Statistical analysis of the vehicle ownership of

American households

1. Introduction

Using the 318 cities from USA in 2015 and 2016 household car ownership and percentage of all household without vehicles, to work on their frequency distribution and cumulative frequency, absolute and relative changes etc. with various charts, contrasted the scatterplot and calculated common description statistics and, tried to analysis the outliers. According to the analysis of the content is given the corresponding conclusion.

1. Analysis
2. Household without vehicles(2015-2016)----(Q2 Q6 Q5 Q4)
3. Vehicles per House(2015-1016)-----(Q7)

**Q2**

Distribution of 2016 percentage of households without vehicles

This is a positive skewed curve，also called a right skewed curve. In 2016, 77 of the U.S. cities listed in the data sheet had a rate of 5%- 7.5% of household without vehicles, which is the most common relative frequency percentage. The curve has a relative high starting point (20 of frequency value) , and surges within the following two units of percentage stratum:67 cities have 2.5% - 5% of household without vehicles. 77 of the U.S. cities listed in the data sheet had a rate of 5% - 7.5% of households without vehicles, which is the most common relative frequency percentage. On the part of more than 5% - 7.5%, the rate is more fragmented. Especially after 12.5%, the curve drops fast below 10. And the cumulative percentage of households without a car in these cities is 13.56%, which is significantly lower than the previous percentage.

Especially, there is one city with 52.5% - 55% of households without a car at the end of the curve, which is far away from the percentage stratum where other cities belong.

**Q6**

The similarities are apparent. Because the data for each city is correlated in time, the curve shape of the two years roughly coincides, which means that the proportion of households without cars has roughly the same distribution. At stratum below 2.5%, the number of cities rose by 50% in 2016 and the number of households without cars declined. Relative frequency in the low percentage range increased in 2016 compared with last year, and more data points were concentrated in the level above 12.5%. Although the lowest relative frequency in 2016 remains to 1, percentage stratum between 5%-7.5% increased slightly in 2016, so the range of relative frequency in 2016 was slightly larger than that in 2015. Also, the relative frequency on the left of the highest point is increasing, while the relative frequency on the right is lower than before, and the curve is more skewed to the right. On the whole, the proportion of households without cars in the city is generally decreasing. But there are still exceptions: the only one city in 52.5% - 55% stratum. Its percentage of household without vehicles stayed nearly the same after the one-year development.

In cumulative frequency, it can be seen that 80% of the data is concentrated below 12.5%, while the data above 12.5% is not significantly changed in the cumulative frequency line plot (except that at the level of 27.5% to 30%, there is a relatively significant improvement in 2016). According to the percentage stratum below 12.5%, 2016 shows an average increase of 3.08% compared with 2015.

**Q5**

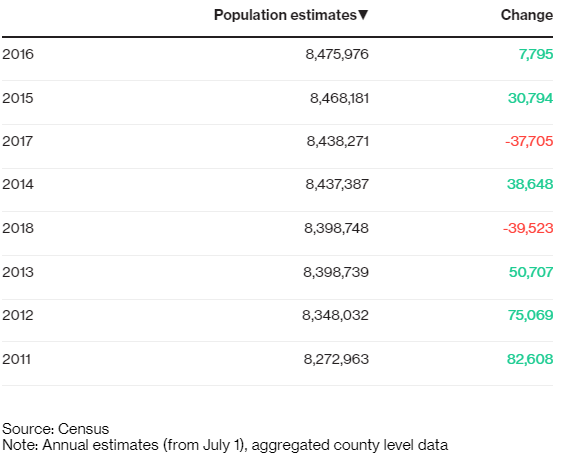
Since the distribution is not symmetrically, the outlier cannot be determined by Sigma Rules on the data points, so I should find them out by comparing and analyzing the original data. Firstly, I will focus on the distance from percentage stratum corresponding to the highest frequency. The farther the distance is, the more likely it is to be the outlier. To be more specific, I will focus on cities where the number of cars owned by families is below zero. In addition, for those the number of cars owned by household shows on decline trend from 2015 to 2016, among these cities, their number is likely to deviate from the normal number due to some undetected reasons.

A total of 6 cities meet the above conditions. I put them in order according to the percentage of families without cars in 2015.



This among them, there are three cities showed a decline, as New York, jersey and Boston. Boston and New York are bustling cities, the most striking features is developed transportation system. Taxi, bus, subway is the main means of transportation, reliance for personal use. For those in need of buying a car is not strong enough.

2. The significant population flow. Their economic structure can also reflected their crowded conditions of the large population flow characteristics. This picture below shows the migration of 2011-2016, New York City. This city has accommodate large population, but not a lot of people living in New York and buy a car, some people from other places came to New York, to the city as a struggle for life, rather than fitting place for a settlement.



**Q4**



This table shows the distribution of data and relative changes between two years. The mean, median, maximum, minimum and sum is greater than 2016 years of data, means that the data presented in 2016 is decreasing. Standard error, standard deviation, variance, range are lower than in 2015, and kurtosis is higher than in 2015, so the data of 2016 is more concentrated than in 2015. Overall, the proportion of household that have vehicles from 2015 to 2016 is on the rise, and focus on a certain percentage.

**Q7**

I think there is a correlation between the two quantities. In the case of a scatter diagram shows, each 2015 data has a 2016 data that is fairly close to it. Concluded that there is some of the same rule, makes them distribute in such a uniform and harmonious way. Second, the average family car ownership of a state in the course of a year is not going to happen very big change, in the second year of ownership depends largely on the continent for the first year the average ownership of every family. Car, for general family, belongs to the durable and consumable good, the demand of which commonly will not change sharply in a year, but would be steady at a relatively fixed value. The change between the two years on average to each family would be little as well.

1. Conclusion

Overall, the data of household without vehicles presented in 2016 is decreasing, and the data of 2016 is more concentrated than in 2015, meaning the proportion of household that have vehicles from 2015 to 2016 is on the rise. The most common relative frequency percentage is at the rate of 5% - 7.5%. And the outliers are six cities that show vehicles per household below 1 in both 2015 and 2016. 80% of household without vehicles is concentrated below 12.5% percentage level, while the data above 12.5% is not significantly changed in the cumulative frequency line plot.

Reference

New York City's Population Is Shrinking: Demographic Trends, Alexandre Tanzi, Retrieved from, https://www.bloomberg.com/news/articles/2019-04-18/new-york-city-s-population-is-shrinking-demographic-trends