**FACULTY OF BUSINESS AND LAW**

**Department of Information Systems and Business Analytics**

**MIS770 – Foundation Skills in Data Analysis**

**Name of the Student :**

1. **An overall summary of the house prices in Kingfisherbay? In addition, the media articles focus on median house prices and not the mean. I have never been able to understand why this is so; surely the “average” is the “average”. Can you clarify that for me?**

The mean and the median are outline estimates used to depict the most "run of the mill" esteem in a lot of qualities. Analysts in some cases allude to the mean and median as proportions of focal inclination. To locate the median, we organize the perceptions all together from littlest to biggest worth. In the event that there is an odd number of perceptions, the median is the center worth. On the off chance that there is a significantly number of perceptions, the median is the normal of the two center values. The mean of an example or a populace is registered by including the entirety of the perceptions and isolating by the quantity of perceptions. As proportions of focal propensity, the mean and the median each have favorable circumstances and impediments. A few upsides and downsides of each measure are outlined beneath.

The median might be a superior pointer of the most normal worth if a lot of scores has an outlier. An outlier is an outrageous worth that contrasts extraordinarily from different qualities.

Be that as it may, when the example size is huge and does exclude outliers, the mean score for the most part gives a superior proportion of focal propensity.

In overall the overall summary of the House price . The largest value is $1761 and the smallest value is $192 . The mean value of the house price is $886.575 and Standard Error is 29.66. The standard error indicates that the term that measures the accuracy with which a sample distribution represents a population by using standard deviation. If we check the Kurtosis , the excess Kurtosis is calculated as

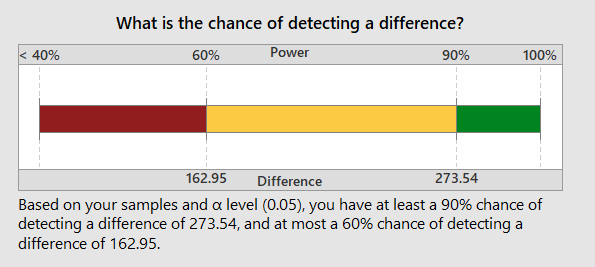
**Kurtosis – 3**

So in our case the Kurtosis is -0.148 -3 = - 3.148. A platykurtic distribution shows a negative excess kurtosis. The kurtosis reveals a distribution with flat tails. The flat tails indicate the small outliers in a distribution. Once we have the outlier in that case the best representative value will be the median in place of the mean. The "mean" is the "average" you're used to, where you add up all the numbers and then divide by the number of numbers. The "median" is the "middle" value in the list of numbers.

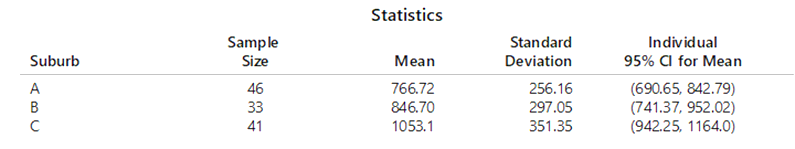
The median value of the house price is $852.

The histogram distribution shows that the as the prices increases of the house the frequency or counts are also increasing which means that the house with more price are more in number. Line the house with price more than $999 are 43.

1. **Does there appear to be any difference in the house prices between the suburbs in Kingfisherbay?**



Based on your sample and alpha level ( 0.05 ) , we have at least 90% chance of detecting a difference of 273.54 , and at most a 60% chances of detecting a difference of 162.95.



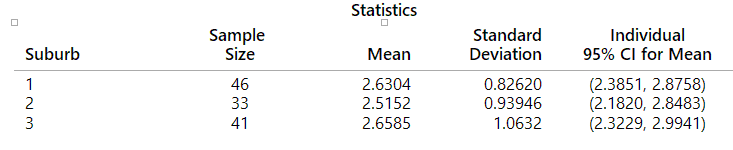
|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Anova: Single Factor |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| SUMMARY |  |  |  |  |  |  |
| *Groups* | *Count* | *Sum* | *Average* | *Variance* |  |  |
| House Price | 120 | 106389 | 886.575 | 105590.3 |  |  |
| Suburb | 120 | 235 | 1.9583333 | 0.729342 |  |  |
|  |  |  |  |  |  |  |
|  |  |  |  |  |  |  |
| ANOVA |  |  |  |  |  |  |
| *Source of Variation* | *SS* | *df* | *MS* | *F* | *P-value* | *F crit* |
| Between Groups | 46952799 | 1 | 46952799 | 889.3328 | 2.414E-82 | 3.8808272 |
| Within Groups | 12565336 | 238 | 52795.53 |  |  |  |
|  |  |  |  |  |  |  |
| Total | 59518135 | 239 |  |  |  |  |

From the Single factor ANOVA, The p – value is 2.414E-82 < 0.05 the significance level which means that there is a significant difference between at least two suburbs.

1. **Can you supply me with a brief summary on the condition of houses in Kingfisherbay? In addition, can you further analyse to see whether there appears to be any differences among the suburbs in terms of condition?**

Correlation examination is a factual technique used to assess the quality of connection between two quantitative factors. A high correlation implies that at least two factors have a solid association with one another, while a feeble correlation implies that the factors are not really related. As it were, it is the way toward contemplating the quality of that association with accessible factual information. This procedure is carefully associated with the straight relapse examination that is a measurable methodology for demonstrating the relationship between a needy variable, called reaction, and at least one logical or autonomous factors. The point of this work is to give a general review of correlation investigation so as to apply it to biomedical applications.

If we do the correlation analysis of the 3 variable namely the House Price, Suburb and condition of the house. The correlation between the Condition and suburb there is no much correlation ( 0.01), the correlation between Suburb and house price of house is 0.37 and correlation between condition and house price has correlation is 0.35.



From the above it is clear that there is no difference between the mean of the conditions of different suburb. • Test: There is not enough evidence to conclude that there are differences among the means at the 0.05 level of significance.

1. **Is house price related to any of the variables, namely: Rooms, Land Size, House Area and Weekly rent that you have collected in the sample? A short description of the relationship would be fine. Is there any one which stands out as having a stronger influence on house prices? Which of these variables has the least influence on house prices?**

From the correlation analysis of the house price with rest of the variables are analysed and results are in the excel sheet. The highest correlation between the House Price and Street is 0,7226 , House Price with BayView is 0.67, with Weekly Rent$ 0.66, with AreaSqm is 0.56, with Storeys is 0.565, with Bedrooms 0.3312. The negative correlation of House price is Rental Return and coefficient is -0.4038. So we can say that Street is one which stands out as having a stronger influence on house prices with value of 07226. The To TrainKm is the variables has the least influence on house prices.

Covariance measures the directional relationship between the returns on two assets. A positive covariance means that asset returns move together while a negative covariance means they move inversely. The Covariance value of the House Price and House rent is 48565 which means . It means both the variables are varying together a lot.

1. **The newspaper article “Buyers smash records” mentions how new land is in high demand and the ridiculous situation of buyers having to camp out for days to purchase land.**

**I’m interested in whether there is the opportunity to make more land available in areas such as Kingfisherbay by subdividing. Generally if a land size is at least 1000m2 or more then it can be subdivided.**

**Can you work out what percentage of all houses in Kingfisherbay are suitable for subdivision? Also, what is the estimated average land size across all of Kingfisherbay?**

The new is VICTORIANS spent more than $950 million on real estate last week alone, with a strong auction day yesterday capping off the biggest spending spree on record. It was also mentioned that the homes are repeatedly sold for up to 20 percent above the quoted price. It also mentioned that the seriousness of the affordability crisis is much more dire than anyone has so far been prepared to say”, advocate Catherine Cashmore warned.

Yes it is possible to subdivide the land having area >= 1000 square meters. There is 47% of the land which is equal or greater than 1000 Square feet. So there is the opportunity of subdividing the land.

1. **As well as housing affordability, the REIV is also concerned about the lack of rental properties available right across Melbourne including areas such as Kingfisherbay.**

**At a recent meeting it was claimed by a senior manager that the proportion of rental properties that were vacant had dropped below 25%.**

**Furthermore, it was recently reported that the average selling price for bay side houses, that is, for all cities and suburbs located around the Bay in Melbourne, had now exceeded $825,000.**

**Can you check whether these two claims are correct for all Kingfisherbay?**

There are 94 homes which are either vacant or owned by the owners. On the other hand we can say that 78% of the homes are not rented. Only 22% are rented. We did the hypothesis testing for the proportion just to check whether the proportion is less than 75% who are renting the home. If the result shows the 75%, it means there is a significance in the result. Hypothesis testing in measurements is a route for you to test the aftereffects of an overview or trial to check whether you have significant outcomes. You're fundamentally trying whether your outcomes are substantial by making sense of the chances that your outcomes have occurred by some coincidence. In the event that your outcomes may have occurred by some coincidence, the test won't be repeatable thus has little use.

In the hypothesis the result found that the p-value is 0.5247. Since the p -value is greater than the 0.05 the significance level it means that we will not be able reject the null hypothesis due to lack of the evidence. So we can say that there not sufficient homes which are rented. So the claim by a senior manager that the proportion of rental properties that were vacant had dropped below 25% is right.

To check whether , as it was recently reported that the average selling price for bay side houses, that is, for all cities and suburbs located around the Bay in Melbourne, had now exceeded $825,000. For this I did the hypothesis testing and found the lower and upper limit

|  |  |
| --- | --- |
| Interval Lower Limit | **828440** |
| Interval Upper Limit | **944710** |

So from the above upper and lower control limit it is clear that even the lower CI limit is more than $825000. So the claim of the rent being exceeded more than $825000 is right.

1. **For next year’s study, we would like to be able estimate the true proportion of houses of land size 1000m2 or more to within 2% and the average house price to within $20,000 with a high level of confidence**

**How many houses would be needed to be included in next year’s survey to achieve these two requirements?**

**Finally, assuming we take a big enough sample from Kingfisherbay next year, would it be possible to generalise the results to all of Melbourne? This will save us a lot of work duplicating the study in other suburbs**

From the calculations of the sample size for a proportion, we have taken 3% estimate of true proportion for the houses having the land area 1000 square meters. For calculating the sample size we have taken the confidence interval 95%. The sample size required is 1118. So we can say that we need 1118 sample size of the houses to be able to estimate the true proportion of the houses of land size of 1000 square meters or more to withing 2% and the average price to within $20,000 with high level of confidence.

The number of houses would be needed to be included in next year’s survey to achieve these two requirements will be 1118 and price would be $405636. This we calculated as explained in the Excel sheet by taking the standard deviation of the house price 324.95 and also taken the sampling error of 1 and sample required are calculated using 95% confidence level. The sample size is 1118 and (house price $405636 ) is required to be included in next year’s survey to achieve these two requirements.

The size of our sample directs the measure of data we have and, in this manner,, to some extent, decides our exactness or level of certainty that we have in our sample gauges. A gauge consistently has a related degree of vulnerability, which relies on the hidden fluctuation of the information just as the sample size. The more factor the populace, the more prominent the vulnerability in our gauge. Likewise, the bigger the sample size the more data we have thus our vulnerability lessens.

As our sample size expands, the trust in our gauge builds, our vulnerability diminishes, and we have more noteworthy accuracy. Expanding our sample size can likewise give us more prominent influence to identify differences. So, bigger sample sizes give increasingly dependable outcomes with more noteworthy accuracy and influence, however they additionally cost additional time and cash. That is the reason you ought to consistently play out a sample size computation before directing a review to guarantee that you have an adequately enormous sample size to have the option to make important inferences, without squandering assets on inspecting more than you truly need. So , it is possible to generalise the results to all of Melbourne. So we can say that with the sample size of 1118 and Standard deviation of 324 and sample mean of $886575 the housing price will have the upper selling price limit will be $944710 and Lower limit would be $828440.