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.

Achievement

TOTAL 10

QUESTION ONE

(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.

96 = innie 0.04 = outie

(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.

Binomial P.D

$P(x=0)$ P ~~= 0.6648~~ $= 0.29385$

$N = 30$ $P(x=0) = 0.2939$

$P = 0.04$

- Independant success or failure
- probability of success stays the same of 0.04
- number of trials stays the same

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

Reason one: The events are independant. The probability of one student NOT having an outie belly button doesn't affect the probability of another student NOT having an outie belly button. Therefore Binomial distribution is appropriate.

Reason two: There are only two out comes. A student has an outie belly button or they don't. so Binomial distribution model is appropriate.

(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.

66% or 0.66 = fluff

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.

$n = \cancel{70}\ 30 \quad x = 1 \quad p = 0.04$

$P(x=1)$ outie belly button $= \cancel{0.2710} = 0.3673$

$P(x=1)$ belly button fluff $= 0.3411$

$x = 1$

$\lambda = 0.66$

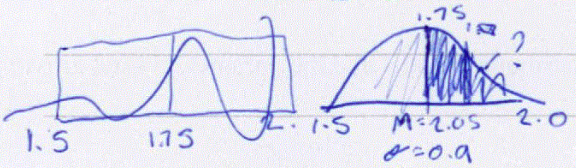
$0.3411 \times 0.3673 = 0.12528$

$= 0.1253$

(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.

$\mu = 2.05 \quad \sigma = 0.9$

(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.



~~1.75 = x - 2.05~~

$z = \dfrac{1.75 - 2.05}{0.9}$ $\quad z = 0.20728$
$= 0.2073$

$z \times 0.9 = 1.75 - 2.05$

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

Not suitable parametere especially ~~the~~ with the maximum being 2.0cm when the mean for adult belly button width is over that being 2.05cm.

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.

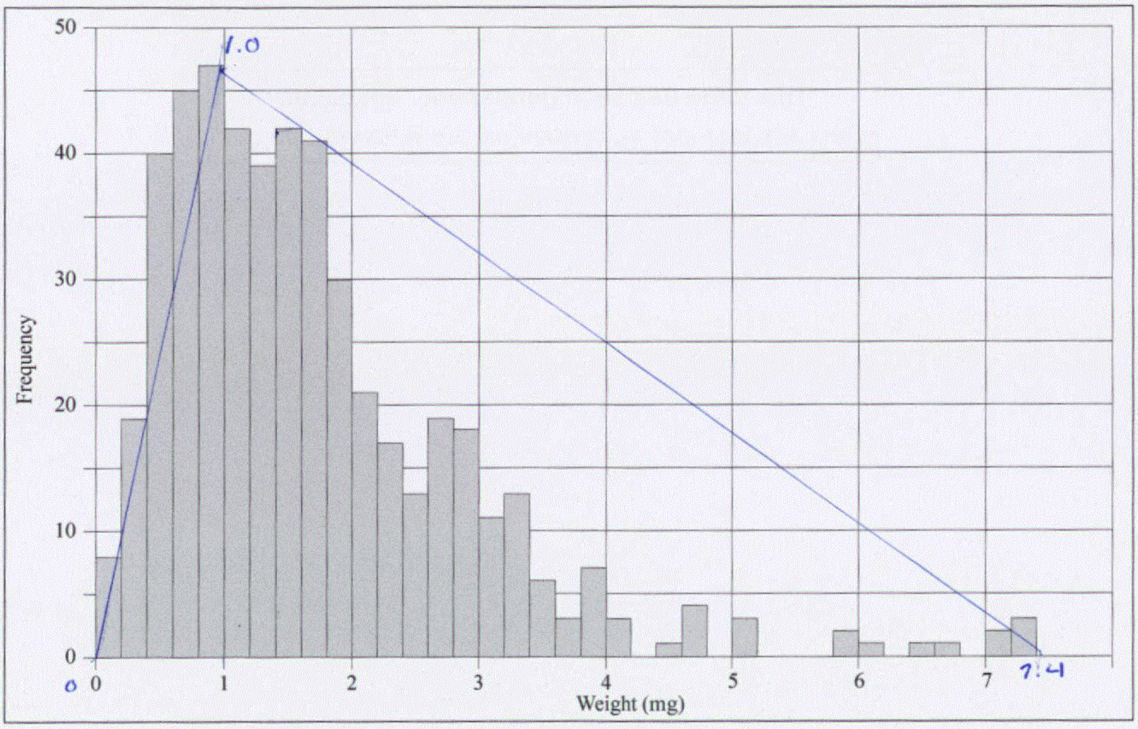


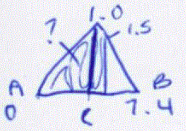
Figure 2(a): Weight distribution of 502 pieces of daily belly button fluff

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.



$P(x < 1.5)$ $\qquad \frac{2(b-x)}{(b-a)(b-c)}$

$\cancel{\frac{2}{7.4}} = 0$

$\frac{2}{7.4} = 0.27$ $\qquad \frac{2(7.4-1.5)}{(7.4-0)(7.4-1.0)}$ $\qquad \frac{2(5.9)}{(7.4)(6.4)} = 0.24915$
$= 0.2492$

$b = 1.5 - 0 = 1.5$ $\qquad 0.5 \times 1.5 \times 0.2492 = 0.1869$

(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.

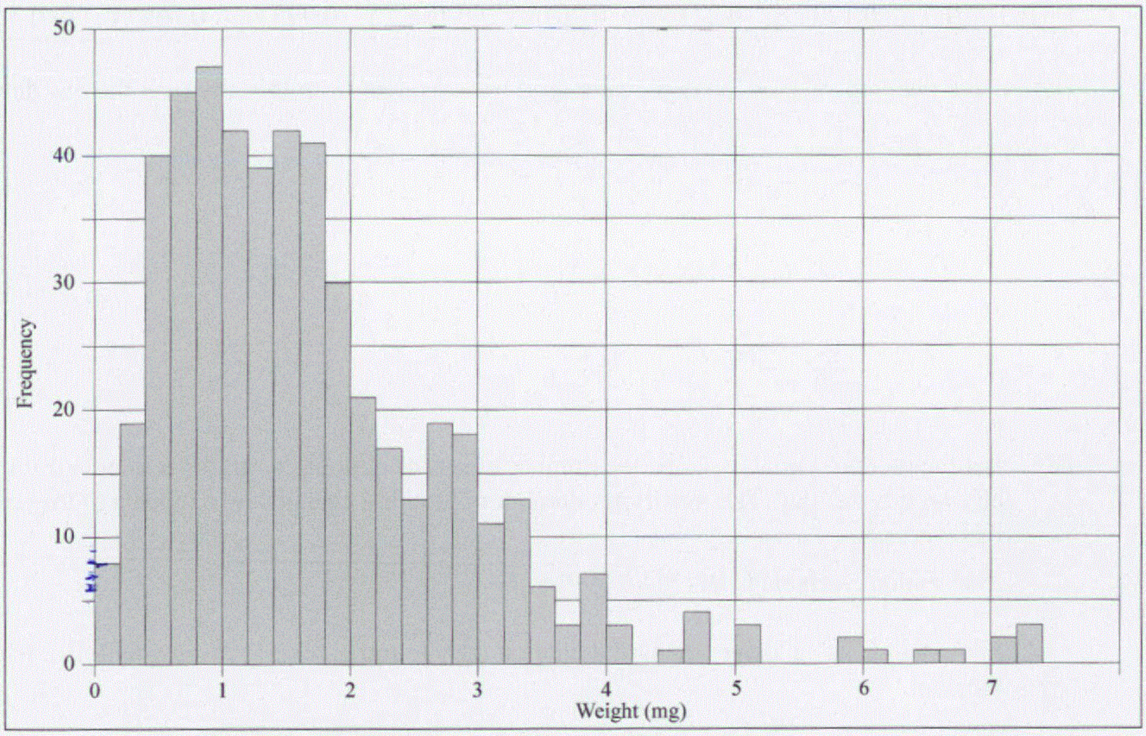


Figure 2(b): Weight distribution of 502 pieces of daily belly button fluff

~~minimum = 8 Maximum = 7.4 mode~~

(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?

$\mu = 2.031$

2 or 3 towels were used by a resident from this city over the one day.

(ii) Let the random variable, C, be the number of sets of clothing worn for residents from this city on this one 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.

(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).

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).

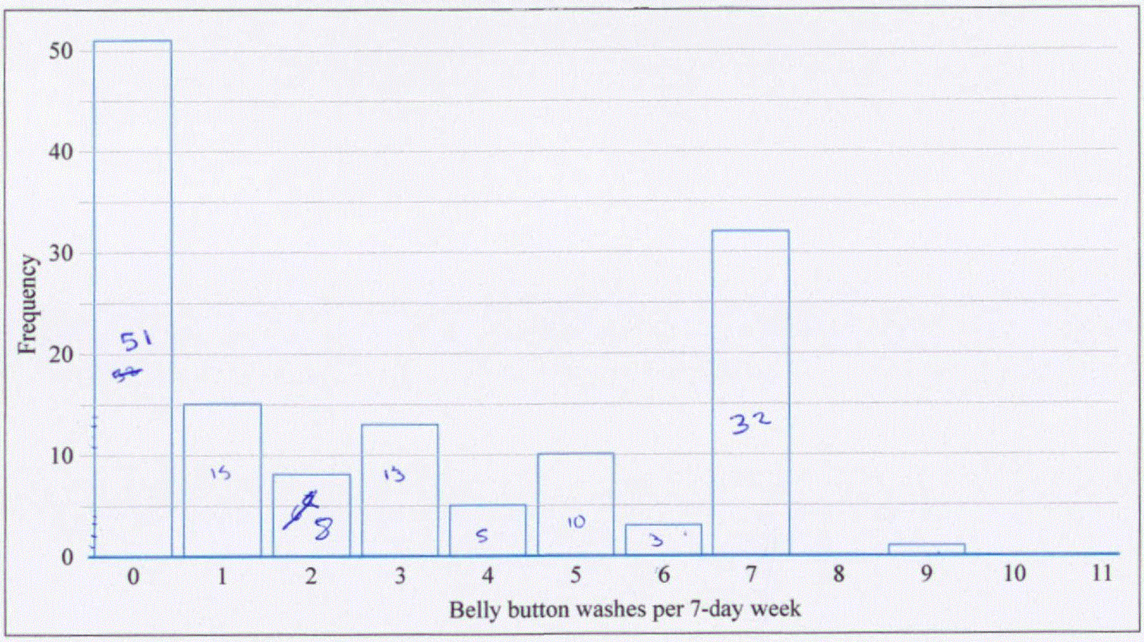


Figure 3(a): The number of reported times survey respondents washed their belly buttons per week

(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".

Comment on whether this headline is correct.

Support your answer with statistical calculations and reasoning.

138 ÷ 4 = 34.5 This head line is correct because 33

n=33 US citizens wash their belly button

~~x=2.8333~~ daily which is just under a quater of

x=2.8333 34.5 so the statement 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.

$P(x=1)$ $\lambda = 2.8$

$= 0.23107$

$= 0.2311$

$1 - 0.2311 = 0.7689$

(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.

~~An assumption is that it is randomly selected. This assumption may be invalid as~~ ~~the population~~ ~~the data is from the general population where you~~ An assumption is independence. that the probability of one person washing their belly button at least once a day doesn't affect the probability of another person washing their belly button at least once a day. This assumption may be invalid as the person is randomly selected.

???

Poisson
- random
- independant
- Cannot occur ~~sihu~~ simultanesly
- rate must be constant

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).

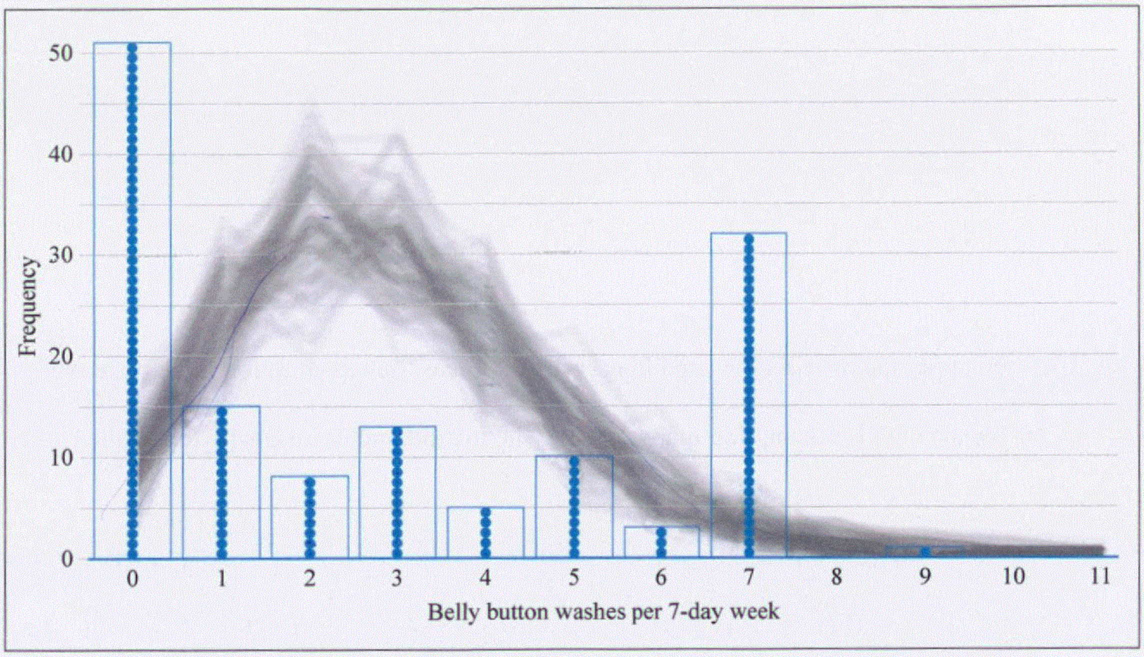


Figure 3(b): Results of the simulation model and original observed data

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

The ~~grey~~ grey ~~bad~~ band is showing ~~that~~ the data being normally distributed underneith a bell-shaped curve.

(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.

(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.

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

QUESTION NUMBER

1aii The binomial probability distribution model is appropriate because the probability of success of having an outie belly button ~~is~~ of 0.04 ~~%~~. (4%) stays the same for all 30 students.

Another reason why the Binomial probability distribution model is appropriate because the number of trials of 30 students stays the same for all of the trials that no students have an outie belly button.

Achievement

Subject: Statistics

Standard: 91586

Total score: 10

Q	Grade score	Marker commentary
One	A4	(a)(i) The binomial distribution and the correct parameters have been clearly stated and the probability correctly calculated. This is u. (a)(ii) Only Reason Two is sufficient. To reach r both reasons need to be correct. In Reason One the candidate has stated probabilities are independent. This is incorrect, the events are independent. To reach r both part a(i) and a(ii) need to be 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 in this part has been calculated correctly (P(1.5<X<2)). To reach r the correct conditional probability needs to be calculated. (b)(ii) Evidence that the given parameters: μ=2.05cm and $\sigma = 0.9$ cm are not 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 would not be suitable is required e.g. *with these parameters 99% of people would have belly buttons between -0.65 and 4.75cm. Since a belly button can't have a negative width, the parameters wouldn't be suitable.* To reach t both (b)(i) and (b)(ii) need to be correct.
Two	A4	(a)(i) A graph with correct minimum, mode and maximum has been drawn. (a)(ii) Only the height at X= 1.5 is correct. This is u. To reach r the correct probability for P(X< 1.5), with working, needs to be given. (a)(iii) This part has not been attempted. To reach u a new model with appropriate minimum, mode and maximum is given. 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) This part has not been attempted. The correct SD(T) is enough for a u. To reach r, an explanation of why SD(T) would be less needs to be given. Discussions need to be in terms of the variation (spread/range) not just maximum number of towels or changes of clothing. (b)(iii) This part has not been attempted. To reach r the correct working in terms of VAR(T), VAR (C) and VAR (T + C) needs to be shown. To reach t a statement that T and C are not independent and an explanation of what this means in terms of changes of clothing and towels needs to be given.
Three	N2	(a) 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. (b)(i) Lambda has not been divided by 7 and 1- P(X=0) has not been found. To reach r this needs to happen. (b)(ii) The assumption is incorrect. Probabilities are not independent, events are. The reason for the assumption being invalid is insufficiently linked to when a person washes, impacts when they wash their belly button the next time period. (b)(iii) The student's response is incorrect. They have not stated that the grey band represents the expected outcomes from a simulation using a Poisson distribution, this is required for u. To reach r, the link to the expected variation that would be expected using the model is required. (b)(iv) This part has not been attempted. To reach u, a comment that the grey band (the Poisson model) does not fit the observed data is necessary. To reach r, a visual comparison of the observed data and the tracked over-fitted shape needs to be made (e.g. *overestimates the frequency of washing 1 to 4 times, underestimate the frequency of washing 0 times a week.*) To reach t discussion of what is expected to be seen 'by chance alone' is required. (b)(v) This has not been attempted. The uniform model, with parameters needs to be identified and justified in terms of the how it will fit the observed data better.