Perspective

Understanding Quartiles and Quintiles: Dividing Data for Analysis and Comparison

Jemma Geor[*]

Department of Mathematics, Texas A&M University, Texas, United States of America

DESCRIPTION

In the field of statistics, quartiles and quintiles are essential tools used to split a dataset into meaningful portions. These divisions aid in understanding the distribution of data, identifying outliers, and making comparisons between different groups or data sets. Quartiles divide data into four equal parts, while quintiles divide data into five equal parts [1].

Quartiles: Dividing data into four equal parts

Quartiles are statistical measures that split a dataset into four equal parts, each containing 25% of the data. The three quartiles are denoted as Q1, Q2, and Q3 [2].

Q1 (First quartile): Q1 represents the 25th percentile, meaning that 25% of the data values fall below Q1. It is also known as the lower quartile.

Q2 (Second quartile): Q2 is the 50th percentile and is equivalent to the median of the data. It divides the data into two halves, with 50% of the data values falling below Q2 and the remaining 50% above it.

Q3 (Third quartile): Q3 represents the 75th percentile, indicating that 75% of the data values fall below Q3. It is also referred to as the upper quartile.

Calculation of quartiles

To calculate the quartiles, the data must first be arranged in ascending order. If the number of data points (n) is odd, the median (Q2) is the middle value. If n is even, the median is the average of the two middle values. Then, to find Q1, the data below the median is divided into two equal halves, and the median of the lower half is taken. Similarly, to find Q3, the data above the median is divided into two equal halves, and the median of the upper half is taken [3].

Quintiles: Dividing data into five equal parts

Quintiles are statistical measures that divide a dataset into five equal parts, each containing 20% of the data. The four quintiles are denoted as Q1, Q2, Q3, and Q4 [4].

Q1 (First quintile): Q1 represents the 20th percentile, signifying that 20% of the data values fall below Q1.

Q2 (Second quintile): Q2 is equivalent to the 40th percentile and is also known as the median of the data.

Q3 (Third quintile): Q3 represents the 60th percentile.

Q4 (Fourth quintile): Q4 corresponds to the 80th percentile, indicating that 80% of the data values fall below.

Similar to quartiles, quintiles are calculated by first arranging the data in ascending order. Then, the appropriate percentile values are identified. For example, Q1 corresponds to the value at the 20th percentile, Q2 corresponds to the 40th percentile, and so on [5,6].

Significance in statistical analysis

Data distribution: By dividing data into equal portions, quartiles and quintiles provide insights into the distribution of data. Skewed or concentrated data can be identified by examining the quartiles [7].

Outlier detection: Outliers, which are extreme data points, can be detected by comparing data values to the quartiles and quintiles [8].

Comparing groups: Quartiles and quintiles enable comparisons between different groups or datasets, highlighting differences in their distributions [9].

Quartiles and quintiles are fundamental measures in statistical analysis that aid in understanding the distribution of data and making meaningful comparisons. By dividing data into four or five equal parts, these measures provide valuable insights into the spread and characteristics of datasets.

Whether used in finance, healthcare, social sciences, or any other field, quartiles and quintiles are indispensable tools that enhance our understanding of data and enable us to draw informed conclusions from statistical analysis [10].

REFERENCES

1. Kruger J, Dunning D. Unskilled and unaware of it: How difficulties in recognizing one's own incompetence lead to inflated self-assessments. J Pers Soc Psychol. 1999;77(6):1121.

2. Hyndman RJ, Fan Y. Sample quantiles in statistical packages. Am Stat. 1996;50(4):361-365.

3. Dunning T, Ertl O. Computing extremely accurate quantiles using t-digests. arXiv preprint. 2019 Feb 11.

4. Vardeman SB. What about the other intervals? Am Stat. 1992;46(3):193-197.

5. Amaratunga D, Cabrera J. Analysis of data from viral DNA microchips. J Am Stat Ass. 2001;96(456):1161-1170.

6. Bolstad BM, Irizarry RA, Åstrand M, Speed TP. A comparison of normalization methods for high density oligonucleotide array data based on variance and bias. Bioinform. 2003;19(2):185-193.

7. Ehm W, Gneiting T, Jordan A, Krüger F. Of quantiles and expectiles: Consistent scoring functions, Choquet representations and forecast rankings. J R Stat Soc Series B Stat Methodol. 2016;78(3):505-562.

8. Shaw WT. Sampling Student's T distribution-use of the inverse cumulative distribution function. J Comp Fin. 2006;9(4):37.

9. Karvanen J. Estimation of quantile mixtures *via* L-moments and trimmed L-moments. Comp Stat Data Anlys. 2006;51(2):947-59.

10. Dawson R. How significant is a boxplot outlier? J Stat Edu. 2011;19(2).