Unity University
Computer Science Department
Introduction to Statistics (STAT 2091) worksheet-for Ch1-Ch4

1. For each of the hypothetical research situations described, identify the scale of measurement and indicate whether the data are qualitative, quantitative discrete or quantitative continuous.
   a) Number of complaint letters received by a firm's administrator
   b) You measure the body lengths (in inches) of 10 full term infants as birth and record the following:
      17.5, 19.5, 17.5, 19, 20, 21, 18, 18, 19.5, 10.75
   c) Number of tomato's on each plant in a field
   d) Number of students taking ESLCE each year
   e) Level of customers satisfaction on service given by CBE
   f) Number of cartons of corn oil manufactured each day
   g) Weights of newborn children at Black Lion hospital
   h) Geologists have a "hardness scale" for identifying different rocks, called Mohs'scale. The hardest rock (diamond) has a value of 10 and will scratch all others. The second hardest will scratch all but diamond, and so on, down to a rock such as talc, with a value of one, which can be scratched by any other rock
   i) The volumes of three different cubes are 40, 64, and65 cubic inches.
   j) Republicans, Democrats, Independents, and others are identified on the voters' list with numbers 1, 2, 3, &4
   k) The pages of a book are numbered from 1 through 150
   l) The winner of the Miss America was Miss California; runners-up were Miss Ohio and Miss Pennsylvania
   m) The prices on three items were $3.00, $10.00, and$12.00.
   n) She earned three degrees: B.SC., M.SC., and PhD.

1. In order to assure the quality of wine, the quality control departments of wineries determine the acidity, the correctness of taste, and the aroma. This task in accomplished by opening the seal, determining the quality of the wine, and then discarding the bottle. Which types of survey (i.e. sample or census survey) you have to recommend the department quality control to use? Why?

2. Identify the type of data represented by each of the following variables and confirm your answer by giving your own explanation. (Nominal, ordinal, interval or ratio).
   i) Blood group, temperature, number of heart attacks,
   ii) Calendar year, serum uric acid level, diagnosis of disease, IQ,
   iii) Number of cases of each reportable disease reported by health officer e.g. polio, stage of cancer,
   iv) The weight of one year old dog (with special diet supplement) was 950gm last month, number of accidents in three year period.

3. What is meant by central tendency? Briefly describe the methods of measurement central tendency. Point out the merits and demerits of each method.

4. Arithmetic mean of 50 observations was 100. At the time of calculations two observations 180 and 90 were omitted. Find the corrected mean?

5. A certain treatment is used in two different centers, A and B; patients in center A were 25 and were on average 54 years old; patients treated in center B were 62 and had mean age equal to 58 years. What is the overall mean among all patients who got the treatment?

6. For a distribution, the first four moments about mean are 0,7,38 and 155 respectively. (i) Compute the moment coefficients of skewness and kurtosis and interpret.

7. The mean weight of 50 students in a certain class is 60 kg. The mean weight of boys is 70 kg and that of girls is 55 kg. Find the number of boys and girls in the class.

8. Trust in Internet Information a survey was taken on how much trust people place in the information they read on the Internet. Construct a categorical frequency distribution for the data. A = trust in everything they read, M = trust in most of what they read, H = trust in about one-half of what they read, S = trust in a small portion of what they read. (Based on information from the UCLA Internet Report.)

M M M A H M S M H M S M M M M A M M A M
M M H M M M H M H M A M M M H M M M M M

a) Prepare a frequency distribution.

b) Draw the appropriate chart.

9. Let k denote the standard deviation for a sample of 25 numbers.

(a) If each number is multiplied by 100, what is the standard deviation of the new set of numbers, as a function of k?

(b) Is the coefficient of variation affected by this multiplication? Explain.

10. Compare the arithmetic mean, median, mode, geometric mean as to the manner in which they are affected by extreme values.

11. Marks of 75 students are summarized in the following frequency distribution.

Marks	**Number of students**
40 – 44	7
45 – 49	10
50 – 54	22
55 - 59	f_4
60 – 64	f_5
65 – 69	6
70 - 74	3

If 20% of the students have marks 55 – 59, find the missing frequencies f_4 & f_5.

12. Based the following frequency distribution

Class limit	frequency
11-20	12
21-30	30
31-40	f_3
41-50	65

51-60	f_5
61-70	25
71-80	18
Total	**229**

If the median value is 46, find

i) The missing frequencies

ii) Calculate the mean, variance, S.D,3rd quartile, second quartile, 5th deciles & 75th percentiles and interpret them

iii) Comment on the value of median and second quartiles

13. Write the merits and demerits of arithmetic mean, median, mode, and quartiles.

14. The average salary of 20 individuals working in a small scale industry was Birr 100. But five qualified persons were employed and then increased the average salary into Birr 200. Find the mean salary of the newly employed persons.

15. The average salary of male employees in a firm was 520 birr and that of female was 420 birr. The mean salary of all the employees was 500 birr. Find the percent of male and female employees in the firm.

16. If the arithmetic mean and the geometric mean of two observations are 13 and 12 respectively, find the values of the observations.

17. In a class there are 30 females and 70 males. If females averaged 60 in an examination and boys averaged 72, find the mean for the entire class.

18. The mean of a set of numbers is 500.

a) If 10 is added to each of the numbers in the set, then what will be the mean of the new set?

b) If each of the numbers in the set are multiplied by -5, then what will be the mean of the new set?

19. In a small business firm, two typists are employed-typist A and typist B. Typist A types out, on an average, 30 pages per day with a standard deviation of 6. Typist B, on an average, types out 45 pages with a standard deviation of 10. Which typist shows greater consistency in her output?

20. Mean and standard deviation of 100 items are 50 and 4, respectively. Find the sum of all the item and the sum of the squares of the items.

21. While calculating the mean and variance of 10 readings, a student wrongly used the reading 52 for the correct reading 25. He obtained the mean and variance as 45 and 16 respectively. Find the correct mean and the variance.

22. Some characteristics of annually family income distribution (in Birr) in two regions is as follows:

Region	**Mean**	**Median**	**Standard Deviation**
A	6250	5100	960
B	6980	5500	940

i. Calculate coefficient of skewness for each region
ii. For which region is, the income distribution more skewed. Give your interpretation for this Region
iii. For which region is the income more consistent?

23. The median and the mode of a mesokurtic distribution are 32 and 34 respectively. The 4th moment about the mean is 243. Compute the Peasonian coefficient of skewness and identify the type of skewness. Assume (n-1 = n).

24. Oromia Broadcasting Network (OBN) wants to know the proportion of TV owners in Adama City who watch the broadcasting new program at least once a week. The station asked a group of 600 TV owners in Adama City if they watch the program at least once a week.

a) Identify the individuals in the study.
b) Identify the variable.
c) Do the data comprise a sample? If so, what is the underlying population?
d) Is the variable quantitative or qualitative?

25. The following data shows the Adama City high temperatures in ^{0}c for 50 randomly selected days.

32	38	30	24	24	36	32	34	28	27
23	26	35	34	22	18	16	38	23	17
30	38	25	37	25	34	25	30	27	32
33	30	29	32	33	35	29	19	17	28
22	31	33	15	20	24	32	34	33	31

a) Construct a grouped frequency distribution which comprises CB, CM, L.C.F, M.C.F, R.F, and %R.F
b) Find the median, mode and mean
c) Find variance, standard deviation and coefficient of variation
d) Draw a histogram, frequency polygon and both ogives

26. Two groups of people were trained to perform a certain task and tested to find out which group is faster to learn the task. For the two group the following information was obtained:

Value	**Group one**	**Group two**
Mean	15.2	17.5
Variance	5	7

a) Relative speaking, which group is more consistent in its performance?
b) Suppose a person **X** from group one took 10.5 minutes while person **Y** from group two too k 11.2 minutes, who was faster in performing the task? Why?

27. The following table shows weekly payment given to five workers in two different firms on part time bases.

Workers	**X**	**Y**	**Z**	**U**	**V**
Firm A	47	49	50	51	53
Firm B	45	51	57	49	53

a) In which firm Mr. U's payment relatively better?
b) In which firm Mr. Z's payment relatively better?