

Real Statistics – Real Decisions: Chapter 2 Example by Mr. Clinch

General Description: Tenants of Area A of a city are claiming that their monthly rental fees are much higher compared to other parts of the city. As a member of the local apartment association, you investigate the situation.

Key Given Information: The table to the right.

Area A	Area B	Area C	Area D
1275	1124	1085	928
1110	954	827	1096
975	815	793	862
862	1078	1170	735
1040	843	919	798
997	745	943	812
1119	796	756	1232
908	816	765	1036
890	938	809	998
1055	1082	1020	914
860	750	710	1005
975	703	775	930

1a) In order to begin to investigate the complaints, I would first begin by determining the average cost for each area. If Area A has a lower average cost than the others, we really don't need to go much further.

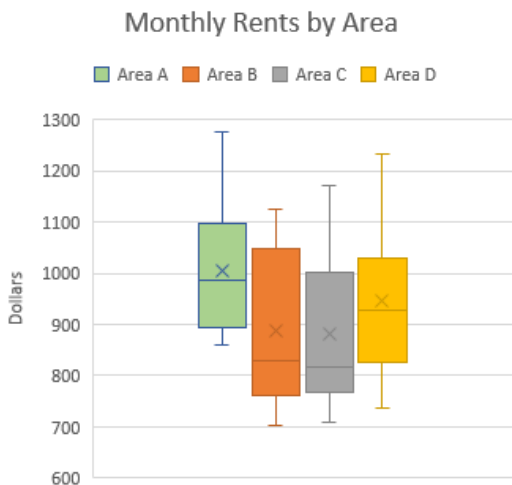
1b) As stated above, I would first compute the mean for each area. It is also helpful to calculate the median value as well, although there does not appear to be any outliers at first glance.

1c) In calculating both the mean and median cost for each area, the results are shared below.

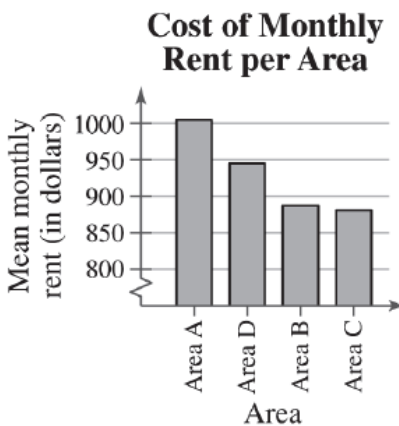
	Area A	Area B	Area C	Area D
Mean	\$1,005.50	\$887.00	\$881.00	\$945.50
Median	\$986.00	\$829.50	\$818.00	\$929.00

2a) It is always wise to first start with a box plot to observe the skewness and if there are any outliers. If there are outliers or serious skewness, then we already have the box plots and median values to rely on. If not, we might follow up with a bar chart and rely mostly on the mean.

2b)



The box plots reveal that there is some positive skewness for each area, but no outliers. It would also be helpful to construct a bar chart, specifically a Pareto chart. This is because the data are quantitative and a Pareto chart positions data in order of decreasing height, with the tallest bar positioned at the left.



2c) Based on the box plots, Pareto chart, and tables, we see that the median and mean monthly rent for apartments in Area A appear to be more expensive than those in other areas.

3a) In order to determine not just the “typical” value through the mean and median, we should also investigate the variability of costs in each area. Perhaps A has a larger variability in cost. To do this, we would compute the standard deviation and range for each area.

3b) Area A: $s \approx \$123.07$
range = \$415.00

Area B: $s \approx \$144.91$
range = \$421.00

Area C: $s \approx \$146.21$
range = \$460.00

Area D: $s \approx \$138.70$
range = \$497.00

3c) The box plots and the values found in part (3b) agree. If you look at the box plots, the box plot for Area A has the smallest interquartile range. This implies Area A has the least amount of variability, which we also see in part (3b) as area A has the smallest standard deviation.

4a) It appears that the complaints from Area A are legitimate in that they are paying a higher average rent amount. Also, they have a smaller standard deviation, so they are consistently paying a higher amount, which also supports their case. However, we can not conclude WHY this might be the case. All we as statisticians can say is they mathematically have a case in stating their mean is higher. The association might address this by either adjusting their rent or addressing why they may be paying a higher amount per month.

4b) Perhaps the apartments in Area A are located in a better part of the city in terms of safety, schools, traffic, or entertainment options.