UNIVERSITY OF ESWATINI

RE SIT EXAMINATION PAPER 2018/2019

TITLE OF PAPER:	DESCRIPTIVE STATISTICS
COURSE CODE:	STA 131/ STA 132/ IDE 132
TIME ALLOCATED:	2 (TWO) HOURS
REQUIREMENTS:	CALCULATOR
INSTRUCTION:	ANSWER ALL PARTS OF SECTION A AND ANY 1 (ONE) QUESTION OF YOUR CHOICE FROM SECTION B. SECTION A IS **COMPULSORY AND THE ANSWER SHEET IS ATTACHED AT THE END OF THE QUESTION PAPER.** ALL QUESTIONS CARRY THE MARKS AS INDICATED WITHIN THE PARENTHESIS.

THIS PAPER IS NOT TO BE OPENED UNTIL PERMISSION HAS BEEN GRANTED BY THE INVIGILATOR.

SECTION A : Compulsory Questions *(Multiple Choice Questions)* [60 Marks]

1. Find the least value, greatest value, mean, median, mode, and range of the data set.

Stems	Leaves
2	2 4 8 8 8 9 9
3	0 2 2 4 6 6 8
4	1 1 4 6
5	0 2

a. least value: 22
greatest value: 52
mean: 33
median: 35
mode: 28
range: 30

b. least value: 22
greatest value: 52
mean: 35
median: 33
mode: 28
range: 30

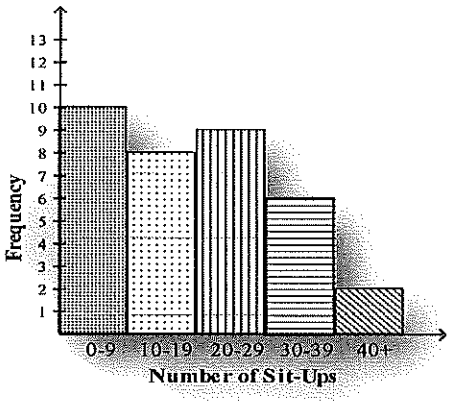
c. least value: 22
greatest value: 52
mean: 28
median: 30
mode: 33
range: 35

d. least value: 0
greatest value: 9
mean: 4.5
median: 3
mode: 8
range: 9

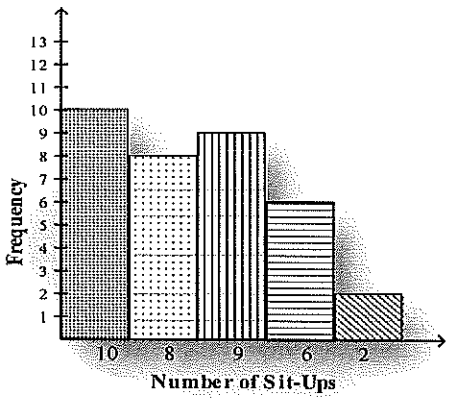
2. Use the frequency table to make a histogram.

Number of Sit-ups Students Can Do in 1 Minute					
Number	0 – 9	10 – 19	20 – 29	30 – 39	40 +
Frequency	10	8	9	6	2

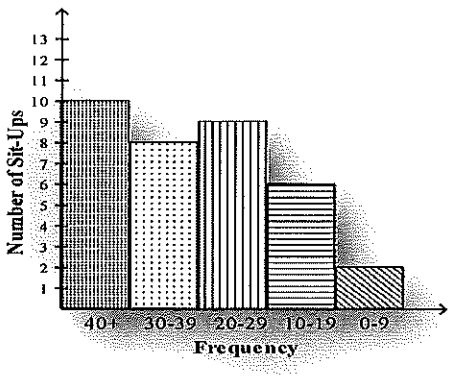
a.



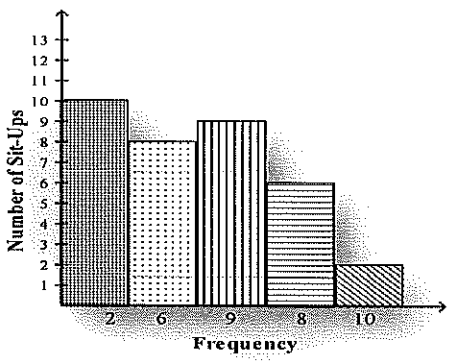
c.



b.



d.



3. The sum of frequencies for all classes will always equal
   a. 1
   b. the number of elements in a data set
   c. the number of classes
   d. a value between 0 and 1

4. A tabular summary of a set of data showing the fraction of the total number of items in several classes is a
   a. frequency distribution
   b. relative frequency distribution
   c. frequency
   d. cumulative frequency distribution

5. If several frequency distributions are constructed from the same data set, the distribution with the widest class width will have the
   a. fewest classes
   b. most classes
   c. same number of classes as the other distributions since all are constructed from the same data

6. The sum of the percent frequencies for all classes will always equal
   a. one
   b. the number of classes
   c. the number of items in the study
   d. 100

7. A tabular method that can be used to summarize the data on two variables simultaneously is called
   a. simultaneous equations
   b. crosstabulation
   c. a histogram
   d. an ogive

8. 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 mean
e. None of the above answers is correct.

9. 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 never be zero
   d. can never be smaller than the population parameter
   e. None of the above answers is correct.

10. μ is an example of a
   a. population parameter
   b. sample statistic
   c. population variance
   d. mode
   e. None of the above answers is correct.

11. The mean of a sample
   a. is always equal to the mean of the population
   b. is always smaller than the mean of the population
   c. is computed by summing the data values and dividing the sum by (n - 1)
   d. is computed by summing all the data values and dividing the sum by the number of items
   e. None of the above answers is correct.

12. 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 median
   d. the 25th percentile
   e. the mean

13. 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. must have a value of at least two
   e. None of the above answers is correct.

Exhibit 2-2

A survey of 800 college seniors resulted in the following cross tabulation regarding their undergraduate major and whether or not they plan to go to graduate school.

	Undergraduate Major			
Graduate School	Business	Engineering	Others	Total
Yes	70	84	126	280
No	182	208	130	520
Total	252	292	256	800

14. Refer to Exhibit 2-2. What percentage of the students does not plan to go to graduate school?
    a. 280
    b. 520
    c. 65
    d. 32

15. Refer to Exhibit 2-2. Of those students who are majoring in business, what percentage plans to go to graduate school?
    a. 27.78
    b. 8.75
    c. 70
    d. 72.22

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

16. 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
    e. None of the above answers is correct.

17. Refer to Exhibit 3-1. The number of students working 19 hours or less
    a. is 40
    b. is 50
    c. is 90
    d. cannot be determined without the original data
    e. None of the above answers is correct.

18. Refer to Exhibit 3-1. The relative frequency of students working 9 hours or less
    a. is 0.2
    b. is 0.45
    c. is 40
    d. cannot be determined from the information given
    e. None of the above answers is correct.

19. Refer to Exhibit 3-1. The cumulative relative frequency for the class of 10 - 19
    a. is 90
    b. is 0.25
    c. is 0.45
    d. cannot be determined from the information given
    e. None of the above answers is correct.

20. The 50th percentile is the

a. mode
b. median
c. mean
d. third quartile
e. None of the above answers is correct.

21. The lower hinge is essentially the same as the
a. 10th percentile
b. third quartile
c. second quartile
d. 25th percentile
e. None of the above answers is correct.

22. In computing the hinges for data with an odd number of items, the median position is included
a. only in the computation of the lower hinge
b. only in the computation of the upper hinge
c. both in the computation of the lower and upper hinges
d. None of the above answers is correct.

23. 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
e. None of the above answers is correct.

24. 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
e. None of the above answers is correct.

25. The value which has half of the observations above it and half the observations below it is called the
a. range
b. median
c. mean
d. mode
e. None of the above answers is correct.

26. The most frequently occurring value of a data set is called the
a. range
b. mode
c. mean
d. median
e. None of the above answers is correct.

27. 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
e. None of the above answers is correct.

28. The median of a sample will always equal the
a. mode
b. mean
c. 50th percentile
d. all of the above answers are correct
e. None of the above answers is correct.

29. The variance of a sample of 81 observations equals 64. The standard deviation of the sample equals
a. 9
b. 4096
c. 8
d. 6561
e. None of the above answers is correct.

Exhibit 3-2
A researcher has collected the following sample data

5 12 6 8 5 6 7 5 12 4

30. Refer to Exhibit 3-2. The median is
a. 5
b. 6
c. 7
d. 8
e. None of the above answers is correct.

31. Refer to Exhibit 3-2. The mode is
a. 5
b. 6
c. 7
d. 8
e. None of the above answers is correct.

32. Refer to Exhibit 3-2. The mean is
a. 5
b. 6
c. 7
d. 8
e. None of the above answers is correct.

33. Refer to Exhibit 3-2. The 75th percentile is
a. 5
b. 6
c. 7
d. 8
e. None of the above answers is correct.

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	5

34. Refer to Exhibit 3-4. The mean is
    a. 35
    b. 670
    c. 10
    d. 67
    e. None of the above answers is correct.

35. Refer to Exhibit 3-4. The variance is
    a. 6.969
    b. 7.071
    c. 48.570
    d. 50.000
    e. None of the above answers is correct.

36. Refer to Exhibit 3-4. The standard deviation is
    a. 6.969
    b. 7.071
    c. 48.570
    d. 50.000
    e. None of the above answers is correct.

37. Which of the following symbols represents the variance of the population?
    a. σ^2
    b. σ
    c. μ
    d. $\bar{x}$
    e. N

38. Which of the following symbols represents the size of the population?
    a. σ^2
    b. σ
    c. μ
    d. $\bar{x}$
    e. N

39. The statement that "$P(A|B)=P(B|A)$ whenever A and B are independent events" is: Please select the best answer of those provided below.
    a. Always True

b. Never True
c. Not Enough Information; we would need to know if A and B are disjoint events
d. Not Enough Information; we would need to know if the events are equally likely

40. If two events (both with probability greater than 0) are mutually exclusive, then:
a. They also must be complements.
b. They also could be complements.
c. They cannot be complements.

41. In a class, there are 15 boys and 10 girls. Three students are selected at random. The probability that 1 girl and 2 boys are selected, is:
a. 21/46
b. 25/117
c. 1/50
d. 3/25

42. From a pack of 52 cards, two cards are drawn together at random. What is the probability of both the cards being kings?
a. 1/15
b. 25/57
c. 35/256
d. 1/221

43. A card is drawn from a pack of 52 cards. The probability of getting a queen of club or a king of heart is:
a. 1/13
b. 2/13
c. 1/26
d. 1/52

44. A bag contains 4 white, 5 red and 6 blue balls. Three balls are drawn at random from the bag. The probability that all of them are red, is:
a. 1/22
b. 3/22
c. 2/91
d. 2/77

45. A numerical measure of linear association between two variables is the
a. variance
b. coefficient of variation
c. correlation coefficient
d. standard deviation
e. None of the above answers is correct.

46. The coefficient of correlation
a. is the same as the coefficient of determination
b. can be larger than 1
c. can not be larger than 1
d. can not be negative
e. None of the above answers is correct.

This scenario applies to Questions 45 and 46:
A randomized experiment was done by randomly assigning each participant either to walk for half an hour three times a week or to sit quietly reading a book for half an hour three times a week. At the end of a year the change in participants' blood pressure over the year was measured, and the change was compared for the two groups.

47. This is a randomized experiment rather than an observational study because:
    a. Blood pressure was measured at the beginning and end of the study.
    b. The two groups were compared at the end of the study.
    c. The participants were randomly assigned to either walk or read, rather than choosing their own activity.
    d. A random sample of participants was used.

48. The two treatments in this study were:
    a. Walking for half an hour three times a week and reading a book for half an hour three times a week.
    b. Having blood pressure measured at the beginning of the study and having blood pressure measured at the end of the study.
    c. Walking or reading a book for half an hour three times a week and having blood pressure measured.
    d. Walking or reading a book for half an hour three times a week and doing nothing.

49. In regression, the equation that describes how the response variable (y) is related to the explanatory variable (x) is:
    a. the correlation model
    b. the regression model
    c. used to compute the correlation coefficient
    d. None of these alternatives is correct.

50. Regression modeling is a statistical framework for developing a mathematical equation that describes how
    a. one explanatory and one or more response variables are related
    b. several explanatory and several response variables response are related
    c. one response and one or more explanatory variables are related
    d. All of these are correct.

51. In regression analysis, the variable that is being predicted is the
    a. response, or dependent, variable
    b. independent variable
    c. intervening variable
    d. is usually x

52. If the correlation coefficient is a positive value, then the slope of the regression line
    a. must also be positive
    b. can be either negative or positive
    c. can be zero
    d. can not be zero

53. If the coefficient of determination is 0.81, the correlation coefficient
    a. is 0.6561
    b. could be either + 0.9 or - 0.9
    c. must be positive

d. must be negative

54. A fitted least squares regression line
   a. may be used to predict a value of y if the corresponding x value is given
   b. is evidence for a cause-effect relationship between x and y
   c. can only be computed if a strong linear relationship exists between x and y
   d. None of these alternatives is correct.

55. The strength (degree) of the correlation between a set of independent variables X and a dependent variable Y is measured by
   a. Coefficient of Correlation
   b. Coefficient of Determination
   c. Standard error of estimate
   d. All of the above

56. The percent of total variation of the dependent variable Y explained by the set of independent variables X is measured by
   a. Coefficient of Correlation
   b. Coefficient of Skewness
   c. Coefficient of Determination
   d. Standard Error or Estimate

57. A coefficient of correlation is computed to be -0.95 means that
   a. The relationship between two variables is weak
   b. The relationship between two variables is strong and positive
   c. The relationship between two variables is strong and but negative
   d. Correlation coefficient cannot have this value

58. Which of the following statements about the method of exponential smoothing is *not* true?
   a. it gives greater weight to more recent data
   b. it can be used for forecasting
   c. it uses all earlier observations in each smoothing calculation
   d. it gives greater weight to the earlier observations in the series

59. The following table contains the number of complaints received in a department store for the first 6 months of last year.

Month	Complaints
Jan	36
Feb	45
Mar	81
Apr	90
May	108
Jun	144

If a 3-term moving average is used to smooth this series, what would be the second calculated term?
   a. 36
   b. 40.5
   c. 54
   d. 72

60. A weighted aggregate price index where the weight for each item is its current-period quantity is called the
a. Aggregate index
b. Consumer Price Index
c. Laspeyres Index
d. Paasche Index

SECTION B : Choose Any One Question *(Either Question One or Question Two)*

QUESTION ONE [2+8+3+4+3]

a) The following table gives the frequency distribution of the number of days to expiry date for all containers of yogurt in stock at a local grocery store. Containers that had already expired but were still on the shelves were given a value of 0 for the number of days to expiry date.

Number of Days	Number of Containers
0 to 5	32
6 to 11	67
12 to 17	44
18 to 23	20
24 to 29	11

i. Prepare a cumulative frequency distribution.
ii. Calculate the cumulative relative frequency and cumulative percentage for each class.

b) The following table gives the distribution of the amounts of rainfall (in inches) for July 2012 for 50 cities.

Rainfall	Number of Cities
0 to less than 2	6
2 to less than 4	10
4 to less than 6	20
6 to less than 8	7
8 to less than 10	4
10 to less than 12	3

Find the
i. Mean,
ii. Variance
iii. Standard deviation

QUESTION TWO **[1+2+2+2+2+3+8]**

a) Let A and B be events in a sample space S, with P(A) = 0.40, P(B) = 0.25 and $P(A \cap B) = 0.10$. Find each of the following.

i. $P(\overline{A})$
ii. $P(A \cup B)$
iii. $P(A \cap \overline{B})$
iv. $P(\overline{A} \cap B)$
v. P(B|A)
vi. Are A and B independent?

b) Select the appropriate word from below that correctly completes each of the following statements

i. A list of the entire population from which a sample is to be drawn is called the _____
ii. A sample that selects every nth item from a list is called a _____ sample.
iii. _____ sampling attempts to adequately represent differing groups and population.
iv. _____ sampling is usefully employed when the population is scattered over a wide geographical area.

stratified, sampling frame, cluster, systematic

END OF EXAMINATION

UNIVERSITY OF SWAZILAND

FINAL EXAMINATION ANSWER SHEET

STUDENT ID.....

PROGRAM.....YEAR.....

INSTRUCTIONS

1. Cross out all your correct answers. For example 1. (A) (B) (C) (D) (E)

2. In the event that more than one answer is crossed out that is deemed automatically wrong.

1. (A) (B) (C) (D) (E)	17. (A) (B) (C) (D) (E)
2. (A) (B) (C) (D) (E)	18. (A) (B) (C) (D) (E)
3. (A) (B) (C) (D) (E)	19. (A) (B) (C) (D) (E)
4. (A) (B) (C) (D) (E)	20. (A) (B) (C) (D) (E)
5. (A) (B) (C) (D) (E)	21. (A) (B) (C) (D) (E)
6. (A) (B) (C) (D) (E)	22. (A) (B) (C) (D) (E)
7. (A) (B) (C) (D) (E)	23. (A) (B) (C) (D) (E)
8. (A) (B) (C) (D) (E)	24. (A) (B) (C) (D) (E)
9. (A) (B) (C) (D) (E)	25. (A) (B) (C) (D) (E)
10. (A) (B) (C) (D) (E)	26. (A) (B) (C) (D) (E)
11. (A) (B) (C) (D) (E)	27. (A) (B) (C) (D) (E)
12. (A) (B) (C) (D) (E)	28. (A) (B) (C) (D) (E)
13. (A) (B) (C) (D) (E)	29. (A) (B) (C) (D) (E)
14. (A) (B) (C) (D) (E)	30. (A) (B) (C) (D) (E)
15. (A) (B) (C) (D) (E)	31. (A) (B) (C) (D) (E)
16. (A) (B) (C) (D) (E)	32. (A) (B) (C) (D) (E)

33.	A	B	C	D	E	47.	A	B	C	D	E
34.	A	B	C	D	E	48.	A	B	C	D	E
35.	A	B	C	D	E	49.	A	B	C	D	E
36.	A	B	C	D	E	50.	A	B	C	D	E
37.	A	B	C	D	E	51.	A	B	C	D	E
38.	A	B	C	D	E	52.	A	B	C	D	E
39.	A	B	C	D	E	53.	A	B	C	D	E
40.	A	B	C	D	E	54.	A	B	C	D	E
41.	A	B	C	D	E	55.	A	B	C	D	E
42.	A	B	C	D	E	56.	A	B	C	D	E
43.	A	B	C	D	E	57.	A	B	C	D	E
44.	A	B	C	D	E	58.	A	B	C	D	E
45.	A	B	C	D	E	59.	A	B	C	D	E
46.	A	B	C	D	E	60.	A	B	C	D	E