Assessment Schedule – 2018
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

Tasks Q1(a)
Evidence:

- Rural –
  - Trend: Linear, non-linear or piecewise
  - Strength: Very weak or weak or moderate
  - Direction: positive or changing direction.

  Eg for direction and piecewise trend: Initial increase in sales for advertising spend up to $400 to $600, then no further increase in sales for additional advertising spend.

  Urban – linear, strong and, positive correlation between advertising and sales for most data but with outlier at (12, 0.5).

Task Q1(b)
Evidence:

- For rural, predicted sales = 3.808 (000$) $3800 [Using S = 0.0825A + 3.3955 Due to small sample, considerable variation and no reason to expect a decrease in the context, the simpler linear model may be appropriate.]

  OR 4.141 (000$) $4140 [Using S = $-0.0356A^2 + 0.4783A + 2.6392$ Due to the fact that the trend is non-linear, a non-linear model might be more appropriate even though the parabola will not be a good fit for higher advertising spends].
- For urban, predicted sales = 4.51 (000$) $4500. [Using S = 0.1263A + 3.8788. Since the sample is small and we do not know the reason for the outlier, it is may not be valid to remove it in this context. When considering return on investment, information about losses shouldn't be ignored]

  OR 4.57 (000$) $4600 [Using S = 0.29A + 3.12 Since the outlier is so far from the other points, it will have a large effect on the trendline and it might be better to remove it].

Task Q1(c)
Evidence:

- For rural there is a change in the trend at about $500, which means that it is difficult to model reliably for $500.
- For rural, there is considerable variation in the scatter about the trend so the prediction is suspect. *Accept weak correlation for rural.*
- For urban (if outlier included) the effect of the outlier has been included in the model so the prediction may not be valid if the outlier turns out not to be valid data.
- For urban (if outlier excluded) the effect of the outlier has not been included in the model so the prediction may not be valid if the outlier turns out to be valid data.
- The small sample size is an issue for rural because of the variability in the data since variability may be due to sampling variation rather than representative of the underlying relationship. *(no mark for small sample size without explanation in terms of variability and sampling variation).*
- The small sample size is an issue for urban as the outlier will have a big effect on the trend for a small sample. We do not know whether the outlier represents an error or a real problem which means that sometimes the advertising spend does not result in sales. *(no credit for small sample size without explanation in terms of effect of outlier).*
- If discussing parabolic model, it does not make sense in terms of the context that an increase in advertising would result in a decrease in sales.

Task Q1(d)
Evidence:

- Number of competitors – expect a negative correlation.
- Score location from 1 = least favourable to 7 = most favourable – expect a positive correlation.
- Income of the population or economic changes, positive correlation as more sales when times or income are better, fewer sales when times/income are worse.
- Number of customers or size of store or number/total salary of staff positively correlated with sales.

  Note: *Accept other reasonable numeric variables with comment on direction of correlation.*

Task Q1(e)
Evidence:

- With the strong relationship between advertising spend and sales for most data from the urban outlets and $1,400 being not too far from the given data it, is possible further sales could be expected, though the outlier indicates that the prediction may not be reliable. Other factors like location and demand come into play. The claim of no increase is not supported.

  There is unlikely to be any further increase in sales in the rural sector since there is no increase in the sales for spends over $600 in the scatterplot. Claim of no increase would be supported in the rural sector.

  Note: *Sufficient to discuss claim for one with discussion of forecast for both. Discussion must be based on observation of the graph, not on the model.*

QUESTION TWO

Task Q2(a)
Evidence:
The probability of the difference in the mean scores of V2 – V1 of at least 9.92 occurring by chance is only 0.006. So V2 is significantly preferred, as 0.006 < 0.05. *Accept "strong evidence" or "highly unlikely" or equivalent instead of <0.05.*

Task Q2(b)
Evidence
The groups need to be as similar as possible with respect to their likes and dislikes regarding cheese. *Other answers possible but must relate* ***in context to effect on this experiment*** *(e.g. generous people who tend to give everything high scores would tend to be evenly distributed between the two groups).*

Task Q2(c)
Evidence:
The mean score for that particular characteristic over the other 24 scores for that group could be used or the mean score for that person over the other 9 characteristics, or a regression of known values for other characteristic(s) for this group is used to predict the missing value.

Task Q2(d)
Evidence:
Use a multiple bar chart with scores on the vertical axis and separate bars; one for V1 and the other for V2, side-by-side for each characteristic (*title not needed). Allow correct graph or description of graph missing some detail, or for other sensible graph type that displays the information but less clearly (e.g. bar chart of differences).*

Task Q2(e)
Evidence:
Use a time series by assessing the taste of V1 out of nine at regular time intervals.
Other possible answers will be accepted but must involve measures at more than two times, e.g. repeated randomisations test between cheeses of different ages, including an indication of how the results will be compared/analysed.

Task Q2(f)
Evidence:
Calculate the **differences in mean/medians** between the taste scores of V1 and V2. Record the differences in means at regular time intervals. *Evidence may be in Q2(e). Apply consistency with 2e. Accept a more complex comparison which enables differences to be compared.*

QUESTION THREE

Task Q3(a)(i)

Evidence:

- The sentiment score is bounded by 0 and 1, while the normal distribution is unbounded.
- Considering two standard deviations either side of the mean would result in an expected central 95% for sentiment scores of (–0.036,1.112), which include scores below 0 and above 1 *(or other probability calculation with interpretation*

  e.g. P(X<0)=0.0304) and P(X>1)= 0.0537, so over 8% of the model is outside the observed distribution, which is a big proportion to exclude

  OR P(X<0.8)= 0.8194, which is 7 percentage points higher than 0.749 observed, which is a large % difference.

Task Q3(a)(ii).

Evidence: *(Must include some comparison of theoretical vs observed probabilities or comparison of graph to Poisson distribution – could be overall evaluation of model).*

- Sample distribution is positively skewed (similar shape to Poisson distribution model for small λ)
- Calculate the mean number of letters per word as $(1 \times 0.05 + 2 \times 0.07 \ldots\ldots 12 \times 0.0025) = 4.9975$ *(CAO)*
- Use the mean of 4 as an estimate of the Poisson parameter λ (since we are modelling $N - 1$) to calculate a probability.
- Create a correct table of expected probabilities (at least 8 out of 12 correct).

E.g. calculating table or probabilities of events

Number of letters (N)	1	2	3	4	5	6	7	8	9	10	11	12
N – 1	0	1	2	3	4	5	6	7	8	9	10	11
Observed proportion	0.05	0.0875	0.145	0.1825	0.1325	0.1375	0.13	0.055	0.045	0.0225	0.01	0.0025
Expected proportion	0.0183	0.0733	0.1465	0.1954	0.1954	0.1563	0.1042	0.0595	0.0298	0.0132	0.0053	0.0019

- Specific comparison observed vs theoretical: e.g. Under a Poisson distribution with $\lambda = 4$, $\Pr(X = 3) = 0.1954$ (using tables) and the observed proportion is 0.18 which is similar. e.g. Similarly, we would expect 0.156 of words to have 6 letters but the observed is 0.14 which is 12% less than expected so fairly different.
- General comparison of the observed proportions with the expected proportions: e.g. (under the proposed model), we get a close fit for most numbers of letters. This would suggest the Poisson distribution as being a reasonable model
- OR (under the proposed model), we get do not observe a close fit for most numbers of letters. This would suggest the Poisson distribution is not an appropriate model or justified by difference in probabilities for X = 5 or discussion of X<3 or X>7.
- Discussion of conditions for Poisson e.g. the rate is number of letters in 1 word, so probability of occurring is proportional to the number of words being measured, letters do not occupy the same space, so are not simultaneous, the occurrence of each letter after the first is unpredictable, so may be considered random. While letters are not independent, the word length is unpredictable, so letters may be considered to occur independently.

Task Q3(b)

Evidence:

1. The mean reading speed for the Website A sample was 11.62 words per minute higher than for the Website B sample.
2. The bootstrap distribution shows that the mean reading speed for articles from Website A is likely to be between 2.99 and 19.40 words per minute higher than that for Website B.
3. Zero is not included in interval; there is evidence to establish that Website A has a higher mean reading speed than the Website B.
4. Therefore, can conclude that the mean reading speed for website A is significantly different from the mean reading speed for website B.

Note: Mean reading speed must be mentioned in at least one place. Only accept points 3 or 4 if point 2 is correct. The full context may be considered holistically over the student's answer.

QUESTION FOUR

Tasks Q4(a)(i)

Evidence:

- The women who test positive for pregnancy will include those who are not pregnant (false positives).
- Calculation with interpretation (may be algebraic) For example, if overall 10% of women who take this test are pregnant,

then you would expect $0.1 \times 0.98 = 0.098$ true positives, and $0.9 \times 0.04 = 0.036$ false positives, and the proportion of women who are pregnant given a positive test result.

$= \dfrac{0.098}{0.098 + 0.036} = 0.731$, which is not 98%.

- The claims made by the developer would be based on the results from a study (or several studies). Therefore, the claims are based on estimates for the true specificity (true positive rate) and true sensitivity (true negative rate) of the pregnancy test. The estimates may differ from the true positive and negative rates because of sampling variation (or randomness or chance).

Task Q4(a)(ii)

Evidence:

- Pr(pregnant | positive test) = $\dfrac{44}{55}$
- Pr(positive test | pregnant) = 0.94
- Pr(negative test | not pregnant) = 0.81
- Number of pregnant women in study = $\dfrac{44}{0.94} = 47$
- Number of not pregnant women in study = $\dfrac{11}{0.19} = 58$
- Therefore, 47/105 = 0.4476 of the women in this study were pregnant.

Note:

Alternate methods are acceptable e.g. two-way table.

Task Q4(b)(i)

Evidence:

- People were surveyed over a period of four months to reduce bias since the views are not overly affected by a particular event in the short term.
- Stratified sampling was used to assist representativeness of sample for different regions in the population.
- Random sampling of telephone numbers to make it likely that the sample is representative of the population.
- CATI used to help with consistency of interviews *(do not accept if response implies that questions were asked by or answered on a computer)*
- Use of multi-choice on a scale to ensure consistency of answers and ability to be analysed.
- Sample size was sufficiently large to reduce variability in responses due to chance/randomness/sampling variation.
- Do not accept sample weighting as a strength since it is explained in the article.

Note: Any two strengths accepted.

Task Q4(b)(ii)

Evidence: *(must describe issue and explain how it can cause bias in this context)*

- Use of telephone directory (white pages) excludes those who do not have landlines or who have made their phone number private (potential selection bias). People with landlines may be richer or more conservative and have different attitudes to alcohol consumption than people without.
- The nature of the questions asked (drinking during pregnancy) could lead respondents to not be honest with their responses (potential behavioural considerations).
- With the nature of the questions asked (drinking during pregnancy), the gender of the interviewer could influence respondent's answers, for example a female interviewer asking the question could result in more or less honest answers than a male interviewer (interviewer effect).
- Telephone surveys tend to have high non-response, and those who are home to answer the phone may have a different lifestyle than those not home (may have different attitudes to alcohol consumption).
- Question effects such as not defining what "small amounts of alcohol" means, so non-drinkers might interpret it differently from heavy drinkers (do not accept criticism of standard questionnaire responses such as "agree" and "strongly agree").
- The time period of the survey included the Christmas season and the summer holidays when many people drink more alcohol, which might have affected their attitude to drinking compared to other times of year.

Note: *Any two points above or other reasonable non-sampling errors are acceptable. Must relate specifically to this survey, not be general points about possible leading questions etc.*

Task Q4 (b)(iii)

Evidence:

The percentage of respondents who disagreed with the statement ‘During pregnancy drinking small amounts of alcohol is OK’ may have varied between the three survey years (i.e. they may not all have been 84%), but the differences between these percentages could be due just to chance / sampling variability.

Task Q4(b)(iv)

Evidence:

- The group size is smaller for “Other” than for “Degree”, which means the margin of error will be larger for the “Other” group.
- The survey percentage (85%) is closer to 50% for the “Other” group, which means the margin of error will be larger for the “Other” group.

QUESTION FIVE

Task Q5(a)(i)
Evidence:
- The unemployment rate for females was higher than that for males (except in the period June 2009 to June 2010).
- The unemployment rates peaked between 6.5% and 6.8%, for males in June 2010 and females in Sept 2012.
- Between the Dec 2009 or Sept 2012 quarters and Sept 2017 quarters the unemployment rates trended downwards for both males and females, (with males decreasing by 2.5% and females decreasing by 1.6%).

Note:
1. Graph shows seasonally adjusted data so no comments accepted about seasonal effects.
2. Must include correct time periods and at least one % from the graph.

Task Q5(a)(ii)
Evidence:
- Downward trend in unemployment from 2012: economic conditions improve causing more employment on offer.
- Fluctuations in employment / changes in government policy due to elections / impact of world markets.
- Women are often responsible for children so may have more difficulty finding work that fits around childcare.

Note:
1. Graph shows seasonally adjusted data so no comments accepted about seasonal effects.
2. Accept other reasonable explanations for a graph feature.

Task Q5(b)(i)
Evidence:
Choose Model B as it is more appropriate to the recent behaviour consisting of a downward trend in the number unemployed.

Qtr. since Dec 2007	Trend Estimate	Actual	Estimate of Seasonal
17	145.7	156	10.3
21	141.9	141	-0.9
25	138.2	141	2.8
29	134.4	143	8.6
33	130.7	140	9.3
37	127.0	139	12.0
			42.1

Estimate of seasonal effect for Quarter 1 = 42.1/6 = 7.0
Forecast = $-0.9351 \times 45 + 161.56 + 7.0 = 126.5$(000), i.e. 126 500 unemployed.

Task Q5(b)(ii)
Evidence:
- There is a high variation in seasonal values with inconsistent peaks and troughs from one year to the next.
- Conditions may change in the future as they changed in 2008.

Note: *A variety of answers accepted, but must be based on observation of the graph.*

Task Q5(c)
Evidence:
- The Holt-Winters model uses a smoothing constant technique where more emphasis is placed on recent observations. Also a 95% interval estimate is provided for the forecast between the 2.5% and 97.5% tails, which gets wider (fan) as we go further into the future.
- Comparing 126 300 with 126 500, they are different by 200 due to more weight being placed on recent behaviour by the Holt-Winters model. The method employed in (b) (i) used data equally from the last six first-quarters.

Task Q5 (d)
Evidence:
Take 2014 = 100 to get:

Year	Male	Female	Male % change from one year to next	Female % change from one year to next
2014	100.0	100.0		
2015	104.2	99.2	+ 4.2%	–0.8%
2016	101.1	90.5	–3.0%	–8.7%
2017	91.7	86.7	–9.3%	–4.2%
Overall change			–8.3%	–13.3%
Average change			–2.7%	–4.6%

Claim is justified, as the index number for females is lower than that of males for all three years 2015 to 2017 inclusive.

Notes:
1. Accept comparing 2014 to 2017 only for both.
2. Other methods are acceptable, eg percentage change, compared or averaged.

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
21 – 33	34 – 40