

CHAPTER THREE
DESCRIPTIVE STATISTICS: NUMERICAL MEASURES

MULTIPLE CHOICE QUESTIONS

In the following multiple choice questions, circle the correct answer.

1. Which of the following provides a measure of central location for the data?
 - a. standard deviation
 - b. mean
 - c. variance
 - d. rangeAnswer: b

2. A numerical value used as a summary measure for a sample, such as sample mean, is known as a
 - a. population parameter
 - b. sample parameter
 - c. sample statistic
 - d. population meanAnswer: c

3. Since the population size is always larger than the sample size, then the sample statistic
 - a. can never be larger than the population parameter
 - b. can never be equal to the population parameter
 - c. can be smaller, larger, or equal to the population parameter
 - d. can never be smaller than the population parameterAnswer: c

4. μ is an example of a
 - a. population parameter
 - b. sample statistic
 - c. population variance
 - d. modeAnswer: a

5. The hourly wages of a sample of 130 system analysts are given below.
mean = 60 range = 20
mode = 73 variance = 324
median = 74
The coefficient of variation equals
 - a. 0.30%
 - b. 30%

2 Chapter Three

c. 5.4%

d. 54%

Answer: b

6. The variance of a sample of 169 observations equals 576. The standard deviation of the sample equals

a. 13

b. 24

c. 576

d. 28,461

Answer: b

7. The median of a sample will always equal the

a. mode

b. mean

c. 50th percentile

d. all of the above answers are correct

Answer: c

8. The median is a measure of

a. relative dispersion

b. absolute dispersion

c. central location

d. relative location

Answer: c

9. The 75th percentile is referred to as the

a. first quartile

b. second quartile

c. third quartile

d. fourth quartile

Answer: c

10. The p th percentile is a value such that at least p percent of the observations are

a. less than or equal to this value

b. less than this value

c. more than or equal to this value

d. more than this value

Answer: a

11. The difference between the largest and the smallest data values is the

a. variance

b. interquartile range

c. range

d. coefficient of variation

Answer: c

12. The first quartile
- a. contains at least one third of the data elements
 - b. is the same as the 25th percentile
 - c. is the same as the 50th percentile
 - d. is the same as the 75th percentile
- Answer: b
13. Which of the following is not a measure of central location?
- a. mean
 - b. median
 - c. variance
 - d. mode
- Answer: c
14. If a data set has an even number of observations, the median
- a. cannot be determined
 - b. is the average value of the two middle items
 - c. must be equal to the mean
 - d. is the average value of the two middle items when all items are arranged in ascending order
- Answer: d
15. Which of the following is a measure of dispersion?
- a. percentiles
 - b. quartiles
 - c. interquartile range
 - d. all of the above are measures of dispersion
- Answer: c
16. The most frequently occurring value of a data set is called the
- a. range
 - b. mode
 - c. mean
 - d. median
- Answer: b
17. The interquartile range is
- a. the 50th percentile
 - b. another name for the variance
 - c. the difference between the largest and smallest values
 - d. the difference between the third quartile and the first quartile
- Answer: d
18. If index i (which is used to determine the location of the p th percentile) is not an integer, its value should be
- a. squared
 - b. divided by $(n - 1)$
 - c. rounded down

d. rounded up

Answer: d

19. When the data are skewed to the right, the measure of Skewness will be

a. negative

b. zero

c. positive

d. one

Answer: c

20. When data are positively skewed, the mean will usually be

a. greater than the median

b. smaller than the median

c. equal to the median

d. positive

Answer: a

21. Which of the following is **not** a measure of dispersion?

a. the range

b. the 50th percentile

c. the standard deviation

d. the interquartile range

Answer: b

22. The interquartile range is used as a measure of variability to overcome what difficulty of the range?

a. the sum of the range variances is zero

b. the range is difficult to compute

c. the range is influenced too much by extreme values

d. the range is negative

Answer: c

23. If the variance of a data set is correctly computed with the formula using $n - 1$ in the denominator, which of the following is true?

a. the data set is a sample

b. the data set is a population

c. the data set could be either a sample or a population

d. the data set is from a census

Answer: a

24. In computing descriptive statistics from grouped data,

a. data values are treated as if they occur at the midpoint of a class

b. the grouped data result is more accurate than the ungrouped result

c. the grouped data computations are used only when a population is being analyzed

d. None of these alternatives is correct.

Answer: a

25. The measure of dispersion that is influenced most by extreme values is
- the variance
 - the standard deviation
 - the range
 - the interquartile range
- Answer: c
26. When should measures of location and dispersion be computed from grouped data rather than from individual data values?
- as much as possible since computations are easier
 - only when individual data values are unavailable
 - whenever computer packages for descriptive statistics are unavailable
 - only when the data are from a population
- Answer: b
27. The descriptive measure of dispersion that is based on the concept of a deviation about the mean is
- the range
 - the interquartile range
 - the absolute value of the range
 - the standard deviation
- Answer: d
28. The measure of location which is the most likely to be influenced by extreme values in the data set is the
- range
 - median
 - mode
 - mean
- Answer: d
29. The most important statistical descriptive measure of the location of a data set is the
- mean
 - median
 - mode
 - variance
- Answer: a
30. The numerical value of the standard deviation can never be
- larger than the variance
 - zero
 - negative
 - smaller than the variance
- Answer: c
31. The sample variance
- is always smaller than the true value of the population variance

- b. is always larger than the true value of the population variance
- c. could be smaller, equal to, or larger than the true value of the population variance
- d. can never be zero

Answer: c

32. The variance can never be
- a. zero
 - b. larger than the standard deviation
 - c. negative
 - d. smaller than the standard deviation

Answer: c

33. If two groups of numbers have the same mean, then
- a. their standard deviations must also be equal
 - b. their medians must also be equal
 - c. their modes must also be equal
 - d. None of these alternatives is correct

Answer: d

34. The sum of deviations of the individual data elements from their mean is
- a. always greater than zero
 - b. always less than zero
 - c. sometimes greater than and sometimes less than zero, depending on the data elements
 - d. always equal to zero

Answer: d

35. Which of the following symbols represents the standard deviation of the population?
- a. σ^2
 - b. σ
 - c. μ
 - d. \bar{x}

Answer: b

36. Which of the following symbols represents the mean of the population?
- a. σ^2
 - b. σ
 - c. μ
 - d. \bar{x}

Answer: c

37. Which of the following symbols represents the variance of the population?
- a. σ^2
 - b. σ

- c. μ
- d. \bar{x}

Answer: a

38. Which of the following symbols represents the size of the population?

- a. σ^2
- b. σ
- c. μ
- d. N

Answer: d

39. Which of the following symbols represents the mean of the sample?

- a. σ^2
- b. σ
- c. μ
- d. \bar{x}

Answer: d

40. Which of the following symbols represents the size of the sample

- a. σ^2
- b. σ
- c. N
- d. n

Answer: d

41. The symbol σ is used to represent

- a. the variance of the population
- b. the standard deviation of the sample
- c. the standard deviation of the population
- d. the variance of the sample

Answer: c

42. The symbol σ^2 is used to represent

- a. the variance of the population
- b. the standard deviation of the sample
- c. the standard deviation of the population
- d. the variance of the sample

Answer: a

43. A numerical measure of linear association between two variables is the

- a. variance
- b. covariance
- c. standard deviation
- d. coefficient of variation

Answer: b

44. Positive values of covariance indicate
- a. a positive variance of the x values
 - b. a positive variance of the y values
 - c. the standard deviation is positive
 - d. positive relation between the independent and the dependent variables

Answer: d

45. A numerical measure of linear association between two variables is the
- a. variance
 - b. coefficient of variation
 - c. correlation coefficient
 - d. standard deviation

Answer: c

46. The coefficient of correlation ranges between
- a. 0 and 1
 - b. -1 and +1
 - c. minus infinity and plus infinity
 - d. 1 and 100

Answer: b

47. The coefficient of correlation
- a. is the same as the coefficient of determination
 - b. can be larger than 1
 - c. cannot be larger than 1
 - d. cannot be negative

Answer: c

48. When the smallest and largest percentage of items are removed from a data set and the mean is computed, the mean of the remaining data is
- a. the median
 - b. the mode
 - c. the trimmed mean
 - d. any of the above

Answer: c

49. In a five number summary, which of the following is **not** used for data summarization?
- a. the smallest value
 - b. the largest value
 - c. the mean
 - d. the 25th percentile

Answer: c

50. During a cold winter, the temperature stayed below zero for ten days (ranging from -20 to -5). The variance of the temperatures of the ten day period
- a. is negative since all the numbers are negative

- b. must be at least zero
 - c. cannot be computed since all the numbers are negative
 - d. can be either negative or positive
- Answer: b
51. Which of the following is **not** a measure of dispersion?
- a. mode
 - b. standard deviation
 - c. range
 - d. interquartile range
- Answer: a
52. Since the mode is the most frequently occurring data value, it
- a. can never be larger than the mean
 - b. is always larger than the median
 - c. is always larger than the mean
 - d. None of these alternatives is correct.
- Answer: d

Exhibit 3-1

The following data show the number of hours worked by 200 statistics students.

Number of Hours	Frequency
0 - 9	40
10 - 19	50
20 - 29	70
30 - 39	40

53. Refer to Exhibit 3-1. The class width for this distribution
- a. is 9
 - b. is 10
 - c. is 11
 - d. varies from class to class
- Answer: b
54. Refer to Exhibit 3-1. The cumulative relative frequency for the class of 10 - 19
- a. is 90
 - b. is .25
 - c. is .45
 - d. cannot be determined from the information given
- Answer: c

Exhibit 3-2

A researcher has collected the following sample data

5	12	6	8	5
6	7	5	12	4

55. Refer to Exhibit 3-2. The median is

- a. 5
- b. 6
- c. 7
- d. 8

Answer: b

56. Refer to Exhibit 3-2. The mode is

- a. 5
- b. 6
- c. 7
- d. 8

Answer: a

57. Refer to Exhibit 3-2. The mean is

- a. 5
- b. 6
- c. 7
- d. 8

Answer: c

58. Refer to Exhibit 3-2. The 75th percentile is

- a. 5
- b. 6
- c. 7
- d. 8

Answer: d

Exhibit 3-3

A researcher has collected the following sample data. The mean of the sample is 5.

3 5 12 3 2

59. Refer to Exhibit 3-3. The variance is

- a. 80
- b. 4.062
- c. 13.2
- d. 16.5

Answer: d

60. Refer to Exhibit 3-3. The standard deviation is

- a. 8.944
- b. 4.062
- c. 13.2
- d. 16.5

Answer: b

61. Refer to Exhibit 3-3. The coefficient of variation is

- a. 72.66%
- b. 81.24%
- c. 264%
- d. 330%

Answer: b

62. Refer to Exhibit 3-3. The range is

- a. 1
- b. 2
- c. 10
- d. 12

Answer: c

63. Refer to Exhibit 3-3. The interquartile range is

- a. 1
- b. 2
- c. 10
- d. 12

Answer: b

Exhibit 3-4

The following is the frequency distribution for the speeds of a sample of automobiles traveling on an interstate highway.

Speed Miles per Hour	Frequency
50 - 54	2
55 - 59	4
60 - 64	5
65 - 69	10
70 - 74	9
75 - 79	<u>5</u>
	35

64. Refer to Exhibit 3-4. The mean is

- a. 35
- b. 670
- c. 10
- d. 67

Answer: d

65. Refer to Exhibit 3-4. The variance is

- a. 6.969
- b. 7.071
- c. 48.570
- d. 50.000

Answer: d

66. Refer to Exhibit 3-4. The standard deviation is
- a. 6.969
 - b. 7.071
 - c. 48.570
 - d. 50.000

Answer: b

PROBLEMS

1. In 2002, the average age of students at UTC was 22 with a standard deviation of 3.96. In 2003, the average age was 24 with a standard deviation of 4.08. In which year do the ages show a more dispersed distribution? Show your complete work and support your answer.

Answer:

C.V. for 2002 = 18%, C.V. for 2003 = 17%

Therefore 2002 shows a more dispersed distribution.

2. A private research organization studying families in various countries reported the following data for the amount of time 4-year old children spent alone with their fathers each day.

Country	Time with Dad (minutes)
Belgium	30
Canada	44
China	54
Finland	50
Germany	36
Nigeria	42
Sweden	46
United States	42

For the above **sample**, determine the following measures:

- The mean
- The standard deviation
- The mode
- The 75th percentile

Answers:

- 43
- 7.56
- 42
- 48

3. The following frequency distribution shows the ACT scores of a **sample** of students:

Score	Frequency
14 - 18	2
19 - 23	5
24 - 28	12
29 - 33	1

For the above data, compute the following.

- a. The mean
- b. The standard deviation

Answers:

- a. 24
- b. 3.77

4. The following is a frequency distribution for the ages of a sample of employees at a local company.

Age	Frequency
30 - 39	2
40 - 49	3
50 - 59	7
60 - 69	5
70 - 79	1

- a. Determine the average age for the sample.
- b. Compute the variance.
- c. Compute the standard deviation.
- d. Compute the coefficient of variation.

Answers:

- a. 54.5
- b. 117.65
- c. 10.85
- d. 19.91%

5. The number of hours worked per week for a sample of ten students is shown below.

Student	Hours
1	20
2	0
3	18
4	16
5	22
6	40
7	8
8	6
9	30
10	40

- Determine the median and explain its meaning.
- Compute the 70th percentile and explain its meaning.
- What is the mode of the above data? What does it signify?

Answers:

- 19; approximately 50% of the students work at least 19 hours
- 26; at least 70% of the students work less than or equal to 26 hours per week
- 40; the most frequent data element

6. For the following frequency distribution,

Class	Frequency
45 - 47	3
48 - 50	6
51 - 53	8
54 - 56	2
57 - 59	1

- Compute the mean.
- Compute the standard deviation. (Assume the data represent a population.)

Answers:

- 50.8
- 3.06

7. The following observations are given for two variables.

y	x
5	2
8	12
18	3
20	6
22	11
30	19
10	18
7	9

- a. Compute and interpret the sample covariance for the above data.
- b. Compute the standard deviation for x.
- c. Compute the standard deviation for y.
- d. Compute and interpret the sample correlation coefficient.

Answers:

- a. 19.286 (rounded). Since the covariance is positive, it indicates a positive relationship between x and y.
 - b. 6.32
 - c. 8.83
 - d. 0.345. There is a positive relationship between x and y. The relationship is not very strong.
8. Paul, a freshman at a local college just completed 15 credit hours. His grade report is presented below.

Course	Credit Hours	Grades
Calculus	5	C
Biology	4	A
English	3	D
Music	2	B
P.E.	1	A

The local university uses a 4 point grading system, i.e., A = 4, B = 3, C = 2, D = 1, F = 0. Compute Paul's semester grade point average.

Answer:

2.6

9. The population change between 1990 and 2000 for several small cities are shown below.

City	Population Change (number of residents)
Chattanooga	3083
Collegedale	1466
East Ridge	-461
Lakeside	1113
Ridgeside	-11
Signal Mountain	395
Soddy-Daisy	3290
Walden	437

For the above **sample**, determine the following measures.

- a. The mean
- c. The standard deviation
- d. The median

Answers:

- a. 1,164
- b. 1,385.51
- c. 775

10. Paul, a freshman at a local college just completed 18 credit hours. His grade report is presented below.

Course	Credit Hours	Grades
Chemistry	5	C
Calculus	5	A
English	4	C
Music	3	F
P.E.	1	A

The local university uses a 4 point grading system, i.e., A = 4, B = 3, C = 2, D = 1, F = 0. Compute Paul's semester grade point average and standard deviation.

Answer:
2.33, 0.81