**Condo**

**Condo**

**Measure of Central Tendancy**

The most appropriate measure of central tendency for the variables would depend on the distribution of the data that has been provided. For instance;

The variable *Price* is a ratio scale since the scale has a true zero point and there are equal intervals between the values. For the variable *Price,* the mean is 206.91, median is 200.00 while the mode is 175. The median is close to the mean hence there could be an apparent skewness present in the data for this variable hence having a lower mode value. Therefore, the median will be more appropriate to use the median for this variable. The variable *floor* uses an ordinal measurement scale it assists in ranking of floors in terms of their position. For the variable *Floor,* the mean is 4.013422819, the median is 4 while the mode is 1. From this variable, the mean and the median are not equal to the mode hence the distribution of this variable is not symmetrical. The mode and median are not the best measures to use since they are heavily influenced by outliers. The median will be the best measure of central tendency to be used to represent this data.

The variable *bedroom* has an ordinal scale. For the variable *Bedroom,* the mean is 2.187919463, the median is 2 while the mode is 2. There is a right skew suggested in this data since the mean is greater than the mode. In this case, the mean is the most appropriate measure of central tendency to be used since it will assist in knowing the overall average of the group of houses that got identified. The variable *insurance* has a nominal scale since they do not have a specific order or rank. Before calculations got done using this variable, the value *A* representing amended coverage got converted to 1 and *F* representing full coverage got converted to 2 to allow statistical analysis. For the variable *Insurance,* the mean is 1.624161074, the median is 2 while the mode is 2. There is not a heavy skew on the data since the mode and median are equal, therefore, both can get used as the most appropriate measure for central tendency. The mean for this variable is significantly different from the other two measures hence the data may not be normally distributed or there may be large and influential outliers present in the data.

The variable age has an ordinal scale since the values represent different age ranges. For the variable *Age,* the mean is 18.13422819; the median is 17 while the mode is 20. There is a right skew present in this variable since the mode is greater than the median and the mean. Hence, the median would be the best measure of central tendency to get used as opposed to the mean since it is less influenced by the extreme values that are present in this variable. The variable *Life Q* has an interval scale. For the variable *Life Q,* the mean is 62.54362416, the median is 63 while the mode is 63. There is not a heavy skew present in this variable since the mean and mode are equal. Therefore, either the mode or median are appropriate measures of central tendency that can get used. Finally, the variable *stability* has an ordinal scale. For the variable *Stability,* the mean is 83.97315436, the median is 96 while the mode is 96. The distribution of the variable is not heavily skewed since the median and mode are equal hence both can be an appropriate measure of central tendency that gets used. On the other hand, since the mean is significantly different from the median and mode, there might be significant outliers present in the variable.

**Outliers**

To be able to find outliers in the data, the quartile ranges together with the upper and lower limits were used to find the outliers, with values of the variables outside a specific range being considered outliers. The variable *price* might has an outlier that may be due to market forces, specifically external factors such as supply and demand in the area that the condo got located. The variable *floor* does not have outliers. The variable *bedroom* has outliers with the possible reason being that those are condos bigger condos with five or more bedrooms. The variable *insurance* has all their values as outliers. The variable *age* has outliers that may be caused by some building being older than other buildings. The variable *life Q* has outliers since some places have a higher life quality that is higher than other areas. Finally, the variable *stability* has outliers that may be caused by a difference in the economic factors that affect the countries that the condos are located.

**Measures of Variability**

For *Price,* the range is 195, variance is 1665.896517, standard deviation is 40.81539559, and Inter Quartile Range (IQR) is 55. The best measure of variability that should be used is standard deviation since the data is approximately normally distributed. For *floor*, the range is 7, variance is 5.013332124, standard deviation is 2.239047146 and IQR is 4. The measure of variability that should be used is standard deviation since the data is almost normally distributed. For *bedroom*, the range is 5, variance is 0.81579902, standard deviation is 0.903215932 and the IQR is 1. The best measure of variability that should be considered is standard deviation since the data is approximately normally distributed. For *insurance*, the range is 2, variance is 0.236169055, standard deviation is 0.485972278 and the IQR is 1. The best measure of variability that should be used is range since the data is nominal and only two values are present. For *age*, the range is 68, variance is 91.40077997, standard deviation is 9.560375514 and the IQR is 8. The standard deviation should get considered as the measure of variability. For *life Q*, the range is 26, variance is 15.31734083, the standard deviation is 3.91373745 and the IQR is 0. The measure of variability that should be used is range or the IQR because there are outliers and also presence of skewness in the distribution of the data. For *stability,* the range is 135, variance is 831.7425177, the standard deviation is 28.83994656 and the IQR is 5. The standard deviation should be used since the data is approximately normally distributed.

**Coefficient of Variation (CV)**

To obtain the Coefficient of Variation (CV), the values are obtained by dividing the standard deviation to the mean of the variable hence making a ratio that is then multiplied by 100. Therefore, a higher value of CV shows that there is more variability in the data.

The values of the CV for *Price, Bedroom* and *Age* are:

* Price= 19.72653652
* Bedroom= 41.28195519
* Age= 52.72005743

*Age* has the highest value at 52.72005743 therefore being the most variable. This situation could have arisen due to factors such as the maintenance and upkeeps that lead to higher CV for the buildings if they were well maintained giving them a longer lifespan.

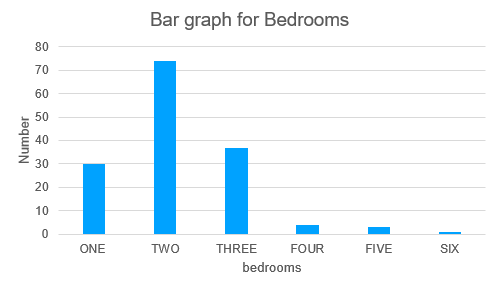
**Discussion**

From the analysis, the *Price* histogram has a right skew because the median is close to the median is closer to the mean, i.e. the mean is 206.91 and the median is 200.00.



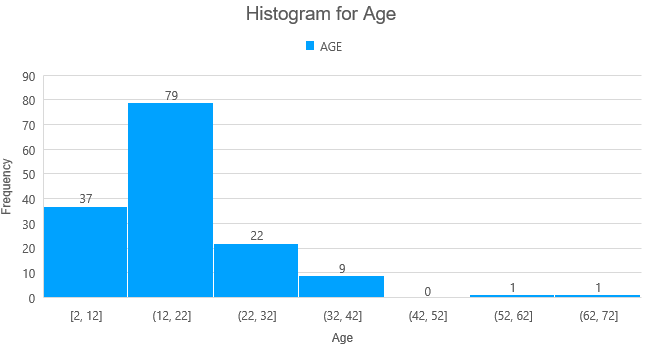
*Figure1: Histogram of Price showing a right skew*

Therefore, the median will be the better measure of central tendency to be utilised since it is not affected by outliers (de Nijs & Klausen, 2013). For the *Bedroom* variable, there is a right skew in the distribution of the data suggesting that the mean is greater than the mode. This situation occurs since there are outliers that increase the value of the mean i.e. the mean is 2.187919463 while the median is 2 (Corwin, 2020). Therefore, the mean will be the better option since it considers the outliers and assist in finding the overall average of the group of houses that got identified and assist the people the people to make the right decision when looking for a condo.



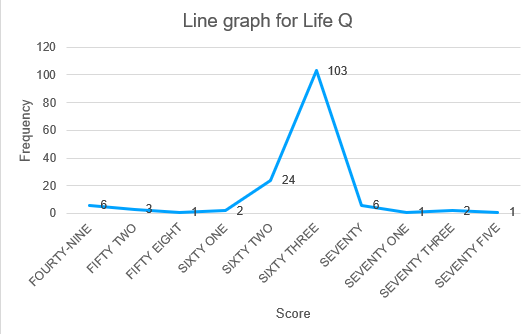
*Figure 2: Bar graph of Bedrooms showing a right skew*

For *Age,* there is a right skew present in the data caused by the presence of outliers affecting the mean. Therefore, the median is the better measure of central tendency to get used because it is not affected by the extreme values.



*Figure 3: Histogram of Age showing a right skew*

For *Life Q,* the mean is 62.54362416, the median is 63 while the mode is 63. There is not a heavy skew present in this variable since the mean and mode are equal.



*Figure 4: Line graph of life Q showing an approximately normal distribution*

The line graph of this variable is approximately normally distributed. Due to this situation, either the mode or median can get used as a measure of central tendency because they are equal.

**References**

Corwin, S. (2020). *Lecture 5: Mean, Median, Mode; Outliers*. Sites.radford.edu. https://sites.radford.edu/~scorwin/courses/200/LectureNotes/Lecture5.pdf

de Nijs, R., & Klausen, T. L. (2013). On the expected difference between mean and median. *Electronic Journal of Applied Statistical Analysis*, *6*(1), 110–117. https://doi.org/10.1285/i20705948v6n1p110