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.

Excellence

TOTAL 19

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.

(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 distribution $X \sim B(30, 0.04)$

$P(X=0) = 0.2939$

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

Reason one: there is a fixed number of trials, in this case 30 students

Reason two: there are two possible outcomes, innie or outie belly button

(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.

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.

binomial distribution $X \sim B(30, 0.04)$

$P(X=1) = 0.3673$

$P(\text{one student outie} \cap \text{fluff}) = 0.3673 \times 0.66$

$= 0.2424$

(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.

$X \sim N(2.05, 0.9)$

$P(1.5 < X < 2.0) = 0.2073$

$P(1.75 < X < 2.0) = 0.1084$

$\cancel{P(1.5 < X < 2.0 / X > 1.75) =}$

$P(X > 1.75 / 1.5 < X < 2.0) = 0.1084 \div 0.2073$

$= 0.5229$

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

very little research has been done so the hypothesis is unlikely to be accurate also the standard deviation of 0.9 implies that some belly button widths are negative, which is impossible, so the distribution may not be suitable for modelling adult belly button width

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.

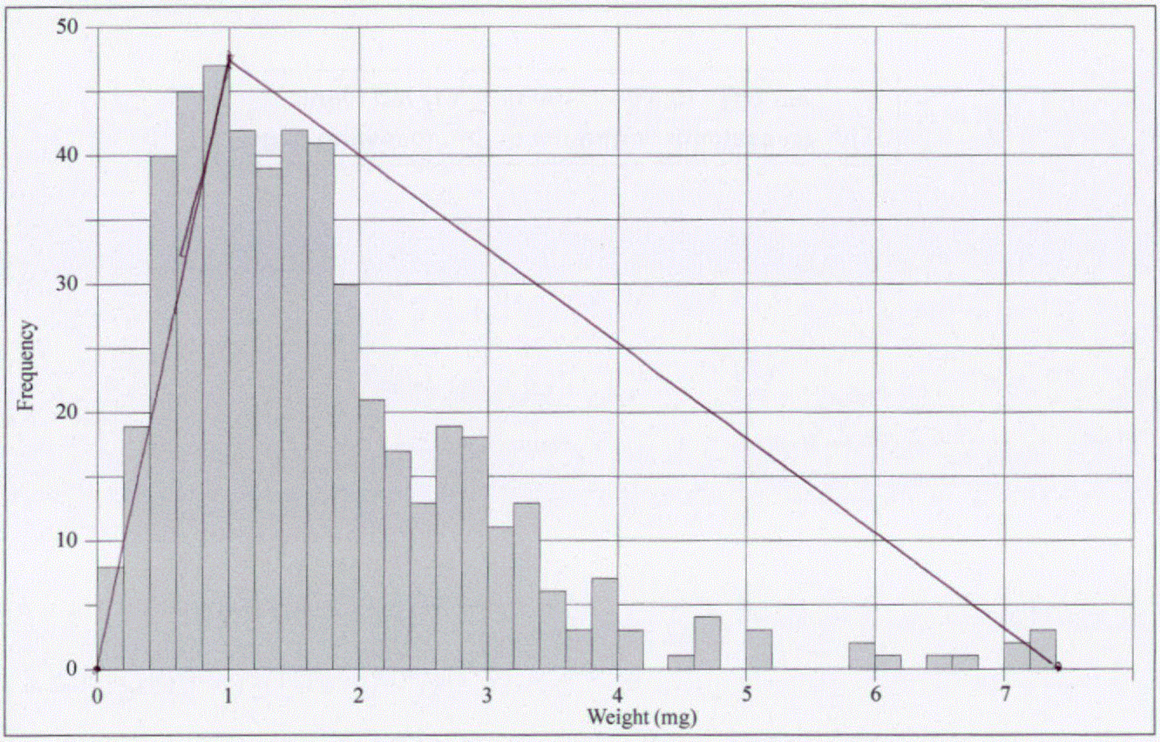

Figure 2(a): Weight distribution of 502 pieces of daily belly button fluff

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.

$f(1.5) = \frac{2(b-x)}{(b-a)(b-c)} = \frac{2(7.4-1.5)}{(7.4-0)(7.4-1)} = 0.2492$

$P(X>1.5) = \frac{1}{2} \times 5.9 \times 0.2492 = 0.7350$

$P(X<1.5) = 1 - 0.7350 = 0.2650$

(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.

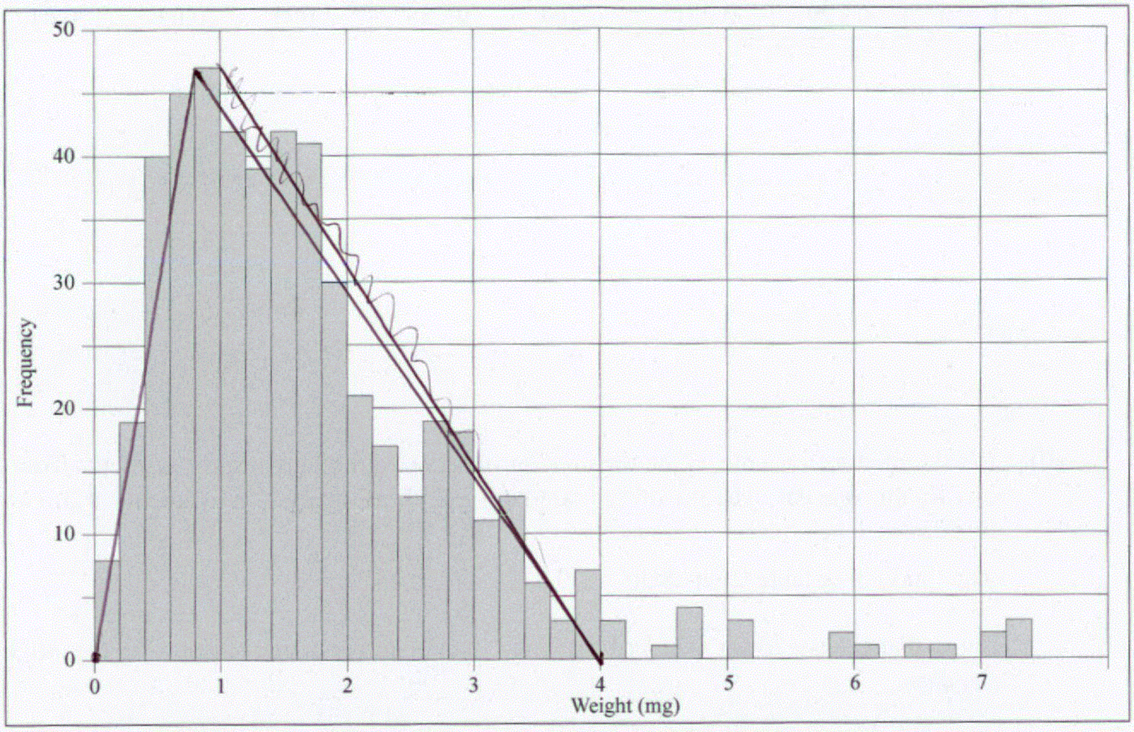

Figure 2(b): Weight distribution of 502 pieces of daily belly button fluff

min = 0mg max = 4mg Mode = 0.8mg
these parameters are a better fit for the majority of the data
the frequency of data between 4mg and 8mg are so small that they can be counted as outliers

(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?

$(0\times0.129)+(1\times0.271)+(2\times0.264)+(3\times0.185)+(4\times0.095)$
$+(5\times0.039)+(6\times0.017) = 2.031$

2 towels on average.

(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.

$SD(T) = \sqrt{(0^2\times0.129)+(1^2\times0.271)+(2^2\times0.264)+(3^2\times0.185)}$
$+(4^2\times0.095)+(5^2\times0.039)+(6^2\times0.017)-2.031^2$

$SD(T) = 1.405$

$0.5 < 1.4$

C probably has a lower standard deviation than T because it has a smaller range as the max num of clothing changes is 3 whereas the max num of towels used is 6.

(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).

if T and C are independent $SD(T+C) = SD(T) + SD(C)$
$1.754 \neq 1.905$ so T and C are not independent. This suggests that residents who ~~change their clothes~~ use more towels are more likely to change their clothes more, which makes sense as they might change clothes after having a shower.

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).

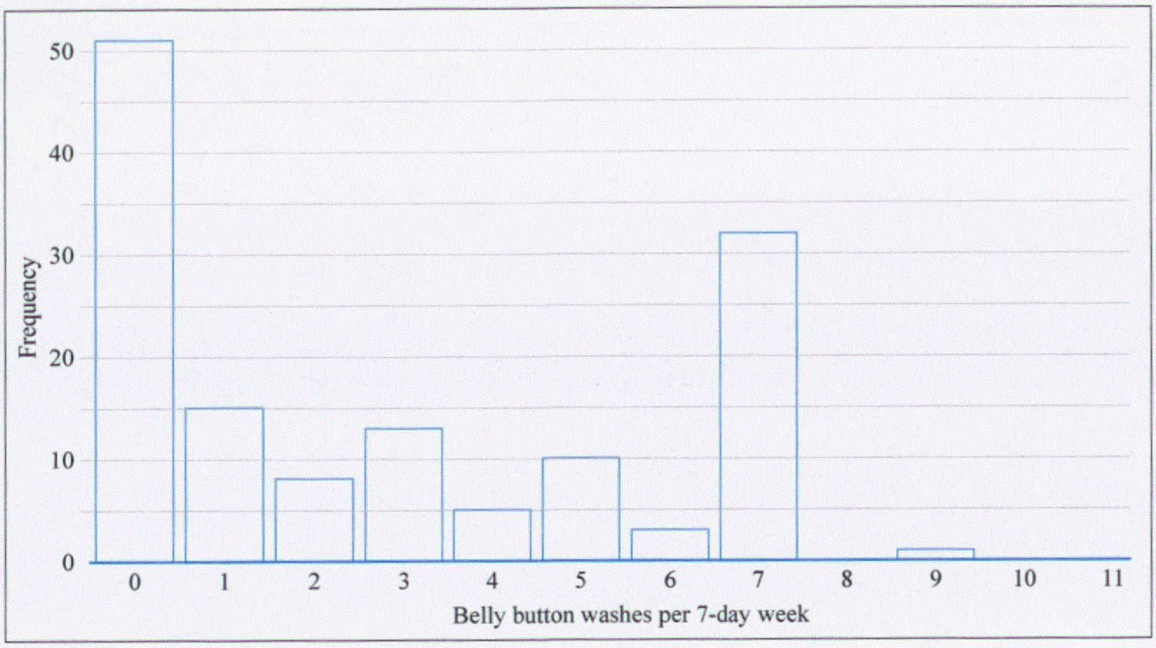

Figure 3(a): The number of reported times survey respondents washed their belly buttons per week

(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.

P(~~≥~~7 washes per week) = $\frac{32+1}{138}$ = 0.2391

0.2391 < 0.25

this headline is correct if the proportion that wash daily in the data is the same nationwide

(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.

$X \sim P(\cancel{2.8}\ 0.4)$ $P(X \geq 1) = 1 - P(X=0)$

$P(X=0) = 0.6703$

$P(X \geq 1) = 1 - 0.6703 = 0.3297$

(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.

probability of washing the belly button may not be independent of washing it again.
someone who washes their belly button at least once a week may be more likely to wash it every day (they are just a clean person)

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).

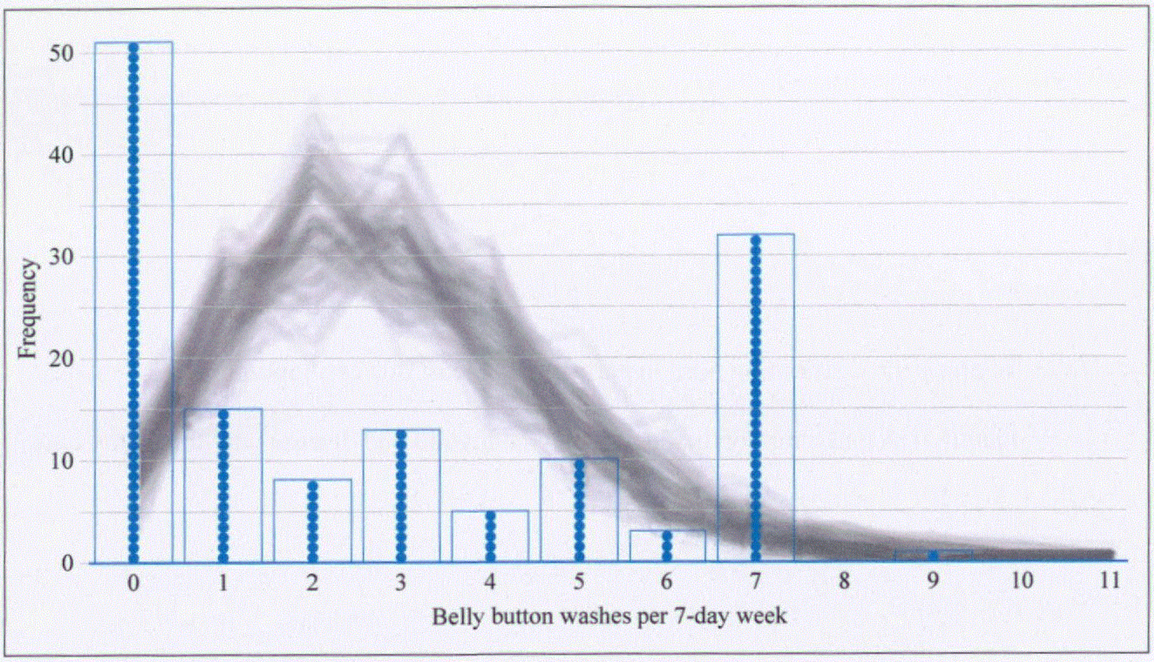

Figure 3(b): Results of the simulation model and original observed data

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

the variation in the simulated frequencies assuming a poisson distribution with $\lambda=2.8$ as ~~the~~ it was run 1000 times

(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.

a poisson distribution model does not appear to be appropriate for modelling the number of belly button washes per 7-day week as the simulation model does not fit the original observed data very well. the simulated model peaks at 2, and ~~then~~ decreases. whereas the original observed data peaks at 0, and then has another peak at 7.

(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.

no probability distribution models fit very well even though the data is discrete, a uniform distribution might be appropriate with a minimum of 0 and a maximum of 7, it is the only model that would still have a relatively high probability of 7 belly button washes per 7-day week.

Excellence

Subject: Statistics

Standard: 91586

Total score: 19

Q	Grade score	Marker commentary
One	E7	(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 correct. This is r since both part (a)(i) and (a)(ii) are correct. (a)(iii) The correct probability that one student has an outie and there is belly button fluff has been calculated correctly. This is t as all three parts (a)(i), (a) (ii). and (a) (iii)) are correct. (b)(i) The correct conditional probability has been calculated for r. (b)(ii) Insufficient detail as to how/ why the given parameters: μ = 2.05 cm 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 may not 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 which is not possible so the model is not 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) A new model given with suitable parameters. The parameters have not been sufficiently 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 0 cm.* To reach r, one of the parameters chosen needs to be justified. 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	M6	(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) the correct probability using lambda = 0.4 has been calculated for r. To reach r, this needs to happen. (b)(ii) The assumption is incorrectly stated. Probabilities are not independent, events are. An assumption about independence or randomness or rate would be appropriate. (b)(iii) This is sufficient for u. To reach r, reference to the sample size is required. (b)(iv) A visual comparison of the observed 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) The uniform model and its parameters has been suggested. The parameters need to be justified for r. To reach t, this model, needs to be clearly linked to the context/ observed data.