

Descriptive Statistics and Its Applications

Yihang Dong*

Department of Mathematics, Boston University, Boston, MA 02215, USA

*Corresponding author: yihangd@bu.edu

Abstract. This review paper examines the topic of descriptive statistics, which is a study of statistics, deriving from basic mathematics. The paper discusses common descriptive statistics using a combination of numerical examples and graphical demonstrations. The paper introduces the definitions of some basic descriptive statistics and their corresponding formulas. The paper then reviews several research articles published in recent years in which researchers have used descriptive statistics to help analyze housing prices and the relationship between the stock market and trust. This review paper consults tables from other research articles with descriptive statistics and uses specific data to demonstrate how these statistics help researchers draw meaningful conclusions about the subjects of the analysis. The final conclusion is that although descriptive statistics is a simple method of summarizing data, it is fundamental and meaningful in statistical analysis and that can be used as a building block on which further analysis can be based.

Keywords: Descriptive Statistics, Housing Price, Stock Market.

1. Introduction

Since the birth of human civilization, mathematics has always been a part of human history, from the most basic numbers to simple algebra and geometry, the progress of human civilization cannot be separated from mathematics. Statistics is one of the branches of mathematics. Census counts date back thousands of years, but statistics as a unique mathematical field first emerged in the early nineteenth century as a way to examine populations, economies, and moral behavior; In later centuries, statistics also emerged as a mathematical instrument for interpreting such kinds of information [1]. Modern statistics offers a quantitative method for empirical science; it is a logical method for measuring uncertainty and for looking at how this uncertainty affects the design of experiments and the interpretation of observations [2]. Statistics can be widely applied to various fields, and there exists a unity to the methods of statistics: the same computer program that analyzes geophysical scientist data can potentially be employed by an economist, chemist, sociologist, psychologist, or political scientist [2]. Essentially, statistics deal with variabilities in our life [3]. For instance, people have different heights and weights, different IQs, different political attitudes, etc, and all of those characteristics can be categorized and analyzed with the support of statistics [3]. Statistics has become an indispensable technique in scientific research.

Descriptive statistics is the simplest form of statistics: it is a tool to help people organize and summarize the inevitable variability in collections of actual observations or scores [3]. This review article aims to introduce descriptive statistics and analyze several specific examples of the use of descriptive statistics in academic research papers in analyzing housing prices and analyzing the stock market. Housing is a significant social indicator of development in a number of ways, including the availability of affordable housing for various socioeconomic levels, the danger of poverty and deprivation, and geographic disparities such as housing segregation in urban regions [4]. Housing price is a basic index that can be analyzed using statistical methods. In finance, the stock market is an essential way to help companies raise money to fund operations, which is an important part of today's global economy, serving as a "barometer" and "alarm" in the global economy and finance [5]. Both housing prices and stock markets involve a lot of fragmented data that need statistical help to be converted into readable data tables and charts for analysis. In such cases, descriptive statistics become crucial. The following sections provide a fundamental overview of descriptive statistics and examine its applications.

2. Methodology

Descriptive statistics is a basic area of study in statistics. It demonstrates the relationship between variables in a given sample, and it is often used to clean up and summarize scattered data, which is crucial for making inferential statistical comparisons and conducting research [6]. Descriptive statistics are often used in statistical modeling. A statistical model expresses the previous knowledge of the probability experiment that produced the observed data [7]. The model assumes that the observation X is created by one of the model's probability distributions [7]. There are multiple measurements in descriptive statistics. There are measures of central tendency, frequency, position, and dispersion, and they also contain categories of variables such as ratio, intervals, and nominal and ordinal variables [6]. Many concepts within descriptive statistics can also be explained using basic calculus, and those descriptive statistics can be put into analyzing real-life data such as housing prices and the stock market. In sections 3 and 4, there are discussions about different kinds of descriptive statistics and their respective functions when analyzing specific data for housing prices and stock markets. Some common descriptive statistics that are used are listed below.

2.1. Mean, median, and mode

2.1.1 Mean

The mean is the arithmetic average of the sum of values in the data divided by the quantities of observations, which is one of the measures of central tendency [6].

For example, given a sample A (students' ages): 15,17,19,14,13, the mean equals the sum of the observations: $15+17+19+14+13$ divided by the number of observations: 5, which equals $(15+17+19+14+13)/5 = 15.6$

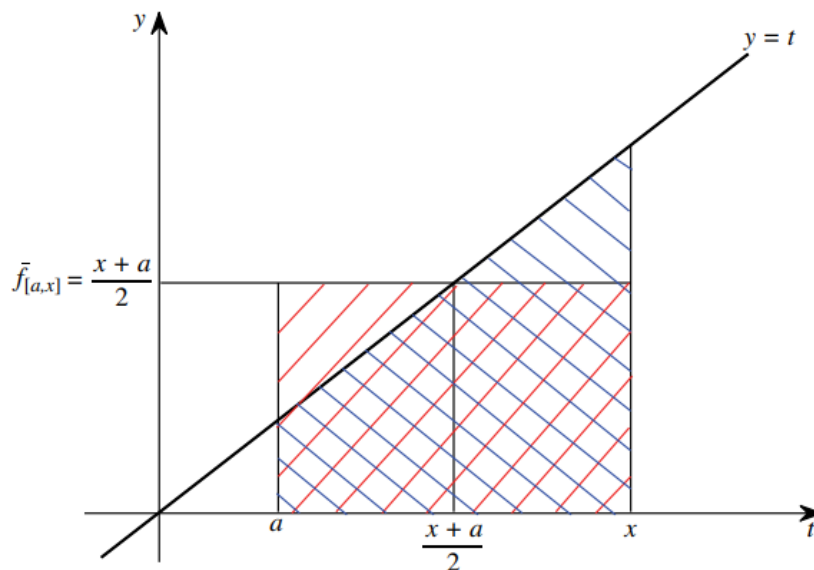


Fig. 1 Mean on a linear function [8]

Looking at Fig. 1, in the function $y = x$, the shaded area under the interval $[a,x]$ can be calculated by adding two trapezoids, which can be transformed into the sum of two rectangles whose width is $\frac{x+a}{2}$, and the height is the same because $y = x$. Therefore, the average y value, or the mean, within the interval $[a,x]$ is $\frac{x+a}{2}$ [8].

2.1.2 Median and mode

The median is the value in the center of the statistics arranged from either lowest to highest or vice versa [6]. If the number of values is even, the median is the average of the two middle values; however, if the number of values is odd, the median is just the middle value [6].

Using sample A above, by rearranging them in sequential order: 13, 14, 15, 17, 19, the median is 15.

Mode is the observation that appears most frequently in the sample [6].

2.2. Variance and standard deviation

A distribution's variance and standard deviation can be used to illustrate its dispersion [7]. In other words, variance and standard deviation reveal how close to the mean is each observed value [6].

The sample variance is calculated by the following formula:

$$S^2_X = \frac{1}{n-1} \sum_{i=1}^n (X_i - \underline{X})^2$$

In the formula, \underline{X} is the mean value of all observations, S^2_X is the sample variance, n is the number of observations, and X_i is the value of one observation [8].

The sample standard deviation equals the square root of the sample variance as shown in the formula: $S_X = \sqrt{\frac{1}{n-1} \sum_{i=1}^n (X_i - \underline{X})^2}$

2.3. Percentile, deciles, and quartile

Using percentiles, deciles, and quartiles, we can divide a dataset into 100 equal sections, ten equal sections, and four equal sections for an ordered dataset [6]. Q1 represents the lower quartile, which is the 25th percentile. Q2 is the second quartile, corresponding to the 50th percentile, and Q3 is the third quartile, which is the 75th percentile.

The formula to calculate the percentile of values in a data set is the following: $n = (\frac{P}{100}) * N$ [9]

In the formula, N is the number of values in the data set, P is the percentile, and n represents the ordinal rank of a given value

Usually, box plots are helpful for visualizing descriptive data by representing percentiles as shown in Fig. 2 [6].

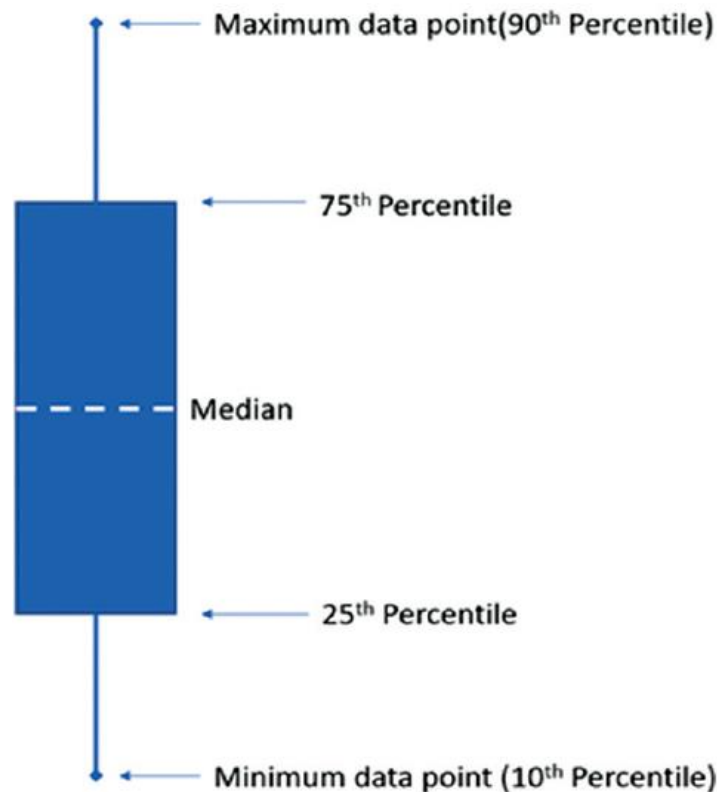


Fig. 2 Box plot sample [6]

Above are some of the descriptive statistics that are widely employed in analyzing data. In the following discussion of the use of descriptive statistics in housing prices and the stock market, these measurements are fundamental and crucial for data analysis.

3. Descriptive Statistics in Housing Price Analysis

3.1. Housing prices in Warsaw

Housing prices have always been a concerning issue for most people. Especially during COVID-19, house prices in many areas were impacted. Trojanek al. analyze the influence of the pandemic on housing prices in Warsaw, more specifically, the Airbnb prices [10]. The team utilizes the hedonic quantile regression model on the collected data, and they employ symmetric and asymmetric weighting, which yields a table of descriptive statistics with quantiles of 0.25, 0.5, and 0.75, as shown in Table 1 [10].

Table 1. Partial estimation from the quantile regression model [10]

	From January 2017 to December 2020			From January 2017 to March 2020			From April 2020 to December 2020		
	0.25Q	0.50Q	0.75Q	0.25Q	0.50Q	0.75Q	0.25Q	0.50Q	0.75Q
Airbnb	0.0303	0.0310	0.0244	0.0308	0.0322	0.0247	0.0183	0.0219	0.0233
characteristics	✓		✓	✓		✓	✓		✓
controls		✓			✓			✓	
area effects	✓		✓	✓		✓	✓		✓
		✓			✓			✓	
time effects	✓		✓	✓		✓	✓		✓
		✓			✓			✓	

In Table 1. Airbnb is the number of active units on Airbnb per month in the neighborhood where the apartment was advertised for long-term lease [10]. By comparing the median quantiles from January 2017—March 2020, which is 0.0332 (when the rent growth is higher before the pandemic), to April 2020—December 2020, which is 0.219, Trojanek al. are able to draw the conclusion that there is a positive relationship between Airbnb and rent supply in 2017 - 2020 [10].

It is also worth noting that in Table 1, Trojanek al. also list out the variables that are controlled in order to minimize the error when yielding experimental results like quantiles [10]. Controlling variables is a crucial method for researchers to satisfy one (among many) requirements for causal inference: the elimination of alternative explanations for the observed connection [11].

As demonstrated by Table 1., quantile values allow researchers to easily compare and analyze the relationship between Airbnb prices and its supply during COVID-19.

3.2. Housing prices in Beijing

In the early 21st century, housing prices in China have been growing substantially, especially in large cities like Beijing, Shanghai, and Guangzhou, due to urbanization [12]. In order to find out the determinants of changes in house prices from 2012 to 2015 in Beijing, Zhang al. use Oaxaca-blinder decomposition, combined with variance decomposition and quantile decomposition to yield a table of descriptive statistics as shown in Table 2 [12].

Table 2. Descriptive statistics about the potential variables that determine house prices in Beijing [12]

Variables	2012				2015			
	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum
Price	336.727	408.996	14.911	12803.321	477.521	497.334	20.609	9256.482
Bedrooms	2.126	0.974	0	21	2.415	1.107	0	21
Dining Rooms	1.34	0.608	0	6	1.559	0.649	0	6
Area	102.523	84.225	25	2600	123.94	84.914	6	1500
Floor	14.099	8.560	0	101	15.342	9.645	0	63
North	0.041	0.198	0	1	0.04	0.196	0	1
West	0.04	0.196	0	1	0.031	0.173	0	1
East	0.049	0.216	0	1	0.043	0.204	0	1
South	0.159	0.365	0	1	0.165	0.371	0	1

In Table 2, the rightmost column is the potential variables that may affect the housing price, and Zhang al. calculated the mean, standard deviation, and minimum and maximum values in 2012 and 2015 respectively [12]. From these descriptive statistics, Zhang al. make an analysis that the typical price of homes has increased significantly between 2012 and 2015 [12]. The group compares the mean real-estate index in 2012, which is 336.727, to the price in 2015, which is 477.521, indicating a large percentage of increase [12].

Analyzing the variables, from 102.523, 2.126, and 1.34, respectively, the average living space size, number of bedrooms, and living room count grew to 123.94, 2.415, and 1.559 [12]. By comparing other variables such as the floor and orientations, the research group draws the conclusion that the average floor number is an indication of trending taller buildings because its increase corresponds to the increase in housing prices [12]. In addition, since from the data, there are no significant changes in the orientation of the housing, the research group is able to conclude that housing orientation is not a determinant of the growth of housing prices[12].

Simply from looking at the descriptive statistics, Zhang al. are able to draw convincing conclusions on the specific factors that affect the housing prices in Beijing.

4. Descriptive Statistics in Stock Market Analysis During COVID-19

Not only does COVID-19 affect the housing price, but it also affects the global stock market. In their research, Engelhardt al. examines the relationship between global stock market volatility and trust [13]. The research group obtains data about new COVID-19 cases in every country each day from Johns Hopkins University [13]. The data collection process is about half a year starting from January 22 to July 29 in 2020 [13]. The research group also collects the closing prices of the corresponding major stock market indices from other sources like Yahoo, World Values Survey, and

Trading Economics [13]. . Using those data, they construct a table of descriptive statistics as shown in Table 3.

Table 3. Collected Descriptive Statistics from Collected data on the Stock Market and Trust [13]

	Observations	Mean	Median	Min	Std.	Max
Market Volatility	5479	0.0178	0.0138	0.0000	0.0148	0.1312
Cases growth rate	5479	0.0511	0.0125	- 11.1132	0.3310	2.0794
Trust	5479	31.0891	26.6000	4.2000	19.7591	73.9000
Trust (Government)	5479	39.3122	37.6000	9.6000	20.8197	94.6000
Trust dummy	5479	0.6404	1.0000	0.0000	0.4799	1.0000
Trust dummy (Government)	5479	0.5085	1.0000	0.0000	0.5000	1.0000

From Table 3, compared to trust in government, which has a mean score of 39.31, the average for social trust is 31.09. 64.04% of the observations are categorized by the research group as countries with high trust when analyzing societal trust, and 50.85% of the observations are treated as countries with high trust while examining governmental trust [13].

In order to find out the relationship between volatility in the stock market and trust, using descriptive statistics, the research group performs a valuation of market volatility using univariate regression model analysis, which yields Table 4 [13].

Table 4. Descriptive Statistics from Univariate Analysis [13]

	Countries with high trust		Countries with low trust		Difference
	Observations	Mean	Observations	Mean	
Market Volatility	3509	0.0179	1970	0.0176	0.0003
Cases growth rate	3509	0.0420	1970	0.0674	-0.0254
Trust	3509	41.9518	1970	11.7403	30.2115
Trust (government)	3509	44.3401	1970	30.3564	13.9836

Based on the data from Table 4, even though from the mean values, there is not a big difference between the two countries' variables (high trust and low trust), the case growth rate of countries with high trust is positive 0.0420, while the case growth rate of countries with low trust is 0,0674, in which the latter is higher than the former, indicating that countries with high trust are more likely to have less stock market's volatility [13].

5. Limitations and Future Outlooks Suggestions

From the reviews of the research articles above, it is clear that using descriptive statistics is a fundamental method for examining any numerical data. Descriptive statistics effectively lessen and synthesize a large data set down to a few significant summary statistics. The summary statistics always come in the form of a table of data with labeled descriptive statistics. Within the table, the

data is straightforward and can be easily manipulated for further comparison or analysis (univariate, bivariate, and multivariate analysis). Descriptive statistics give a holistic picture of the data set, which is helpful for drawing conclusions. To be specific, the measure of mean reflects the central tendency, which can be used to make comparisons among multiple variables. In addition, the quartiles, which measure the position of a set of data, can also be used between variables to make conclusions about the correlations. There are more kinds of descriptive statistics like skewness, frequency, and range, that are not discussed in this review article but are also significant in analyzing data.

Even though descriptive statistics provide a summary of data, they are not enough for statistical analysis. For example, descriptive statistics cannot make any predictions about the future values of variables because it is unable to demonstrate trends. It can only be used to compare existing variables or the same variable in the past. The data collection process of descriptive statistics is also time-consuming. For example, in section 4.1., to obtain the stock market data, the research group spends months collecting data from multiple sources.

Therefore, a single descriptive statistics table is never enough for a comprehensive analysis. Usually, descriptive statistics can be used for inferential statistics, which can extrapolate conclusions from the data set regarding broader scenarios [14]. In the articles that are reviewed in sections 3 and 4, some research groups also employ regression models like hedonic quantile regression and univariate regression to elaborate on the existing descriptive statistics. Accordingly, descriptive statistics can be employed as a basic summarizing data for more specific and oriented statistical analysis.

6. Conclusion

As evident from several reviews on the use of descriptive statistics to analyze housing prices and stock market volatility and trust, descriptive statistics is a useful tool that helps summarize data and provide straightforward results. While the data collection process can be complex, with proper processing of the data, the resulting descriptive statistics tables can save researchers a great deal of time when analyzing tons of data. When analyzing Airbnb prices in Warsaw during COVID-19, descriptive statistics show that higher rental availability is positively correlated with the number of Airbnb. When analyzing housing prices in Beijing, descriptive statistics help to confirm the sharp increase in housing prices. When examining the relationship between stock markets and trust, descriptive statistics prove that volatility is lower in countries with high levels of trust. The examination of the use of descriptive statistics reflects the significance of this basic statistical concept and its versatile applications. This paper not only emphasizes the importance of descriptive statistics but also touches on indexes of housing and stock markets, both of which are closely related to people's life.

References

- [1] Porter, T. M. (2022, September 27). probability and statistics. Encyclopedia Britannica. <https://www.britannica.com/science/probability>
- [2] Stigler, S. M. (1986). *The history of Statistics: The measurement of uncertainty before 1900*. Harvard Univ. Press.
- [3] Witte, R. S., & Witte, J. S. (2017). *Statistics*. Wiley.
- [4] Causa, O., & Woloszko, N. (2020). Policy Considerations on Housing, Wealth, and Inequality. *Cityscape*, 22(3), 339–364. <https://www.jstor.org/stable/26967207>
- [5] Bhowmik, R., & Wang, S. (2020). Stock Market Volatility and Return Analysis: A Systematic Literature Review. *Entropy (Basel, Switzerland)*, 22(5), 522. <https://doi.org/10.3390/e22050522>
- [6] Kaur P, Stoltzfus J, Yellapu V. Descriptive statistics. *Int J Acad Med* 2018;4:60-3

- [7] Bijma, F., Jonker, M., & van der Vaart, A. (2017). Descriptive Statistics. In *An Introduction to Mathematical Statistics* (pp. 21–44). Amsterdam University Press. <http://www.jstor.org/stable/j.ctt1v2xsxr.6>
- [8] Kimberly Leung, Chris Rasmussen, Samuel S. P. Shen, and Dov Zazkis. (2014). Calculus from a Statistics Perspective. *The College Mathematics Journal*, 45(5), 377–386.
- [9] Taylor, Courtney. "Definition of a Percentile in Statistics and How to Calculate It." ThoughtCo, Aug. 26, 2020, [thoughtco.com/what-is-a-percentile-3126238](https://www.thoughtco.com/what-is-a-percentile-3126238).
- [10] Trojanek, Radoslaw & Gluszak, Michal & Hebdzynski, Michal & Tanaś, Justyna. (2021). The COVID-19 Pandemic, Airbnb and Housing Market Dynamics in Warsaw. *Critical Housing Analysis*. 8. 72-84. 10.13060/23362839.2021.8.1.524.
- [11] Klarmann, M., & Feurer, S. (2018). Control Variables in Marketing Research. *Marketing: ZFP – Journal of Research and Management*, 40(2), 26–40.
- [12] Zhang, Lei, and Yimin Yi. "What Contributes to the Rising House Prices in Beijing? A Decomposition Approach." *Journal of Housing Economics*, vol. 41, 2018, pp. 72–84.
- [13] Engelhardt, N., Krause, M., Neukirchen, D., & Posch, P. N. (2021). Trust and stock market volatility during the COVID-19 crisis. *Finance Research Letters*, 38, 101873.
- [14] Guetterman T. C. (2019). Basics of statistics for primary care research. *Family medicine and community health*, 7(2), e000067.