Assessment Schedule – 2024

Scholarship Statistics (93201)

Evidence Statement

General Principles:

1 Ignore incorrect answers if alongside correct answers. The exception is contradictory statements.

2 Ignore minor copying errors.

3 When required in evidence, answers need to be contextual.

QUESTION ONE

Task Q1(a)(i)

Evidence:

- In 2019 and 2022 there is a large peak in June / July, around 4% in 2019 and around 3% in 2022.
- For all years where there is data (2021, 2022, and 2023), there are low points in late January or early February, around 0.5%.
- In 2020, 2021, and 2023, there was less variation with a range of less than 1.5%, compared to 2019 and 2022 with a range of over 2.5%.
- Overall, 2019 and 2022 had higher rates of symptoms than other years, peaking around July at 4% and 3% respectively.
- In every year there was a decrease in the rate of reported symptoms between August and October.

Task Q1(a)(ii)

Evidence:

- Sample is not randomly selected, so cannot make a sample-to-population inference.
- Participants are self-selected and choose to fill in the survey, so may not be representative of the population – they may be more health aware than the general adult population; therefore they may give an over-estimate of symptoms.
- Participants may not accurately remember their symptoms from the previous week, and symptoms may be over- or under-reported.
- Symptoms from non-influenza related illnesses may be recorded, so cannot infer that the data indicates influenza.

Task Q1(b)

Evidence:

N.B. the data set for Australia is for 2021–2023, while New Zealand is for 2023.

- For both New Zealand and Australia, the age-group with the highest incidence of fever and cough is under 5 years, and the age group with the lowest incidence of fever and cough is 65+.
- In 2023, under 5 in both countries had a peak in June. New Zealand's incidence was higher at around 12%, while Australia's incidence was around 8%.
- In 2023, the 5–17 age group peaked in June in both countries. New Zealand's incidence was higher at around 7% compared to Australia's incidence at around 4%.
- The 18–64 age group in New Zealand, peaking at around 3% in July 2023, peaks later than the under 5 and 5–17-year groups. In Australia this group has been split into 18–39 and 40–64, both of which peak at around 1.9% in May / June 2023.
- There is a similar overall decreasing incidence of illness for both countries as the age groups become older.

Task Q1(c)(i)

Evidence:

- $279 \times 0.71 = 198.09$ – there were 198 index patients and household contacts who were under 15 that met study criteria.
- $279 + 615 = 894$ – there were 894 index patients and household contacts that met study criteria.
- $\frac{198}{894} \times 131 = 29.01$ – so an estimated 29 children were recorded as secondary cases in this study.

Task Q1(c)(ii)

Evidence:

- Contacts with missing infection status might not be missing at random, e.g., children may be more likely to be identified with influenza as they have caregivers closely monitoring their health. This may increase the proportion of children recorded as secondary cases in this study.
- Influenza vaccine rates may be different for children and adults – children may be more or less likely to become secondary cases compared to adults.
- Infection may be more likely to spread in households with more children, leading to a higher incidence of secondary cases in those households.

Scholarship level	Outstanding level
• Describing trend(s) for time series data in context. • Describing key features of time series data in context. • Describing non-sampling errors in context. • Demonstrating understanding of the need for random sampling when making a sample-to-population inference. • Comparing trends of time series data from two graphs. • Comparing features of time series data from two graphs. **Max 6 marks**	• Discussing potential issues with the study design. • Calculating the proportion of children who were recorded as secondary cases in the study. **Max 2 marks**

QUESTION TWO

Task Q2(a)(i)

Evidence:

- Normal distribution with mean of 3.4–3.8 and standard deviation of 1.4–1.6.
- Normal distribution – the model may not be a good fit as it assumes a symmetrical distribution of data, whereas the observed data is not symmetrical. The model also may not be a good fit because it does not have fixed end points, giving non-zero probability for negative days of symptoms.
- Triangular distribution with parameters: a:0 b:8 c:3 or a:1 b:7 c:3.
- Triangular distribution – the model may not be a good fit because it requires fixed endpoints, which gives zero probability of experiencing symptoms beyond these points.

Task Q2(a)(ii)

Evidence:

- This model overestimates P(Sore throat < 2.361 days), giving P(Sore throat < 2.361 days) = 0.2827
- This model underestimates P(Sore throat > 4.521 days), giving P(Sore throat > 4.521 days) = 0.2048
- The model slightly overestimates P(2.361 days < Sore throat < 4.521 days), giving P(2.361 days < Sore throat < 4.521 days) = 0.5125

N.B. Accept either position as to whether the distribution is suitable, supported with appropriate calculations.

Task Q2(b)(i)

Evidence:

- P(exceeds 2012) = 0.2801 P(exceeds 2022) = 0.0282.
- $\frac{0.2801}{0.0282} = 9.93$ the probability that an employee exceeded their annual sick leave entitlement in 2012 was 10 times the probability that an employee exceeded their annual sick leave entitlement in 2022.

Task Q2(b)(ii)

Evidence:

- We cannot assume independence of events, taking one day off sick may mean an increased probability of taking the next day off sick.
- The probability that an employee uses their sick leave may not be constant, it may change over time, e.g. if an employee is nearing the limits of their entitlement, they may be less likely to take time off; higher likelihood of using sick leave in winter months.

Task Q2(c)

Evidence:

- Difference: 2.79%
- Estimate margin of error: $1.5 \times \frac{\frac{1}{\sqrt{2104}} + \frac{1}{\sqrt{794}}}{2} = 4.3\%$
- CI for the difference of two independent proportions: (–1.5%, 7.1%).
- With approximate 95% confidence, the proportion of New Zealanders who took time off from their regular duties due to an influenza-like illness in July 2023 was between 1.5 percentage points less and 7.1 percentage points more than the proportion of New Zealanders who took time off from their regular duties in December 2023.
- Because the confidence interval contains zero, the claim is not supported.

N.B. Because this was a sample *from one month in summer and one month in winter across just one year, we would not be able to make any claim about New Zealanders' behaviour in summer or winter months.*

Scholarship level	Outstanding level
• Providing suitable parameters for a probability distribution model. • Evaluating a probability distribution model. • Evaluating the goodness-of-fit of a probability distribution model. • Correctly calculating and comparing probabilities from a Poisson distribution model. • Evaluating TWO models and discussing issues with modelling approaches. • Calculating an appropriate estimate for the margin of error. **Max 6 marks**	• Evaluating the goodness of fit of a given model with appropriate probability calculations. • Communicating inference for comparison of two independent proportions using appropriate language. **Max 2 marks**

QUESTION THREE

Task Q3(a)(i)

Evidence:

- The design is a comparison of three independent groups.
- The experimental units were adults with mild asthma who had used a reliever only for the past 12 months.
- The treatment variable was the type of asthma medication: using only a reliever inhaler when symptomatic, using a preventer inhaler twice a day and a reliever inhaler when symptomatic, and using a combined preventer-reliever inhaler when symptomatic.
- The experimental units were randomly allocated into one of the treatment groups.
- The response variable was the rate of asthma attacks. *(Accept risk of asthma attacks.)*

Task Q3(a)(ii)

Evidence:

- Participants were randomly allocated in order to reduce the possible effects of related factors, such as general health, exposure to environmental factors that may trigger asthma symptoms, being more or less likely to take reliever medication when experiencing mild symptoms.
- Random allocation meant researchers may be able to make a causal claim from their data, providing there was a significant enough difference between the groups.

Task Q3(a)(iii)

Evidence:

- To ensure that the usage more closely matched the clinical use of the medications.
- If it were blinded, the advantage of using a single inhaler compared to two different inhalers would have been uncertain.

Task Q3(b)

Evidence:

- Sample mean number of asthma-related healthcare visits for those who live in rented homes without acceptable insulation is 0.42 higher than for those who do live in rented homes with acceptable insulation.
- The variation in the sample number of asthma-related healthcare visits is greater for those who live in rented homes with acceptable insulation than those without acceptable insulation.
- Based on the bootstrap confidence interval, we can be fairly confident that the mean number of asthma-related healthcare visits for those who live in rented homes without acceptable insulation is somewhere between 0.05 visits and 0.78 visits per year higher than for those who live in rented homes with acceptable insulation.
- Because this confidence interval is entirely positive, there is enough evidence to conclude that the mean number of asthma-related healthcare visits for people in Otago, Canterbury, and the West Coast in rental property without acceptable insulation is higher than for those with acceptable insulation.
- Because this sample was only taken from rental properties from Otago, Canterbury, and the West Coast, we cannot apply our findings to non-rental properties or to rental properties outside of these areas.
- This sample only used the number of asthma-related healthcare visits in 2023. It is possible if samples were taken across other years, that results would differ.
- Because the sample data for acceptable insulation is skewed to the right, using the median as the sample statistic may be more appropriate than the use of mean.

Task Q3(c)(i)

Evidence:

- $P(\text{prior asthma diagnosis} \mid \text{wheeze in the past 12 months} \cup \text{currently used asthma medications}) = \dfrac{53}{318}$.

Task Q3(c)(ii)
Evidence:

- The sample proportion of asthmatic Ugandan people aged 12 years or older is 0.0946. This is below 30%, so will result in an overestimate of the size of the margin of error.
- Using the rule of thumb: margin of error $\frac{1}{\sqrt{3416}} = 0.017$ with confidence interval: 0.095 $\pm$ 0.017 = (0.08, 0.11).

  Alternatively, an exact confidence interval could be used using $Z \times \sqrt{\frac{p(1-p)}{n}}$ *as the margin of error.*

- Margin of error $= 1.96 \times \sqrt{\frac{0.0946 \times 0.9054}{3416}} = 0.0098$ generating confidence interval: 0.095 ± 0.0098 = (0.09, 0.10).

Scholarship level	**Outstanding level**
• Identifying and describing key features of an experiment. • Discussing the purpose of random allocation in context. • Comparing relevant features of the sample distributions in context. • Interpreting the confidence interval in context. • Calculating a conditional probability. • Discussing potential issues with sampling strategies. • Discussing how the rule-of-thumb calculation for margin of error will overestimate the confidence interval. *(Or accept a description of the use of the margin of error formula.)* **Max 6 marks**	• Discussing the purpose of an open-label trial in context. • Discussing the suitability of the mean as a sample statistic given the skew of the data distribution. **Max 2 marks**

QUESTION FOUR

Task Q4(a)(i)

Evidence:

- Asthmatics had similar rates of pneumonia to non-asthmatics. Rates of pneumonia were 3.5% higher for asthmatics than non-asthmatics.
- Asthmatics were less likely than non-asthmatics to have severe outcomes with a relative risk of 0.56.

Task Q4(a)(ii)

Evidence:

- Asthmatics were 1.14 times as likely to use antibiotics prior to admission compared to non-asthmatics. There was a 14% higher rate of antibiotic use for asthmatics than non-asthmatics.
- Asthmatics were 1.01 times as likely to use antivirals prior to admission compared to non-asthmatics. There was a 1% higher rate of antiviral use for asthmatics than non-asthmatics.
- Asthmatics were 11.73 times as likely to use corticosteroids prior to admission compared to non-asthmatics. There was a 1173% higher rate of corticosteroid use for asthmatics than non-asthmatics.
- As antibiotic and antiviral use is similar for asthmatics and non-asthmatics prior to admission, but corticosteroid use is almost 12 times as likely, it can be concluded that pre-admission corticosteroid use is the best explanation for the negative association of asthma with severe outcomes.

Task Q4(a)(iii)

Evidence:

- 19% × 1520 = 289 patients had short hospital stays.
- 289 – 71 = 218 patients with short hospital stays were non-asthmatic.
- Relative likelihood of short hospital stays for asthmatics compared to non-asthmatics $\dfrac{\frac{71}{385}}{\frac{218}{1135}} = 0.96$.
- Asthmatics and non-asthmatics were approximately equally likely to have short hospital stays, with asthmatics having only 4% lower probability of short hospital stays compared to non-asthmatics.

Task Q4(a)(iv)

Evidence:

- Let n = number of children

  $$\frac{n}{385} \times \frac{n-1}{384} = 0.0356$$

  $$n = 73.05$$

- There were 73 children in the study.
- Assuming the proportion of children in both groups is the same, there are 288 children in the study.

Task Q4(b)

Evidence:

- There is a moderate, negative then positive non-linear relationship between age in years and the number of colds in previous months.
- For ages 0–60 years as age increases, the number of colds tends to decrease at a decreasing rate. For ages > 60 years as age increases the number of colds tends to increase at an increasing rate.
  (Accept descriptions of linear piecewise relationships.)
- There are two individuals (approximately 3 and 6 years) who had a high number of colds. These unusual values improve the strength of the apparent decreasing trend for < 60 years.
- As there is very little data for individuals in their late 30s / early 40s and 80-plus, using a bivariate model with decreasing, followed by increasing trend, may underestimate the true number of colds for people in the 45–80-year age group, and overestimate the true number of colds for people in the over-80-year age group.

Scholarship level	Outstanding level
• Correctly comparing appropriate rates. • Reasoning with evidence that the use of corticosteroids is associated with the observed difference in severe outcomes for asthmatic patients. • Calculation of correct proportions. • Appropriate comparison of proportions. • Describing the relationship in context. **Max 6 marks**	• Calculating the number of children in the study. • Describing impact of outliers on strength of relationship. **Max 2 marks**

Sufficiency Statement

For each question:

Score 1–4 No award	5–6 Scholarship level	7–8 Outstanding Scholarship level
Shows understanding of relevant statistical and probability concepts and methods, and some progress towards applying this in context.	Application of high-level statistical analysis and critical thinking, knowledge and skills, to complex situations. Shows logical development, precision, and clarity of ideas.	In addition to the requirements of Scholarship, demonstration of perception and insight, sophisticated integration and abstraction of ideas, independent reflection and extrapolation, and convincing communication.

Cut Scores

Scholarship	Outstanding Scholarship
17–25	26–32