93201A

S

SUPERVISOR'S USE ONLY

OUTSTANDING SCHOLARSHIP EXEMPLAR

NZQA

NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

Scholarship 2014
Statistics

9.30 am Wednesday 12 November 2014
Time allowed: Three hours
Total marks: 40

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 all your answers in this booklet.

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

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

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

QUESTION NUMBER | Start each question on a new page. | ASSESSOR'S USE ONLY

1a. Diabetic rate vs obesity rate

The trend for the relationship between the diabetic rate and obesity rate is a linear trend. This trend has a positive association. It is a fairly weak relationship as there is a relatively large amount of scatter about the regression line. The scatter is non-constant. as there The graph shows us that for every ~~increase in the~~ 0.0344% increase in the diabetic rate in the data gathered by WHO and World fact book from 70 countries, there is an average increase of 1% in the obesity rate.

- Diabetic rate vs GDP per Capita

The trend for the relationship between the diabetic rate and GDP per capita is a linear trend with a positive association. The relationship is ~~very weak~~ moderate as, although there is some scatter, it is nowhere near as much as graph 1. There is one atypical point at $10,000 GDP per capita and 7% Diabetic rate. If we looked into this and found that it was in fact an outlier, removing this point would increase the strength of our relationship. The Scatter is non-constant. At small values of GDP Per capita there is a lot more scatter of data points compared to at higher values of GDP. The regression equation shows that for a

QUESTION NUMBER | Start each question on a new page. | ASSESSOR'S USE ONLY

0.6845% increase in the diabetic rate, there is an average of an $10,000 increase in the GDP per capita of the 70 countries

Life expectancy vs Diabetic rate

The trend for the relationship between the life expectancy and diabetic rate is a linear trend. This trend has a positive association. The relationship is the strongest of the three graphs as it has the least amount of scatter of data points. The graph has a constant scatter across the regression line. The graph shows us that for an increase in life expectancy of 2.0677 years, there is an average of a 1% increase in the Diabetic rate.

bi) Using the regression equation $Y = 0.6845x + 2.1082$ for GDP per capita of $5000.

$Y = 0.6845 \times 0.5 + 2.1082 = 2.43\%$

Using regression equation $Y = 0.0346x + 2.3718$ for obesity of 71%

$Y = 0.0346 \times 71 + 2.3718 = 4.81\%$.

The average diabetic rate for this country is $\frac{2.43 + 4.81}{2} = 3.62\%$

The prediction for of 2.43% diabetic rate for $5000 GDP per capita is not very precise due to there being large amounts of scatter. By looking at the graph, the range of values at $5000 GDP per capita is from about 0.5% to 4%. However, I am more confident with this prediction than I am with my prediction of 4.81% when

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

there is an obesity rate of 71%. This is because 71% is outside the data range meaning I have to extrapolate, giving me less confidence. it is likely that the graph will begin to flatten off outside data range as there is a limit to the diabetic rate % therefore giving me little confidence with the prediction of 4.81% ~~obesity~~ diabetes rate for a country with 71% obesity. I am more confident with the prediction of 2.43% than 4.81%

b(ii) Although the GDP per capita is only $5000, and the percentage of the population is obese is 71%, according to the percentages calculated the diabetic rate in Nauru will still be relatively low. The country has a very small population 9434 with a large unemployment rate 90% indicating that the Standard of living in Nauru is very low. The average life expectancy is only 25.3 years. ~~Which could indicate why the predicted diabetic rate calculated above is still fairly low when compared to other~~ Although we would expect a relatively low ~~than~~ diabetic rate of 2.43% calculated above, their actual diabetic rate is 22%. The ~~A~~average diabetic rate in Nauru is very large when compared to the expected diabetic rate. The small population size will result in decreased variation

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

within the population which may ~~mean that~~ explain this large diabetic rate. There are other factors then just obesity rate of 71% and gdp of $5000 in Nauru which will impact this diabetic rate, indicating that there are some confounding variables which have not been looked at, but actually affect diabetic rate as well. These include median age, unemployment rate, population, Location etc. These could also be affecting diabetic rate so therefore there are underlying factors present which are also affecting diabetes rate ~~rate~~ which explains Nauru's relatively large diabetes rate of 22%

fe

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

c) (2.6) Although at first glance it may appear that the obesity rate has increased by 2.1%, ✓1 this may not actually be the case due to the confidence intervals overlapping. ✓1 (25.5 - 27.5 overlaps with 27.4 - 29.8) ✓2 meaning that ✱ it is possible that in fact the obesity rate has remained the same or even decreased slightly. S S

Q1 SP05 Q.1. 8

S05 5

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

2ai) Random assignment was used to form the 2 groups as it will help balance out the traits of the students. It will result in their being less sampling variability between the groups, meaning we will be presented with more accurate results.

ii) From this re-randomisation output, we can see that these results result of a mean difference of at least 0.137 proportion of overweight students will only occur by chance 8.4% of the time. 8.4% is less than 10% so therefore we can make the claim that those who had non-curricular activities were more likely to not be overweight/obese compared to those without extra curricular activities as this only occurs due to chance 8.4% of the time meaning we can be somewhat confident with this claim as the tail proportion is relatively small

P

P

bi) Both of these box and whisker plots are roughly represented by normal distributions, as the mean and median are very similar.

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

more similar for survey 1 with a difference of 0.1 than survey 2 ~~with~~ with 0.3. There are tails at either end of the graphs, and they are both roughly symmetrical apart from both having a unusual amount of data at 0 sugar containing drinks. This is due to some participants not drinking those types of drinks at all. In the first survey, the mean ~~was~~ number consumed ~~was~~ ~~low~~ of 4.9 drinks was larger than in the second survey after the advertisement of 4.3 drinks per secondary school student on average. The lower quartile in both surveys was 3 drinks whereas the upper quartile in survey one was 7 but the upper quartile in survey 2 was only 5.5 this indicates the range of the middle 50% of data points for ~~survey group~~ 1 was larger than that of group 2 by 1.5 drinks. The overall range of ~~group~~ survey 1 was also 1 drink larger than that of group two. By looking at the graph we cannot make a call using the $\frac{1}{2}, \frac{3}{4}$ rule as the mean of survey 1 (4.9 drinks) is not larger than the upper quartile of ~~group~~ survey 2 (5.5 drinks) so we cannot make the claim whether or not the advertising worked. In order to do this we would

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

need inferred confidence intervals, or to bootstrap our data.

c) Bootstrapping works by ~~taking resampling from~~ taking smaller samples from the sample taken, and recording the mean ~~fake~~ of difference a large number of times. The samples are done with replacement so the same person could be selected twice within the same sample. After this has been done many times, we cut 2.5% of data off each side of the graph we are given. In this case students would have been randomly picked and recorded, then the mean difference recorded, and then repeat the process many times. ~~After doing that~~ This is done to reduce sampling variation in the initial sample. It gives us a 95% confidence interval of by how much more ~~gro~~ sample 1 drinks than sample 2 which we can use to infer about the whole secondary school student population

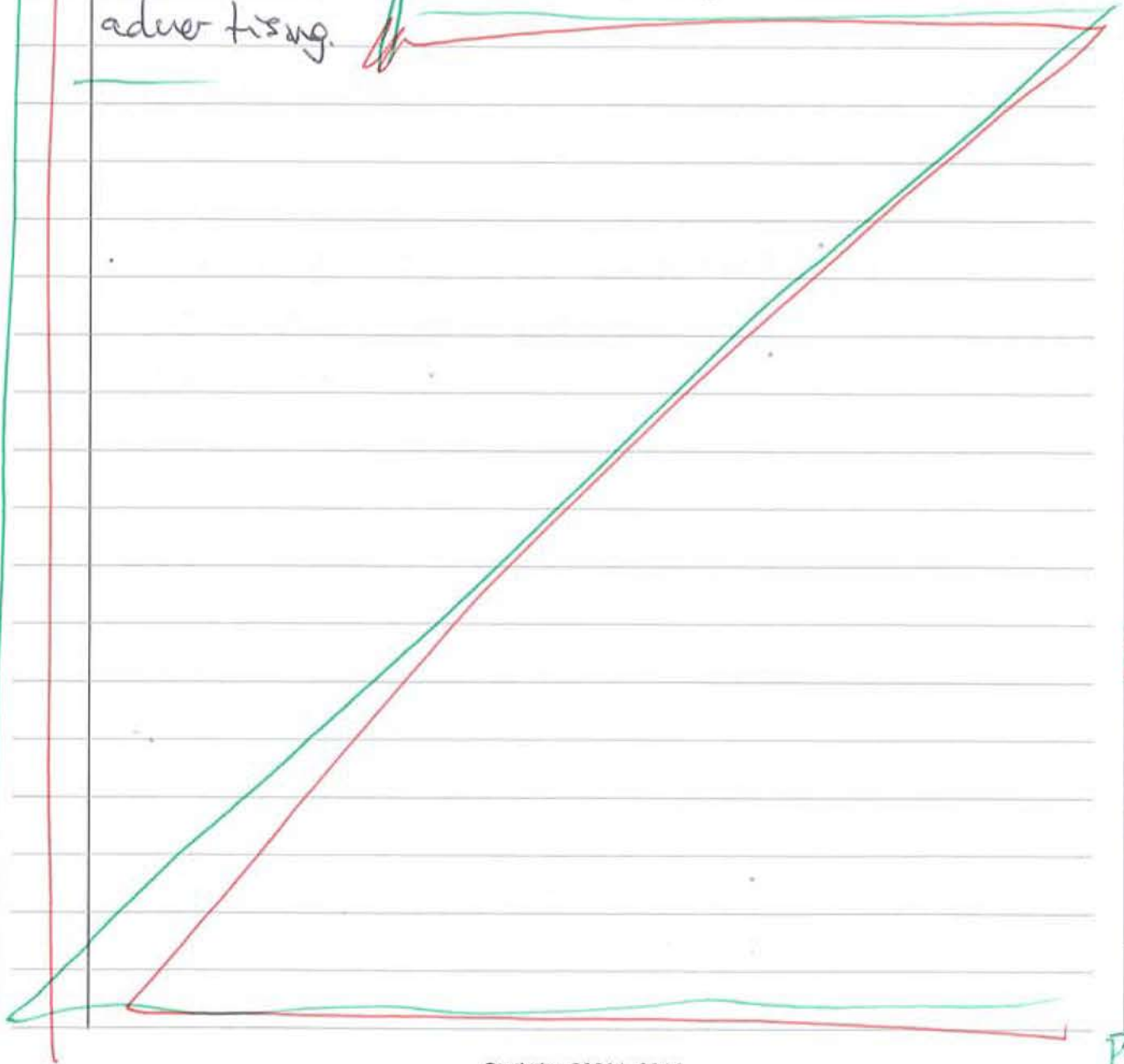
iii) From this output, we are given a 95% confidence interval of 0.14 to 1.24. This shows us that sample 1 could drink up to 1.24 more sugar containing drinks than sample 2, or as little as

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

0.19 sugar containing drinks more than sample 2. Therefore we can claim that the advertising has worked as, with 95% confidence, we can say the reduction of sugar drinks for all secondary students in New Zealand is between these values of 0.19 and 1.24 drinks. We can say this as the interval does not straddle zero meaning in no case does the number of drinks on average increase after advertising.



Q2

O+P

7

P+S+P

5

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

3a) Obesity rates between male and female.
Over the 2004-2013 period, the obesity rate for male and female has been increasing. Females tended to have an average of about 0.8% more obese females when compared to males, however in 2012 both males and females had obesity rates of 28.6%.. Males have been steadily increasing at a roughly constant rate of about 0.4% per year, with a slight dip in 2009. females however have been much more volatile and, although overall has increased by 0.3%, they keep increasing and decreasing in obesity rate. ~~Both males and~~ females have their largest decrease in 2009 ~~(females fell by~~ of a 0.8% decrease, and this is also the year when the rate of increase in obesity of males fell slightly to 0.2%. This is likely due to the Global financial crisis making households spend less on food and luxury items.

For age group over these 10 years, there has been steady increases for those 15-24 and 35-44, and a steady decrease for those 55-64. The difference between 35-44 and 15-24 at any given time is on average 14.3% more for the 35-44 group. Up until the 2012 year, 55-64

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

year olds were consistently higher than 35-44, but due to the decreasing trend, ~~in 2012~~ the between 2011 and 2012, the obesity rate in 35-44 became larger than 55-64 at 32.6% compared to 32.4% for 55-64 in 2012. The gap between the obesity rates of 15-24 and 55-64 is also narrowing. There was a large spike in 2008 obesity levels for 35-44 year olds. This is most likely due to the Global financial crisis putting stress on this age group, increasing diabetes rate. This age group was affected the most by this as young people who do not work and ~~treat~~ have, and older people who are retired/have saved up will not be impacted too much whereas those who are in the middle of their careers will be.

b) For men:

~~predict the equation~~ By extending the male curve, I would expect males to have an obesity rate of 30% in 2015. CSO

For Women:

By extending the female curve I would expect females to have an obesity rate of about 30.5% CSO in 2015.

Predicting into the future is always ~~risky~~

QUESTION NUMBER | Start each question on a new page. | ASSESSOR'S USE ONLY

going to contain some accuracy as we have not yet got any concrete data points to go by. My prediction of 30% for males however, is much more valid than the 30.5% obesity rate for females. This is due to males increasing at an almost completely linear rate over the past 10 years, meaning this trend is likely to continue giving me confidence in this prediction. My women prediction however is much less valid as across the course of the 10 years, there have been many fluctuations making it hard to predict into the future as it is very possible these fluctuations will continue to occur, making it much harder to ~~predict~~ predict in the future, and therefore making my 30.5% obesity rate in 2015 somewhat invalid 2P 2P.

ci) ~~One reason~~ One reason why the obesity rates for women have fluctuated more than men over these 10 years could be attributed to the ~~global~~ economy. In times when the economy has been worse like in 2008/2009, the obesity rates for women has fallen, and at booms in the economy it has risen. This ~~can~~ means that, when in a downturn of economic crisis like in 2008/2009, couples will be much less

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

likely to have babies as their future outlook will be bad, so less babies will be born, and as babies give their mothers baby fat and less are being born, the baby fat gained by women will decrease in these times decreasing women obesity. The opposit occurs when the economy is in a good position

ii) The surveyors would have taken samples of the population and recorded their results to get these numbers. The reason for the 0.6 % points difference between the two is likely due to sampling variation. Both surveyors would have taken ~~same~~ different people from the ~~45~~ 45-54 age group which means that due to the nature of sampling these groups will have different characteristics. This is why the results were conflicting. The more people you take to sample, the less sampling variability that will occur, so it is likely that are,

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

if not both surveyors took relatively small sample sizes resulting in high sampling variability

d) For the 15-24 age group, the obesity rate has been increasing ~~at about~~ steadily over the 10 years at about 0.9% per year. Overall there has been an increase of 9.7% across the 10 years
For the 35-44 age group, the obesity rate has been increasing over the five years at about 0.4% per year, with a spike in 2008/2009. Overall there has been an increase of 5.2%.
For the 55-64 age group, the obesity rate has been decreasing by an average of 0.6% per year by 6%.
Overall, 15-24 age group increased the most by 9.7%, 35-44 age group increased, but by much less (5.2%) which can mainly be attributed to 2008/2009, and the 55-64 age group were the only group to decrease and they decreased by 6% obesity rate.

5
5

Q3
&3
5+5+2p

6

5+5+2p 6

QUESTION NUMBER | Start each question on a new page. | ASSESSOR'S USE ONLY

(a) Using binomial:

- There is a discrete number of people (cannot have half-people)
- Each either are obese or are not obese
- Fixed number of people between 65-74
- Assume one person being obese is independant to another person being obese
- There is a constant probability that a person is obese - this probability is 0.386.
- $\lambda = 0.386$ $x = 360$

using binomial distribution, the probability that 360 people or less out of 1000 are obese is: 0.0484

S S

(b)

	Obese	not obese	total
65-74	0.386		
75 cont. 65			
total	0.313		1

number of people in 65-74 age group is ~~0.14 × 1115000 = 156100~~

From those 156100 × 0.386 = 602345 are obese. total number of obese people

Population size is $\frac{1115000}{0.313} = 3562300$ people

From those people, 0.14 × 3562300 = 498722 are between ages of 65-74 group.

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

From this 498722, 38.6% are obese.
498722 × 0.386 = 192507 people
192507 are 65-74 and obese.

$\frac{192507}{1115000} = 0.1727$

0.173 randomly chosen is the probability that an obese person is in the 65-74 category is 65-74

c)

	obese	not obese	total
65-74	0.054	0.086	0.14
>65 but ≤15	0.259	0.601	0.86
total	0.313	0.687	1

the percentage of the population that is obese and 65-74 is $\frac{192507}{3562300} = 0.054$

the likelyhood a person 65-74 is obese is $\frac{0.054}{0.14} = 0.385$

the likelyhood a person not in that age is obese is $\frac{0.259}{0.86} = 0.301$.

The relative risk is therefore $\frac{0.3857}{0.381} = 1.28$. This means that it is 1.28 times more likely that a person who is 65-74 is obese when compared to a person who is not in this age group.

Q4 2s+0

2s+c 8

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

5a(i) Overweight

- ~~Adipting in Brazil the over for~~ when comparing Brazil to China for 6-9 year olds, the brazil obesity rate increased by far more as shown in the table. Brazil increased from 4.9 ± 0.18^3 to 17.4 ± 1.35^{88} so increased by ~~about~~ an average 12.5% over the years whereas China 6-9 year olds increased from 10.5 ± 1 to 11.3 ± 1.11 by about 0.8% over the 23 years. This shows that ~~China's to~~ Brazil's obesity rate for 6-9 year olds has increased by about 16.7% more when compared to China ✓½
- Overall, for ~~6-18 year olds~~ 10-18 year olds, the obesity rate in USA has increased ~~from~~ from 16.8 ± 0.86^{89} to 27.3 ± 1.47^{94} so an increase of about 10.5% over the years. In Russia the obesity rate has ~~increased~~ decreased from 11.5 ± 0.46^{8} to 8.5 ± 0.74 giving about a 3% decrease ✓ This shows for USA, the obesity rate in young people has increased whereas in Russia it has decreased ~~by~~ by 3%. ~~If we were to~~ ~~create~~
- If we were to create confidence intervals, for B.P are, ~~USA's~~ Brazil would not

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

~~overlap~~ overlap, however China's would mean it is possible that the obesity rate did not actually increase (hence only a small increase in report)

For Bullet point 2, Russia's confidence interval would not overlap and neither would USA's meaning it is likely that the changes explained above have occurred

- For Children 6-9 in China, the confidence interval before is $9.5 < \mu < 11.5$ and after is $10.19 < \mu < 12.41$ so there is some overlap meaning it may not have increased at all, or only slightly. When compared to Russia's interval of $25.35 < \mu < 27.45$ to $8.89 < \mu < 11.51$ which shows ~~very~~ we are confident that overweight has decreased. So overall, China's overweight for 6-9 has remained similar whereas Russia's has decreased

Underweight

- For 6-9 year olds, the amount of underweight in Brazil has gone from $12.03 < \mu < 12.57$ to $5.31 < \mu < 6.89$. No overlap so confident it has decreased. Compared to Russia which has gone from $6.48 < \mu < 7.72$ to $6.81 < \mu < 9.19$ so there is overlap so no concrete evidence, however, it is likely to have increased in Russia for this age group, but not certain when compared

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

to brazil

When comparing USA to Russia for 6-9 year olds, USA has gone from $3.56 < \pi < 4.64$ to $2.73 < \pi < 4.07$ there is overlap so cannot be confident underweight has actually fallen. Compared to Russia which has gone from $6.48 < \pi < 7.72$ to $6.81 < \pi < 9.19$ overlap here too so cannot say underweight has increased/decreased for these two countries with confidence for 6-9 year olds

10-18 year olds in Brazil compared to 10-18 year olds in China. For ~~Brazil~~ China underweight went from ~~$15.84 < \pi < 16.36$~~ to $8.95 < \pi < 10.25$. No overlap so confident that it decreased. For 10-18 Y.O's for ~~China~~ Brazil it went from $15.84 < \pi < 16.36$ to //

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

b) For ~~Figure~~ table 1, it could be improved by actually displaying confidence intervals which will allow them to be read more easily in order to identify whether or not the percentage has actually changed with confidence, or if there is overlap meaning we cannot be confident that a change has occurred. 1/6

For ~~Figure 2,~~ & figure 1, it would have been good if they had told us what they classify (low medium and high earners as (what income level they are receiving) so we can better see the and understand the changes occurring. 2s

They could also have an extra bar showing the change making it more easy to read. 2s

c) For figure 1, if they were to introduce another bar for ~~the~~ the change in weight, it would ~~be more fit~~ support the argument in each case, as it will give us actual figures for the changes

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

allowing us to identify the changes and compare to other countries in order to make claims like "tripling in Brazil, doubling in USA, increased by $\frac{1}{5}$ China" for example as the figures will be readily available.

- If confidence intervals were displayed in table 1, it would make it easier to support the argument, as we would be ~~not~~ able to see the actual changes occurring in the 4 countries and give us confidence with our ~~predictions~~ arguments. It will make ~~the~~ comparisons easier. Likewise with figure 1.

Qs

Q5.

6

Outstanding Scholarship

1. An O was score here due to a correct prediction, a justification that the best choice of line to make this prediction had been chosen, a correct comment about the precision of the prediction along with an appropriate comment about Nauru.
2. Both the increase in the point estimate was stated along with the interval overlap leading to a correct conclusion.
3. Despite Q2 (a) (ii) being correct, the answer to why random assignment was used to form the two groups was incomplete.
4. This question part was moderated down from an S to a P due to an insufficient comparison of the difference in spread between the first and second surveys.
5. The key word "mean" was missed out so this answer was moderated downwards from an O to a P.
6. The prediction for the male obesity rate was correct along with the justification for making that prediction. There was no account taken in the fluctuations of the women obesity rates in making the other prediction.
7. Both points in (c) were correct however the percentage differences were stated rather than the percentage changes in the two end points for each age range.
8. All conditions for the Binomial were stated correctly in context along with a correct answer.
9. Four comparisons were deemed correct for comparisons of over-weights (2) and under-weights (2) between two countries all observed from the Table.
10. In (b) and (c), the improvement brought about by constructing confidence intervals for the differences along with an appropriate reason gave two further points so 2S was scored for the complete question.