3

9 1 5 8 4 R

Level 3 Mathematics and Statistics (Statistics), 2020 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 1

Sorry everyone, but you should probably keep flossing!

Studies haven't confirmed if flossing prevents tooth decay and disease, and the ones that have shown some sort of benefit have been flawed, relying on too few subjects over too little time. The evidence just isn't there. It has been claimed that flossing does indeed reduce inflammation and bleeding of the gums, indications that it could theoretically prevent gum disease.

A three-month study was conducted based on a random sample of 1000 people that had not been flossing. Half of these people were randomly selected and then they were instructed to floss daily. Before the study began, everyone's oral hygiene score is recorded. At the end of the study, each person from the sample was given a new oral hygiene score. As expected, there wasn't a significant difference in the change in oral hygiene score between the two groups. Nonetheless, the researchers added that this study only lasted a few months, not nearly enough time to track the development of long-term disease.

In order to get more reliable data, researchers would have to run a study for years, with groups of both flossers and non-flossers. Conducting such an extensive study is costly, and it seems that this study not a high priority for the researchers.

Why force the public to stick their hands in their mouths? Well, because while the effectiveness of flossing isn't proven, it's also not disproven, at least not yet. After all, it's a low-risk, low-cost addition to a dental hygiene regimen. Even if flossing turns out to be not so effective, it doesn't hurt to do it - though in rare cases overzealous flossers can injure their gums or break dental work. Meanwhile, the risk of not recommending flossing is relatively high, if it is indeed an effective defence against gum disease and tooth decay.

So at the very least, talk to your dentist before chucking out your floss.

Adapted from: https://www.wired.com/2016/08/hey-sorry-everyone-probably-keep-flossing/

REPORT 2

Call for more pay transparency

Almost half of New Zealanders say they've been paid less for doing the exact same job as another person, but most didn't complain to their boss because they didn't trust it would help.

According to a recent survey, women, younger workers and those earning between $40,000 and $80,000 per annum were more likely to experience unequal pay - usually finding out they were being under paid from a colleague, not their boss.

Employees said sometimes the discrepancy was because of their skill level - but age, gender and ethnicity were also believed to be key drivers of pay gaps.

The survey findings, in new Human Rights Commission research about workers' pay experiences, have prompted fresh calls to end remuneration "secrecy" in New Zealand, by passing new transparency laws that would bring us in line with other nations.

Negotiation was rare - with 26 per cent of men and 20 per cent of women saying they negotiated for higher pay. Those on more than $100,000 per annum were three times as likely to negotiate as those earning less than $40,000.

About two in three said having more information about pay rates should be mandatory, and around the same amount said that large employers should have to make data public, and that it would help them see if their pay rate was fair.

Figure 1

Source: Human Rights Commission.

Adapted from: https://www.nzherald.co.nz/business/news/article.cfm?c_id=3&objectid=12344730

3

9 1 5 8 4

Level 3 Mathematics and Statistics (Statistics), 2020 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) Report 1 states that 1 000 people were involved in the research.

(i) Explain why this study is an experiment.

(ii) Describe the control group and treatment group for this experiment.

Control group:

Treatment group:

(b) The report states that "... there wasn't a significant change in oral hygiene score between the two two groups".
Discuss how the researchers would have established this conclusion.

QUESTION TWO

ASSESSOR'S USE ONLY

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

(a) Report 2 states that the Human Rights Commission research surveyed 500 people.
Show that the margin of error for this survey is 4.5%, and explain why a margin of error should be included in reports on statistical surveys.

(b) Identify one of the survey percentages shown in **Figure 1**, and explain why it would not be appropriate to use a margin of error based on this survey percentage to construct an approximate 95% confidence interval for the population proportion for the percentage of people who held this view on the issue of unequal pay.

(c) Assume that the people the Human Rights Commission surveyed are representative of all New Zealanders who are currently employed.

(i) Based on the survey done by the Human Rights Commission research, can it be concluded that half of working New Zealanders have had an experience where they believe they were paid less than they should have been?

QUESTION THREE

ASSESSOR'S USE ONLY

*Refer to **Report 3** in the resource booklet to answer the following questions.*

(a) Identify and describe the explanatory and response variables in the study in Report 3.

Explanatory variable: _____

_____

_____

Response variable: _____

_____

_____

(b) (i) Describe the type of the study conducted.

_____

_____

_____

_____

_____

(ii) The heading for the article is "Need to lose weight? Running and skipping will do it!". Criticise the validity of this heading in relation to the type of the study used.

_____

_____

_____

_____

_____

_____

_____

Q	Expected Coverage	Achievement (c)	Merit (j)	Excellence (i)
TWO (a)	$\text{MOE} = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{500}} = 0.045$ or approximately 4.5%. MOEs are needed to take into account the variation in survey percentages due to sampling.	Correct calculation of MOE.	Correct calculation of MOE. AND Correct explanation why the MOE is required.	
(b)	Either 1% claiming "Don't know" or 25% claiming having been paid more. These survey percentages are outside of the 30% to 70% range that the rule of thumb $\frac{1}{\sqrt{n}}$ can be applied to. The rule of thumb MOE will overestimate the size of the MOE.	The confidence interval is constructed for the survey percentage. AND The CI constructed is used to explain that the claim is not supported.		
(c)(i)	A 95% confidence interval for the survey percentage of 44% is 44% ± 4.5% or [39.5%, 48.5%]. Based on this confidence interval, the claim is not supported, as the upper end of the interval is less than 50%.	The confidence interval is constructed for the survey percentage. AND The CI constructed is used to explain that the claim is not supported.		

Q	Expected Coverage	Achievement (c)	Merit (j)	Excellence (i)
(c)(ii)	Comparison within one group: $2 \times \text{MOE} = 2 \times 4.5\% = 9.0\%$ Difference in poll percentage = 44% –25% = 19% Confidence interval generated as a result [10%, 28%] I'm pretty sure that for currently employed New Zealanders, the percentage of the people who have been paid less is somewhere between 10% to 28% higher than those who have been paid more. Both limits of the confidence interval are positive, therefore a higher percentage of people have been paid less than having been paid more.	ONE confidence interval for the difference of two proportions correctly calculated. OR TWO confidence intervals correctly calculated and used in context to discuss claim about higher percentage.	ONE confidence interval for the difference of two proportions correctly calculated. AND ATTEMPTED TO EITHER Interpret in context. OR Used to discuss claim in context.	ONE confidence interval for the difference of two proportions correctly calculated. AND Interpreted in context. AND Used and justified to discuss claim in context.
(d)	**Mode/time/venue of survey** The surveyed employees would want discretion, as the survey involves disclosing sensitive information that could upset their employers. Traditional methods such as calling, emailing, asking in person could easily result in a decrease in response rate. **Salary range** The Human Rights Commission would have needed a large volume of market rate salary data for different occupations, qualifications and duration of experience. Failure to provide the surveyed employees with their market rate salary could result in getting an incorrect/biased response or even no responses. *Accept other valid potential issues with study design.*	Identifies a feature of the study or report that could cause a potential issue.	Describes a potential issue with the design of the survey. AND Attempts to explain how it could affect responses.	Describes a potential issue with the design of the survey. AND Describes how it could affect responses 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

3

9 1 5 8 5

Level 3 Mathematics and Statistics (Statistics), 2020 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) There are currently 135 adult members at the Southern Cross Gymnastics Club.
Each member is on one of the three different levels of membership: Starter, Intermediate or Extreme.

Of the adult members:

- 65 members are on the Extreme membership.
- 60.7% of members are male.
- Of the members who are on the Starter membership, 15 are male.
- Of the male members, 30.5% are on the Intermediate membership.
- 18.5% of members are on the Starter membership.

One of the members is chosen at random.

(i) Calculate the probability that the member is male or on an Extreme membership.

(ii) How many times more likely is it that a member on an Extreme membership is a male than a female?
Support your answer with calculations.

ASSESSOR'S USE ONLY

(b) The following information summarises how the adult members of the club usually travel to the gymnasium and whether they are students or not.

	Public Transport	**Drive**	**Walk**
Students	20	15	12
Not students	5	80	3

Three of the members are randomly selected.
Calculate the probability that each of these members travelled to the gym in a different way.
State any assumptions made and justify why they are necessary.

QUESTION TWO

ASSESSOR'S USE ONLY

(a) City Hand Therapy is a clinic that specialises in treating people with any finger, hand, wrist and arm injuries.
When patients visit the clinic, they are usually given at least one of three treatments: a splint, exercise, or a surgery referral (to a surgeon).
In the case when the patient's injury is minor, they are not given any of the three treatments.

The following tables are created based on the treatment given to 170 patients from last year.

Exercise	**Surgery**	
	Y	**N**
Y	15	97
N	20	38

Surgery	**Splint**	
	Y	**N**
Y	15	20
N	57	78

Exercise	**Splint**	
	Y	**N**
Y	29	83
N	43	15

In addition, the number of patients that were given a splint only is 36.

One of these patients is chosen at random.

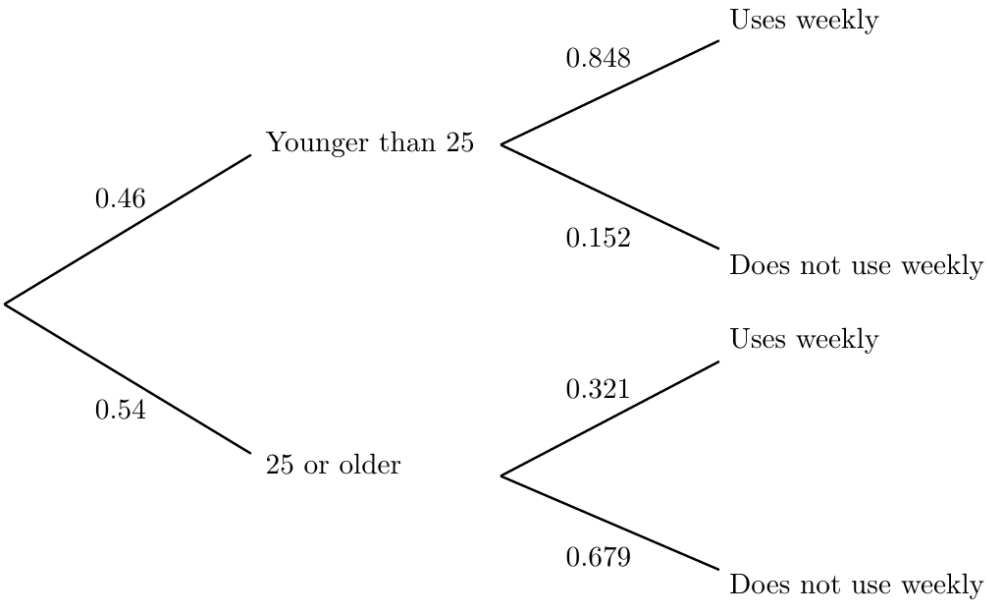
(i) Calculate the probability that the patient is treated with exercise or given a splint.

(ii) Are the events "given a surgery referral" and "given a splint" mutually exclusive?
Support your answer with appropriate statistical statements.

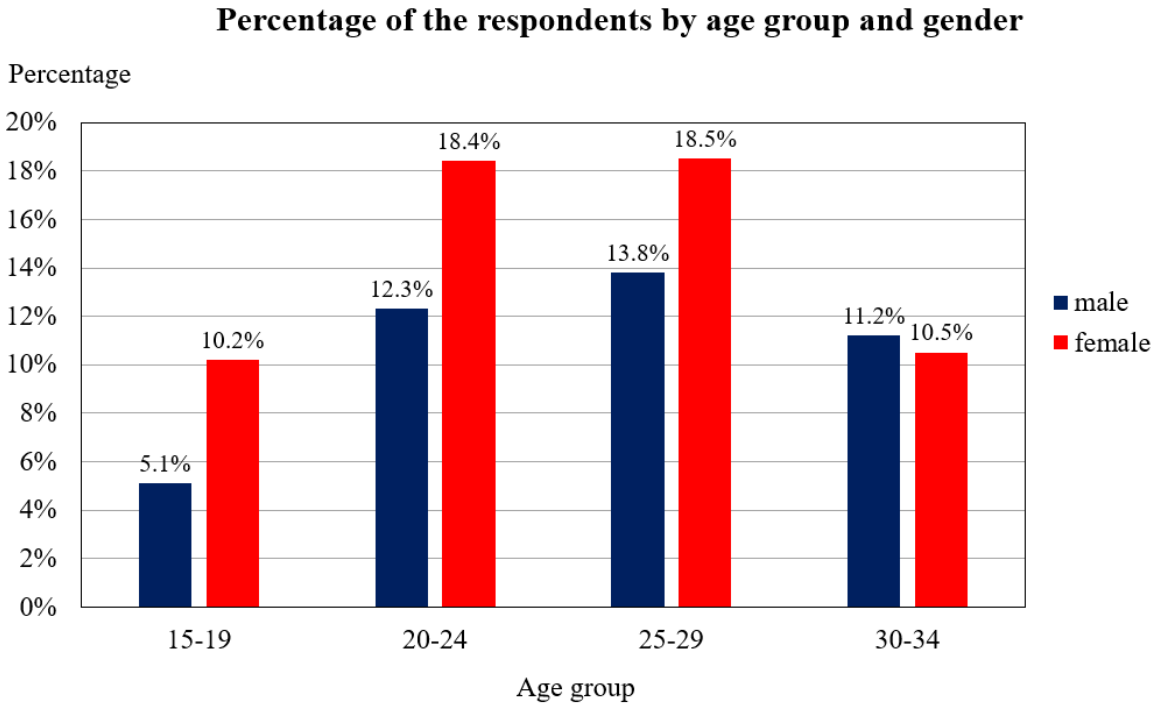
QUESTION THREE

ASSESSOR'S USE ONLY

(a) 300 people who downloaded a particular social media app were surveyed about whether they use it weekly.
It has also been recorded whether the surveyed people are younger than 25 years old or 25 years or older as well as their gender.
The results are shown in thee tree diagram below.



The graph below has also been constructed.



ASSESSOR'S USE ONLY

(i) Calculate the proportion of people who use the social media app weekly.

(ii) Find the probability that a randomly chosen person is in the 15-19 age group, given that he is male.

(iii) Use the information given to complete the table with the appropriate frequencies. Round all frequencies to the nearest whole number.

		Age	
		Younger than 25	25 or older
Use weekly	Yes		
	No		

ASSESSOR'S USE ONLY

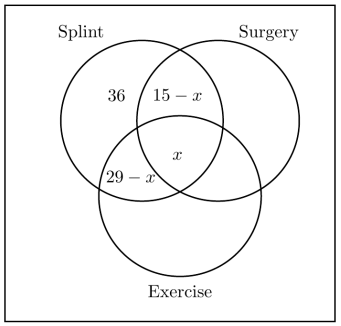
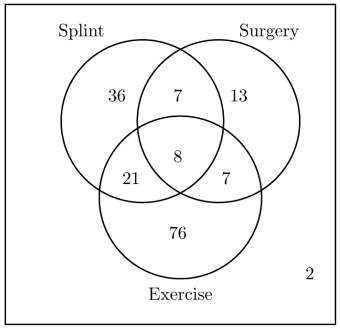
(b) A cyber security check program is run for a banking database on a daily basis.
The number of security breaches identified by the program varies depending on the day and the performance of the computers.
Let the random variable T represent the actual number of security breaches.
The following table is the probability distribution of T.

t	none	one	two	three	four	five	More than five
P(T = t)	0.91	0.05	0.03	0.004	0.003	0.002	0.001

(i) The program has been designed to make sure it does not give a false alarm.
If there is one security breach, the program identifies the breach about 50% of the time.
If there are two to four security breaches, the program identifies all of them 42% of the time.
The program identifies five or more security breaches 8% of the time.

Use the model to estimate the probability of the program identifying one or more security breaches.

(ii) State TWO assumptions in order for the model to be valid.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
TWO (a)(i)	$\text{P(exericse} \cup \text{splint)} = \frac{29+43+83}{170} = \frac{31}{34} \approx 0.9118$	Correct probability calculated.		
(a)(ii)	$\text{P(surgery} \cap \text{splint)} = \frac{15}{170} \neq 0$ Therefore, the events are not mutually exclusive.	Correct probability calculated.	Applied the definition of mutual exclusivity to show the desired results.	
(a)(iii)	  $\text{n(given splint} \cap \text{no surgery ref)} = 57$ $36 + 29 - x = 57$ $-x = -8$ $x = 8$ Therefore, $\text{P(none of treatment)} = \frac{2}{170} = \frac{1}{85}$ Accept 0.0118	Correct number (8) found in the Venn diagram.	Correct Venn diagram completed and value of 2 found.	Correct probability.

<table>
<tr><th>Q</th><th>Expected Coverage</th><th>Achievement (u)</th><th>Merit (r)</th><th>Excellence (t)</th></tr>
<tr><td>THREE
(a)(i)</td><td>P(use weekly)
$= 0.46 \times 0.848 + 0.54 \times 0.321$
≈ 0.5634</td><td>Correct probability.</td><td></td><td></td></tr>
<tr><td>(a)(ii)</td><td>P(15 – 19 years old | male)
$= \dfrac{5.1}{5.1 + 12.3 + 13.8 + 11.2}$
≈ 0.1203</td><td>Correct conditional probability.</td><td></td><td></td></tr>
<tr><td>(a)(iii)</td><td>
<table>
<tr><th colspan="2" rowspan="2"></th><th colspan="2">Age</th></tr>
<tr><th>Younger than 25</th><th>25 or older</th></tr>
<tr><td rowspan="2">Use weekly</td><td>Yes</td><td>$300 \times 0.46 \times 0.848$
≈ 117</td><td>$300 \times 0.54 \times 0.321$
≈ 52</td></tr>
<tr><td>No</td><td>$300 \times 0.46 \times 0.152$
≈ 21</td><td>$300 \times 0.54 \times 0.679$
≈ 110</td></tr>
</table>
</td><td>Two numbers are correct.</td><td>All of the numbers are correct.</td><td></td></tr>
<tr><td>(a)(iv)</td><td>• The experimental estimate, 0.5634, was found from a sample of 300 users
• The model estimate, 0.325, takes into account a range of factors.
• The true probability is unknown and cannot be found - the whole population of users of the app would need to be available.
• There is not enough information to determine whether the model or experimental estimate is closer to the true probability. Perhaps, it can be said that neither are exactly the true probability.</td><td>The experimental estimate is calculated and described.</td><td>At least two of the three types of probability are described in the response.</td><td>All three types of probability are described in the response.</td></tr>
</table>

3

9 1 5 8 6

Level 3 Mathematics and Statistics (Statistics), 2020 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) Motor vehicles arrive at a petrol station at an average rate of 1.6 vehicles every 20 minutes.

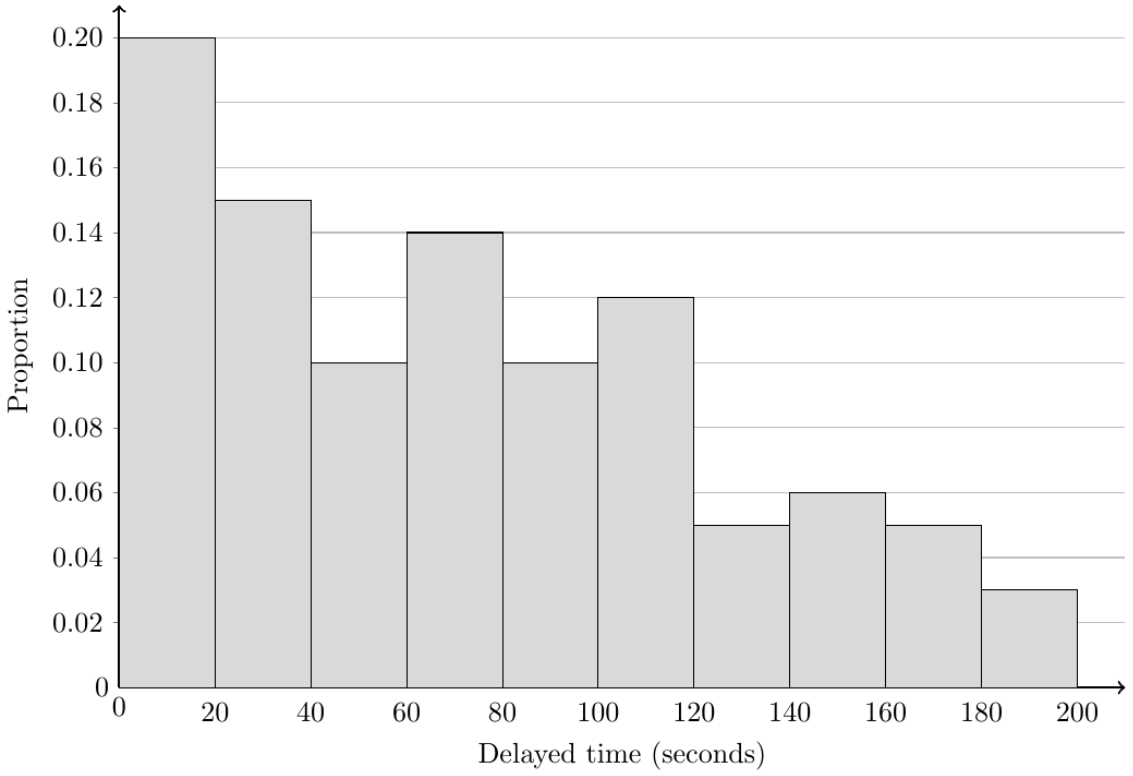
(i) Using an appropriate probability distribution model, calculate an estimate for the probability that no vehicles arrive during a 20-minute interval.

(ii) Using an appropriate probability distribution model, calculate an estimate for the probability that at least two vehicles arrive during a one hour interval.

(iii) To apply the distribution used in parts (a)(i) and (a)(ii), at least one assumption needs to be made.
Identify ONE such assumption that may be invalid and discuss why this is the case.

ASSESSOR'S USE ONLY

(b) The departure time for a particular intercity coach service is often delayed.
Based on a random sample of 1000 departure times, the following graph has been prepared.



The intercity coach drivers are given a punctuality score for each trip they make, according to the information shown in the table below.

Delayed time (seconds)	Punctuality score
0 to 40	10
40 to 80	9
80 to 120	8
120 to 160	7
160 to 200	6

(i) Use this information and the graph above to complete the probability distribution model for the random variable S, the punctuality score.

s	6	7	8	9	10
P($S = s$)		0.11	0.22	0.24	

ASSESSOR'S USE ONLY

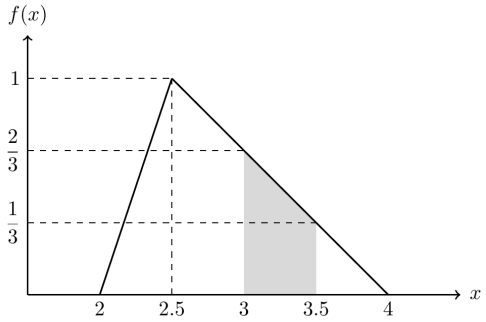
(ii) Use the probability distribution in (b)(i) to calculate the mean punctuality score of the coach service per departure.
Give any assumption(s) that need to be made.

(iii) Suppose that the industry standard punctuality scoring system changes, so that the existing scores, S, are mapped to the new scores, T, using the relationship $S = \frac{25 - T}{2}$.

Calculate the new mean punctuality score of the coach service per departure.
Briefly describe how practical the new punctuality scoring system would be in practice.

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
(b)(i)	Binomial distribution with parameters $n = 10$, $p = 0.45$. $\mathrm{P}(X \geq 3)$ $= 1 - \mathrm{P}(X \leq 2)$ $= 1 - 0.0995$ $= 0.9005$ Accept 0.9004.	Parameters for binomial distribution correctly stated.	Parameters for binomial distribution correctly stated AND Probability correctly calculated.	
(b)(ii)	Using a normal distribution with parameters $\mu = 1000 \times 0.45 = 450$ and $\sigma^2 = 1000 \times 0.45 \times 0.55 = 247.5$. As the number of customers is a discrete variable but the normal distribution is a continuous variable, continuity correction is used. $\mathrm{P}(X > 480.5) \approx 0.0263$ **Limitation** The minimum number and the maximum number of customers should be restricted to 0 and 1000 respectively to make it clear what values are allowed. This is because the normal distribution is unbounded.	Parameters for normal distribution correctly stated.	$P(X > 480)$ or $P(X > 480.5)$ correctly calculated. AND Correct identification of the normal distribution with the correct parameters as an appropriate model.	Applied normal distribution correctly to find the required Probability AND ONE correct limitation stated.

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Reasonable start / attempt at one part of the question.	1 of u	2 of u	3 of u	1 of r	2 of r	1 of t	2 of t

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
TWO (a)(i)	 $\mathrm{P}(X>3)=0.5\times1\times\frac{2}{3}=\frac{1}{3}$ $\mathrm{P}(X>3.5)=0.5\times0.5\times\frac{1}{3}=\frac{1}{12}$ $\mathrm{P}(3<X<3.5)=\frac{1}{3}-\frac{1}{12}=\frac{1}{4}$ Also accept calculating the required probability using trapezium formula.	Either Correct $\mathrm{P}(X>3)$ calculated OR Correct $\mathrm{P}(X>3.5)$ calculated.	Correct probability calculated.	
(a)(ii)	Using the triangular distribution, $\mathrm{f}(2.5)=1$. But since $\mathrm{P}(2<X<2.5)=0.5\times0.5\times1=0.25$, it can be concluded that 2.5 hours is the lower quartile time.	Shown that $\mathrm{f}(2.5)=1$.	Shown that the lower quartile is 2.5 hours with justification.	

Q	Expected Coverage	Achievement (u)	Merit (r)	Excellence (t)
(c)	The normal distribution model would not be an appropriate model. • The shape of the distribution appears to be not quite symmetrical about the mean (35.9 minutes). The shape seems unimodal but seems skewed to the right. There also appears to be unequal proportions of times above and below the mean. • Comparison of theoretical probability and observed probability e.g. Model $P(X < 30) = 0.2722$ which is not that close to $P(X < 30) = (5+16+46)/200 = 0.335$. • Discussion or calculation of standard deviation (by inverse normal) or range or middle 50% or similar. • Ruling out any sampling variation. Even though 200 could well be a sufficient sample size, the observed distribution evidently deviates from the normal distribution.	ONE relevant feature of the distribution is described in context and compared to the feature of the normal distribution. OR A relevant probability is calculated using the normal distribution and compared to the given distribution.	ONE relevant feature of the distribution is described in context and compared to the feature of the normal distribution. AND A relevant probability is calculated using the normal distribution and compared to the given distribution.	TWO different relevant features of the distribution are described in context and compared to the features of the normal distribution. AND A relevant probability is calculated using the normal distribution and compared to the given distribution.

NØ	N1	N2	A3	A4	M5	M6	E7	E8
No response; no relevant evidence.	Reasonable start / attempt at one part of the question.	1 of u	2 of u	3 of u	1 of r	2 of r	1 of t	2 of t

Cut Scores

	Not Achieved	Achievement	Achievement with merit	Achievement with Excellence
Score range	0 – 7	8 – 12	13 – 18	19 – 24