93201A

932011

S

12 · 14

SUPERVISOR'S USE ONLY

TOP SCHOLAR

NEW ZEALAND QUALIFICATIONS AUTHORITY
MANA TOHU MĀTAURANGA O AOTEAROA

Scholarship 2015
Statistics

9.30 a.m. Thursday 12 November 2015
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

1 a. From March 2010 to ~~December 2014~~ March 2015, monthly median house prices for Auckland have ~~generally increased~~ increased from \$475,000 to \$675,000. This is an increase of \$200,000 over a period of 60 months or an average increase of \$3,333.33 a month. The trend is however, not a linear one. As can be seen from the 12 month moving average, prices increased fairly slowly from March 2010 to March 2012 and then increased much more rapidly from March 2012 to March 2015. There was only an average increase of around \$20,000 using the moving average line from March 2010 to 2012 compared with an increase of \$170000 over March 2012 to March 2015 (in 12 month MA).
There are also regular dips in the ~~house~~ median house prices around the January mark of each year (particularly obvious in Jan 2013, Jan 2014) and regular peaks around ~~November~~ Oct/Nov of each year (particularly at ~~Sep~~ Oct/Nov 2012, 2013) so a clear seasonal pattern is evident.
Sales volume in the Auckland region fell from ~~April~~ a 12 month MA of 2000 in March 2010 to around 1,650 in March 2011, before climbing steadily to a 12 month MA of 2,850 in December 2013 and then falling again to just under ~~20~~ 2,500 in March 2015.
There is a ~~clear~~ somewhat cyclical pattern evident similar to what is expected as part of the natural business cycle however even though median prices are increasing ~~as~~ from Dec 2013 to March 2015, sales volume is falling. There is also a clear seasonal trend in

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

sales volume as regular troughs occur ~~in~~ around January of each year, down to a low ~~of~~ of around 1,100 houses in Jan 2014. Regular peaks also occur in March of each year, followed by a trough in April and another peak in May, so there is strong evidence to suggest a seasonal pattern. The highest peak is at March 2013 with sales of 3,400 houses.

a 4 a 4

1. b. ~~~~ For ~~~~ North Shore median house prices have increased 17.6% from $744,000 in Feb-14 to $875,000 in Feb-15. Volume of houses sold however only increased from 378 to 392, a 3.7% increase. This resulted in an estimated total sales value increase ~~~~ from $281,232,000 to $343,000,000, an increase of 21.96%.

For ~~~~ Waitakere median house prices increased 17% from $530,000 to $620,000, volume of houses sold rose 16.7% from 293 to 342 giving an ~~~~ estimated total sales value ~~of~~ ~~~~ increase from $155,290,000 to $212,040000 an increase of 36.54%.

For Manukau median house prices increased 9.6% from $575,000 to $630,000, volume of houses sold rose 6.1% from 440 to 467, giving an estimated total sales value increase from $253000000 to $294,210,000, an increase of 16.29%. So while ~~North~~ Shore had

b 3 b 3

MAX 6 max 6

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

the biggest % increase in median price Waitakere had the biggest % increase in estimated total sales value from Feb-14 to Feb-15.

1. c. i. The overall trend must be taken into account, as should the seasonal component. If possible the cyclical component and residuals should also be quantified and form part of the forecast. ci ci 1

ii. Extrapolating the overall trend would be the first step to get to ~~the desired year and month~~ each month in 2016. Then, the variation of each month from the trend should be added on or taken off. This can be calculated by averaging the seasonal variation of each month over the last 5 years to get an average seasonal variation. The 12 forecasts for each month can then be added together to give the forecast for total sales volume in 2016. cii 2 cii 2

Q1. 8*

Q1. 8*

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

2. a. Regional markets showing strength can be seen by the phrase "Compared to April 2014 all regions recorded increases in sales volume, with Northland recording the largest of 67.8% followed by Waikato/Bay of plenty with 53.2% and Central Otago Lakes with 41.1%", Therefore regional sales volumes are increasing compared to the same month last year. Secondly, median house ~~prices~~ prices also seen to be increasing based on the Housing Price Index. ~~The~~ Despite Wellington & Christchurch HPI falling other North Island and other South Island rose 3.2% and 7% respectively over 12 months, showing strength in regional markets.

a 2 a 2

2. b. ~~a~~ Median prices are used as they are not ~~aw~~ affected by outliers as mean prices are, otherwise the selling of a few expensive houses would change the ~~[illegible]~~ findings.

b 1 b 1

2. c. Seasonally adjusted values adjust for buying patterns that vary monthly and so allow us to see if there are more ~~people~~ houses being sold than the previous month. If raw data was used it may just be due to natural reasons that people ~~[illegible]~~ ~~sell~~ buy more houses in March than April (eg due to Easter). So a fall in raw values may not necessarily mean the trend is decreasing, and using seasonally adjusted values lets us see the trend.

c 1 c 1

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

2. d. Because there is so much more demand for houses in Auckland (seen by the high increase in ~~price~~ median prices) ~~a~~ buyers are more likely to put a house on auction in Auckland, where they are likely to receive the maximum ~~cost~~ amount for their property. d 1 d 1

2. e. House prices ~~along~~ of different ~~regions~~ regions tend to vary differently from each other as they are influenced by different factors. A stratified index allows us to see which regions are experiencing the most and least growth. e ϕ e ϕ

2. f. The percentage changes are based on a census of houses sold, not a sample so all possible data points are taken into account so there is no sampling error and hence no margin of error. f 1 f 1

MAX 6
MAX 6

2. g. Both these measures median price and sales volume show growth and strength in the market but some regions experience more of one than the other. Quoting both also allows us to see if there is an excess of supply or demand in the market. g 1 g 1

2. h. The HPI for Auckland in 2014 April would be $\frac{612{,}000}{450{,}000} \times 100 = 136$. The ~~da~~ extract then says the HPI for Auckland rose 18.9% from April 2014 to 2015 so the new value is $(136 \times 0.189) + 136 = 161.7$. HPI for ~~eg~~ Auckland in April 2015 is 161.7. h 2 h 2

8 8*

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

B. a. i. For both Auckland and Christchurch the number of people per occupied house is distributed with a right skew, and the skew for Auckland is more extreme. Neither distribution is symmetrical. The ~~de~~ number of people per occupied house is also much more spread out for Auckland than it is for Christchurch with Auckland having a larger interquartile range than Christchurch ~~by~~ ~~to a~~ (Auckland's ~~IQR is~~ interquartile range is 3, Christchurch's is 2). Auckland also has a much larger standard deviation than Christchurch with a standard deviation of 2.519 compared to Christchurch's 1.605 giving more strength to the claim that the data for Auckland is more spread out. Finally Auckland ~~~~ has higher point estimates for both the mean and median number of occupants; a mean of 3.947 ~~comp~~ and a median of 4 compared to Christchurch's mean of 3.246 and median of 3, showing us that there tends to be more occupants per house in Auckland. Auckland also has a significantly higher maximum, 18 compared to Christchurch's 11.

ai
2
ai
2

~~a~~ ii.

The distribution of the data in Figure 5/Table 3 is approximately normally distributed and symmetrical compared to the asymmetrical, ~~~~ right skewed distribution of Figure 3/Table 2. The data for mean number of people for each Auckland suburb is also much less spread out, with an interquartile range of only 0.5 and a standard deviation of 0.563 compared to Table 2's IQR of 3 and standard

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

deviation of ~~9~~ 2.519. The central measures are also much lower, ~~than~~ when looking at the means for each suburb, with a mean of only 3.0 and a median of 2.9, compared to Table 2's mean of 3.947 and median of 4. aii 1 Aii 1

3. a. iii The mean of 3.0 from Table ~~2~~ 3 has reliability in that ~~a~~ census data was used as opposed to a sample, however a mean was calculated for each suburb and then a mean of the means was calculated. The suburbs were ~~are~~ ~~not~~ weighted equally, whereas it is completely plausible that certain suburbs are much bigger and have bigger households, and these suburbs would be under-represented with only one data point as in figure 5. ~~Figure~~ 3 The mean of 3.947 in table 2 ~~is likely to be a much better~~ would be likely to be a good estimate, however it is unknown how the data was chosen and whether the houses chosen were representative of all of Auckland based on the size of Auckland, 488 is also a ~~very~~ small sample size. ~~so large sample~~ Because of this, table 3 is actually a better estimate of the mean as it uses a census and so represents all ~~the~~ suburbs fairly. While suburbs are not ~~ever~~ weighted by size there are so many of them that any size difference would be insignificant, therefore 3.0 is a better estimate of the mean. Aiii 1 Aiii 1

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

3\. b. i. There is a weak ~~moderate~~ positive correlation between asking price and number of bedrooms however there is also a lot of scatter in the graph and a lot of variability. There appears to be a moderate positive correlation between floor area and asking price from a floor area of 50 - 250 m^2 but there is still a fair bit of scatter. After 250 m^2 there is little correlation ~~betw~~ and a higher degree of scatter. There is a very strong positive correlation between ~~and~~ asking price and capital value with only one ~~point clear~~ possible outlier at \$500 000 capital value and \$1 250 000 asking price. There is ~~almost~~ no correlation between asking price and land area, although most properties tend to have less than 10 000 m^2 of land area with only 3 data points above this.

bi 3

bi 3

MAX 6

MAX 6

3\. b. ii As there is a lot of scatter for the asking price vs floor area graph, and a weaker correlation, I would use the asking price vs capital value graph ~~and draw~~ create a regression line and substitute 1,000,000 into the equation of it to get an estimate of the asking price. If a regression line was unavailable the price can be estimated by the eye at about \$1,250,000 ~~\$~~ based off a \$1,000,000 capital value. This ~~also~~ choice is due to the ~~high~~ strong correlation and low degree of scatter.

bii 1

bii 1

~~3c~~ b ~~iii~~

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

3. b iii The age of the house is a ~~data~~ numerical variable that ~~might~~ would affect selling price and the location of the house is a categorical ~~statistic~~ variable that would affect selling price.
As the age of the house increases you would expect the asking price of the house to decrease. Therefore there would be a negative correlation (likely weak or moderate) between asking price and the age of a house.
The location of the house would likely affect the selling price as bigger cities (eg Auckland) or wealthier suburbs (eg North Shore) would be expected to have a higher ~~sel~~ asking price than smaller cities and poorer suburbs.

b iii 2

b iii 2

Q3 8*

Q3 8*

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

4. a. i. $P(\text{1 or 2 bedrooms} \mid \text{Free-standing}) = \dfrac{P(\text{1 or 2 bedrooms} \cap \text{Free-standing})}{P(\text{freestanding})}$

$P(\text{1 or 2 bedrooms}) = 0.28$

~~P(1 or 2 be~~

$P(\text{1 or 2 bedrooms} \cap \text{Apartment}) = 0.28 \times 0.74 = 0.2072$

So 0.2072 of houses are 1/2 bedroom apartments and 0.28 of houses in total are 1/2 bedrooms so proportion of free-standing 1/2 bedroom houses is $0.28 - 0.2072 = 0.0728 = P(\text{1 or 2 bedrooms} \cap \text{freestanding})$

$P(\text{1 or 2 bedrooms} \mid \text{Freestanding}) = \dfrac{0.0728}{1-0.28}$

$= 0.1011$

ai 2

ai 2

4. a. ii. $P(\text{Apartment} \mid \text{1 or 2 bedrooms})$ required
$P(\text{Freestanding} \mid \text{1 or 2 bedrooms})$ required

Probability table

	1 or 2 bedrooms	3 bedrooms	4+ bedrooms	Total
Apartment	0.2072	0.056	0.0168	0.28
Freestanding	0.0728	0.6192	0.028	0.72
Total	0.28	0.6752	0.0448	1

Probability required for how many times more likely that a 1/2 bedroom house is an apartment rather than free standing is: $\dfrac{\dfrac{P(\text{Apart} \cap \text{1/2 bedrooms})}{P(\text{1/2 bedrooms})}}{\dfrac{P(\text{Freestanding} \cap \text{1/2 bedrooms})}{P(\text{1/2 bedrooms})}}$

$= \dfrac{0.2072}{0.0728} = 2.846$ times more likely

aii 2

aii 2

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

Therefore it is 2.846 times more likely that a 1 or 2 bedroom house is an apartment rather than a free-standing house.

4. b.i. Estimating that half of the houses in each bracket sold ~~from~~ in the top 50% of the bracket:
~~We want P(~~ Let X = selling price
We want $P(450{,}000 < X < 750{,}000)$

$$= \frac{\frac{68}{2} + 86 + 61 + \frac{23}{2}}{325} = 0.5923$$

So 0.5923 of houses sold for between \$450,000 and \$750,000.

bi 2

bi 2

4. b.ii ~~Two~~ Two different models that could be used are the triangular distribution model and the normal distribution model. The triangular distribution model works as we have a minimum and maximum value as well as a mode and the data is continuous. The parameters would be a minimum of \$200,000 a maximum of \$1,000,000 and a mode of \$550,000. For the normal distribution we must calculate a mean and standard deviation.

$\mu = \sum x \times P(X = x)$ using midpoints:

$$\mu = 250{,}000 \times \tfrac{32}{325} + 350{,}000 \times \tfrac{48}{325} + 450{,}000 \times \tfrac{68}{325} + 550{,}000 \times \tfrac{86}{325} + 650{,}000 \times \tfrac{61}{325} + 750{,}000 \times \tfrac{23}{325} + 850{,}000 \times \tfrac{5}{325} + 950{,}000 \times \tfrac{2}{325}$$

$\mu = \$510{,}000$

$\sigma = \sqrt{E(X^2) - (E(X))^2}$

$\sigma = \$149{,}460.57$

So the mean is \$510,000 and the standard deviation is \$149,460.57

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

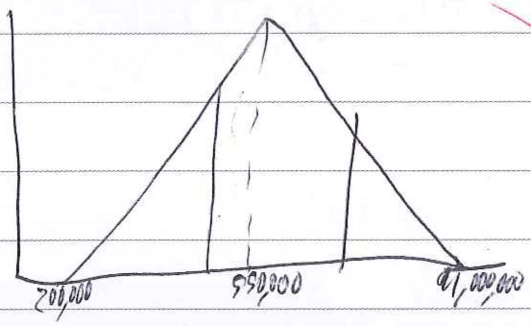
For the normal distribution,

~~From the triangular distribution~~

Using the triangular distribution:

~~Ple~~ let X be selling price.

$P(450{,}000 < X < 750{,}000) = P(450{,}000 < X < 550{,}000)$
$+ P(550{,}000 < X < 750{,}000)$.



Using the equations: $a = 200{,}000$, $b = 1{,}000{,}000$
$c = 550{,}000$.

$P(X < 550{,}000) = \frac{1}{2} \times 350{,}000 \times \frac{2}{800{,}000} = 0.4375$.

$P(X < 450{,}000) = \frac{1}{2} \times 250{,}000 \times \frac{2(450{,}000 - 200{,}000)}{(800{,}000) \times (350{,}000)} = 0.2232$

$P(X > 550{,}000) = 1 - 0.4375 = 0.5625$.

$P(X > 750{,}000) = \frac{1}{2} \times 250{,}000 \times \frac{2(250{,}000)}{800{,}000 \times 450{,}000} = 0.1736$.

$P(450{,}000 < X < 750{,}000) = \big(P(X < 550{,}000) - P(X < 450{,}000)\big)$
$+ \big(P(X > 550{,}000) - P(X > 750{,}000)\big)$.

$= (0.4375 - 0.2232) + (0.5625 - 0.1736)$
$= 0.6032$.

Using the normal distribution now:

let X be selling price

$X \sim N(510{,}000, 149{,}460.57)$

$P(450{,}000 < X < 750{,}000) = P(X < 750{,}000) - P(X < 450{,}000)$

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

$P(X<750{,}000)$: $z = \frac{x-\mu}{\sigma}$

$z = \frac{750{,}100 - 510{,}000}{149\,460.57} = \cancel{1.6064}\ 1.606$

$P(z<1.606) = P(X<750{,}000) = 0.9458$

$P(X<450{,}000)$: $z = \frac{x-\mu}{\sigma} = \frac{450000-510000}{149{,}460.57} = -0.401$

$P(z<-0.401) = P(z>0.401) = 1 - P(z<0.401)$

$= 1 - 0.6558 = 0.3442$

bii 3

$P(X<750{,}000) - P(X<450{,}000) = 0.9458 - 0.3442 = 0.6016$

bii 3

So $P(450{,}000 < X < 750{,}000) = 0.6016$ ✓

Both the triangular distribution and the normal distribution appear to fit the model well as 0.6016 and 0.6032 are close to the value of 0.5923 obtained by ~~a~~ calculation.
The triangular distribution was $\frac{0.6032-0.5923}{0.5923} \times 100 = 1.84\%$ off the calculated probability while the normal was $\frac{0.6016-0.5923}{0.5923} \times 100 = 1.57\%$ of the calculated probability, making the normal distribution have a slightly better fit. //

Q4
8*

Q4.
8*

QUESTION NUMBER

Start each question on a new page.

ASSESSOR'S USE ONLY

5. a, The real estate agents who went on course X appear to have ~~so much large~~ an approximately normally distributed difference in before/after ~~times~~ scores whereas the distribution for those who went on course Y has a left skew. The differences in ~~times~~ scores for those that went on course X was far less spread out than those who went on course Y, with an interquartile range of just 3.5 for those ~~are~~ on course X and 7.25 for those on course Y. Additionally, the standard deviation for course X differences was only 3.66 while for course Y ~~it~~ was 5.03, showing as the data is more spread out for course Y. ~~Both~~ Course X appears to have made more of a difference to the scores than course Y, with a mean increase of 6.5 and a median of 5.5 compared to course Y's mean of 3.35 and median of 4.5. Some who did course Y also worsened by up to 6 points.

a 4 a 3

5. b. i. Random assignment was used to form the two groups in an effort to eliminate any confounding variables and hope to achieve an ~~even~~ equal balance of these other characteristics that could affect score improvements, such as motivation, age, ~~ability and~~ natural ability to learn and experience.

bi R bi

5. b. ii By looking at both scores ~~the can~~ the improvement caused by the course can be analysed. ~~or without~~ ~~as oppose~~ A high score can be caused by an agent already being good at their job or a

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

low score by an agent being terrible, so by looking at before and after we can see the change caused by the course.

bii 2

MAX 6

5. b. iii. The randomisation test output shows that those who went on Course X had ~~a~~ mean improvement in score that was 3.25 points greater than those who went on Course Y. This is justified by the re-randomisation which has a tail probability of just 0.017. This means that a mean ~~a~~ difference of 3.25 points occurs due to chance alone in only 17 out of 1000 re-randomisations, making it very unlikely that chance is acting alone. Therefore we can conclude that Course X is more effective than Course Y at improving real estate agent scores.

biii 3

5. b. iv. It is likely that there is some positive correlation between before and after scores for the agents on Course X but it is unlikely to be strong, more probably weak/moderate. This is because there was a fair bit of variation in the differences in scores in table 6, and the correlation would reflect this. Similarly, for those on Course Y, there would be even less of a correlation as there is much more variation in the data for Course Y. A positive correlation, if any existed would be very weak.

biv 1

Q5 8*