

Workbook



Table of Contents

Theoretical Statistics.....	3
Introduction	3
Categorization of Variables and Measuring Scales.....	4
Summation.....	5
Measure of Central Tendency.....	7
Measures of Spread; Range, Variance and Standard Deviation.....	11
Data Presentation	14
Boxplot	19
Weighted Average and Combined Variance.....	20
Coefficient of Variation.....	21
Analysis of Printout.....	22

Categorization of Variables and Measuring Scales

Questions

- 1) Given the following list of variables, state the measuring scale of the variable listed (nominal, ordinal, interval, ratio):
- Height of a person in cm.
 - Number of children in a family.
 - Degree of anxiety before an exam.
 - Satisfaction with customer service on a scale of 1 to 7.
(1 means completely dissatisfied and 7 means very satisfied)
 - Education.
 - Bus numbers
 - Location of residence.
 - Gender (1 = male, 2 = female).
 - Shoe size.

- 2) The following is the probability distribution of the number of times that employees of the Star Company were late getting to work. The company has 200 employees.

- What is the variable involved?
- Is the variable qualitative or quantitative?

Number of Times Late	Number of Employees
0	17
1	23
2	85
3	50
4	25

- 3) The following is a list of quantitative variables. State next to each variable whether it is continuous or discrete.
- Wages in \$.
 - Mark on a matriculation exam.
 - Results of tossing a die.
 - Running speed in a competition.
 - Proportion of support for the government.

Answer Key

- 1) a. Ratio. b. Ratio. c. Ordinal. d. Ordinal.
e. Cannot be determined without the units measured. f. nominal. g. nominal.
h. nominal. i. Ordinal.
- 2) a. Number of late occurrences. b.
- 3) a. Continuous. b. Discrete. c. Discrete. d. Continuous. e. Continuous.

Summation

Questions

- 1) There are five apartments in a building. The number of rooms in each apartment (X) and the number of people living in the apartment (Y) are listed in the following table.

Calculate:

$$\sum_{i=1}^3 X_i, \sum_{i=1}^5 Y_i, \sum_{i=1}^4 X_i, \left(\sum_{i=1}^4 X_i \right)^2,$$

$$\sum X_i, \sum X_i Y_i, \sum (X_i) \cdot \sum (Y_i).$$

Number of Apartments	X	Y
1	2	1
2	3	1
3	2	2
4	4	3
5	3	2

- 2) The following table gives the value of variables X_i and Y_i , where $i = 1, 2, \dots, 6$:

i	1	2	3	4	5	6
X_i	3	2	4	-2	1	4
Y_i	2	0	0	1	-5	2

Assuming the constants $a = 2$, $b = 5$, calculate all of the following formulas:

a. $\sum_{i=1}^4 y_i$ b. $\sum_{i=1}^6 a$ c. $\sum_{i=1}^6 x_i y_i$ d. $\sum_{i=1}^6 (x_i + y_i)$ e. $\sum_{i=1}^6 x_i + a$

- 3) State whether the following equalities are true or false:

a. $\sum_{i=1}^n bX_i = b \cdot \sum_{i=1}^n X_i$ b. $\sum_{i=1}^n a = a \cdot n$ c. $\left(\sum_{i=1}^n X_i \right)^2 = \sum_{i=1}^n (X_i)^2$

- 4) Given: $\sum_{i=1}^{10} X_i = 80$, $\sum_{i=1}^{10} X_i^2 = 1640$. Calculate: $\sum_{i=1}^{10} (X_i - 4^2)$.

Answer Key

- 1) a. 7 b. 9 c. 11 d. 121
 e. 14 f.27 g.126
- 2) a. 3 b. 12 c. 7 d. 12 e. 14
- 3) a. True b. True c. False
- 4) 1160

Measure of Central Tendency

Questions

- 1) The following is a list of marks of 20 students on a reading comprehension test:
7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6.
Calculate the median, mode, and average of the marks.

- 2) A check of the number of rooms in each apartment in a building with five apartments showed an average of 3.8. The number of rooms in four of the apartments is as follows:
4, 3, 4, 5.
- How many rooms are there in the fifth apartment?
 - What are the mode and the median?

- 3) The following table displays the probability distribution of the number of television sets counted for each family in a given community.

- Calculate the average, median and mode of the probability distribution.
- Explain (without any calculation) how each measure you calculated in a. would change, if some of the families (not all) that did not have a television set up until now, were to buy one.

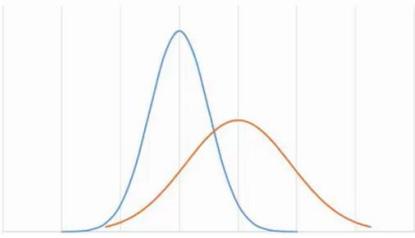
Number of Families	Number of Television Sets
22	0
28	1
18	2
22	3
10	4

- 4) The following table displays the number cars owned by family in the a community

- How many families are there in the community?
- What proportion of families in the community has at most two cars?
- Calculate the average, the median, and the mode.
- Explain in each question the significance of the result you obtained.

No. of cars in the family	1	2	3	4	5
Frequency	65	150	220	140	55

- 5) A teacher teaches two classes. He describes on the same set of coordinate axes the probability distributions of each class' mark as follows (Class 1 in blue, and Class 2 in orange):

<p>Select the correct answer:</p> <ol style="list-style-type: none"> a. The mode of Class 1 is higher than the mode of Class 2. b. The mode of Class 2 is higher than the mode of Class 1. c. Both classes have the same mode. d. There is not enough information to determine which class has the higher mode. 	
---	---

- 6) In a certain community, the number of television sets in each family's home was examined. There are 200 families in the community. The average number of television sets per family is 1.5.
- a. Complete the table.
 - b. Find the mode, the midrange, and the median.
 - c. Some of the families that had exactly one television set took it out of their homes. How will each measure change (increase, decrease, unchanged). Explain your answer without any calculation.

Number of Television Sets	Number of Families
0	28
1	62
2	
3	

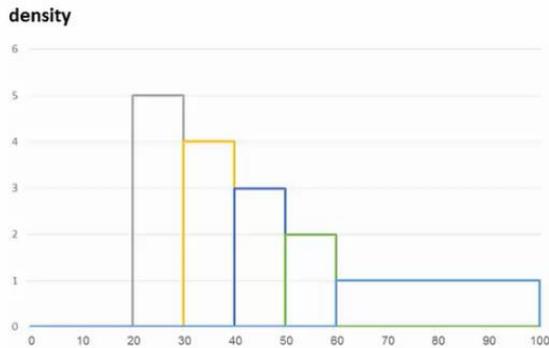
- 7) The following table displays weight in kg for a certain group: What are the average and median of the probability distribution?

Weight	No. of Cases
45-40	10
50-45	20
60-50	30
65-60	20
70-65	10

- 8) The following table displays the probability distribution of the height in cm in a certain group: Calculate the average, the median, and the mode of height in this group.

Frequency	Height in cm
30	160-150
40	170-160
60	175-170
70	180-175
40	190-180

- 9) In a certain faculty, the hourly wage of working students was examined. The results were as follows:



- Find the mode of the probability distribution.
- Find the median of the probability distribution.
- Explain (without any calculations) whether the average is larger/smaller/equal to the median.
- It is learned that a number of students in the 20–30\$ category must be removed from the population. How will this affect the average, the median, and the mode? Explain without any calculations.

Answer Key

1) Median: 7 Mode: 6 Average: 6.9

2) a. 3 b. Mode: 3 and 4; Median: 4

3) a. Average: 1.7; Median: 1.5; Mode: 1

b. The average will increase, and the other measures will not change.

4) a. 630 b. 34.14% c. Mode and Median: 3; Average: 2.95

5) b

6)

a.

# TV Sets	# Families
0	28
1	62
2	92
3	18
Total	200

b. Median: 2; Mode: 2; Midrange: 1.5

c. Mode: unchanged; Mid-Range: unchanged;

Median: No Change; Average: Decrease

7) Median and Average: 55

8) Average: 172.6; Median: 174.17; Mode: 177.5

9) a. Mode: 25 b. Median: 40

c. The average is greater than the mode because the distribution is right asymmetric (long tail to the right).

d. Will not affect the Mode, and will Increase the Median and the Average

Measures of Spread; Range, Variance and Standard Deviation

Questions

- 1) The following are the marks of 20 students from a reading comprehension test:
7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6.
Calculate the variance, standard deviation, and range of the marks.

- 2) The following table displays the number of cars owned by family in a certain community:

Number of cars in the family	1	2	3	4	5
Frequency	65	150	220	140	55

- Calculate the standard deviation.
 - Calculate the range of the data.
 - Explain in each question the meaning of the results that you obtained.
- 3) A telemarketing company examined the seniority (in years) of its employees. The average seniority is four years, and the standard deviation is two years.
- Will the average increase/decrease/remain unchanged, and will the standard deviation increase/decrease/remain unchanged if two employees with four years of seniority each are added to the probability distribution?
 - Will the average increase/decrease/remain unchanged, and will the standard deviation increase/decrease/remain unchanged if two employees are added to the probability distribution: one with 0 years of seniority and one with eight years of seniority?
- 4) A list of five observations is given, but the deviation from the average is given for only four of them: $-1, 2, 3, 2$.
Calculate the variance of the five observations.

- 5) The number of rooms in each apartment is examined in a certain neighborhood. There are 200 apartments in the neighborhood.

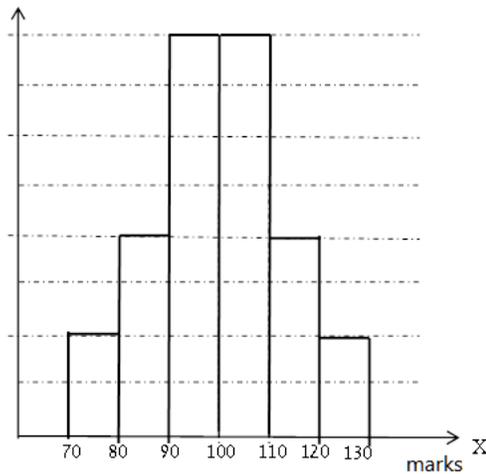
# Rooms	Proportion
1	0.1
2	0.2
3	0.4
4	0.15
5	

- What is the average number of rooms per apartment in the neighborhood?
 - Calculate the standard deviation of the number of rooms.
 - Some of the owners of a 2 room apartment knocked out a wall and converted their apartment to 1 room.
 - How will this affect (increase/decrease/no change) the average and standard deviation?
- 6) The following table displays the probability distribution of weight in kg for a certain

group:
What is the standard deviation of the weight probability distribution?

Weight	No. of Cases
45-40	10
50-45	20
60-50	30
65-60	20
70-65	10

7) The following diagram displays the probability distribution of marks on an intelligence test:



- a. What are the average and median of the probability distribution?
- b. Calculate the standard deviation of the marks.
- c. It is learned that 20 observations must be added to both the 90-100 and 100-110 categories. How will this affect the average and standard deviation?

Answer Key

- 1) Variance: $S_x^2 = 2.19 \text{ Marks}^2$; Standard deviation: $S = 1.479$ Range: 6
- 2) a. Standard deviation: 1.106 b. Range:4
- 3) a. The average will not change, and the standard deviation will decrease.
b. The average will not change, and the standard deviation will increase.
- 4) 10.8.
- 5) a. 3.05 b. 1.16
c. The Average will decrease, and the Standard Deviation will increase.
- 6) 7.73
- 7) a. $\bar{X} = Md = 100$ b. 12.96
c. The Average will not change, and the Standard Deviation will Decrease.

Data Presentation

Questions

- 1) In a television rating survey, the following results were obtained: 25 people watched Channel 1, 10 watched Channel 10, 75 watched Channel 2, 50 watched one of the cable TV channels, and 25 did not watch television at the time of the survey.
- Write the frequency table and the relative frequency.
 - Describe the data graphically.

- 2) The following table displays figures for the favorite subject of 6th grade students at the City High School:

- What is the variable involved?
- What proportion of students prefers Bible?

Number of Students	Subject
44	Mathematics
20	Bible
12	English
26	History

- 3) The following table displays figures for education at a specific place of work

- What is the variable involved?
What type of variable is it?
- Describe the data graphically.

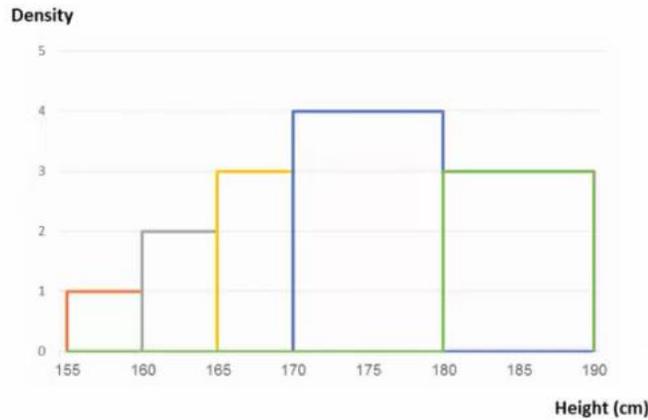
Number of Employees	Education
60	Low
120	High-school
20	University

- 4) The following is a list of marks of 20 students tested for their reading skills:

7, 6, 8, 9, 10, 6, 4, 5, 8, 7, 6, 7, 6, 8, 9, 6, 7, 8, 5, 6.

- What is the variable involved? Is it discrete or continuous?
- Describe the list in a frequency table.
- Add relative frequencies to the table.
- Describe the data graphically.

- 5) The following is a histogram describing the probability distribution of height in cm of a certain group:



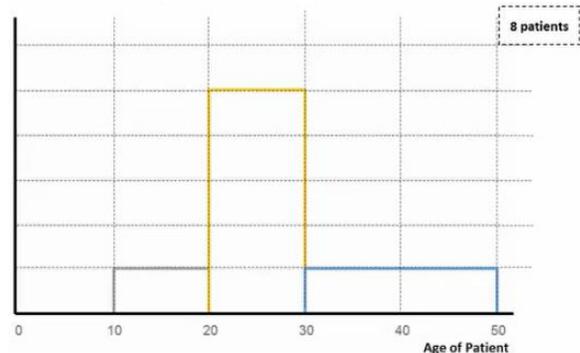
- What is the variable involved? Is it discrete or continuous?
 - Describe the data in a frequency table by categories.
 - Add relative frequency to the table.
 - Add the density of each category to the table.
 - What is the form of the probability distribution of height?
- 6) The following table contains the probability distribution of a certain group's weight in kg:

- Describe the probability distribution graphically.
- What can be said about the form of the probability distribution?

Weight	Number of cases
40-45	10
45-50	20
50-60	30
60-65	20
65-70	10

- 7) The following chart describes the age of Dr. Schwartz's patients, in years:

- What is the variable involved? Is it discrete or continuous?
- What is the group that is involved?
- Translate the histogram into a frequency table.
- What is the proportion of Dr. Schwartz's patients aged 20-30?



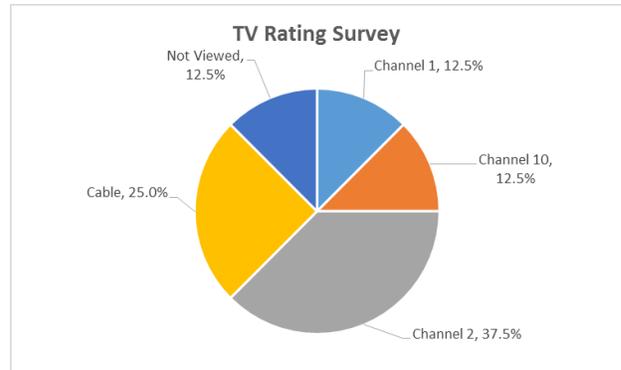
Answer Key

1)

a.

Channel	$f(x)$	$\frac{f(x)}{n}$	%
Channel 1	25	$\frac{25}{200}$	12.5%
Channel 10	25	$\frac{25}{200}$	12.5%
Channel 2	75	$\frac{75}{200}$	37.5%
Cable	50	$\frac{50}{200}$	25%
Not Viewed	25	$\frac{25}{200}$	12.5%
Total	200	1	100%

b.

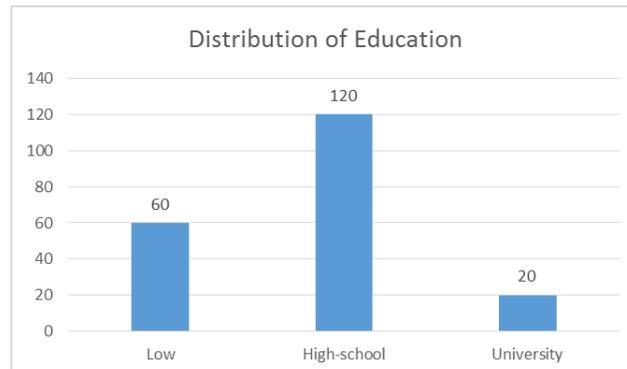


2) a. Subject

b. 19.6%

3) a. The variable is the level of education; It is Qualitative and measured by Order.

b.



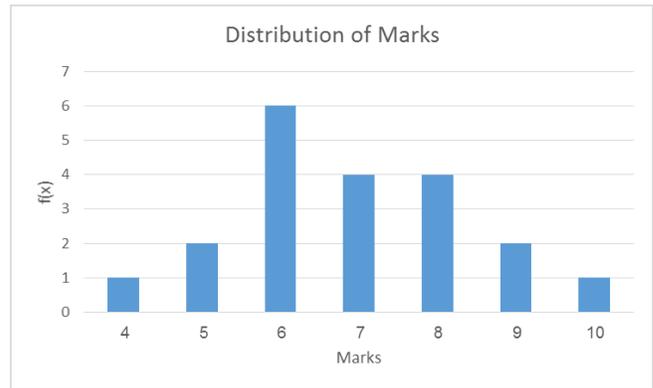
4)

a. The variable is the grade; It is discrete

b.

x	$f(x)$	$\frac{f(x)}{n}$	%
4	1	$\frac{25}{20}$	5%
5	2	$\frac{25}{20}$	10%
6	6	$\frac{25}{20}$	30%
7	4	$\frac{25}{20}$	20%
8	4	$\frac{25}{20}$	20%
9	2	$\frac{25}{20}$	10%
10	1	$\frac{25}{20}$	5%
Total	20	$\frac{25}{20}$	100%

c.



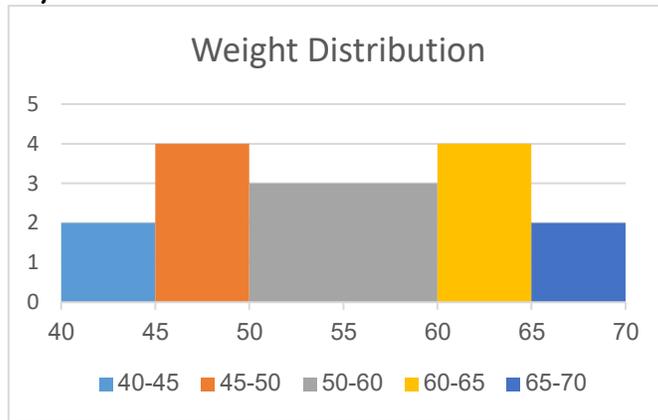
5) a. Height; Continuous variable

b+c+d:

x	L (Width of bar)	Density (d)	$f(x) = d*L$	$f(x)/n$
155-160	5	1	5	5%
160-165	5	2	10	10%
165-170	5	3	15	15%
170-180	10	4	40	40%
180-190	10	3	30	30%
Total			100	100%

e. Left Asymmetrical distribution

6) a.



b. It is symmetrical and has 2 maximums.

7) a. Age (in years); Continuous
c. Table:

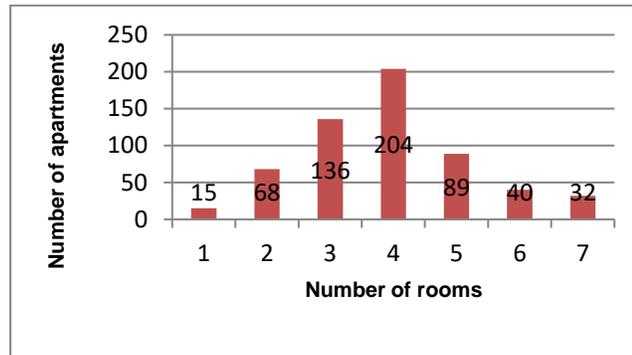
x	f(x)
10-20	8
20-30	40
30-50	16
Total	64

b. Patients
d. 62.5%

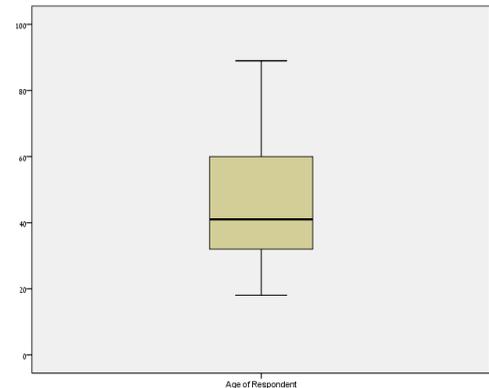
Boxplot

Questions

- 1) The following chart displays the probability distribution of the number of rooms in apartments built in 2009 in Boston:



- Calculate the median, the lower quartile and the upper quartile of the probability distribution.
 - Sketch a boxplot diagram of the probability distribution.
 - What can be said about the form of the probability distribution?
- 2) The following is a boxplot diagram describing the probability distribution of age (in years) in a certain population:
- What is the approximate median age in the population?
 - What is the approximate range of the ages?
 - What can be said about the form of the probability distribution?



Answer Key

- 1) a. Median: 4 Lower quartile: 3 Upper quartile: 5 c. Almost symmetric
 2) a. Median: 40 b. Range: 70 c. Right asymmetric probability

Weighted Average and Combined Variance

Questions

1) The following table displays the marks on an English exam for three 10th grade classes at a high school:

- a. Calculate the weighted average for all of the 10th grade.
- b. Calculate the combined variance for 10th grade.

Class	Average Mark	No. of Students	Standard Deviation
1	76	40	12
2	68	20	15
3	82	30	10

2) You are given two groups:

Group I has twice as many observations as Group II.

The average of each groups is 70.

The variance in Group I is 100.

The variance in Group II is 400.

- a. Find the average of the observations after the two groups have been consolidated into a single group.
- b. Find the standard deviation of the observations after the two groups have been consolidated into a single group.

Answer Key

- 1) a. 76.22 b. 173.5
2) a. 70 b. 14.14

Coefficient of Variation

Questions

- 1) The following table displays the marks on an exam in English for three 10th grade classes:

Class	Average	No. of Students	Standard Deviation
1	76	40	12
2	68	20	15
3	82	30	10

- a. Calculate the coefficient of variation for each class.
 - b. Which is the most heterogeneous group?
- 2) Two groups are given:
 The average in Group *A* is 100 and the variance is 100.
 The average in Group *B* is 500 and the variance is 400.
 In which group is the measure of dispersion smaller?

- 3) In a car battery factory, we count the number of employees per day, and measure their daily output (the number of batteries – in hundreds) for 40 continuous days.
 The following table shows the information collected about the two variables:

	Output	No. of Employees
Average	48	15
Standard deviations	10	2

Select the correct answer according to the coefficient of variation criterion:

- a. The relative dispersion is greater for the daily output than for the number of employees in a day.
- b. The relative dispersion is greater for the number of employees in a day than for the daily output.
- c. The data is not sufficient to calculate the coefficient of variation.
- d. The relative dispersion is equal to the daily output per the number of employees working that day.

Answer Key

- 1) a. Class 1: 0.158 Class 2: 0.221 Class 3: 0.112 b. Class 2
 2) Group B 3) b

Analysis of Printout

Questions

- 1) The following table displays a printout of the probability distribution of ages in a certain population:

- a. Find the values of **a** and **b** in the table above.
- b. It is given that the probability distribution is asymmetric. Is it right or left asymmetric?

		Statistic
Age of Respondent	Mean	45.63
	Median	41.00
	Variance	317.140
	Std. Deviation	a
	Minimum	18
	Maximum	b
	Range	71
	Inter-quartile Range	28

- 2) The following table displays the probability distribution of the employees' level of education at a specific company:

Fill in the values marked with question marks.

		Statistic
Years of Education	Mean	?
	Median	12.0000
	Variance	?
	Std. Deviation	2.54786
	Minimum	?
	Maximum	?
	Range	?
	Inter-quartile Range	?

Answer Key

- 1) **a** = 17.81 **b** = 89 b. Right asymmetric
- 2) Average: 11.909 Variance: 6.492 Minimum: 8 Maximum: 18
 Range: 10 Inter-quartile range: 3