No part of the candidate's evidence in this exemplar material may be presented in an external assessment for the purpose of gaining an NZQA qualification or award.

S

93201A

932011

SUPERVISOR'S USE ONLY

OUTSTANDING SCHOLARSHIP EXEMPLAR

NZQA
NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

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

Tick this box if you have NOT written in this booklet

Scholarship 2022
Statistics

Time allowed: Three hours
Total score: 32

ANSWER BOOKLET

Check that the National Student Number (NSN) on your admission slip is the same as the number at the top of this page.

Write your answers in this booklet.

Make sure that you have Formulae Booklet S–STATF.

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

Check that this booklet has pages 2–24 in the correct order and that none of these pages is blank.

Do not write in any cross-hatched area. This area may be cut off when the booklet is marked.

YOU MUST HAND THIS BOOKLET TO THE SUPERVISOR AT THE END OF THE EXAMINATION.

Question	Score
ONE	
TWO	
THREE	
FOUR	
TOTAL	

ASSESSOR'S USE ONLY

© New Zealand Qualifications Authority, 2022. All rights reserved.
No part of this publication may be reproduced by any means without the prior permission of the New Zealand Qualifications Authority.

Q1

Total farm area in New Zealand decreased from 15 600 000 hectares in 2000 to 13 600 000 hectares in 2020, with a percentage decrease of 87.2 %. This is larger than the percentage decrease in total farm counts in New Zealand of 71.2%, where farm counts decreased from 69500 in 2000 to 49500 in 2020. This shows that total farm area exhibits a higher rate of decrease in comparison to the rate of decrease of total farm counts. The percentage decrease rates are both over the time interval of 20 years, thus can be directly compared.

Nitrogen fertilisers and phosphorus fertilisers sold both increased from 1990 to 1995, however the number of phosphorus fertilisers sold plateaued since 1995, before increasing sharply again at 2000 to 2005, and declined sharply until 2009. Since 2009, the ~~number~~ tonnes of phosphorus fertilisers sold had a generally increasing trend with numbers reaching almost 160 thousands tonnes in 2020. However there is a large amount of fluctuation in ~~number~~ tonnes of phosphorus fertiliser sold as compared to the small fluctuation in the tonnes of nitrogen fertiliser sold.

Additionally, sheep numbers farmed in New Zealand in 1970 (60 000 000) was 7.5 times ~~higher than~~ the total cattle farmed in 1970 (8 000 000). This ratio has decreased since, where sheep numbers farmed in 2020 (28 000 000) is 2.8 times the total cattle farmed in 2020 (10 000 000). This could indicate an increase of methane emissions per livestock numbers, as cattle ~~generate~~ have a higher environmental impact than sheep. Although the decrease in total farm area and total farm count indicates less livestock farming, the generally increasing trend of fertilisers sold since 1990 suggests intensification of agriculture. As more fertilisers are used on a smaller farm area, this implies that the concentration of fertilisers used has increased, which would lead to higher percentage of fertilisers per land area.

b). A graphical technique used to visualise the positive correlation between human modified land cover in upstream catchment and the total nitrogen concentrations is the use of color to group ~~e~~ and classify percentages. Purple was used to identify low human modified land cover (≤25%) and At or under DGV nitrogen concentrations. Both categories are the lowest in their respective groups. In contrast, yellow was used to indicate the highest category in each group (more than 200% over for total nitrogen concentrations by percent over ANZG DGV and ≥76%. for human modified land cover in upstream catchment). Similarly, intermediate ranges were also classified based on their ranks to associate with the same colour in both groups. The classification based on color is used to allow visualisation of the ~~overlap~~ positive correlation between the two groups, where higher proportions of human modified landcover in upstream catchment area correlated to ~~more than 200% over~~ higher concentration of nitrogen. This is the same principle for lower proportions. Using yellow for the highest proportions and concentrations and purple for the lower proportions is successful as both are opposite each other on the color wheel. The yellow colour on both New Zealand maps quickly draws the reader to notice the positive correlation pointed out above.

c). The trend for the value of fertiliser imports is an increasing one, with the value per quarter in 1988 of approximately 20 million of dollars increasing to the value per quarter in ~~2022 of~~ 2021 of approximately 400 millions of dollars. The seasonal variations have appeared to increase since 1988 (*). Seasonal fluctuations include toughs in Q1 and peaks in Q3. Additionally, there is an unexplained spike in ~~fertiliser im~~ value of imported fertiliser ~~values~~ at 2008 to 2009, with the value of imported fertilisers reaching approximately 300 millions of dollars.

(*) There is minimal seasonal fluctuations ~~since~~ before 2000, although since 2000 onwards there appears to be seasonal variations in ~~each year~~ each year differing by each quarter.

ii)

The prediction interval for last quarter of 2022 ranges from around 275 millions of dollars to around 450 millions of dollars. This is a large prediction interval, and the prediction is suspect for the following reasons.

1) The data ends on a high value at the end of 2021, which has the effect of increasing the prediction value due to Holt-Winters model placing highest weightings on recent data.

2). The additive Holt-Winters model may not be suitable considering the increasing ~~seas~~ magnitude of seasonal variations. A multiplicative model may be more suitable which models with seasonal variations increasing in magnitude.

3). The fitted data for previous years do not provide close match to the raw data. This is likely due to many unexplained troughs and peaks like the peak at 2008-2009, and ~~wide~~ fluctuations, and ~~especially~~ the lack of seasonality prior to 2000

4). The forecast ~~a~~ could be made more accurate if data is used from 2010 onwards. Although more weighting is given to the recent data in Holt-Winters, the inclusion of data since 1988 means there will still be weighting placed on the older values. As the trend and seasonality ~~increase~~ since 2010 is markedly different from data before 2010, ~~it~~ the prediction may be more accurate, and the prediction interval smaller if data since 2010 onwards is used.

5). ~~Data conc~~ It is unknown if the current trend will continue into 2022, ~~as~~ Since the model predicts future data based on past behaviour, it does not take into account any unexpected events which may occur. For example, global economy recession could lead to decrease in value of imported fertilisers, thus the ~~pred~~ prediction would be an overestimate.

Q11
80
MAX

Q2

a)i) Each cow would act as a control for itself. This would allow researchers to isolate ~~compare~~ the effect of adding seaweed ~~with~~ through comparison of the treatment groups (0.5% or 1% seaweed) with the control group (0% seaweed). Through establishing each cow as a control for itself, the researchers minimize the effects of natural variation amongst treatment groups (for example, cows that have naturally higher methane emissions could be allocated to the control groups. This would lead to wrong conclusion made that seaweed consumption reduced methane emissions). Therefore, a causal conclusion can be made regarding the effects of seaweed as other confounding variables are minimized.

ii). Firstly, the researchers would calculate ~~can retrieve~~ difference in methane emission of each cow ~~to~~. ~~To~~ between ~~the mean/median me~~ its methane emission in the low additive diet and in the control diet. Then, a statistical summary of the differences can be plotted and analysed. To test for statistical significance, a ~~ran~~ randomisation test can be undertaken. The ~~different~~ values of methane emission would be randomly allocated to ~~each~~ treatment groups (control and low additive diet). A difference in mean/median of both groups would be found. This process is repeated up to 1000 times to construct a rerandomisation distribution for the difference in mean/median. The tail proportion would show the probability of obtaining the experimental or more difference in methane emission of the low additive diet and the control diet if chance was acting alone. If the tail proportion is smaller than 0.1, there is evidence to support the claim that methane emission in the low additive diet is significantly lower than the methane emission in the control diet. This process would be repeated for comparison between high additive diet and the control to evaluate the claim that the methane emission in the high additive diet is significantly lower than in the control. The choice of mean or median depends on the distribution of ~~sample~~ values in the sample.

If the distribution of methane emission in each diets are approximately ~~symmetrical~~ ~~u~~ symmetrical, unibnodal, and does not have outliers (value outside IQR×1.5), then the mean would be a suitable measure when carrying out the randomisation test. Additionally, other more complex statistical test can be carried out such as ~~to~~ the Wilcoxon ranked sign test which ~~us~~ takes into account the difference of each cow's measurement in different treatments. This would be the non-parametric alternative to a paired t-test. which would be also suitable if the underlying distribution of methane emissions can be assumed to be normal.

b) Problem: What is the difference in mean/median ~~nitrogen~~ nitrogen content in cows' urine between cows that ~~are~~ live on soils fertilised with seaweed (karengo) fertilisers and cows that do not?

Plan: Determine experiment parameters: The overall design would be comparison of two independent groups. ~~Random allocation would be carried out to allocate cows into each group~~. The first group would be cows living on soil fertilised with karengo fertiliser, and the second group would be the control, ~~with the same characteristics as the treatment group except that they do not live on soil fertilised with Karengo~~. The sample units would be cows, ~~and~~ they are to be randomly selected and they should be represntative of the cow population of NZ. A sample size of at least 30 should be used.

Data: Record ~~methane emission in both with sensors that measure methane emissions~~ nitrogen content in cow's urine for both ~~treatment grou~~ groups.

Analysis: Obtain ~~statistical~~ statistical summary of ~~cows with~~ the nitrogen content in urine. difference in ~~methane emission~~ between the two groups. There is likely to be a difference if confidence intervals for both means do not overlap.

To ~~sh~~ estimate the difference, ~~show statistical different significant difference~~, conduct a bootstrap test. ~~re randomisation test to find the probability of the experimental difference or more occurring~~. If the tail proportion is less than

~~0.1, there is evidence that~~ chance is not acting alone. The claim ~~would be supported that nitrogen in cows' urine is decreased when karengo is used~~.

~~The mido~~

~~b)~~.

Bootstrap test involve sampling with replacement to generate a confidence interval of the middle 95%. ~~of the~~ The confidence interval can be used to estimate where the population parameter lies. If the confidence interval for the difference of mean methane emissions ~~are~~ in cows ~~feed~~ living on karengo fertilised soil and ~~cows the~~ mean methane emissions in cows ~~living on it~~ not living on karengo fertilised soil ~~includes 18% and is entirely posite, then~~ is entirely higher than 18%, ~~there is evidence to support the claim that nitrogen in cow's urine only in karengo fertilisers~~ with the ~~conten~~ mean of nitrogen content ~~methane emissions~~ in cows urine ~~living on~~ not living on karengo soils exceeding mean of nitrogen content in cows urine living on karengo fertilised soils, then there is evidence to support that the mean nitrogen content in cow's urine living on karengo fertilised soils is lower than mean nitrogen content in cow's not living on karengo fertilised soils by 18%.

(Cont. on Pg 13) → seen

c). The percentage of articles containing the word "climate" is highest in September, with ~~about 12%~~ almost 12% out of all months. In contrast, ~~the this percento~~ August is the ~~for~~ month with lowest articles published containing the word "climate" in the headline at about 5%. October, November and December all have similar percentages of ~~stags~~ 8.5% of articles published containing the word climate in the headline. The sentiment score for these articles appear to be around 0.55, which is slightly over half of the sentiment scale. Out of the articles containing the word climate, 55% of these headlines also contain the word change. The median sentiment score of headlines ~~words~~ containing 'change' and 'climate' (0.58) is ~~bo~~ approximately 0.06 ~~0.067~~ higher than the median sentiment score of headlines not containing 'change' (0.52). The sentiment score distribution for ~~both~~ all articles containing the word climate is negatively skewed and could ~~p~~ be ~~bimod~~ bimodal, with a peak at 0.55 and one at 0.63. The sentiment score distribution of ~~my~~ headline which also contains the word 'change' and the sentiment score distribution of headlines which do not are similar. Both are negatively skewed and have IQR of around 0.14.

Qu2
70
see pg13

Q3

a)i) The survey would have undertaken stratified sampling. This is where the population is divided into stratas such as gender, education, occupation, income. Here, the survey would have divided the population into stratas and sampled based on the representation of NZ adult population at each of these stratas. For example, if 20% of NZ population at the most recent census received ~~a~~ PhD qualification, then 20% of the survey sample would have received a PhD qualification to represent the NZ adult population accurately.

5

ii). $MOE = \frac{1}{\sqrt{1097}} = 0.03019$

~~$\mu_{2006} - \mu_{2007} =$~~ ~~$p_{2006} - p_{2007} = 0.01$~~

Confidence interval for difference in proportion who consider climate change a problem in 2006 and 2007 is $[-0.02019, 0.04019]$. ~~Since~~ Since this interval contains 0, the claim is not supported.

~~CI for difference in proportion who considered c~~ Similarly, this can be carried out for subsequent years.

Difference in proportion who consider climate change a problem in	Confidence interval	
2006 and 2007	[-0.02019, 0.04019]	
2007 and 2008	[0.4898, 0.5502]	*
2008 and 2010.	[0.02019, 0.04019]	*
2010 and 2012	[-0.2702, ~~20~~ -0.2098]	
2012 and 2014	[-0.01019, 0.05019]	
2014 and 2018	[0.06981, 0.1302]	*
2018 and 2019	[0.01981, 0.08019]	*
2019 and 2021	[~~0.00010~~ -0.00192, 0.06019].	

~~As or~~ The confidence interval for difference in proportion of NZ adults who considered climate change a problem is only entirely positive for between

2008 and 2007, 2010 and 2008, 2015 and 2014 and 2019 and 2018.
Therefore, it can be claimed that this proportion has increased ~~in~~ in these intervals and not for ~~others as their confidence~~ the other years.

~~b). This is a comparison within one group.~~
~~The MoE is $\frac{1}{\sqrt{291}} \times 2 = 0.117$.~~

~~Point estimate for difference in proportions:~~
~~$(0.256 + 0.414) \times 0.457 - [(0.302+0.190)0.217 + (0.526+0.211)0.326] = 0.30619-$~~

b1) Of those that considered climate change a problem, $\frac{89}{190}$ were Y9-13s and only $\frac{31}{190}$ were Y1-6 and $\frac{70}{190}$ were Y7-Y8. Of those that did not consider it a problem, $\frac{7}{16}$ were Y9-13 and $\frac{3}{16}$ were Y7-Y8 and $\frac{6}{16}$ were Y1-6, so the proportions were similar.
It is 2.87 times likelier out of those who consider climate change a problem to be Y9-13.
It is 1.36 times likelier for a Y9-13 to consider climate change a problem than a Y1-6.

b) ii). No. It is likely that number of students at Y7 to Y8 is much less than students from Y1 to Y6 as less year levels are covered. ~~However,~~ Because the sample is taken from an online survey, it is likely that more senior students were chosen as they are able to complete the survey with online ~~too~~ computers or phones. ~~In addition~~

bi). ~~P~~ Let ~~P~~ (consider climate change a problem) be X

$P(X) = 0.457 \times (0.414 + 0.256) + 0.326 \times (0.526 + 0.211) + 0.217(0.302 + 0.190)$
$= 0.6532$

$P(Y9-Y13|X) = 0.4688$

~~$P(Y9-13|P) = \frac{89}{190}$~~
~~$P(Y1-6|P) =$~~

$P(Y7-Y8|x) = 0.3678$

$\frac{P(Y9-13|X)}{P(Y1-Y6|X)} = 2.87$

$P(Y1-Y6|X) = 0.~~1832~~ 0.1632$

$\frac{P(X|Y9-13)}{P(Y1-Y6|X)} = 1.36$

	Y1-6	Y7-8	Y9-13	T
X ~~P(W)~~	~~410~~ 31	70	89	~~199~~ 190
X' ~~P(W)~~	6	3	7	~~15~~ 16
Don't know	26	22	37	85
T	63	95	133	291

~~Let A be considered as a problem~~

Q3
70 | ~~80~~ all MAX

01068

Q4.

a) The proportion of school aged students can be estimated through counting the number of students present in the picture. This would act as one of the data values. This would be repeated for many photos to obtain a ~~do~~ statistical summary of proportion of school aged students. A margin of error can be calculated through $\frac{1}{\sqrt{n}}$, where n would be the amount of photos analysed.

~~Photos are not~~ ~~&~~ School aged children on photos may be overrepresented as media attention tends to focus on these groups. Therefore, the true proportion of school aged students may be lower. In addition, photos as a source are not representative of marches around the world. Marches at ~~a~~ high population density areas may have more media attention, and therefore the proportion observed would not be an accurate representation of the true population.

bi). This assumes that the number of people occupying a certain area is constant across the event. That is, the crowd is spread homogeneously. Secondly, this method assumes that there is independence between ~~of~~ ~~as~~ number of people in one area and in the next area. This is not likely true as certain areas would be more congregated than others and would lead ~~no~~ to correlation between areas.

ii). Area per square $= \frac{18000}{25} = 720$

No. of crowd $= (20.4 \times 720, 23.2 \times 720)$

$= (14688, 16704)$.

Because the confidence interval ~~does not include~~ for the mean size of crowd is less than 17000, this ~~so~~ gives no evidence to support the claim that over 17000 people attended. It is a fairly safe bet that the mean number of people attended is between 14688 and 16704.

X	0	1	2	3	4	5	6	7	8	9	10
Freq. Raw data	3	3	6	7	6	4	3	2	4	2	1.
Predicted data	0.6148	~~0.5563~~ 2.582	5.423	7.592	7.971	6.696	4.687	2.812	1.476	0.689	0.2894

$Po \sim (4.2)$.

The predicted values do not match the true values well. For example, the predicted data with poisson estimates frequency of 0.6148 ~~ear~~ for 0 weather related natural disasters per annum whereas the true vale is 3. In addition, ~~the mode~~ ~~the data~~ the data may be bimodal as there is a peak ~~at~~ ~~at~~ of 4 events of occurance of 8 weather related natural disasters. Therefore, it is not accurately modelled with Poisson distribution.

Weather events are likely to not be ~~be~~ independent, ~~on each year~~ as extreme events could correlate to high global temperatures for example. This violates the Poisson distribution assumption of independence. Additionally, the rate of occurence per annum is not fixed, and may vary from year to year. So a poisson model is not suitable.

6S

Q2b) (cont.)

This is a bootstrap test, and a form of inferential statistic which ~~and~~ can't infer a causal claim. ~~To infer that a causal claim. Conduct a rerandomisation test to test for~~ To make a causal claim, an experiment needs to be carried out. The treatment groups would be cows living in soils fertilised with karengo and cows which do not. ~~Random~~ (*) Record the difference in cow urine nitrogen levels and plot on a summary statistics for analyzing. To show statistical significance Conduct a rerandomization test, where differences are randomly allocated into each treatment group and a difference in mean is found. This is repeated up to 1000 times. The tail proportion shows how likely the observed experimental difference or more ~~value~~ can occur if chance is acting alone. If the tail proportion is less than ~~0.5~~ 0.1, there is evidence to suggest that the difference is statistically significant. Then, the claim in Report 3 would be supported as the use of karengo causes ~~that~~ a significant decrease in nitrogen content of cows' urine.

(*) Units: cows ~~are~~ randomly selected

Cows randomly allocated into treatment groups.

Dependent variable: Nitrogen concentration in cow's urine.

Overall design: Comparison of 2 independent groups.

93201A

Subject	Statistics	**Standard**	93201	**Total score**	28
Q	**Grade score**	**Annotation**			
1	8	The candidate has succinctly and accurately, using correct units and numerical evidence, described the features in the visualisations. They built into this a meaningful and insightful contextual conclusion about the impact of the use of fertiliser on land intensification. In their time series description they correctly identified in depth features at the outstanding level with features of the data that could impact the type of model used (additive or multiplicative)			
2	7	Although a strong response overall the candidate identified the incorrect tool to analyse the results of the experiment in (a)(ii). For this type of experiment the use of the re-randomisation was not appropriate. For the design of their own experiment to measure the impact of seaweed concentration of methane emissions the candidate has correctly used sufficient experimental design principles and followed this with a detailed description of the analysis required – in this case the re-randomisation test was appropriate.			
3	7	The candidate has described with the extra part of stratification to get a better representation of the population when taking a sample. The candidate did not correctly calculate the margin of error when comparing independent groups for part (a)(ii). All correct probabilities required to make a correct conclusion about older students thinking climate change is a problem compared to younger students were calculated and used to respond to the claim.			
4	6	Limitations and assumptions about the estimate for the total student count were well described and the claim about more than 17000 well responded to. The discussion about the appropriateness of the Poisson model for extreme weather events was not at the appropriate level. More than a comparison of theoretical and experimental probabilities was needed. The contextual descriptions of the requirements for a Poisson were not strong enough and the candidate failed to take into account the role of natural sampling variation in terms of explaining the spike of 8 extreme weather events over what was only 60 years worth of data.			