ISCTE - Statistics 1

September 18, 2023

1   Introduction

Email: pnsoa@iscte-iul.pt
Statistics for business and economics newbold pdf
Khan academy on yt

2   Descriptive statistics

2.1   Topics

- Types of variables

- Frequency

- Central Tendency Measures

- Dispersion measures

- N.C.T measures

2.2   Types of variables

2.2.1   Qualitative

Nominal variable
Type of categorical variable where the categories represent distinct groups with no inherent order or ranking

- Blue

- Brown

- Green

Ordinal variable
Type of categorical variable where the categories have a meaningful order or ranking ordinal variable

- High School Diploma

- Bachelor's Degree

- Master's Degree

- Doctoral Degree

2.2.2   Quantitative

Discrete variable
Variables are those that can only take on specific, distinct values and cannot be divided into smaller units.

- Number of cars in a parking lot (0, 1, 2, 3, ...)

- Number of students in a classroom (0, 1, 2, 3, ...)

Continuous variable
Variables represent data that can take on an infinite number of values within a given range.

- Temperature measured in degrees Celsius or Fahrenheit is a continuous variable. It can have decimal values, and there are infinite possibilities within the temperature scale

- Human height. It can vary continuously from the shortest to the tallest individuals, and there are infinitely many possible height values within a specific range.

2.3 Frequency

2.3.1 Frequency table

A frequency table is a table that displays the number of outcomes of a sample that fall into different categories. The categories are usually specified in the first column, and the frequencies are specified in the second column. The frequency of a particular category is the number of times the data value occurs in the data set.

Legend:

- X_i = Value of the variable

- F_i = Frequency of the value

- f_i = Relative frequency of the value

- $cumF_i$ = Cumulative frequency of the value

- $cumf_i$ = Cumulative relative frequency of the value

Data

X_i	F_i	f_i	$cumF_i$	$cumf_i$
13	3	15%		15%
14	2	10%	5	25%
15	2	10%	7	35%
16	0	0%	7	35%
17	0	0%	7	35%
18	4	20%	11	55%
19	3	15%	14	70%
20	6	30%	20	100%
SUM	20	1		

Data values:
13
13
13
14
14
15
15
18
18
18
18
19
19
19
20
20
20
20
20
20

2.4 Mean, median and mode

2.4.1 Mean

The mean, often referred to as the average, is calculated by adding up all the values in a data set and then dividing the sum by the total number of values. It provides a measure of central tendency and represents the "typical" value in a data set.

Mean = (Sum of all values) / (Number of values)

2.4.2 Mode

The median is the middle value in a data set when it is arranged in ascending or descending order. To find the median, you first arrange the data points and then identify the middle value. If there is an even number of values, the median is calculated as the average of the two middle values. The median is less affected by extreme outliers compared to the mean, making it a robust measure of central tendency.

2.4.3 Mode

The mode is the value that appears most frequently in a data set. Unlike the mean and median, which focus on the central tendency, the mode highlights the most common value(s) in the data. A data set can have one mode (unimodal), multiple modes (multimodal), or no mode at all if all values occur with equal frequency.

2.4.4 Example

Data

13	
13	
13	
14	
14	
15	
15	
18	
18	
18	
18	
19	
19	
19	
20	
20	
20	
20	
20	
20	

X_i	F_i	f_i	$cumF_i$	$cumf_i$
13	3	15%		15%
14	2	10%	5	25%
15	2	10%	7	35%
16	0	0%	7	35%
17	0	0%	7	35%
18	4	20%	11	55%
19	3	15%	14	70%
20	6	30%	20	100%
SUM	20	1		

2.5 Quantiles

2.6 Variance dispersion