S

93201Q

932012

NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

QUALIFY FOR THE FUTURE WORLD
KIA NOHO TAKATŪ KI TŌ ĀMUA AO!

Scholarship 2019
Statistics

2.00 p.m. Wednesday 13 November 2019
Time allowed: Three hours
Total score: 40

QUESTION BOOKLET

There are FIVE questions in this booklet. Answer ALL questions.

Pull out Formulae and Tables Booklet S–STATF from the centre of this booklet.

Write your answers in Answer Booklet 93201A.

Show ALL working. Start your answer to each question on a new page. Carefully number each question.

Check that this booklet has pages 2–15 in the correct order and that none of these pages is blank.

YOU MAY KEEP THIS BOOKLET AT THE END OF THE EXAMINATION.

QUESTION ONE

Data was obtained about buses registered with the New Zealand Transport Agency (NZTA) as at the end of 2018. A sample of 500 buses was initially used to carry out exploratory data analysis.

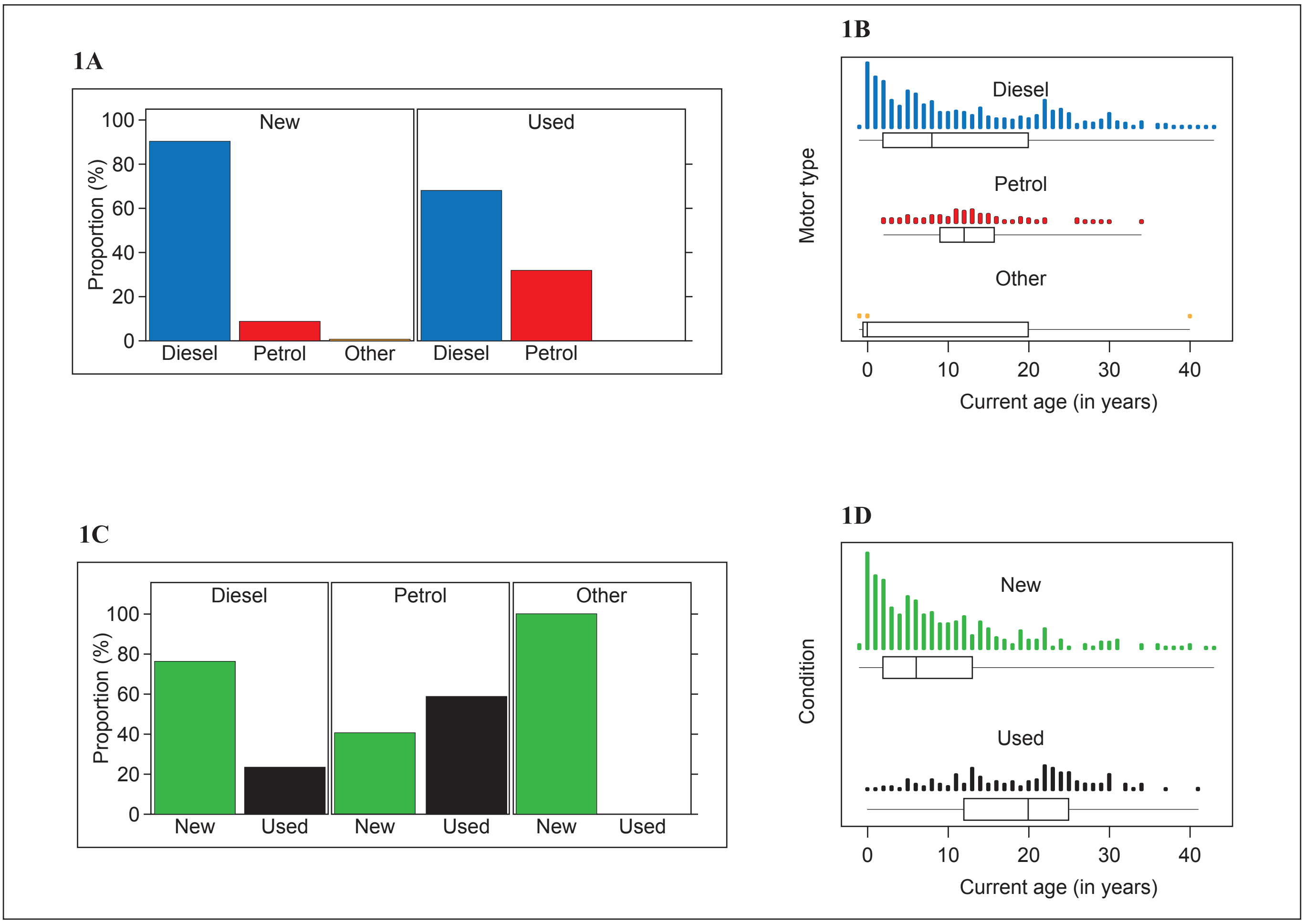
(a) Figure 1 shows four plots produced from the sample.

The following variables were developed and used for these plots:

- current age (difference, in years, between 2018 and the year the bus was built)
- motor type
  - diesel
  - petrol
  - other
- condition (status of the bus when it was first registered with NZTA)
  - new
  - used.

Using each of the plots shown in Figure 1, give FOUR different comparisons between the various bus types.

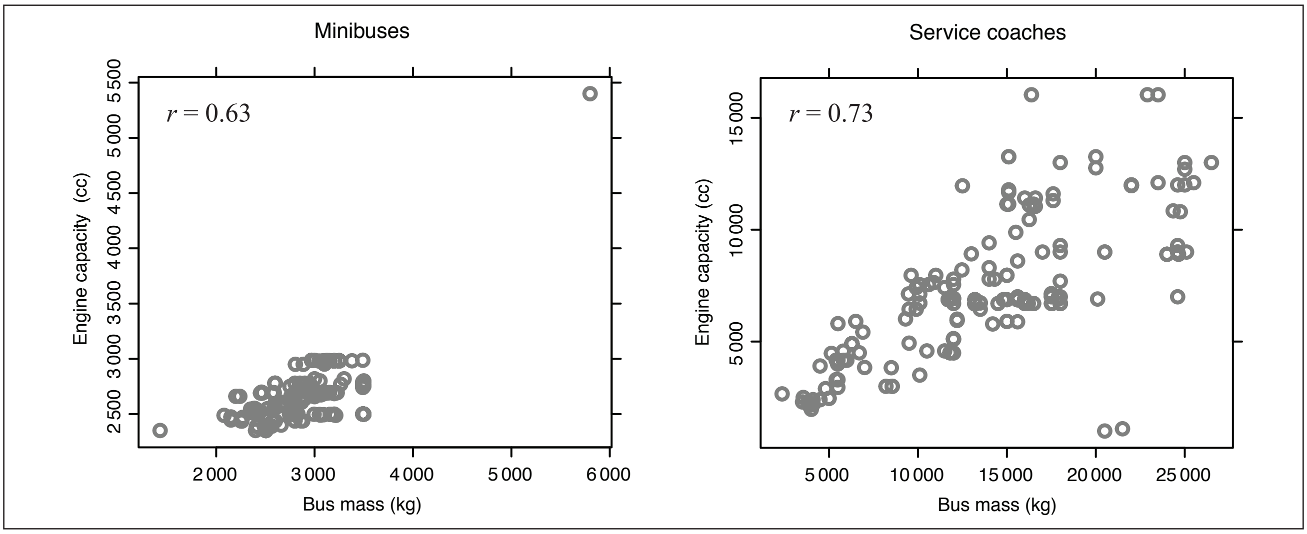
Figure 1: Plots produced during exploratory data analysis



(b) Buses can be classified by their body type as either minibuses or as service coaches.

Figure 2 shows two scatter plots produced for the bus mass (in kg) and the engine capacity (in cc) for buses with engine capacities of at least 1000 cc, for minibuses and service coaches respectively. The correlation coefficient, r, has been provided in each case.

Figure 2: Scatterplots of engine capacity for minibuses and service coaches



(i) Describe the relationship between engine capacity and bus mass, for service coaches.

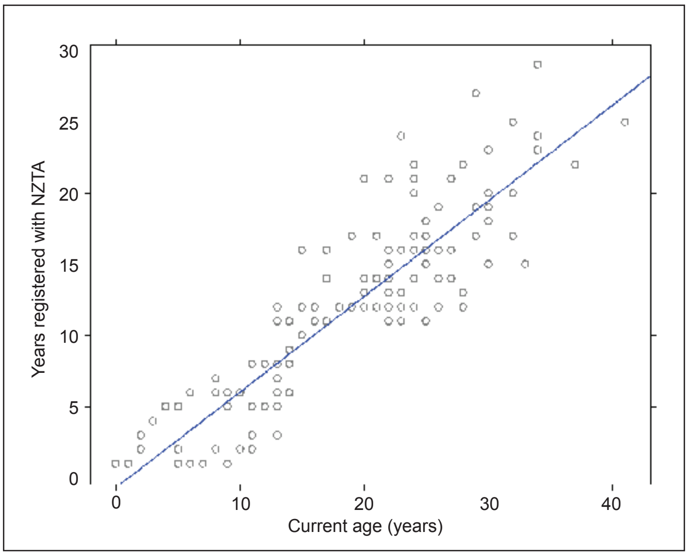
(ii) Suppose the minibus with a mass of 5800 kg and an engine capacity of 5400 cc was removed from the data to investigate its influence on the value of the correlation coefficient for minibuses.

Describe any effect this would have on the correlation coefficient.

(c) The relationship between the current age of a bus and the number of years the bus has been registered with NZTA was explored for buses that were classified as "used" when first registered with NZTA.

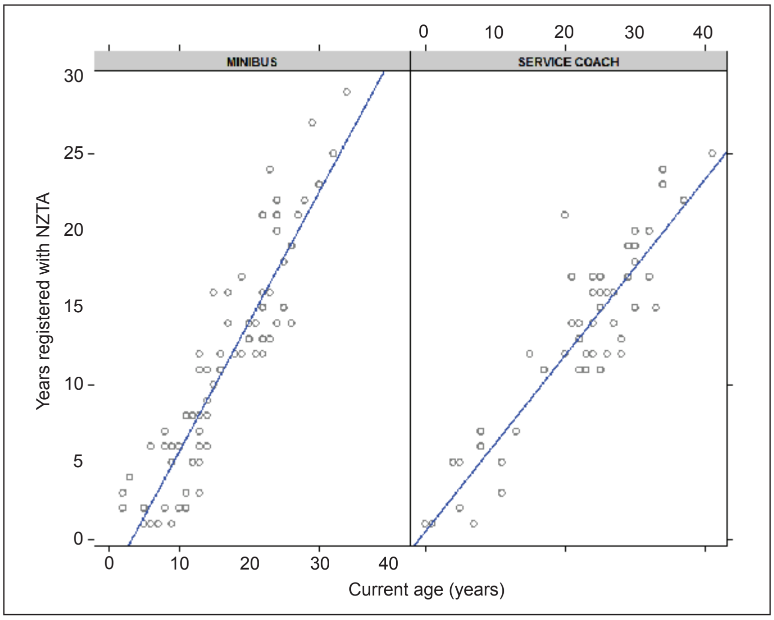
Figures 3 and 4 show plots produced from the sample, each with a possible model fitted to the data.

Figure 3: Used buses



Model 1: Years registered with NZTA = –0.655 + 0.6706 × Current age

Figure 4: Minibuses and service coaches



Minibus

Model 2: Years registered with NZTA = –2.796 + 0.8441 × Current age

Service coach

Model 3: Years registered with NZTA = 0.4702 + 0.5725 × Current age

(i) Using the models given above, calculate TWO different predictions for the number of years a **minibus** that is 40 years old has been registered with NZTA.

(ii) Select which one of the predictions in (i) should be used and justify your selection.

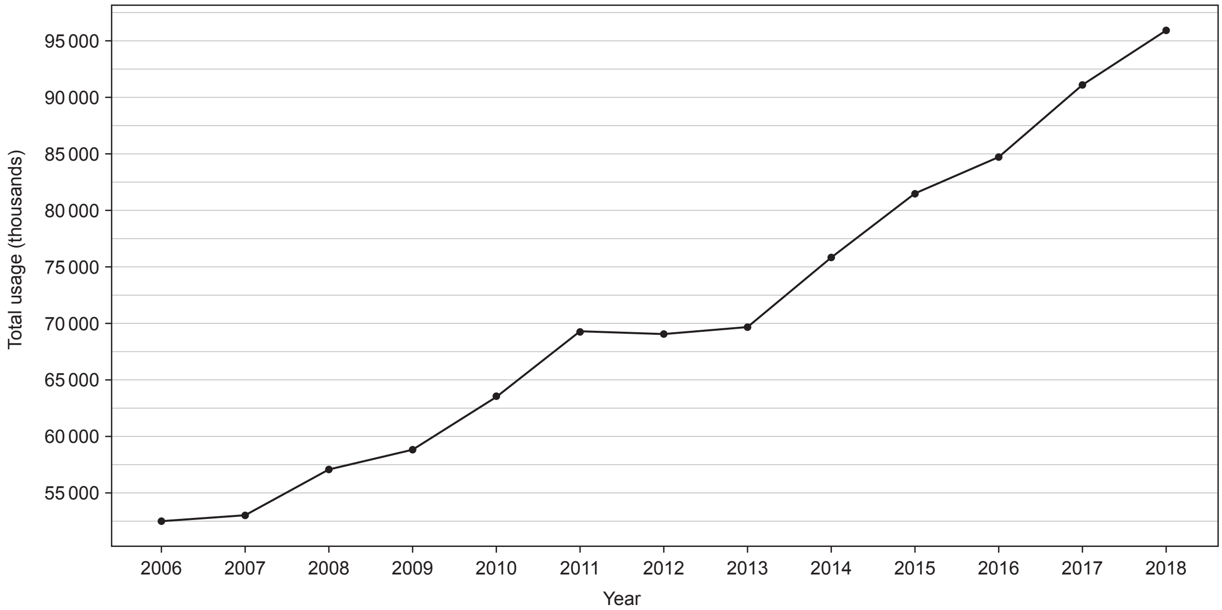
(iii) Give TWO reservations you might have with using the prediction you selected in (ii).

QUESTION TWO

Data was obtained from Auckland Transport on the usage of buses, ferries, and trains per quarter for the years 2006 to 2018. Usage was measured in thousands of trips made.

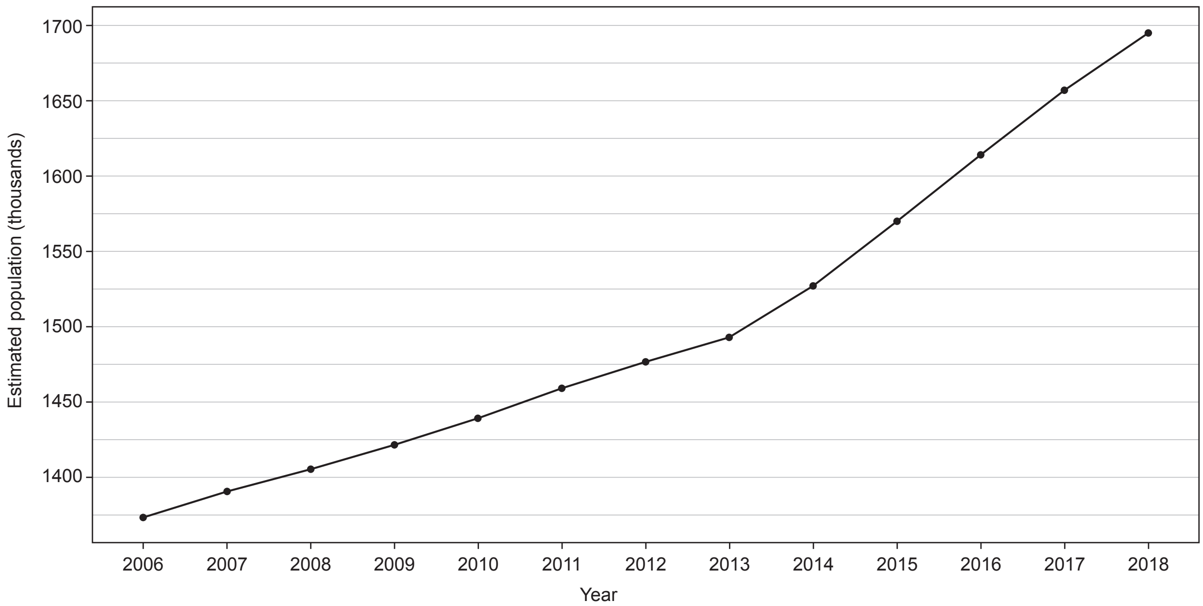
(a) Figure 5 displays the data on total usage of buses, ferries, and trains per year.

Figure 5: Total usage in thousands, 2006–2018



Data was also obtained from Statistics New Zealand on the estimated population of the Auckland Region for the years 2006 to 2018. Figure 6 displays this data.

Figure 6: Estimated population of Auckland region, 2006–2018

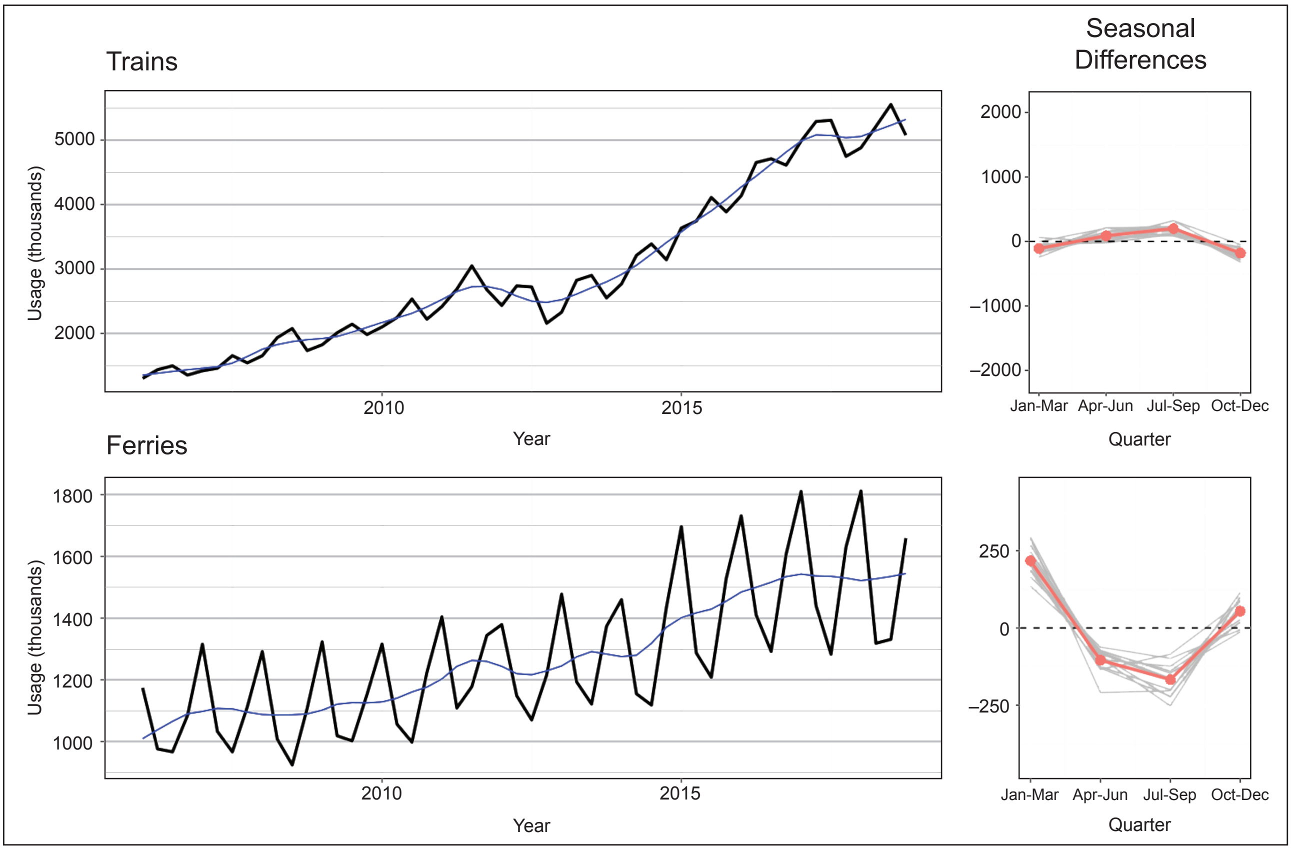


It was claimed that the rate of increase in total usage of Auckland Transport was similar to the rate of increase for the estimated population of the Auckland region.

Using Figures 5 and 6, investigate if this claim is justified.

(b) Figure 7 displays the raw data for the usage of trains and ferries for the years 2006 to 2018, with smoothed trend curves shown in blue. Figure 7 also displays the seasonal differences, with their average (mean) shown in red.

Figure 7: Usage of trains and ferries, 2006–2018



Write two short paragraphs comparing the features of the data for the usage of trains and ferries over the period 2006 to 2018.

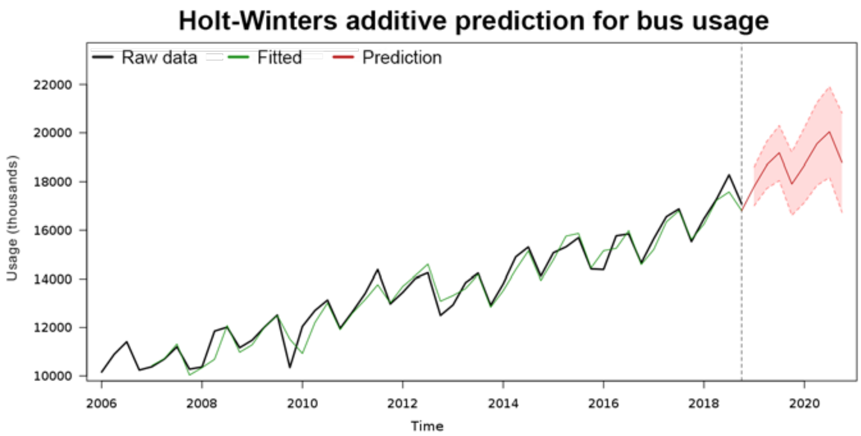
(c) Two different Holt-Winters Additive models were used to obtain forecasts for the usage of **buses** per quarter for the years 2019 and 2020.

Figure 8 shows the raw and fitted data for the years 2006 to 2018, and forecasts produced from Model 1.

Figure 9 shows the raw and fitted data for the years 2016 to 2018, and forecasts produced from Model 2.

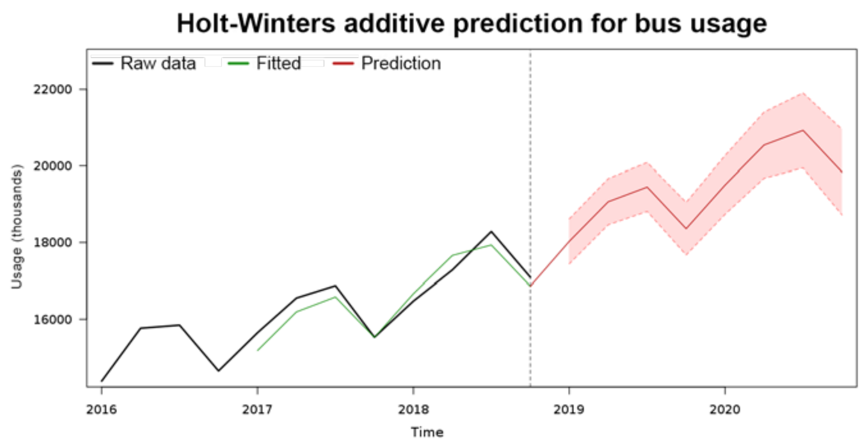
Both figures also include the forecast tables.

Figure 8: Model 1



	fit	upr	lwr
2019 Q1	17806.24	18602.28	17010.21
2019 Q2	18681.31	19643.54	17719.08
2019 Q3	19176.11	20303.50	18048.71
2019 Q4	17905.78	19199.41	16612.15
2020 Q1	18674.83	20209.98	17139.68
2020 Q2	19549.90	21249.01	17850.78
2020 Q3	20044.70	21912.27	18177.13
2020 Q4	18774.37	20814.85	16733.88

Figure 9: Model 2



	fit	upr	lwr
2019 Q1	18029.41	18618.64	17440.17
2019 Q2	19067.40	19671.45	18463.35
2019 Q3	19450.08	20086.19	18813.98
2019 Q4	18361.17	19050.60	17671.73
2020 Q1	19505.04	20270.39	18739.69
2020 Q2	20543.03	21406.06	19680.01
2020 Q3	20925.72	21906.17	19945.27
2020 Q4	19836.80	20952.16	18721.44

(i) Compare the forecasts from Models 1 and 2 for the second quarter of 2020.

(ii) Give reasons for the similarities or differences between the forecasts for the second quarter of 2020 in (i) in terms of the different models.

QUESTION THREE

(a) In a large international survey, students from New Zealand and the United Kingdom were asked about the main mode of transportation they used to get to school, and the time it took them to travel to school.

Data from a random sample of 400 students from each country was used to construct bootstrap confidence intervals for the difference in mean travel times (in minutes) between two common modes of transportation: bus and walking. Other modes of transportation, such as cycling or motor vehicle, were not included in the analysis.

The outputs from these analyses are shown in Figures 10 and 11.

Figure 10: New Zealand (*n* = 198)

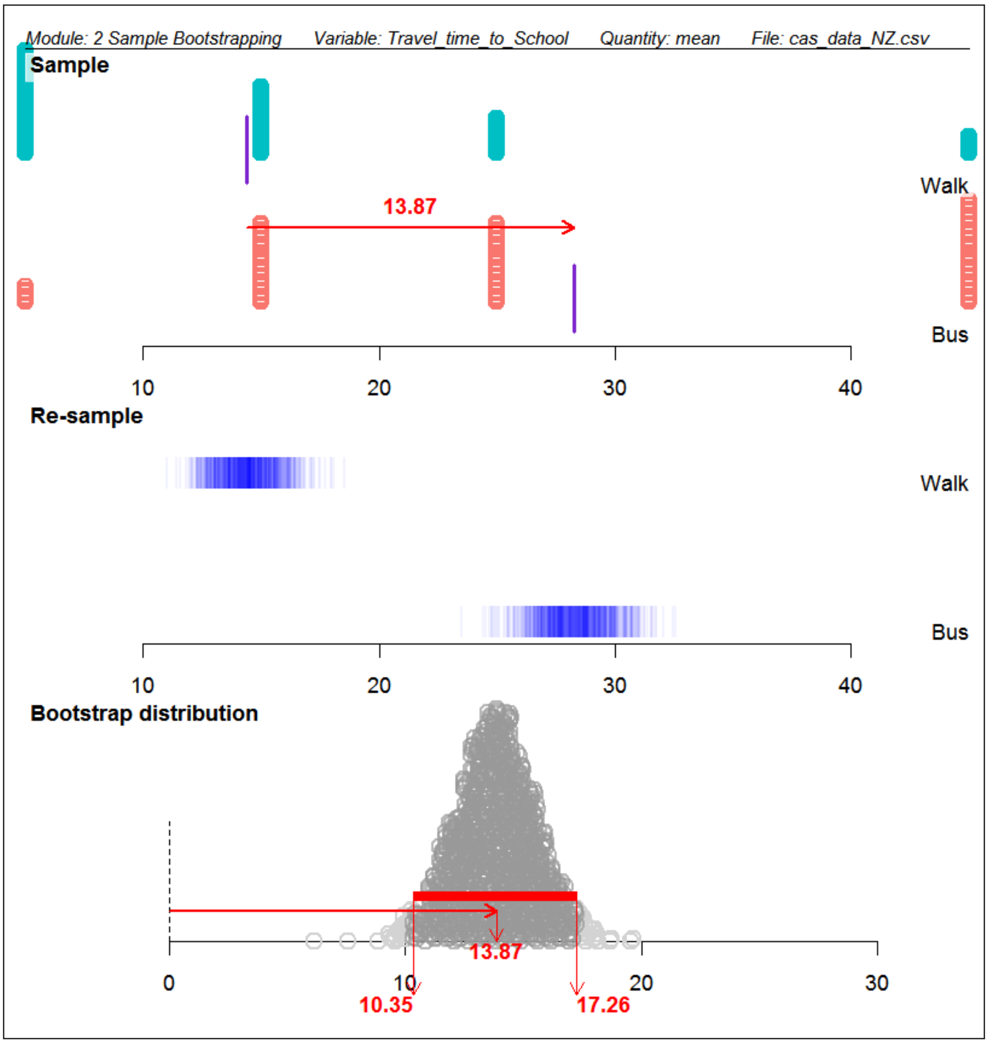
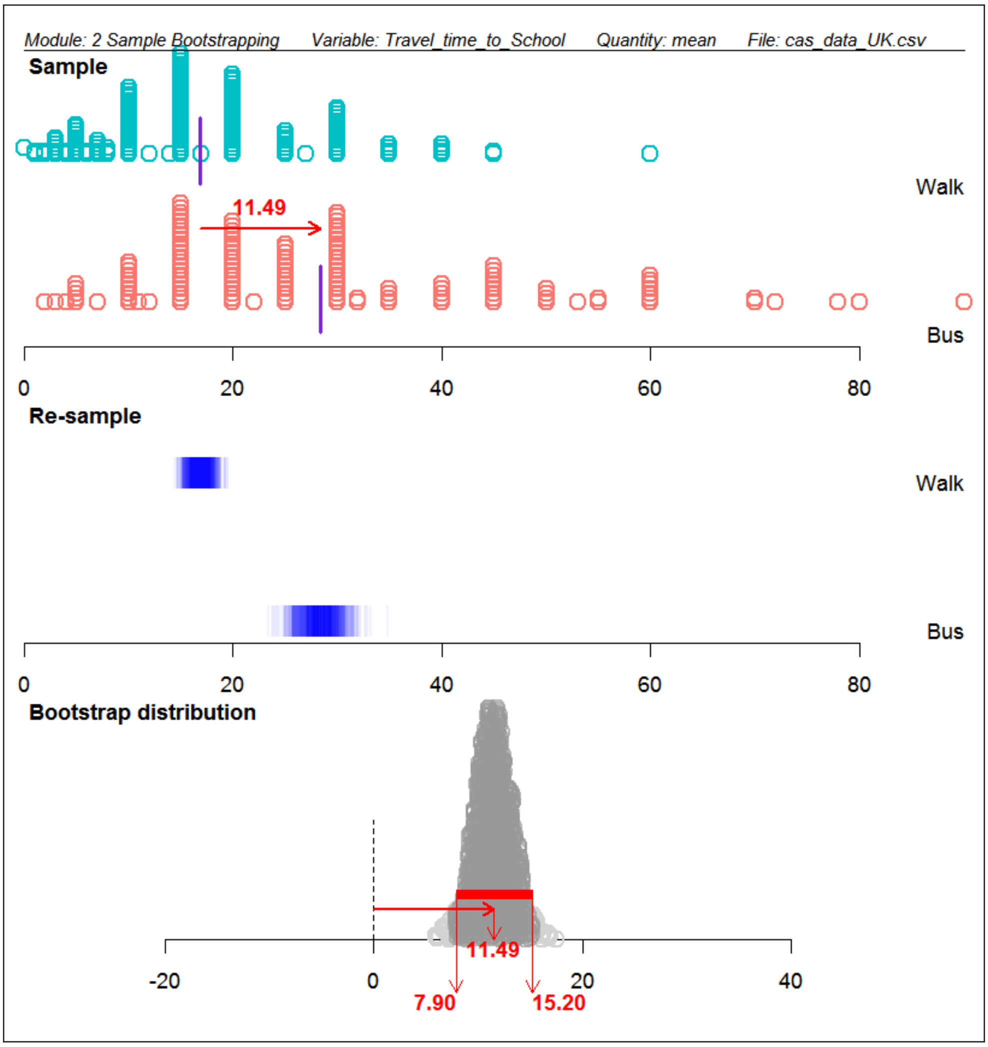


Figure 11: United Kingdom (*n* = 287)



Discuss THREE conclusions that can be drawn from both the features of the sample data distributions and the confidence intervals constructed using the sample data.

(b) An experiment was carried out to investigate if university students could be encouraged to walk or cycle to university more often. The participants were 136 university students who were members of an environment club, all of whom sometimes, but not always, walked or cycled to university. These students were asked to install an app on their phones that requested them to record each evening whether they had walked or cycled to university that day.

The students were randomly allocated into one of two groups of 68. Students in one group were sent daily messages from the app either encouraging them to walk or cycle to university the next day or providing information about the benefits of walking or cycling. Students in the other group were not sent daily messages. After three months, the data collected through the app was used to determine how many days each student had walked or cycled to university.

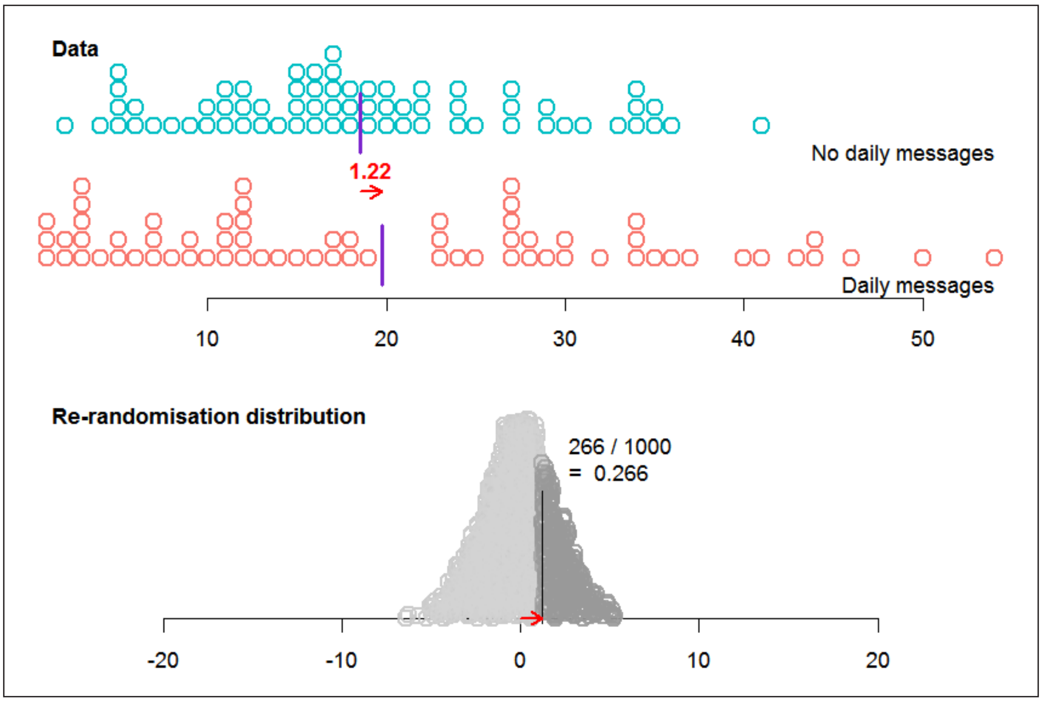
(i) Write a short paragraph that summarises the design of the experiment using appropriate statistical terminology.

A randomisation test was carried out using the difference between the mean number of days participants had cycled or walked to university for the two groups. Table 1 gives some statistics and Figure 12 gives some output from this test.

Table 1: Summary of number of days cycled or walked by group

	Min	Lower quartile	Median	Upper quartile	Max	Mean	Standard deviation	Group size
Daily messages	1	7.75	17	29.25	54	19.72	14.07	68
No daily messages	2	12.00	17	24.00	41	18.50	9.20	68

Figure 12: Summary of number of days cycled or walked, and re-randomisation distribution



(ii) Interpret the randomisation test output and explain why the result could have been expected in this context.

(iii) Give TWO questions you would want answered about the study, data, or participants before accepting the result in (ii).

For each of the questions, explain why it would be important to know each answer.

QUESTION FOUR

Read the following report.

What do passengers do during travel time?

This study reports on 812 adult passengers in Wellington, New Zealand. The aim of this study was to assess the frequency of passenger activities during bus and train travel using structured observations of passengers in a sample of bus and train routes and times in the Wellington area.

Bus and train routes selected were short (20-minute) or long (up to 2-hour) distances, downtown and suburban routes, encompassing wealthier and poorer areas, and included routes where passengers had a clear choice of bus or train mode. Both morning (before 9.00 a.m.) and evening (3.00 p.m. to 6.30 p.m.) peak commuting times were included for observations, as were several night and midday times. Public transport providers were contacted to explain the research and generously provided free passes for the two researchers and a covering letter of support. The two researchers worked together for safety reasons and avoided late night trips.

Researchers recorded passenger characteristics and behaviour over a 4-minute period, on a range of routes and times, using 12 pre-set codes. During the four-minute observation period, a passenger might be recorded as carrying out only one or more than one activity at a time (multitasking), for example, reading a book while wearing headphones or texting while eating. To accommodate this diversity, the data analysis refers to the numbers of passengers who were "ever observed" doing the activity.

Most passengers (65.3%) were "looking ahead / out of the window" at some point in the observation period, more on buses than on trains. About one-fifth of all passengers observed were seen reading, more on trains. Other activities included listening on headphones, talking, texting, and sleeping / eyes closed (see Table 2 for all activities and additional results).

Table 2: Ever-observed activities on bus and train (*N* = 812)

Activities	Bus		Train		Total	
	Number	*% of bus sample*	*Number*	*% of train sample*	*Number*	*% of total sample*
Looking ahead / out of window	270	76.5	260	56.6	530	65.3
Reading	44	12.5	132	28.8	176	21.7
Headphones in	60	17.0	96	20.9	156	19.2
Talking	48	13.6	77	16.8	125	15.4
Texting	29	8.2	46	10.0	75	9.2
Sleeping / eyes closed	15	4.2	57	12.4	72	8.9
Handling wallet, etc.	16	4.5	42	9.2	58	7.1
Other	15	4.2	28	6.1	43	5.3
Eating / drinking	13	3.7	25	5.4	38	4.7
Using computer	1	0.3	34	7.4	35	4.3
Writing	4	1.1	22	4.8	26	3.2
On phone	6	1.7	6	1.3	12	1.5

Gender and broad age group were recorded (young = about 18 to 30 – 34; middle age = 35 to 60; older = over 60). There were 402 women observed. Activities were compared using odds ratios on the basis of gender, age group, mode, and time of day. An odds ratio compares whether the probability of an event is the same for two groups; an odds ratio of 1 means that the event is equally likely for each group. It was found that females were 2.1 times as likely to be observed talking than males and 0.2 times as likely to be observed using a computer than males.

Adapted from: Russell, M., Price, R., Signal, L., Stanley, J., Gerring, Z., & Cumming, J. (2011). What do passengers do during travel time? Structured observations on buses and trains. *Journal of Public Transportation*, 14(3), 7.

(a) Evaluate the design of the study described in the report, including discussion of the following points:

- two strengths of the study design
- two challenges with how the data was collected.

(b) Suppose the data from this study is used to construct confidence intervals for:

- the difference between the percentage of all Wellington bus passengers who look ahead or out the window and the percentage of all Wellington train passengers who look ahead or out the window
- the difference between the percentage of all Wellington bus passengers who look ahead or out the window and the percentage of all Wellington bus passengers who read.

(i) Identify the relevant information from the report that would be needed to construct each confidence interval. **Do not calculate or construct the confidence intervals.**

(ii) Explain why the calculation of the margin of error for each confidence interval will be different.

(iii) Discuss TWO reservations you would have with using the confidence intervals to make inferences about all current bus or train passengers in New Zealand.

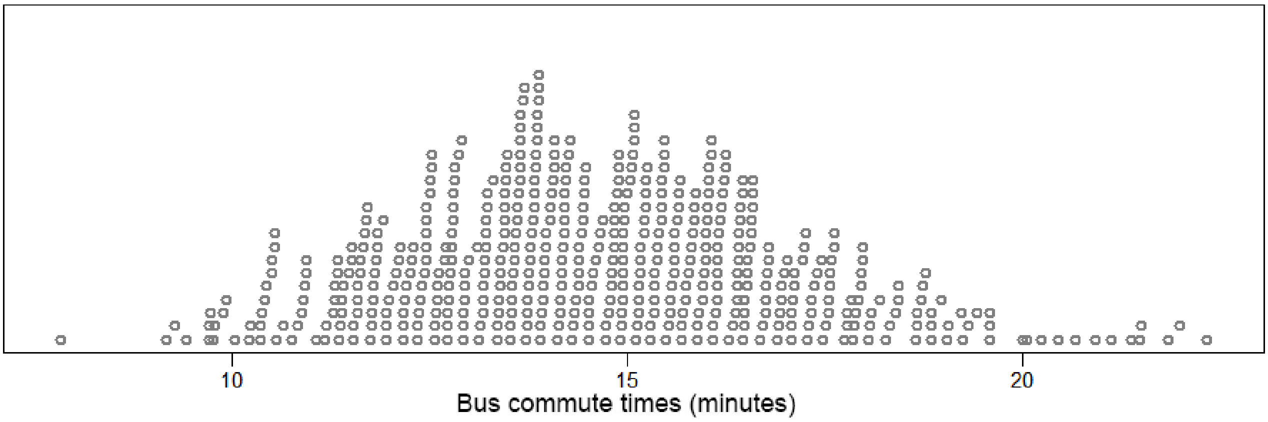
(c) The report states that “females were 2.1 times as likely to be observed talking than males”.

Using the data and results reported from the study, calculate an estimate for the proportion of passengers observed talking in the study who were male.

QUESTION FIVE

(a) Amelia regularly commutes to work by bus. There are 11 bus stops and five intersections with traffic lights along the 3.8 km long journey. Most of her bus commutes are during the morning, but some occur later in the day. The length of each of Amelia's bus commutes (in minutes) was recorded over several years and is displayed in Figure 13.

Figure 13: Amelia's bus commute times



(i) Give TWO potential reasons for the variation in Amelia's bus commute times, and discuss how each could affect the length of her bus commute.

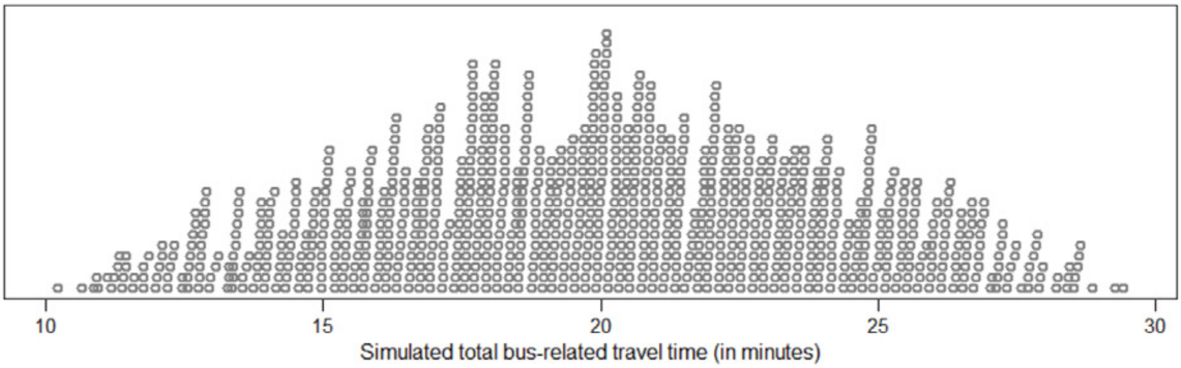
(ii) Discuss the appropriateness of using a normal distribution as a probability model for Amelia's future bus commute times.

(b) Jacob also commutes to work by bus at roughly the same time each day. He developed a model for his total bus-related travel time (the time spent waiting for the bus plus the time spent commuting to work by bus), using one triangular distribution and the following information:

- the time it will take for the next bus to arrive at the bus stop is between 0 and 6 minutes, with all times in between equally likely
- the commute to his work by bus takes between 10 and 24 minutes, with the most likely commute time around 17 minutes.

Jacob then simulated 1000 total bus-related travel times using his triangular distribution model. These times are displayed in Figure 14.

Figure 14: Jacob's simulated travel times

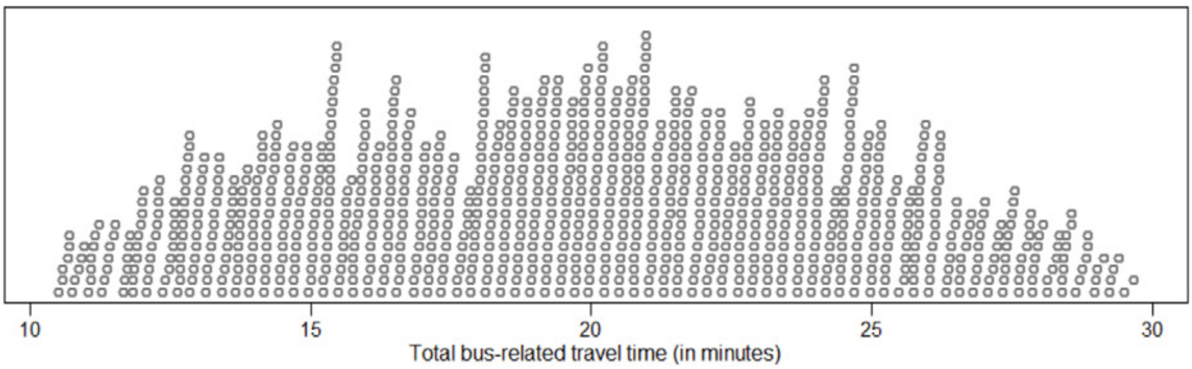


(i) Give the parameters for Jacob's model of his total bus-related travel times, and explain how Jacob appears to have determined these parameters.

(ii) Use Jacob's model to calculate an estimate for the probability that his total bus-related travel time is longer than 28 minutes, given that it is longer than 25 minutes.

(iii) Jacob then recorded his total bus-related travel times over an extended period. These times are displayed in Figure 15.

Figure 15: Jacob's total bus-related travel times



Discuss how the distribution of the data shown above challenges the model Jacob developed, and why assumption(s) he may have made when developing his model may not be valid.

93201Q