SUPERVISOR'S USE ONLY

3

91586

915860

Draw a cross through the box (☒) if you have NOT written in this booklet

Mana Tohu Mātauranga o Aotearoa
New Zealand Qualifications Authority

Level 3 Mathematics and Statistics (Statistics) 2024

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.

Check that the National Student Number (NSN) on your admission slip is the same as the number at the top of this page.

You should attempt ALL the questions in this booklet.

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

Show ALL working.

If you need more room for any answer, use the extra space provided at the back of this booklet.

Check that this booklet has pages 2–16 in the correct order and that none of these pages is blank.

Do not write in any cross-hatched area. This area will be cut off when the booklet is marked.

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

Merit

TOTAL 16

QUESTION ONE

2 options

(a) A person's belly button can be described as an "innie" (navel indented) or an "outie" (navel protruded), as pictured below. Belly-button shape is a cosmetic characteristic and not a significant indicator of health, so data about belly buttons is not routinely collected. It is therefore difficult to make a reliable estimate of the proportion of outie belly buttons in the population.

https://www.pampers.com/en-us/baby/newborn/article/outie-belly-button

Data on belly-button type was collected in the world's first belly button survey. From this data, it was found that 96% of respondents had an innie belly button.

(i) Using an appropriate probability distribution model, calculate an estimate for the probability that, in a class of 30 students, no students have an outie belly button.

State the probability distribution model and the parameters used as part of your answer.

96% = innie

4% = outie = 0.04 = p

n = 30

x = 0

P (no students in a class of 30 have an outie) = 0.2939

(ii) Give TWO reasons why the use of the probability distribution model selected in part (i) is appropriate.

Reason one: Because there are 2 possible options, either they have an innie of they have an outie which binomial distibution

Reason two: Because each result of belly button is independent from the next result. The result of one belly button being innie or outie will not ~~result~~ have a result on the next belly button, result.

(iii) Belly button fluff is the accumulation of fibres, lint, or debris that collect in a person's belly button. A belly button survey found that 66% of all the respondents had fluff in their belly buttons.

34% didn't

Using an appropriate probability distribution model, calculate an estimate for the probability that one student in a class of 30 has an outie belly button, and it contains belly button fluff.

You should state the probability distribution model and the parameters used as part of your answer.

P(1 student has an outie) = 0.3673

P(1 student contains belly button fluff)

= 5.12×10^{-13}

$0.3673 \times 5.12 \times 10^{-3} = 1.88 \times 10^{-13}$

The probability distribution model used is binomial distribution as there are 2 outcomes, either has belly button fluff or doesn't have belly button fluff. There is a fixed number of trials of 30

*

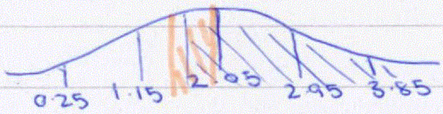
continued page 14

(b) Very little research on the typical size of belly buttons has been undertaken. It is hypothesised that the diameter of adult belly buttons (the distance across at the widest point) could be modelled by a normal distribution with mean of 2.05 cm and standard deviation of 0.9 cm.

(i) Calculate an estimate for the probability that a randomly selected adult belly button has a diameter of at least 1.75 cm, given that its diameter is between 1.5 cm and 2.0 cm.

$\mu = 2.05$ $z = \frac{x - \mu}{\sigma}$ $P(A|B)$

$\sigma = 0.9$



P (diameter between 1.5 and 2cm) = 0.2073

-0.6745

$z = \frac{1.75 - \mu\ 2.05}{0.9}$ $z = 0.3333$

~~0.3333~~ $0.3333 \div 0.2073$

$= 0.06909$

(ii) Comment on the suitability of using the parameters given above for modelling adult belly button width.

The parameters above are suitable for modelling adult belly button width as it provides an suitable range which the belly button could fit in. The parameters above are suitable are the most common probabilities are central.

QUESTION TWO

Fluff collects in the belly buttons (navels) of some people. The Nature of Navel Fluff study (Steinhauser) was published in 2009. The author of the study collected data about his own belly button fluff. Each day, for 502 days, he had a morning shower where he washed his belly button. At the end of each day, he removed and weighed the belly button fluff that had accumulated during the day.

(a) The weight distribution of the amount of belly button lint (fluff) he collected each day is shown in Figure 2(a) below.

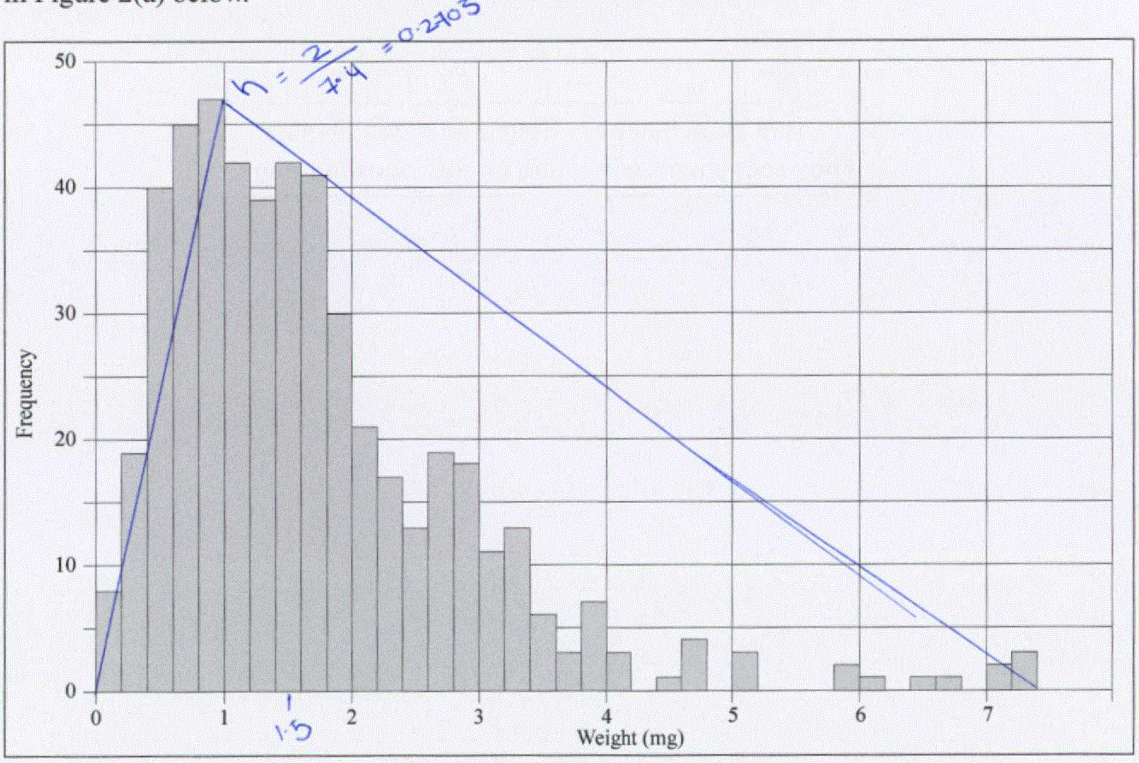


Figure 2(a): Weight distribution of 502 pieces of daily belly button fluff

It is proposed to model daily belly button fluff weight using a triangular distribution with parameters:

minimum = 0 mg, maximum = 7.4 mg, and mode = 1.0 mg

(i) Sketch this model on Figure 2(a).

(ii) Using the given parameters, estimate the probability that on any given day the belly button fluff weight is less than 1.5 mg.

P(less than 1.5) = 1 - > 1.5

~~f(1.5)~~ $f(1.5) = \dfrac{2(7.4-1.5)}{(7.4-0)(7.4-1)} = 0.2492$

Area = $\frac{1}{2} \times (7.4-1.5) \times 0.2492 = 0.73514$

P(less than 1.5) = $1 - 0.73514 = 0.26486$

(iii) Suggest alternative parameters for a triangular distribution that will give a better fit to this data and sketch your new model onto Figure 2(b) below.

As part of your answer, justify your choice of these alternative parameters.

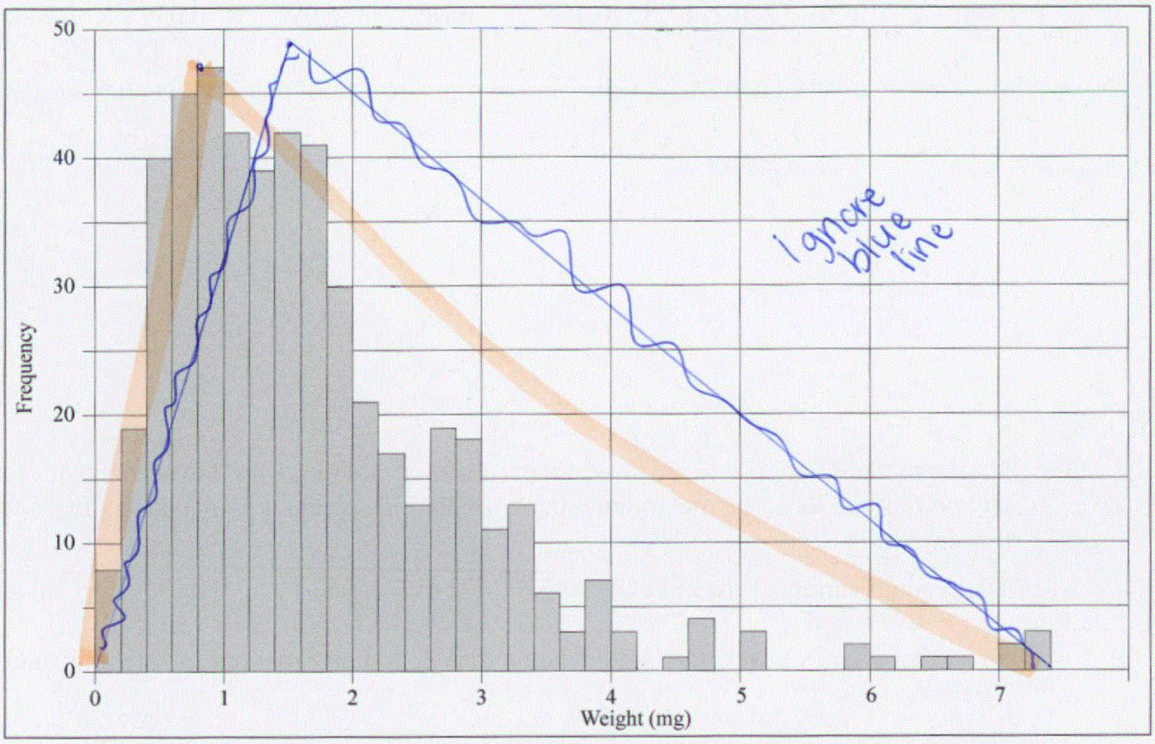


Figure 2(b): Weight distribution of 502 pieces of daily belly button fluff

Minimum = 0

Maximum = 7.4

~~Median~~ Mode = 0.8

As the data is skewed right, majority of the data is concentrated left, therefore moving the mode more left will encounter for the right skew.

(b) The table below shows the probability distribution of the random variable, T, the number of towels used per person for residents from one city over one day.

t	0	1	2	3	4	5	6
P(T = t)	0.129	0.271	0.264	0.185	0.095	0.039	0.017

(i) How many towels, on average, were used by a resident from this city over the one day?

$(0 \times 0.129) + (1 \times 0.271) + (2 \times 0.264) +$
$(3 \times 0.185) + (4 \times 0.095) + (5 \times 0.039) + (6 \times 0.017)$

average = 2.031

(ii) Let the random variable, C, be the number of sets of clothing worn for residents from this city on this day. The maximum number of clothing changes by residents of this city on this day is 3.

The random variable C has SD(C) = 0.5.

Show that C has a lower standard deviation than T, and give one reason why this might be the case.

C has a standard deviation of 0.5 and T has a standard deviation of 1.41. T has a larger standard deviation as the maximum amount of towels used per day is 6, 6 - 0 = 6 so T has a spread of 6, whereas for C the maximum amount of clothes worn each day is 3 3 - 0 = 3, therefore C has a spread of 3. A spread of 6 is larger than a spread of 3 so a larger standard deviation

* Continued Page 14

(iii) SD(T + C) = 1.754

Are T and C independent?

Comment on what this suggests about the number of towels used and number of clothing changes for residents of this city.

Support your answer with statistical statement(s) and calculation(s).

Independence = $P(A \cap B) = P(A) \times P(B)$

$1.754 \neq 0.627$

These evidents are not independent, this suggests the amount of clothes changes is linked to the number of towels used by residents each day. This may be due to as someone takes a shower using a towel they will also change there clothes which are linked events and are not independent.

QUESTION THREE

A study published in 2012, based on data collected from 138 people participating in a nationwide American citizen science project in 2011, collected information about belly button habits. The distribution of washing frequency (defined as the number of belly button washes per 7-day week), is shown in Figure 3(a).

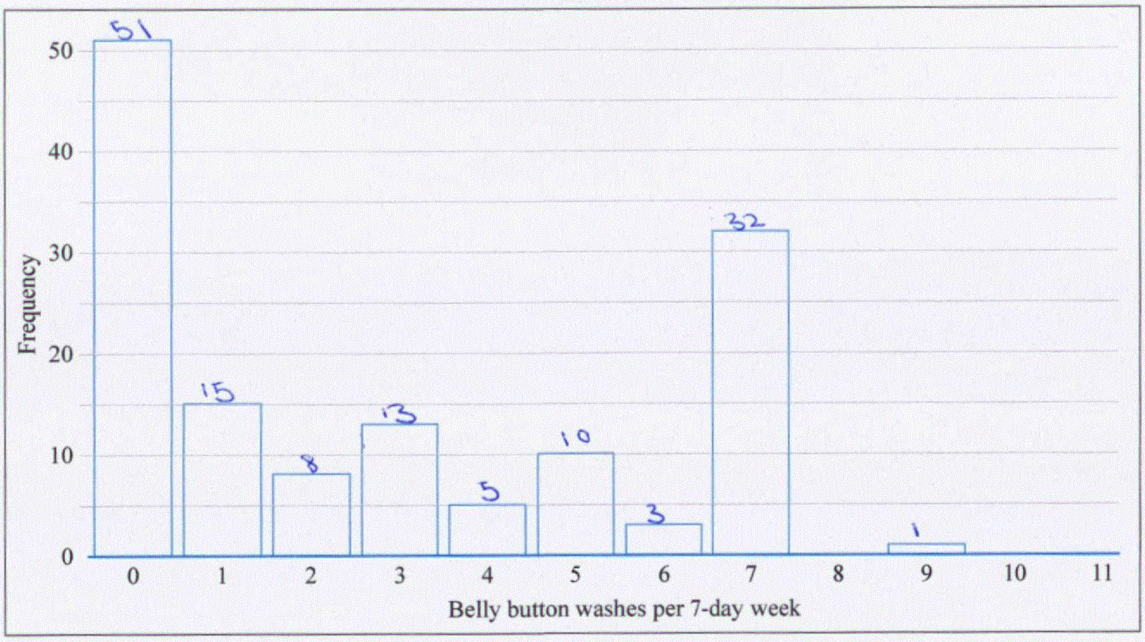


Figure 3(a): The number of reported times survey respondents washed their belly buttons per week

(a) A media release reporting on the study's results had the following headline:

"SHOCKING – less than a quarter of US citizens wash their belly button daily".

1/4

Comment on whether this headline is correct.

Support your answer with statistical calculations and reasoning.

Daily = ≥ 7 days

$32 + 1 = 33$

$33 / 138 = 0.2391$

0.2391 is less than $1/4 = 0.25$ so the headline is correct.

(b) Suppose the number of belly button washes per week for the general population is modelled by a Poisson distribution with $\lambda = 2.8$.

(i) Use this Poisson model to calculate an estimate for the probability that a randomly selected person from the general population washes their belly button at least once in a day.

≥ 1

$P(X \geq 1) = 1 - P(X=0)$

$P(X = 0) = e^{-\lambda}$

$P(X = 0) = e^{-2.8}$

$= 0.06081$

$P(X \geq 1) = 1 - 0.06081 = 0.9319$

(ii) To apply the distribution used in part (b)(i), at least one assumption must be made.

Identify ONE assumption made that may be invalid and discuss why this is the case.

~~Po~~ Poisson models rare events, so the event that at least once a day a selected person washes there belly button is a rare event. Therefore it is uncommon to wash your belly button. This may be invalid as most people shower everyday so it is likey that there stomach and belly button gets washed in some way

*
continued page 14

Question Three continues on the next page.

The sample data shown in Figure 3(a) is run through a simulation model 1000 times, assuming that a Poisson distribution with $\lambda = 2.8$ is used to model the number of belly button washes per 7-day week for the general population. Figure 3(b) shows the results of the simulation model and the original observed data (blue dots).

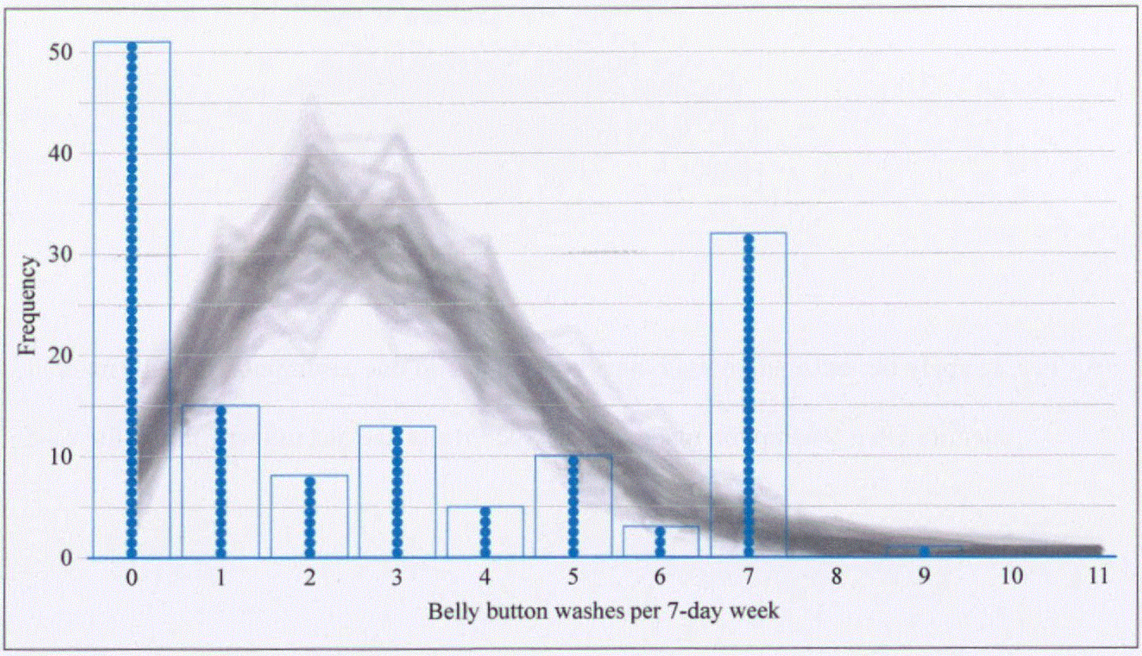


Figure 3(b): Results of the simulation model and original observed data

(iii) Explain what the grey band is showing in Figure 3(b).

The grey band is showing the variation for the simulation and ~~the likhihood~~ is compearing the observed model against the simulated model to see the liklihood of the simulated model occuring.

* continued page 14

(iv) Based on the results of the simulation model and the original observed data (Figure 3(b)), discuss whether the Poisson distribution model (presented opposite) appears to be appropriate for modelling the number of belly button washes per 7-day week.

Support your answer with statistical reasoning.

The simulated model shows the most common amounts to wash your belly button are 2-3 times. The observed model shows the most common amount is 0 times. Poisson model however doesn't follow the shape of the observed model so poisson is not an appropriate model to show the number of belly button washes per day. The mean for the observed model and poisson is both 2.8.

(v) Propose and justify the use of an alternative model that could be appropriate for modelling the number of belly button washes per 7-day week.

You should state the probability distribution model and the parameters used as part of your answer.

Triangular distribution as the poisson simulation models a triangular shape. Bellybutton washing is likely to have fixed endpoints as it isn't a continuous thing. Therefore min = 0, max = 9 and mode = 0. This model shows that mean doesn't equal the mode which is true in triangular distribution. The right skew triangular shape may be a useful alternative model.

Extra space if required.
Write the question number(s) if applicable.

QUESTION NUMBER

Question One (a) (iii)

Students. The probability success stays the same. 0.04 for an outie belly button. 0.66 for belly button fluff. Each result is independent and the likely hood of one belly button fluff having fluff does not affect the result of the other.

Parameters : x = 1 , n = 30

p = 0.66 for belly button fluff

p = 0.04 for an outie belly button

Question Two (b) (ii)

Is required for T than C as T has more variation.

Question Three (b) (ii)

Which means the event is not very rare, as belly buttons are likely to be washed in some sort of way.

Question Three (b) (iii)

the grey band shows a right skew, therefore its more likely for a belly button to be washed 1 - 4 times a week compeared to 7 - 11 times a week. The most common amounts are 2-3 times a week.

Merit

Subject: Statistics

Standard: 91586

Total score: 16

Q	Grade score	Marker commentary
One	M5	(a)(i) The distribution and correct parameters have been clearly stated and the probability correctly calculated. This is u. (a)(ii) Both Reason One and Two are sufficient when read together. This is r since both part a(i) and a(ii) are correct. (a)(iii) Only the probability for one student has been calculated correctly. To reach r, this probability needs to be multiplied by 0.66. To reach t, all three parts: a(i), a(ii) and a(iii), need to be correct. (b)(i) Only one of the three probabilities required has been correctly calculated. ($P(1.5<X<2)$). To reach r, the correct conditional probability needs to be calculated or (b) (ii) needs to be correct. (b)(ii) Insufficient detail as to how/ why the given parameters: μ=2.05cm and $\sigma = 0.9$ cm are suitable has not been provided. To reach r, a comment referencing a feature of the Normal distribution linked to the specific parameters and the context, and stating that they could be suitable is required *e.g. P(X ≤0)= 0.0114 using this model with these parameters. This suggests that 1.1% of people could have a belly button with a negative diameter but since this is such a small % the model could be suitable.* To reach t, both (b)(i) and (b)(ii) need to be correct.
Two	M6	(a)(i) a graph with correct minimum, mode and maximum has been drawn. (a)(ii) This is r as the correct probability for $P(X < 1.5)$, with working, has been given. (a)(iii) The new model is incorrect. The maximum chosen is too large. To reach u a maximum between 4 and 5.2 needs to be chosen. To reach r, one of the parameters chosen needs to be justified in terms of the features of the data and/or context given *e.g. belly button lengths can't be less than 0 so the minimum should stay the same at 0cm.* To reach t, two of the three parameters chosen need to be justified in terms of the features of the data and/or context given. (b)(i) The correct mean has been given. (b)(ii) The correct SD(T) and an explanation of why SD(T) would be less is given in terms of the difference in variation (spread/range)

		between the number of towels used compared to number of changes of clothing. (b)(iii) The correct working in terms of VAR(T), VAR (C) and VAR (T + C) has not been given. To reach r, this is necessary. The contextual explanation of why they what this means in context would be sufficient for t, if the mathematical working was correct.
Three	M5	A correct calculation using frequencies from the graph is made, compared to 25% of the sample and a statement that the headline is correct is made for u. (a)(i) Lambda has not been divided by 7 and 1- P(X=0) using this new lambda has not been found. To reach r, this needs to happen. A grade of u has been given as the probability using lambda= 2.8 of 1- P(X=0) is correct. (b)(ii) The assumption is incorrect. An assumption about randomness, rate or independence is required. (b)(iii) The grey band (tracked over-fitted shape) has been linked to the expected outcomes from a simulation using a Poisson distribution, this is required for u. To reach r, the link to the variation that would be expected using the model is required. (b)(iv) A visual comparison of the real sample data and the tracked over fitted shape has been made along with the conclusion that the model is not appropriate. To reach t, discussion of what is expected to be seen ‘by chance alone’ is required (b)(v) A wrong alternative model has been suggested. The uniform model, with parameters needs to be identified and justified in terms of the how it will fit the observed data better.