

Chapter 2

Describing, Exploring, and Comparing Data

Chapter 2

Describing, Exploring, and Comparing Data

- 2-1 Overview**
- 2-2 Frequency Distributions**
- 2-3 Visualizing Data**
- 2-4 Measures of Center**
- 2-5 Measures of Variation**
- 2-6 Measures of Relative Standing**
- 2-7 Exploratory Data Analysis (EDA)**

1

2 -1

Overview

Important Characteristics of Data

- 1. Center:** A representative or average value that indicates where the middle of the data set is located
- 2. Variation:** A measure of the amount that the values vary among themselves
- 3. Distribution:** The nature or shape of the distribution of data (such as bell-shaped, uniform, or skewed)
- 4. Outliers:** Sample values that lie very far away from the vast majority of other sample values
- 5. Time:** Changing characteristics of the data over time

2 -1 Overview

❖ Descriptive Statistics

summarize or describe the important characteristics of a known set of population data

❖ Inferential Statistics

use sample data to make inferences (or generalizations) about a population

2-2 Frequency Distributions

❖ Frequency Table

lists data values (either individually or by groups of intervals), along with their corresponding frequencies (or counts).

2

Qwerty Keyboard Word Ratings

2	2	5	1	2	6	3	3	4	2
4	0	5	7	7	5	6	6	8	10
7	2	2	10	5	8	2	5	4	2
6	2	6	1	7	2	7	2	3	8
1	5	2	5	2	14	2	2	6	3
1	7								

Frequency Table of Qwerty Word Ratings

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Frequency Table

Definitions

Lower Class Limits

are the smallest numbers that can actually belong to different classes

3

Lower Class Limits

are the smallest numbers that can actually belong to different classes

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Lower Class Limits

are the smallest numbers that can actually belong to different classes

Lower Class Limits

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Upper Class Limits

are the largest numbers that can actually belong to different classes

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Upper Class Limits

are the largest numbers that can actually belong to different classes

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Upper Class Limits

4

Class Boundaries

number separating classes

Rating	Frequency
0 - 2	20
3 - 5	14
6 - 8	15
9 - 11	2
12 - 14	1

Class Boundaries

number separating classes

Rating	Frequency
- 0.5	
0 - 2	20
2.5	
3 - 5	14
5.5	
6 - 8	15
8.5	
9 - 11	2
11.5	
12 - 14	1
14.5	

Class Boundaries

Class Midpoints

midpoints of the classes

Class Midpoints

midpoints of the classes

Class Midpoints

Rating	Frequency
0 - 1 2	20
3 - 4 5	14
6 - 7 8	15
9 - 10 11	2
12 - 13 14	1

5

Class Width

is the difference between two consecutive lower class limits or two consecutive class boundaries

Class Width

is the difference between two consecutive lower class limits or two consecutive class boundaries

Class Width

Rating	Frequency
3 0 - 2	20
3 3 - 5	14
3 6 - 8	15
3 9 - 11	2
3 12 - 14	1

Constructing A Frequency Table

1. Decide on the number of classes .
2. Determine the class width by dividing the range by the number of classes and round (up usually).

$$\text{class width} \approx \frac{\text{Highest value} - \text{lowest value}}{\text{number of classes}}$$
3. Select for the first lower limit either the lowest score or a convenient value slightly less than the lowest score.
4. Add the class width to the starting point to get the second lower class limit, add the width to the second lower limit to get the third, and so on.
5. List the lower class limits in a vertical column and enter the upper class limits.
6. Represent each score by a tally mark in the appropriate class.
Total tally marks to find the total frequency for each class.

Guidelines For Frequency Tables

1. Be sure that the classes are mutually exclusive, that is, do not overlap so each data value belongs to only one class.
2. Include all classes, even if the frequency is zero.
3. Try to use the same width for all classes, although sometimes open-ended intervals are necessary.

The sum of the class frequencies must equal the number of original data values.

6

Relative Frequency Table

$$\text{relative frequency} = \frac{\text{class frequency}}{\text{sum of all frequencies}}$$

Relative Frequency Table

Rating	Frequency	Rating	Relative Frequency
0 - 2	20	0 - 2	38.5%
3 - 5	14	3 - 5	26.9%
6 - 8	15	6 - 8	28.8%
9 - 11	2	9 - 11	3.8%
12 - 14	1	12 - 14	1.9%

$$20/52 = 38.5\%$$

$$14/52 = 26.9\%$$

etc.

Total frequency = 52 Total percentage \approx 100%

Cumulative Frequency Table

Rating	Frequency	Rating	Cumulative Frequency
0 - 2	20	Less than 3	20
3 - 5	14	Less than 6	34
6 - 8	15	Less than 9	49
9 - 11	2	Less than 12	51
12 - 14	1	Less than 15	52

} Cumulative Frequencies

Frequency Tables

Rating	Frequency	Rating	Relative Frequency	Rating	Cumulative Frequency
0 - 2	20	0 - 2	38.5%	Less than 3	20
3 - 5	14	3 - 5	26.9%	Less than 6	34
6 - 8	15	6 - 8	28.8%	Less than 9	49
9 - 11	2	9 - 11	3.8%	Less than 12	51
12 - 14	1	12 - 14	1.9%	Less than 15	52

7

