Assessment Schedule – 2020
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:

- Relationship: Linear trend, positive direction, moderate / weak strength, two outliers in terms of trend, one outlier in terms of x, generally more scatter / variation as the number of pages increases.
- Notable features: Two suspect points e.g. one extreme outlier indicates a book with close to 2000 pages but only just over 150 000 words (average 75 words per page), which seems implausible for a popular fiction book. Similarly, a book with around 300 pages, but with around 420 000 words (average 400 words per page).

Note: Accept discussion of possible groupings.

Task Q1(a)(ii)
Evidence:

- Books can be printed in different sizes, so can fit more or fewer words per page.
- Text can be printed in different font sizes, so can fit more or fewer words per page.
- Books can have more or fewer chapters, which typically start on a new page, with the same number of words.
- Illustrations, if used, have few words but can take up whole pages.

Note: Accept other reasonable examples.

Task Q1(a)(iii)
Evidence:

- The slope of the fitted line is 0.2202, which represents an average increase of 220.2 words for *each additional page*, which is lower than the stated average of 250 words per page. However, the fitted line has a non-zero y-intercept (57 840 words). The slope of the line fitted, however, will be influenced / lowered by the outlier at approximately (2000,150).
- The linear model has been fitted to a sample of 65 fiction books, and so the slope of the fitted line is just an estimate of the average increase in words for each additional page in a book. Additionally, this is a sample of 65 of the most popular fiction books written in English, which could also explain why the relationship between number of words and number of pages is different from what has been stated for all fiction books.

Note: Evidence towards the standard can be considered from across part (a).

Task Q1(b)(i)
Evidence:

Two possible claims to evaluate.

Claim 1: The percentage of NZ adults who had read or started to read at least one book in the past year was lower in 2018 compared to 2017, but this difference may not appear to be statistically significant.

- 86% 2018 (n = 2261), lower than 88% 2017 (n = 2082)
- Difference between survey percentages: 88% – 86% = 2%
- Estimated margin of error for difference: $1.5 \times \frac{\frac{1}{\sqrt{2261}} + \frac{1}{\sqrt{2082}}}{2} = 3.2\%$
- Since both survey percentages are well above 70%, the margin of error for the difference will be smaller.
- If the margin of error was less than 2%, then it could be claimed that percentage of NZ adults who had read or started to read at least one book in the past year was lower in 2018 compared to 2017.

Claim 2: Males made up most of NZ adults who did not read a book in the past year.

- 69% stated in report, but this is out of the respondents who did not read a book in the past year.
- 14% of 2261 = around 317 respondents, so the estimated margin of error is $\frac{1}{\sqrt{317}} = 5.6\%$
- 69% is 19 percentage points above 50%, 69% – 5.6% = 63.4%
- It can be claimed that males made up most of NZ adults who did not read a book in the past year.

Note: Accept the construction and interpretation of confidence intervals.

Task Q1(b)(ii)
Evidence:

- Respondents may state they have read more books than they actually have as reading books is considered to be a socially desirable trait.
- Respondents may not be able to accurately recall how many books they read over an entire year.
- Respondents may have different interpretations of how much of a book needs to be read, or how it was read (e.g. skim read), to have "read it".

Note: Accept other descriptions of relevant non-sampling errors with respect to being asked how many books were read.

QUESTION TWO

Task Q2(a)(i)
Evidence:

- "Most" means over half, so the median age would need to be at least 2020 – 1924 = 96 years.
- With approx. 95% confidence, the median age of all eBooks from Project Gutenberg is somewhere between 108 to 143 years.
- Since all of the plausible values for the median age are higher than 96 years, the claim can be supported.

Task Q2(a)(ii)
Evidence:

- With approx. 95% confidence, the mean age of all audiobooks from LibriVox is somewhere between 14.1 years lower and 22.8 years higher than the mean age of all eBooks from Project Gutenberg.
- Because the plausible values for the difference are both positive (LibriVox audiobooks with a higher mean age than Project Gutenberg eBooks) and negative (Project Gutenberg eBooks with a higher mean age than LibriVox audiobooks), we cannot claim that one library has older books, on average, than the other.
- The sample distributions of book ages are similar for the Project Gutenberg and LibriVox books: positively skewed, with the mean age of books in the sample from LibriVox only 4.6 years higher than the mean age of books in the sample from Project Gutenberg.

Task Q2(a)(iii)
Evidence:

- Model this situation using a Binomial distribution with $n = 24$, $p = 0.5$.
- Binomial distribution appropriate since we have a fixed number of trials (24 books), a fixed probability (assuming 50%), two outcomes (English or not English) and can assume independence (random selection used).
- Five of the 24 books were not written in English.
- $P(X \leq 5) = 0.0033$
- The probability of randomly selecting 5 or fewer books not written in English out of 24, if 50% of all books available from Project Gutenberg are not written in English, is small.
- Yes, we can conclude that more than half of the books available from Project Gutenberg are written in English.

Task Q2(b)
Evidence:

- Problem: What is the difference between the mean time morning visitors spend at the library and the mean time afternoon visitors spend at the library?
- Plan: Use some form of random selection to determine which visitors to observe – for example, select 100 different hours to observe across two weeks and morning and afternoon sessions. Record the time each visitor arrived and the time each visitor left the library during those hours.
- Data: Calculate the time spent in the library using the arrival and leaving times.
- Analysis: Use the data collected to construct a bootstrap confidence interval for the difference between mean time spent at the library for morning and afternoon visitors.
- Conclusion: Use the confidence interval to communicate an interval estimate for how much longer, on average, morning visitors spend at the library compared to afternoon visitors.

Note: Accept other problem statements that involve a comparison of the mean or median time visitors spend at the library, between morning and afternoon visitors.

QUESTION THREE

Task Q3(a)
Evidence:

- Similarities: For both fiction and non-fiction genres, most of the reading takes place between 9am and 2am or 3am each day, with 8pm to 10pm (20 to 22) a popular time to read for both genres.
- Differences: A higher percentage of the non-fiction genre reading takes place between 9am and 1pm (13) than the percentage of fiction genre reading during this same time period. For example, at 11am, non-fiction reading is approximately 1.2 times as likely as fiction reading.

Note: Accept other valid comparisons and numeric comparisons expressed in other forms e.g. as a ratio.

Task Q3(b)
Evidence:

- Normal distribution, $\mu = 150$, $\sigma = ?$
- $P(120 < X < 150) = \frac{3310}{8000} = 0.4138$
- Using the proportion from the study as an estimate for the Normal distribution probability, σ can be estimated to be 22, (e.g. using $z = \frac{x - \mu}{\sigma}$).
- Using a Normal distribution with $\mu = 150$, $\sigma = 22$, the middle 95% is (106.9, 193.1).
- It can be estimated that 95% of the reading speeds for people in the study were between 107 and 193 words per minute.

Note: Do not penalise the use of, nor expect the use of, continuity corrections for modelling this situation.

Task Q3(c)(i)
Evidence:

- Triangular distribution with $a = 0$, $b = 100$, $c = 68$
- Using the model, $P(X < 64) = 0.6024$ compared to the proportion from the study, which was 0.5 (about half).
- Using the model, $P(X > 90) = 0.0313$ compared to the proportion from the study, which was 0.05 (about 5%).
- Given that the study uses several thousands of books, the difference between the model probability of 0.6024 and the proportion from the study of 0.5 is too large to consider the proposed triangular distribution model a good model for the mean percentage completion values for books available from the eBook subscription company.

Task Q3(c)(ii)
Evidence:

- What type of books are being read, for example fiction vs non-fiction books. Different types of books may be associated with lower or higher mean percentage completion rates.
- What time of the year the book was first accessed from the company. Books that people start reading during the Christmas holidays, for example, might have higher mean percentage completion rates than those started during normal work weeks.
- How long the book is. Books that are very long may be associated with lower mean percentage completion rates.

Note: Accept other reasonable factors that discuss the potential association with the mean percentage completion rates.

QUESTION FOUR

Task Q4(a)
Evidence:

- Random allocation used to allocate treatments to units / participants.
- Treatment groups: Text in serif font, text in sans-serif font.
- Overall design: Comparison of two independent groups.
- Units / participants: 238 university medical students.
- Response variable: The number of words read from the supplied text in one minute.

Task Q4(b)
Evidence:

- The tail proportion is large (0.4), which does not give evidence to support a claim that serif fonts are easier to read than sans-serif fonts.
- Looking at the data collected, the results look similar both in terms of centre and spread (the difference between the group means is 1.75 words per minute).
- As the text to read was only 288 words and the response variable was the number of words read in one minute, we could expect to see similar values for the response variable irrespective of the applied treatment. You can also see that a large number of students read all 288 words during the one minute time allocated. We could expect to see a wider variation of reading speeds within groups compared to a smaller variation between groups.

Task Q4(c)(i)
Evidence:

- Nearly 80% of the medical students were female students, and nearly 60% of the medical students were specialising in therapy. Of the medical students specialising in therapy, around 23% were male, whereas of the medical students specialising in paediatrics, around 18% were male.
- The medical students specialising in paediatrics tended to be older and more similar in age, with a median age of 22 years and an IQR of 1 year, compared to the medical students specialising in therapy who had a median of 21 years and an IQR of 3 years. The age distributions for both specialisations are positively skewed.

Task Q4(c)(ii)
Evidence:

- There is an imbalance within the student participants in terms of gender and specialisation. There is also a wide age range for students, from 20 years to 40 years.
- The design of the experiment could be modified to include a blocking variable. For example, half of the females could be randomly allocated to the serif treatment group and half to the sans-serif treatment group, and then half of the males could be randomly allocated to the serif treatment group and half to the sans-serif treatment group. This would ensure a balance of gender within the two treatment groups.

Note: Stratification and sampling ideas are not relevant for design of experiment.

Task Q4(d)
Evidence:

- The experiment was conducted using only one language – Russian. Further studies would be needed to investigate the possible effect of type of font (serif vs sans-serif) for other languages, as the characters used are different (Russian uses a different alphabet).
- The experiment was conducted with medical students, who have to read and memorise a lot of information from written text during their studies. They may have developed techniques for reading text that mitigate any effect from the type of font used (e.g. serif vs sans-serif).
- The participants were all fluent speakers of Russian and had normal or corrected vision. The participants' ability to read text quickly will be different from those people who are still learning how to read Russian or have difficulty reading text due to vision problems, and we don't know how the type of font used (serif vs sans-serif) might affect these readers.
- The topic of the text would be familiar and of interest to the participants, and this could mitigate any effect from the type of font used.

Note: Accept other valid discussion points that are linked to specific information provided in the question about the experiment.

QUESTION FIVE

Task Q5(a)(i)
Evidence:

- The mean number of new articles published in the English language on Wikipedia per month rose during 2001 to 2007, where it reached a mean of around 50 000 new articles published per month, and then declined at a slower rate from 2007 to 2018.
- There was an unusual spike in the number of new articles published just before the end of 2003, of nearly 40 000 new articles in just one month. There was more variation during the years 2006 to 2009.
- There does not appear to be a strong seasonal pattern for the number of new articles published. The plot showing the seasonal differences shows averages (means) for each month close to zero, with large variation within individual seasonal differences for the same month (grey lines). There is a small seasonal increase in the months of July and August, perhaps corresponding to the Northern Hemisphere summer.

Note: At least one of the points must contain numerical evidence.

Task Q5(a)(ii)
Evidence:

- The model has produced a very wide forecast interval for November 2020 (considerably wider than the range of the raw data from 2011 to 2018) that includes negative values. The interval would have been narrower if only the data from 2010 onwards were used for the forecast.
- Due to the lack of strong seasonality and the 'unexplained' spikes and troughs, there is only a weak fit between the raw data and the fitted additive Holt-Winters model.

Task Q5(b)(i)
Evidence:

- The height of each bar represents how far "up" the list the book was for that week, the highest height represents number one on the list.
- The colour of the bar also represents how far "up" the list the book was for that week, the darker the blue colour, the higher up the list.
- Weeks when the book was number one on the list are represented by yellow bars.
- The titles have been arranged in order of when they first appeared on the best-sellers list.
- Icons have been used to represent if the book has been made into a TV series or movie.

Task Q5(b)(ii)
Evidence:

- Most of these 10 books reached number one on the list, but four did not. Of the six books that reached number one during 2011 to 2019, five of them reached number one more than once.
- *A Game Of Thrones* had the longest time span (over eight years) between first appearance and last appearance, presumably because of the television series that ran for many years during this time, and *A Man Called Ove* has the shortest time span.
- *Gone Girl* was on the best-seller list almost continuously for nearly three years, and *The Girl on the Train* seems to have the longest uninterrupted run on the list of over two years.
- Most of these books have been made into movies, and most of these books have had multiple "peaks" in terms of their ranking on the lists, presumably because of movie releases.
- All of these books stayed on the best seller list for an extended period of time (more than one year).
- The three "Fifty Shades" books had very similar patterns of sales, both initially in 2012 and then returning to the list in late 2013 and early 2015.

Accept other valid points.

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
23	31