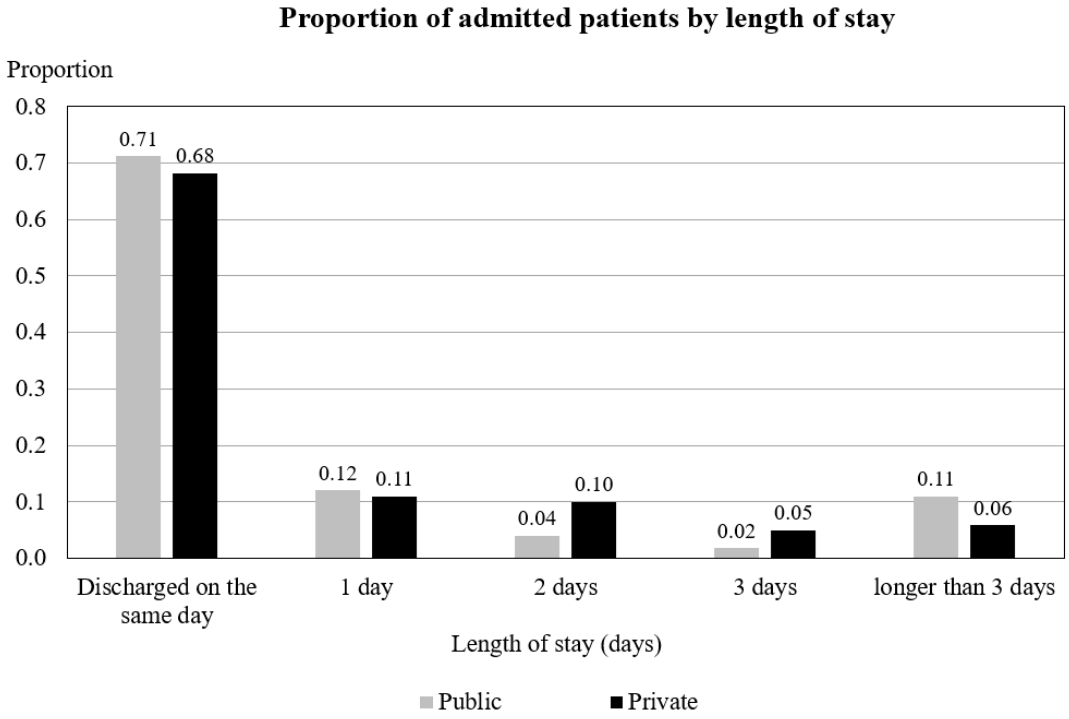
(i) Calculate the probability that a randomly selected individual from this group was diagnosed with only one of asthma, eczema or hay fever.

(ii) The medical researcher also knows that 14.5% of the individuals in this group were diagnosed with eczema.
What is the probability that a randomly selected individual is diagnosed with asthma, given that the individual is diagnosed with eczema?

QUESTION THREE

ASSESSOR'S USE ONLY

(a) When a patient is admitted to a hospital, the patient's admission date is recorded. Once the patient is discharged, the discharge date is also recorded, from which the length of stay, in days, is calculated.
The graph below shows the distribution of admission by length and the type of hospital (public or private), based on the current model used by hospitals.



The table below shows a breakdown of last year's admissions by length of stay and the type of the hospital.

Length of stay (days)	Public	Private
0	1,278,968	631,960
1	490,776	121,623
2	77,496	22,176
3	51,672	121,652
4	25,836	77,609
5 or more	258,312	33,252
Total	2,183,060	1,008,272

ASSESSOR'S USE ONLY

(iii) Suppose the second requirement of the engineering project is changed to the following:

- The generated number can now be anywhere between a single digit number and ***6-digit number***.

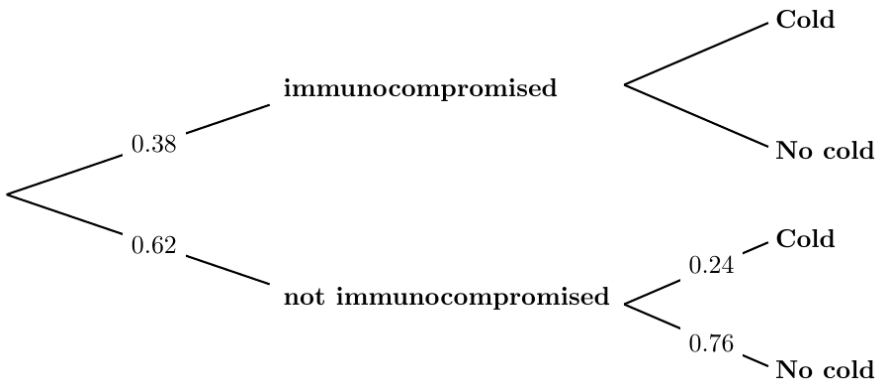
In addition, a weight is assigned to the generated number conditionally as specified below:

- 0 for a single digit number
- 1 for a 2-digit number
- 2 for a 3-digit or a 4-digit number
- 3 for a 5-digit number
- 4 for a 6-digit number

If W represents the random variable for the weight, complete the following probability distribution table.

w	0	1	2	3	4
$\mathbf{P}(W = w)$					

Assessment Schedule – 2021 v1
Mathematics and Statistics (Statistics): Apply probability concepts in solving problems (91585)
Evidence Statement

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
ONE (a)(i)	 P(not compromised and cold) = 0.62 × 0.24 = 0.1488 Number expected = 0.1488 × 108 = 16	Correct expected number calculated.		
(a)(ii)	$P(\text{compromised and cold}) + P(\text{not compromised and cold}) = P(\text{cold})$ $P(\text{compromised and cold}) = 0.411 - 0.1488$ $= 0.2622$	One relevant probability calculated.	Correct probability calculated.	
(a)(iii)	• It is not clear what the total number of patients is (sample size). If this number is small, it is not appropriate to make generalisations. • Only the patients who were seen by this general practitioner are considered. This data could give invalid estimates, if the patients in the data are not representative of patients elsewhere. *Accept other valid reasons with clear links to context.*	ONE issue identified.	ONE issue identified and explained, with clear link to context. OR TWO issues identified.	TWO issues identified with clear links to context.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
THREE (a)(i)	$\frac{121623}{2183060+1008272}=0.0381$	Conditional probability calculated.		
(a)(ii)	Model estimate $1-0.11=0.89$ Experimental estimate $\frac{2183060-(25836+258312)}{2183060}=0.8698$ True probability is not known. It has not been made clear what the current model used by the hospitals (graph) is based on. The table is only based on the last year's numbers. Moreover, the number of hospital admissions generally increases each year, which means that calculating estimates based on the table may not be accurate.	Model estimate is correctly calculated. OR Experimental estimate is correctly calculated.	Model estimate is correctly calculated. AND Experimental estimate is correctly calculated. ONE issue about the validity of either estimate.	Model estimate is correctly calculated. AND Experimental estimate is correctly calculated. ONE issue about the validity of **each** estimate.
(b)(i)	There are 5 possible single digit numbers (excluding 0). There are 5×6 = 30 possible 2-digit numbers. Therefore $\frac{5+3}{35}=\frac{8}{35}=0.2286$	Conditional probability calculated.		
(b)(ii)	There are 5×6×6 = 180 possible 3-digit numbers. There are 5×6×6×6 = 1080 possible 4-digit numbers. There are 5×6×6×6×6 = 6480 possible 5-digit numbers. Therefore, the probability that the generated number is either a 3-digit number or a 4-digit number is $\frac{180+1080}{5+30+180+1080+6480}=\frac{1260}{7775}=0.1621$	Correct number of possible numbers for any of the three cases mentioned.	Correct probability calculated.	

3

9 1 5 8 6

Level 3 Mathematics and Statistics (Statistics), 2021 v1

91586 Apply probability distributions in solving problems

Credits: Four

Achievement	Achievement with Merit	Achievement with Excellence
Apply probability distributions in solving problems.	Apply probability distributions, using relational thinking, in solving problems.	Apply probability distributions, using extended abstract thinking, in solving problems.

You should attempt ALL the questions in this booklet.

Show ALL working.

Make sure that you have the Formulae and Tables Booklet L3–STATF.

If you need more space for any answer, use the page(s) provided at the back of this booklet and clearly number the question.

YOU MUST HAND THIS BOOKLET TO THE SUPERVISOR AT THE END OF THE EXAMINATION.

TOTAL

QUESTION ONE

ASSESSOR'S USE ONLY

(a) According to the Ministry of Health, 45% of middle-aged people (40 – 60 years old) have private health insurance.

(i) Suppose that 9 middle-aged people from the same suburb are randomly selected.

Using an appropriate probability distribution, calculate an estimate for the probability that two of these middle-aged people have private health insurance.

(ii) Consider the following claim:

"The chance that more than 2 out of these 9 middle-aged people havE private health insurance is approximately 72%."

Use an appropriate probability distribution to determine whether this claim is correct.

ASSESSOR'S USE ONLY

(ii) Given that a randomly selected client claims less $3000 per month, the probability that this client claims more than an unknown amount per month, which is less than $2500, is 0.9286.
Use the cumulative distribution table below for the Triangular distribution to find the unknown amount.

***x* values**	**Probability**
$1000 \leq x \leq 2500$	0.5
$1250 \leq x \leq 2500$	0.4861
$1500 \leq x \leq 2500$	0.4444
$1750 \leq x \leq 2500$	0.375
$2000 \leq x \leq 2500$	0.2778

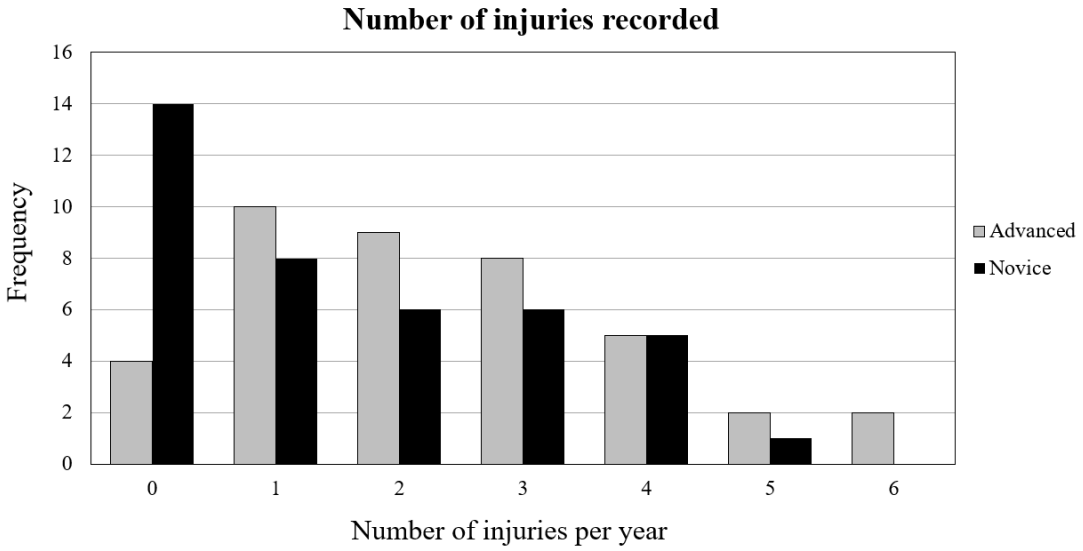
x is the amount claimed per month

ASSESSOR'S USE ONLY

(b) Wesley club runs an adult class that invites participants who are 18 years old or older.
The adult class is taught once a week for 40 weeks per year in two separate groups:

- Novice – complete beginners or those who have done gymnastics for up to 2 years
- Advanced – those who have done gymnastics for at least 3 years.

The graph below displays the number of injuries recorded in the 40 classes by group last year.



(i) The expected number of injuries per year for the Advanced group (2.35) is higher than that for the Novice group (1.575).
Explain why such results would be observed in general.

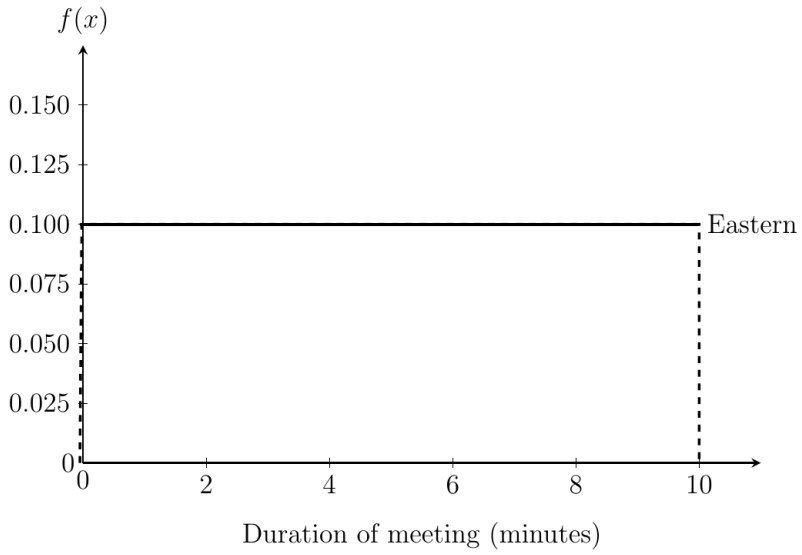
(ii) The manager of the club claims that the variation in the number of injuries occurring in the Novice group is greater than that for the Advanced group.
Comment on the manager's claim with statistical reasoning.

ASSESSOR'S USE ONLY

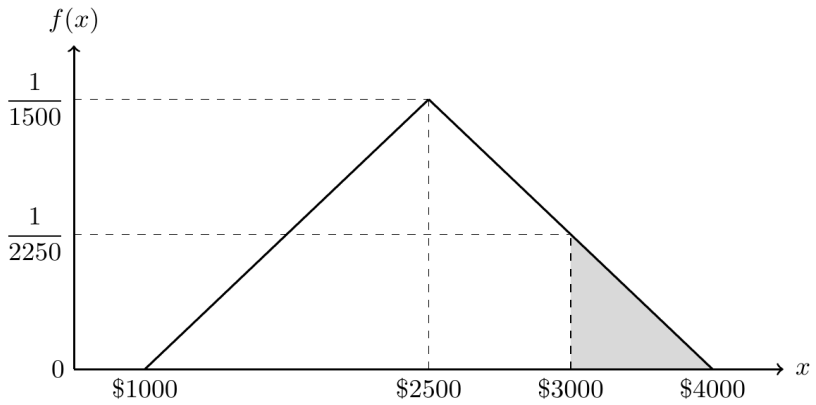
(iii) A training session from Wesley is randomly selected.
Calculate an estimate for the probability that this training session lasted between 110 minutes and 125 minutes, given that it lasted longer than 105 minutes.

(b) The coaches from both Eastern and Wesley have staff meetings that could last anytime up to 10 minutes and 8 minutes, respectively.

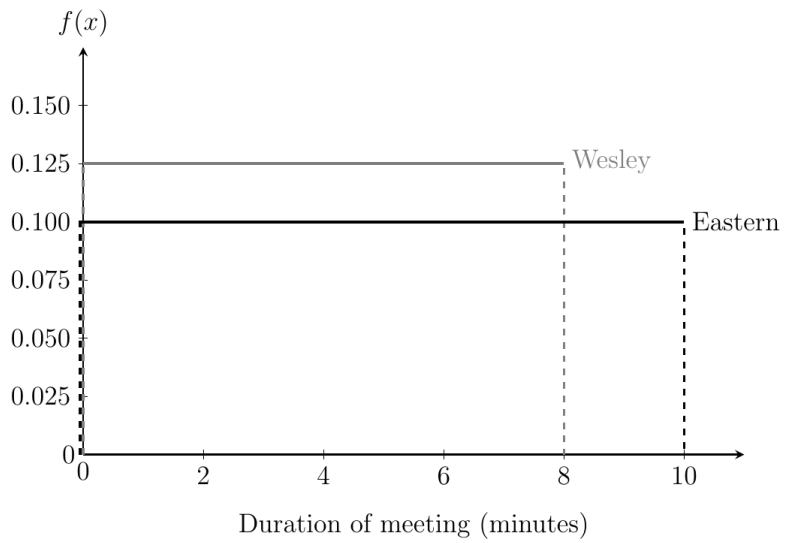
The probability distribution model for staff meeting times at Eastern is shown below.



(i) Sketch the probability distribution model for staff meeting times at Wesley on the set of axes above.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
(b)(i)	 Triangular distribution with $a = 1000$, $b = 2500$, $c = 4000$ Height at X = 3000 is $h = \frac{4000-3000}{1500^2} = \frac{1000}{1500^2} = \frac{1}{2250}$. $P(X > 3000) = 0.5 \times 1000 \times 1/2250 = 0.2222$	Correct probability is calculated.		
(b)(ii)	Let the unknown amount be k. $P(X > k \mid X < 3000) = 0.9286$ $\frac{P(k < X < 3000)}{P(X < 3000)} = 0.9286$ $\frac{1 - 0.2222 - P(1000 < X < k)}{1 - 0.2222} = 0.9286$ $P(1000 < X < k) = 0.05553$ which gives $P(k < X < 2500) = 0.4444$. Therefore, using the given table, the unknown amount k is \$1500.	Correctly equated the conditional probability with the given probability.	Simplified the equation so that it is in terms of $P(1000 < X < k)$.	Found the correct unknown amount.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
(b)(i)	Advanced gymnasts are more likely to try more difficult skills and tend to take more risks in general.	Stating any valid reason.		
(b)(ii)	Let A be the number of injuries occurring in the Advanced group. Let N be the number of injuries occurring in the Novice group. $\text{VAR}(A) = (0^2\times4/40 + \ldots + 6^2\times2/40) - 2.35^2 = 2.478$ $\text{VAR}(N) = (0^2\times14/40 + \ldots + 5^2\times1/40) - 1.575^2 = 2.294$ Since $\text{VAR}(A) > \text{VAR}(N)$, the manager's claim is false. Also accept calculating SD instead. $\text{SD}(A) = 1.57$ $\text{SD}(N) = 1.51$ **Alternatively** Accept the explanation that the number of injuries is more uniform for the Advanced group and much more right-skewed for the Novice group, so the variation in the Novice group is less.	$\text{SD}(A)$ or $\text{SD}(N)$ correctly calculated.	$\text{SD}(A)$ and $\text{SD}(N)$ both correctly calculated and have been compared to conclude that the manager's claim is false. OR Concludes claim is true/false with correct justification of distributions. OR The spread is the same for both groups, based on SDs or VARs.	
(b)(iii)	Let R be the number of reports written per year. $R = 2A + 2N + 12$ $\text{E}(R) = 2\text{E}(A) + 2\text{E}(N) + 12 = 2\times2.35 + 2\times1.575 + 12 = 19.85$ **Potential problems with the model used:** • When gymnasts get injured, they have to rehabilitate for some time during which they would be absent from class, reducing the probability that another injury occurs in the class since the class size is smaller. • It is assumed from the graph that no more than 6 injuries occur in any of the two groups, however more injuries can occur. *Accept any other valid statements*		Correct calculation of the total number of reports written per year. AND Identified ONE problem with the model used.	Correct calculation of the total number of reports written per year. AND Identified and fully described ONE problem.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
(b) (i)		Correct uniform distribution (rectangle) for Wesley drawn.		
(b) (ii)	P(Wesley lasts less than 2 minutes and Eastern lasts less than 4 minutes) $= (2 \times 0.125) \times (4 \times 0.1)$ $= 0.1$ P(Wesley lasts less than 4 minutes and Eastern lasts less than 2 minutes) $= (4 \times 0.125) \times (2 \times 0.1)$ $= 0.1$ Therefore, the required probability is $0.1 + 0.1 = 0.2$.	One probability is correctly calculated.	Correct final probability calculated.	