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 2016
Statistics

9.30 a.m. Thursday 17 November 2016
Time allowed: Three hours
Total marks: 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 (8 marks)

A study was carried out to determine the relationship between the population size of a city and ozone concentration levels in the atmosphere near the city. Data were collected from 10 cities chosen at random from each of three different states X, Y, and Z within a particular country. For each city, the population size was found, and the concentration of ozone in the atmosphere at that city measured (in parts per billion, ppb). Data obtained are tabulated below.

State X		State Y		State Z	
City population (millions)	Ozone concentration (parts per billion)	City population (millions)	Ozone concentration (ppb)	City population (millions)	Ozone concentration (ppb)
0.6	126	0.4	122	0.3	127
2.4	135	2.3	138	0.6	128
0.2	124	2.1	138	0.7	128
0.5	128	1.6	132	0.4	121
1.1	130	1.8	135	0.6	127
0.1	128	1.5	130	0.3	120
1.1	126	1.9	137	1.1	129
2.3	128	0.6	125	1.0	129
0.6	128	0.7	126	0.4	124
2.3	129	2.0	120	0.5	127

This table has been corrected from that used in the examination.

(a) For each state, a scatter plot of the data above was constructed.

Describe, for each state, the relationship between city population size and ozone concentration.

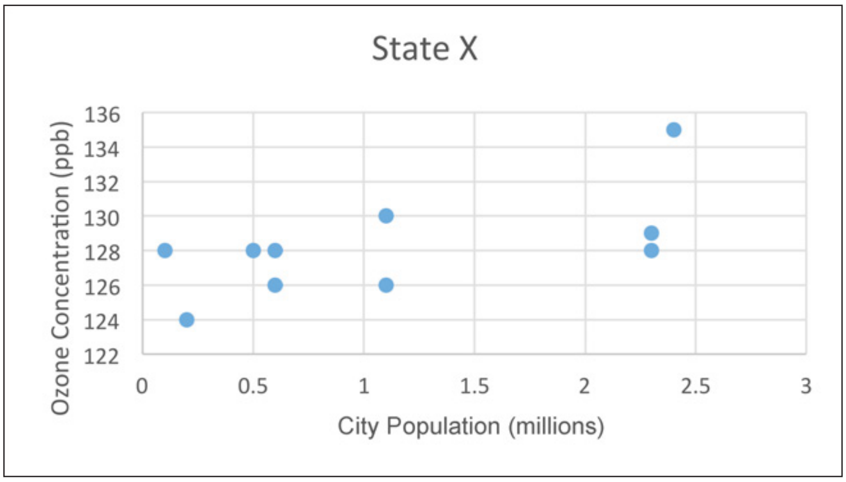


Figure 1: Ozone Concentration versus City Population for State X

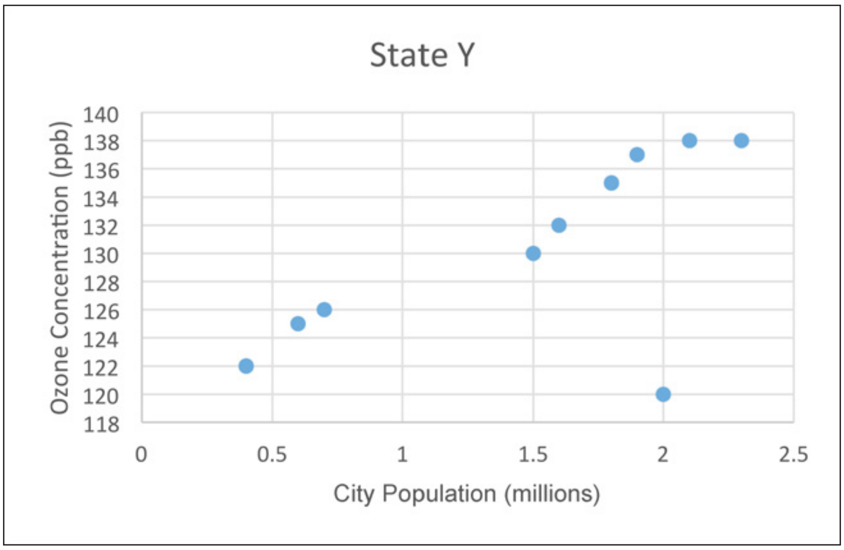


Figure 2: Ozone Concentration versus City Population for State Y

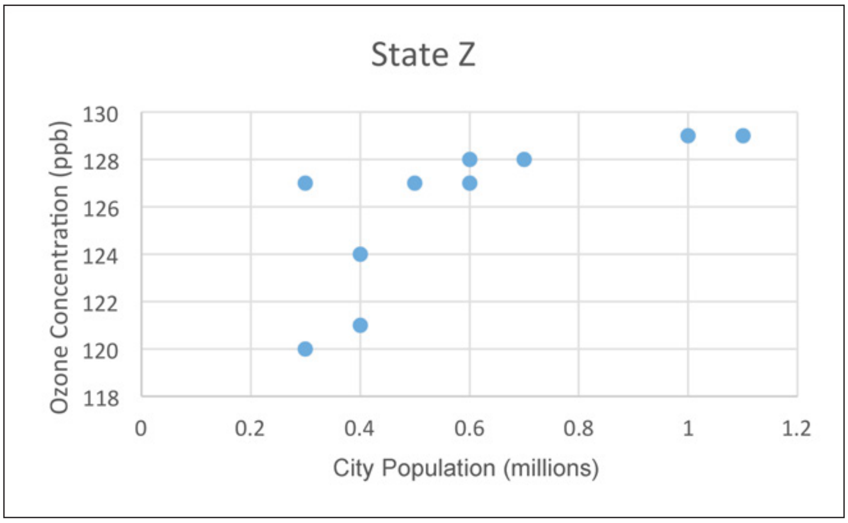


Figure 3: Ozone Concentration versus City Population for State Z

(b) Suppose we wish to predict the ozone concentration for a city of 1.4 million people from any state in that country.

For the three states combined a scatterplot was constructed and a regression line fitted.

Make the prediction from the summary below and comment on its validity.

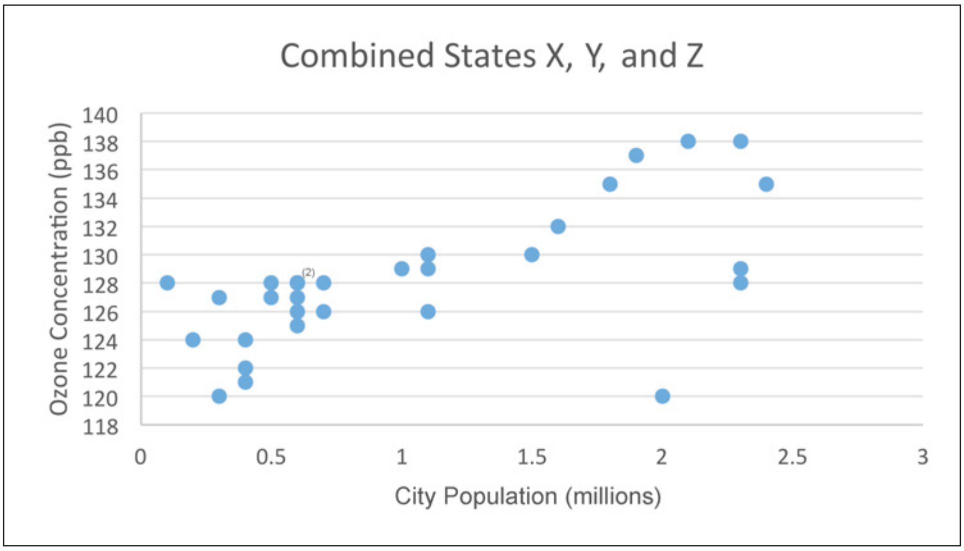


Figure 4: Ozone Concentration versus City Population for States X, Y, and Z

Summary for States X, Y, and Z Combined

Linear Trend: City ozone concentration = 4.1495 * Population + 123.74
Correlation Coefficient = 0.64296
Sample size: 30

(c) Suppose a confidence interval for the mean ozone concentration level was constructed for each state from the data.

What reservation(s) would you have in the use of these confidence intervals?

(d) Explain how you would proceed to ascertain whether there is a difference between the mean ozone concentration for all cities in State X, and the mean ozone concentration for all cities in State Y.

This page has been deliberately left blank.

The examination continues on the following page.

QUESTION TWO (8 marks)

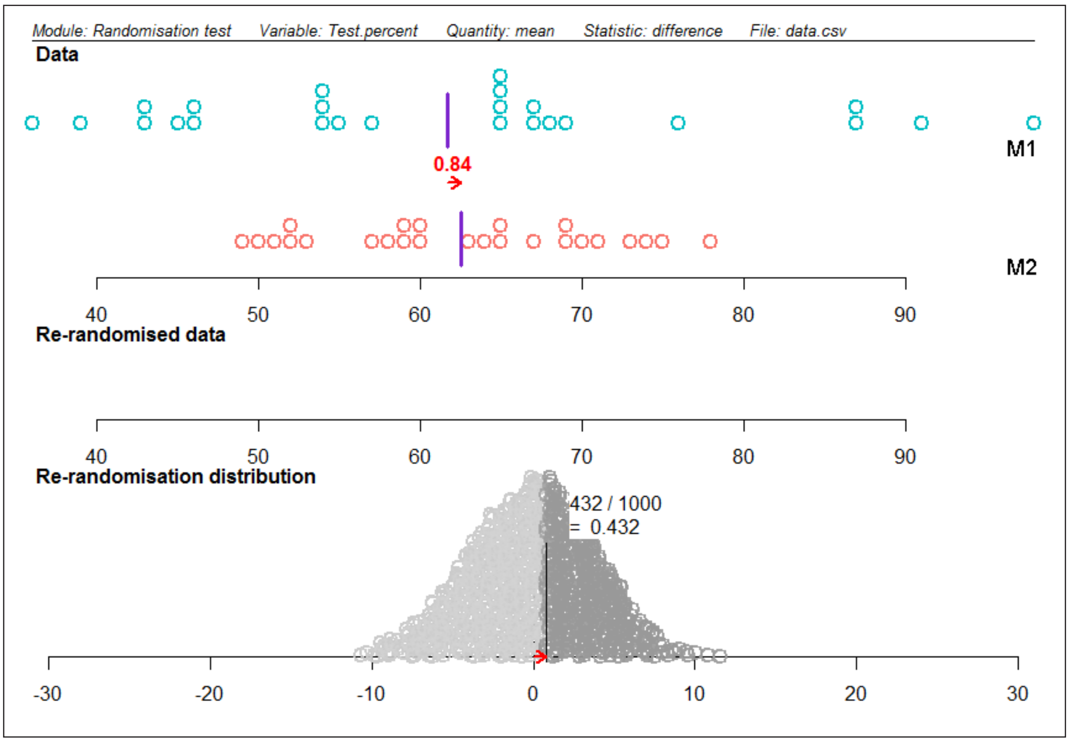
(a) A course on global warming was offered at a university. The course was taught using two different methods, M1 and M2. A study was carried out to see if there was any difference between the effectiveness of the two methods of teaching. Fifty students, who were enrolled for the course, were randomly assigned into two groups, one group to be taught using M1, and the other M2.
After the course was completed, a test of the course content was administered to all fifty students. The following test scores were obtained.

Test scores (%) for M1	
67	57
45	54
65	65
87	76
65	54
87	39
91	55
43	98
36	67
43	54
46	68
46	69
65	

Test scores (%) for M2	
50	58
71	59
67	60
65	60
57	74
59	73
64	78
63	53
70	52
75	52
65	51
69	69
49	

A randomisation test was carried out on the difference of the scores between M1 and M2. The output is shown in Figure 5.

Figure 5



(i) What can be concluded from this output?

Justify your answer.

(ii) Suppose that a particular student did not sit the test.

Suggest how a score could be estimated for that student.

(iii) Suggest a possible factor that could have affected the results of this study.

(b) Suppose the study in part (a) had been to determine the effectiveness of using a single teaching method for all students who may enrol for the course.

Discuss how the study should have been carried out, including how the results could have been analysed.

(c) A new course in global warming is being planned for two large groups of students, G1 and G2. In order to establish if there was any difference between the two groups, G1 and G2, in their background knowledge, a diagnostic test was devised.

A random sample of 25 students was selected from each group to sit the diagnostic test. Their percentage scores are shown below.

Test scores (%) for G1	
66	57
43	56
68	65
83	74
62	54
84	41
88	53
45	90
32	63
42	53
41	67
39	63
65	

Test scores (%) for G2	
40	50
73	42
60	40
59	40
50	70
51	69
61	70
60	50
60	46
72	42
61	41
60	60
41	

Figure 6, on page 9, gives a summary of the test scores for G1 and G2, and the bootstrap distribution of the difference between the test score means.

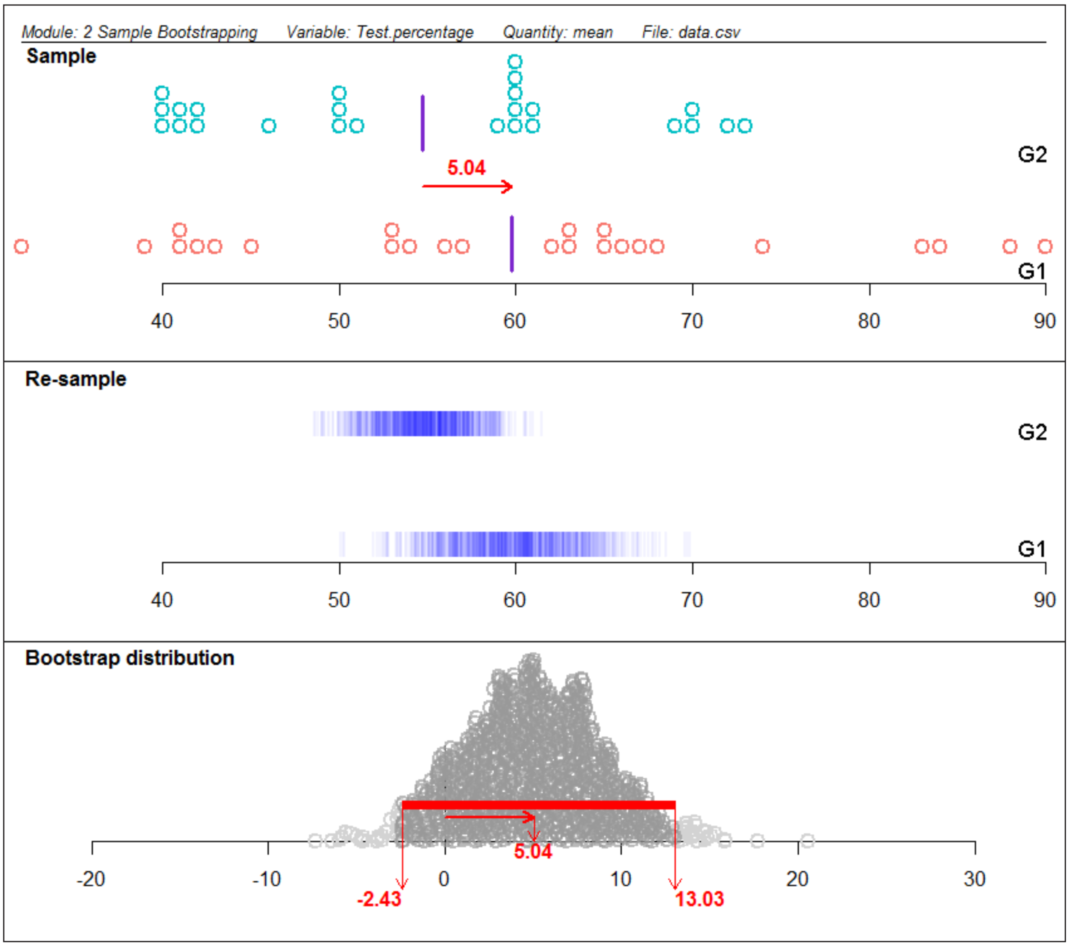
What can be concluded from this output?

Justify your answer.

Figure 6

Summary of Test Percentage by Course:

	Min.	1st Qu.	Median	Mean	3rd Qu.	Max.	Std. dev	Sample Size
G2	40	42	59	54.72	61	73	11.3	25
G1	32	45	62	59.76	67	90	16.0	25



QUESTION THREE (8 marks)

Each year over the last 70 years, there has been an increase in the number of major weather events compared with the previous year.

The percentage increase in the number of major weather events from the previous year, for the last 70 years, had the following frequency distribution:

Percentage Increase Per Year	**Frequency**
0.28 and under 0.40	3
0.40 and under 0.52	5
0.52 and under 0.64	9
0.64 and under 0.76	13
0.76 and under 0.88	24
0.88 and up to 1.00	16

(a) Suppose the percentage increase each year from the previous year was modelled by a normal distribution with mean 0.64% and standard deviation 0.12%.

(i) Find the probability that, over a 10-year period, there has been at least a 0.7% increase in the number of major weather events from the previous year for at least five of those years.

(ii) Suppose that the percentage increase each year was modelled by a triangular distribution with values between 0.28% and 1.00%, and having mode 0.82%.

Recalculate the probability in (i) using this distribution.

(iii) Which of the two models reflects the actual distribution more accurately?

Justify your answer.

(b) Over the last 70 years, the number of floods per annum classified as major weather events were recorded in the following frequency table.

Number of Floods	0	1	2	3	4	5	6	7
Frequency	6	18	22	12	7	3	1	1

(i) Using these data, investigate how closely the number of floods per annum is modelled by a Poisson distribution.

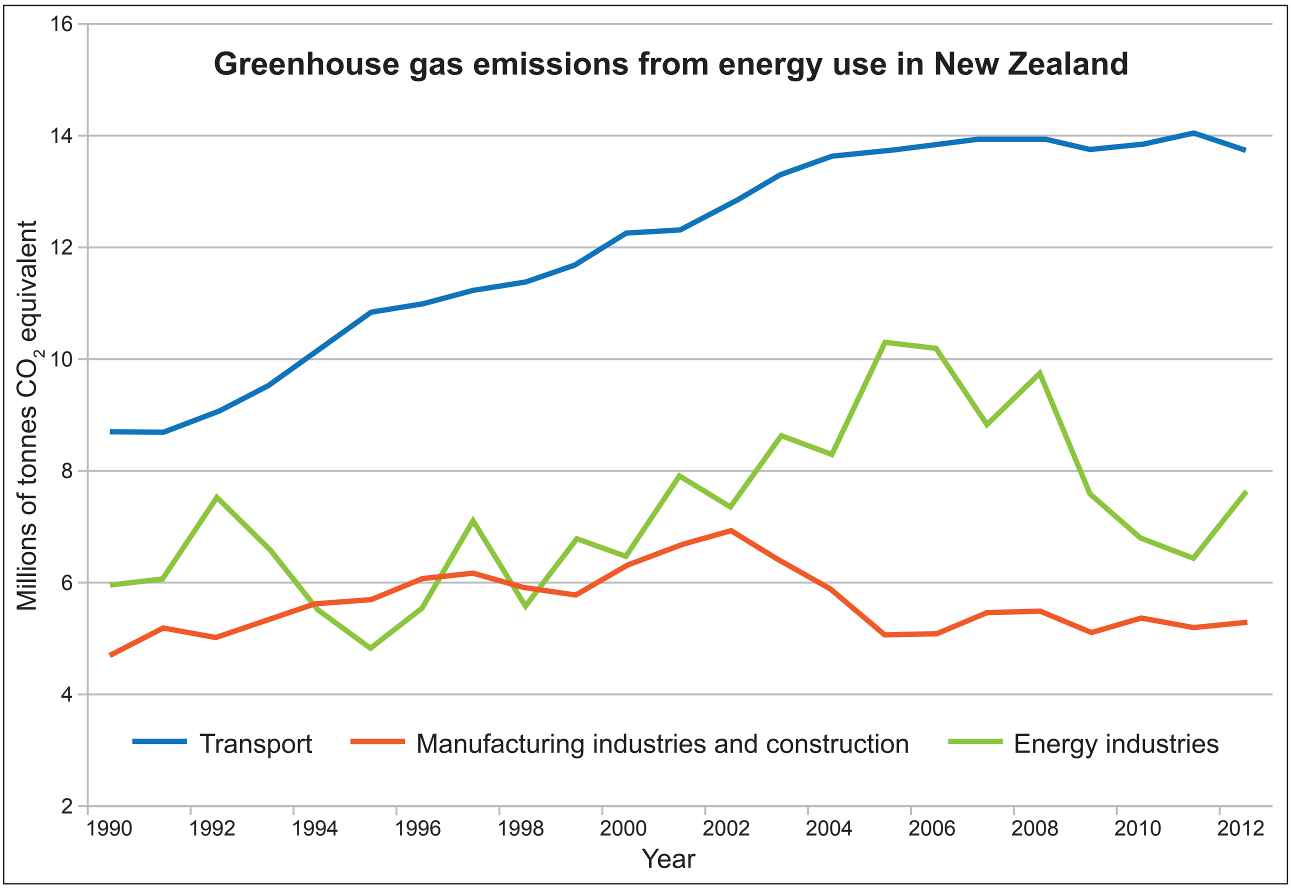
(ii) What assumptions do you need to make in (i)?

QUESTION FOUR (8 marks)

Greenhouse gases in the Earth's atmosphere absorb infrared radiation from both the Sun and the Earth, and re-radiate it. The more of these gases that exist, the more heat is returned to the Earth instead of escaping into space and, consequently, the more the Earth heats. This increase in heat is called the greenhouse effect. Common examples of greenhouse gases are water vapour, carbon dioxide, methane, nitrous oxide, ozone, and fluorocarbons.

(a) Greenhouse gases are emitted when we make or use energy. In New Zealand, major sources of greenhouse gases are transport, and the energy production, manufacturing and construction industries. Figure 7 below shows the equivalent amounts of carbon dioxide (a measure of the greenhouse effect), measured in millions of tonnes, emitted annually from each of these sources, for the years 1990 – 2012.

Figure 7



Source: OECD

Discuss carbon dioxide emissions from transport, and the energy production, manufacturing and construction industries over the period 1990 to 2012.

(b) Figure 8 shows the total amount of greenhouse gases in kg emitted in 2012 per person, for a selection of developed countries.

Figure 9 shows the amount of methane in kg emitted in 2012 per person, for the same countries.

Figure 8

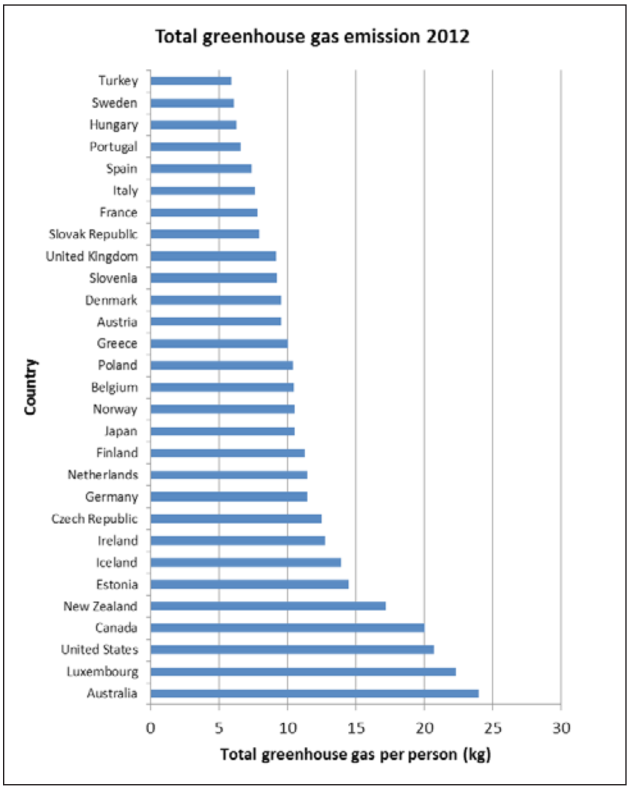
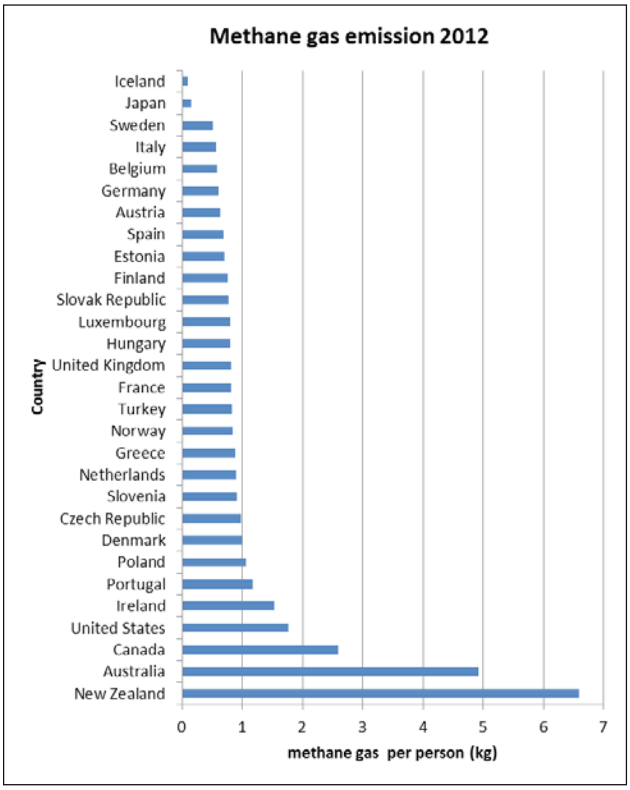


Figure 9



Source: OECD

(i) How could these data be presented to allow for easier comparison of total greenhouse gas emission and methane gas emission for each country?

(ii) In 2012, the population of Australia was approximately 22.7 million and the population of New Zealand was approximately 4.4 million.

Use the graphs to compare the total greenhouse gas and methane emissions of Australia and New Zealand.

(c) Figure 10 shows the amount of carbon dioxide in the air as recorded at Baring Head, Wellington between 1999 and 2005. The air arriving at this site from the south has originated from areas of no human activity, and therefore is not influenced by any local activity.

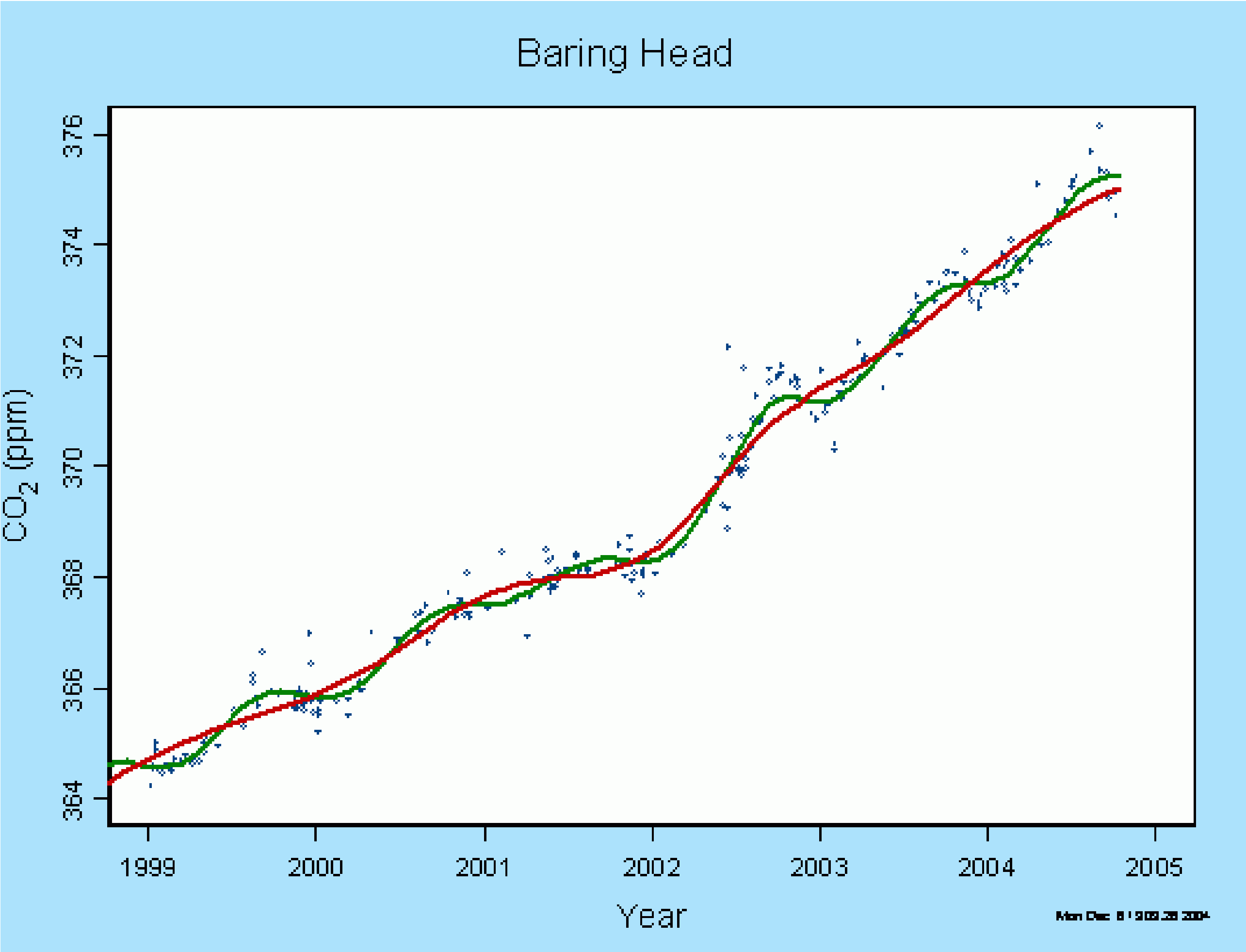
The blue dots show individual data points.

The green curve shows the seasonal variations.

The red line shows the trend.

The vertical scale is measured in parts per million (ppm), i.e. the number of molecules of carbon dioxide in every million molecules in the air.

Figure 10: Carbon Dioxide in the Air at Baring Head



Source: NIWA

(i) Comment on the seasonal component apparent in this graph.

(ii) A news item about future predictions of climate recorded the following:

Before the Industrial Revolution in the mid-1800s, the level of carbon dioxide had been stable at about 280 ppm for most of the last 10 000 years. Computer models predicting the climate in the future refer to the 'doubling of carbon dioxide compared with pre-industrial times'.

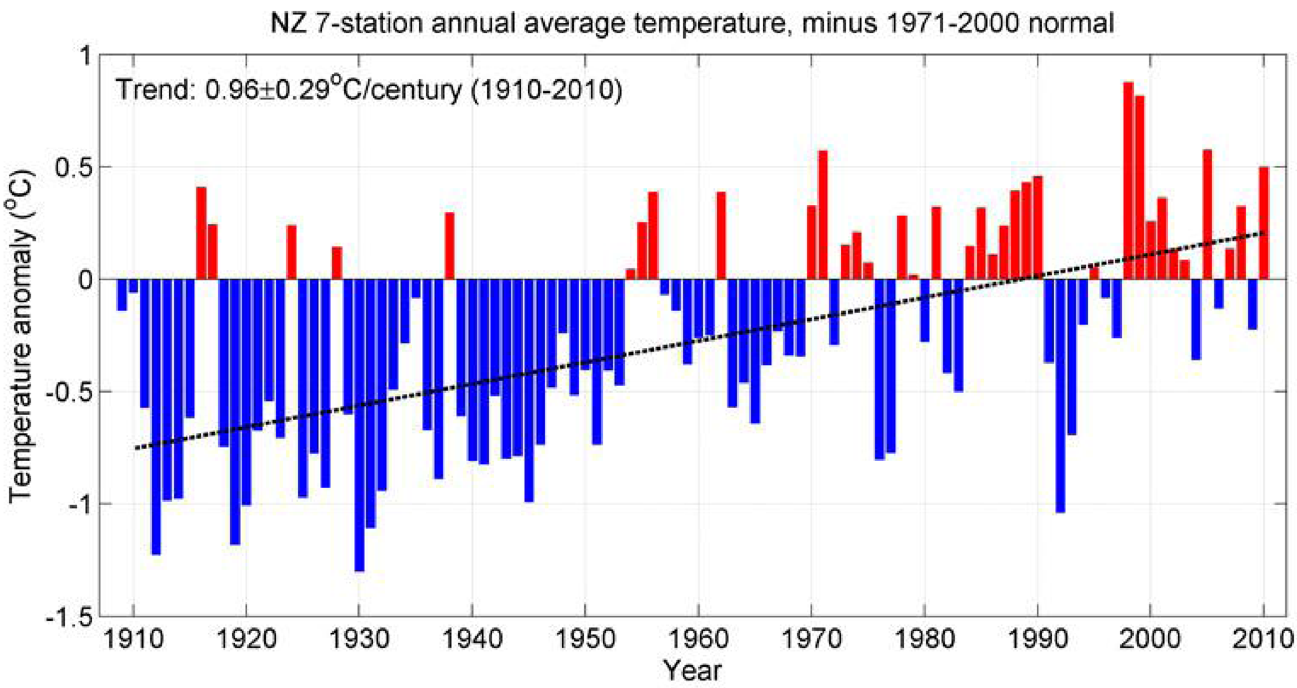
Using the information in the news report, find an upper estimate and a lower estimate for when this doubling of carbon dioxide might occur, assuming that the rate of increase remains roughly as shown in Figure 10.

Justify and comment on the validity of your answer.

QUESTION FIVE (8 marks)

Temperature data were collected for seven geographically representative weather stations in New Zealand. The data were used to give the mean temperature for New Zealand for a particular year. The difference between this mean and the mean temperature for the years 1971 – 2000 was then found. This process was repeated for each of the years 1910 – 2010. The temperature anomalies (these differences) for 1910 – 2010 are shown in Figure 11 below. A trend line is also shown. The trend line shows an increase in annual mean temperature of approximately 0.96°C per century.

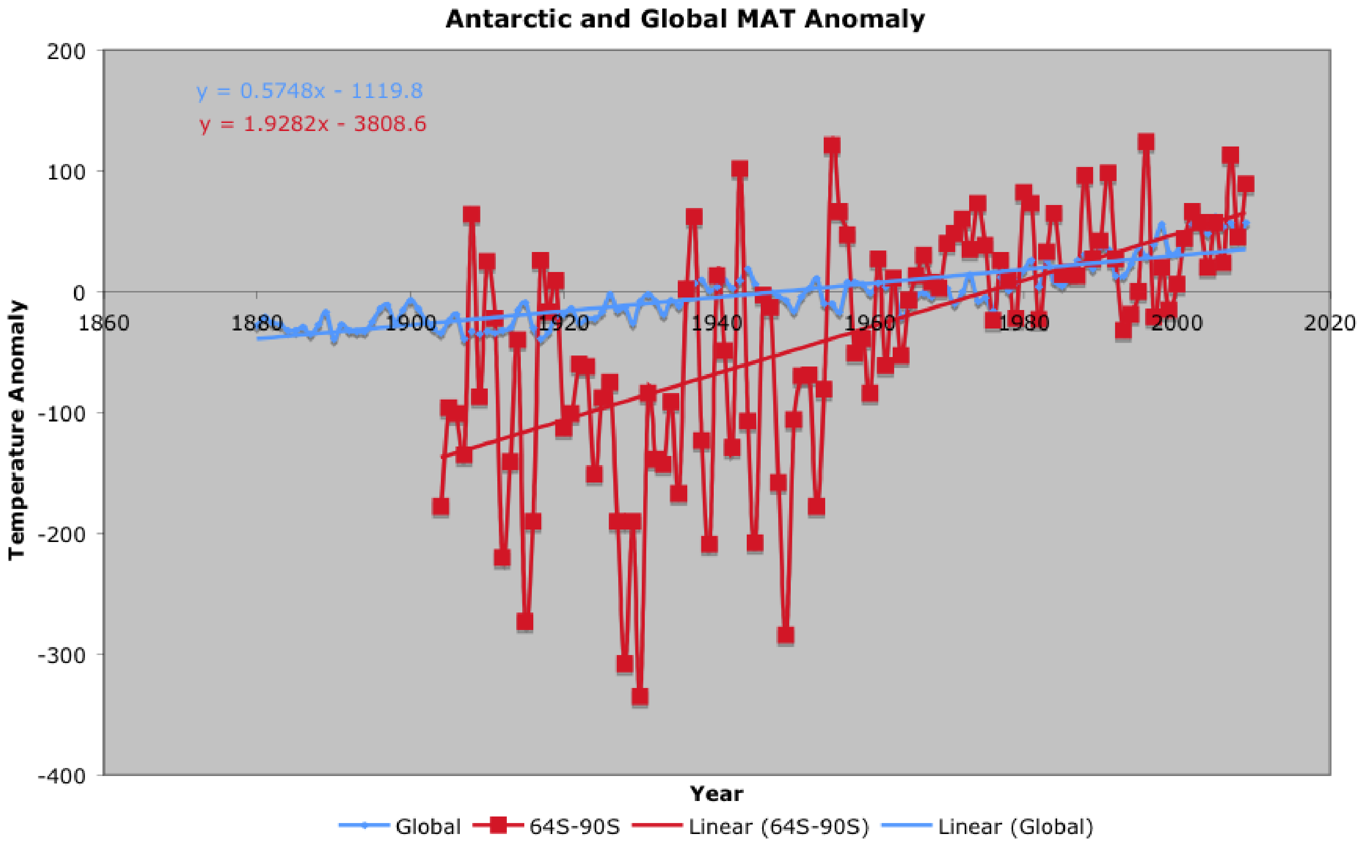
Figure 11



Source: NIWA

Global temperature data and temperature data for the Antarctic region (64°S – 90°S) were collected. Global temperature anomalies were calculated using a similar method as for New Zealand above, for the years 1880 – 2010. The difference between the mean annual temperature (MAT) and the mean temperature for the years 1951 – 1980 was used. The results are shown in Figure 12. The results have been shown on a relative scale with a base of zero rather than in degrees Celsius. A trend line is also shown for the global data and for the Antarctic region data, and the equation of each trend line is shown.

Figure 12



Source: NASA

(a) Give two factors that would have been considered when selecting the weather stations from which temperature data were collected in New Zealand.

Explain why these factors should be considered.

(b) Sam is a member of the Environmental Group at his school and has been asked to give a statistical report on global warming.

Give five main points that Sam could make based on the information in Figures 11 and 12.

Number each point clearly.

(c) In Figure 12 there appears to be a difference in the behaviour of the Antarctic data for the years 1900 – 1960 as compared with the data for the years 1960 – 2010.

Describe this difference and give a possible explanation for the difference.

93201Q