

Part I: Choose the best answer and put the letter of your choice on the space provided. (1 point each)

- _____ 1) A random sample of 300 households in Ambo town was selected and several variables were recorded for each household. Which of the following is not correct?
 - ☒ A) Household total income is a continuous data
 - ☐ B) The primary language used at home is a nominal data
 - ☐ C) Socioeconomic status was coded as 1 =low income, 2=middle income, 3=high income and is an interval data
 - ☐ D) The number of persons in the household is a ratio data.
 - ☐ E) None of the above
- _____ 2) All of the following are the main functions of statistics except one.
 - ☐ A) It presents facts in a definite and precise form.
 - ☐ B) Furnishes a technique of comparison
 - ☐ C) Testing and formulating of hypothesis
 - ☐ D) Study the relationship between two or more variable
 - ☒ E) None
- _____ 3) Which one of the following is not true about the nominal scale of measurement?
 - ☒ A) There is inherent order associated with individual categories
 - ☐ B) Difference is not possible
 - ☐ C) No arithmetic operation is possible
 - ☐ D) None of the above
- _____ 4) _____ can assume any value within specified relevant interval of values.
 - ☒ A) Continuous variable
 - ☐ B) Discrete variable
 - ☐ C) Qualitative variable
 - ☐ D) None of the above
- _____ 5) One of the following is method of obtaining primary data:
 - ☐ A) Direct personal interviews
 - ☐ B) Written questionnaires
 - ☐ C) Computerized data base
 - ☐ D) Published data in journals
 - ☐ E) C and D
 - ☒ F) A and B
- _____ 6) Which of the following is not true?
 - ☒ A) Parameter is any characteristic of a population
 - ☐ B) Statistic is characteristic of a sample
 - ☐ C) Statistic is used to estimate the value a parameter
 - ☐ D) The value of statistic changes from one sample to the other sample
 - ☐ E) None of the above
- _____ 7) Which one of the following is the merit of arithmetic mean?
 - ☐ A) It is affected by extreme values.
 - ☐ B) It can be obtained for open end classes.
 - ☒ C) It cannot be used for qualitative characteristics such as intelligence, honesty, etc
 - ☐ D) It is not suitable for further mathematical treatment.
 - ☐ E) None of the above.
- _____ 8) One of the following does not belong to ratio scale of measurement?
 - ☐ A) Intelligence quotient
 - ☐ B) Weight
 - ☐ C) Time it takes to run 42 kms
 - ☒ D) Number of farmers in West Showa
 - ☐ E) None of the above.

own was selected and several
 following is not correct? (1 point)
 come, 3=high income

$$\frac{5}{78} \times 100 = 6.41\%$$

$$\frac{2}{78} \times 100 = 2.57\%$$

$$\frac{53}{78} \times 100 = 67.95\%$$

Answer questions 9 through 11 based on the following information

Score	10 - 20	20 - 30	30 - 40	40 - 50	50 - 60	60 - 70	70 - 80
Number of students	5	7	10	30	14	3	1

- 9) The number of students who scored less than 40 is
 A) 22 B) 70 C) 5 D) 12 E) None of the above
- 10) The percentage of students who scored less the 50 is
 A) 66% B) 25.7% C) 74.3% D) 94.3% E) None of the above
- 11) The percentage of students who scored above 30 is
 A) 58% B) 82.9% C) 17.1% D) 68.6% E) None of the above

Part II: Short Answer questions.

Write correct answer on the space provided.

- If the mean of the variable x is 25, what will be the mean of $y = 2x - 6$?
 (1 point)
- Frequency distribution of height in cms of students in a DMU is given in the table below
 (2 points)

Height in cms	158	152	154	156	150	160	164	162	164	166
No. of students	60	40	52	100	28	48	20	32	20	7

- The median is _____
 - The mode is _____
3. Define the following terms (2 points)

i) Population
 Population is a complete observation, or measurement of individuals. Its synonyms are "Universal set"

ii) Statistic

Statistic: is the measure of value of statistic data taken from a sample.

$$\bar{x} = \frac{\sum x}{n} = \frac{200}{3} = 66.67$$

$$s = \sqrt{\frac{\sum x^2}{n} - \bar{x}^2} = \sqrt{\frac{13000}{3} - 66.67^2} = 10.54$$

-3-

-3-

Page 2 of 5

4. Write at least four limitations of the statistics (2 point)

- i. Statistics doesn't deal with individuals
- ii. It doesn't deal with qualitative characteristics.
- iii. The conclusion obtained in statistics methods are not universally true.
- iv. Statistics can be misused. It can't be used to full advantage.

5. State and explain briefly the two broad categories of statistics (3 points)

- i. Descriptive Statistics:- deals with describing the data collected without going to further conclusion.
- ii. Inferential Statistics:- This part of statistics is connected with drawing conclusion.

Sheets for qu
The no. of students
we have 10-20
20

III: Work out problem. Solve the question given below showing all the necessary steps clearly and neatly.

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

Marks	No. of students	Class boundary	Class mark	Less than cumulative frequency
40-44	f_1 15	39.5 - 44.5	42	15
45-49	10	44.5 - 49.5	47	15 + 10 = 25
50-54	22	49.5 - 54.5	52	25 + 22 = 47
55-59	7	54.5 - 59.5	57	47 + 7 = 54
60-64	f_5 12	59.5 - 64.5	62	54 + 12 = 66
65-69	6	64.5 - 69.5	67	66 + 6 = 72
70-74	3	69.5 - 74.5	72	72 + 3 = 75

- i. Based on the given information, complete the above table. [6.5 points]
 ii. If 20% of the students have marks between 40 and 44 inclusive, then find the missing frequencies f_1 and f_5 . [3 points]

$$\text{We have } \frac{f_1}{75} \times 100 = 20$$

$$\frac{100f_1}{75} = 20$$

$$100f_1 = 1500$$

$$f_1 = 15$$

$$f_5 = f_1 + 10 + 22 + 7 + f_5 + 6 + 3 = 75$$

$$63 + f_5 = 75$$

$$f_5 = 12$$

- iii. Compute mean, median and mode by showing all the necessary steps. [4.5 points]

$$Md = LCB + \left(\frac{\frac{N}{2} - cf}{f} \right) CW$$

$$= 50 + \left(\frac{37.5 - 25}{22} \right) 4$$

$$= 50 + \frac{50}{22}$$

$$-7- = 50 + 2.27$$

$$Md = 52.27$$

$$\frac{N}{2} = \frac{75}{2} = 37.5$$

The median class is

$$50 - 54$$

$$M_o = \frac{\Delta_1}{\Delta_1 + \Delta_2} \times CW$$

$$= 50 \times \left(\frac{22-10}{(22-10) + (22-7)} \right) \times 4$$

$$50 \times \left(\frac{12}{27} \right) \times 4$$

$$50 \times 1.777$$

$$M_o = \underline{\underline{51.777}}$$

$$\Delta_1 = f_3 - f_2$$

$$\Delta_2 = f_3 - f_4$$

2. The average salary of male employees in a firm was 630 birr and that of female was 530 birr. The mean salary of all the employees was 580 birr. Find the percent of male and female employees in the firm. [5 points]

for male

$$= \frac{630}{580} \times 100$$

$$= \frac{63000}{580}$$

$$\underline{\underline{108.621\%}}$$

for female

$$\frac{530}{580} \times 100$$

$$\frac{53000}{580}$$

$$= \underline{\underline{91.379\%}}$$

Adama Science and Technology University

Probability and Statistics for Engineers Mid Exam.

Name: _____ ID.NO: _____ Group: _____

December 28, 2013

Time: 2:00hrs

Max. Marks: 40%

Set by: Mr. Fekadu Lemessa, Dr. Kamlesh Kumar Shukla and Mr. Dame. K

Only For Instructor's Use

Exam Part	Part I (20%)	Part II (20%)	Total (40%)
Result			

General Instruction:

Attempt all the questions. Your answers should be readable and neat. Write your answers on the space provided. On the workout part, use the back page if the given space is not enough. Check that the examination consists of two parts and six pages including the cover page. Switch-off your mobile before you start the examination.

-9-

Part I: Choose the Best Answer (2pts each).

1. _____ The mean wage of 100 workers working in a factory running two shifts is birr 60. If there are 60 workers in the morning shift with average wage of 62 birr, what must be the mean wage of evening shift?
A. 60 Birr ☒ B. 52 Birr C. 55 Birr D. 62 Birr E. None of the above
2. _____ Fill in the missing words to the quote "Statistical methods may be described as methods for drawing conclusions about _____ based on _____ computed from the _____."
A. statistics, samples, populations ☒ B. populations, parameters, samples
C. statistics, parameters, samples D. parameters, statistics, populations
E. populations, statistics, samples
3. _____ The class width, for distributions having constant class width, is the difference between
A. the upper and lower class boundary of a class
B. the lower limit of any two consecutive classes
☒ C. the lower boundaries of any two consecutive classes
D. any two consecutive class marks
E. All of the above
4. _____ The arithmetic mean of 20 observations is found to be 15. What is the arithmetic mean of the new data set obtained by the linear transformation $Y_i = 2X_i - 5$? A. 30 B. 20 ☒ C. 25 D. 35
5. _____ A number calculated with complete population data and quantifies a characteristic of the population is called
A. datum. B. parameter. ☒ C. statistic. D. population.
6. _____ Graphs are better than tables in the sense that
A. they reveal detailed patterns more quickly.
B. they contain information in columns as well as rows.
C. they are better at conveying information.
D. they are not easily understandable.
☒ E. A and C.

7. _____ Which of the following is true about the data set in Table 1.

Table 1: Frequency Distribution

x_i	6	2	4	3	5
f_i	2	1	2	1	2

- A. Mode doesn't exist. ☒ B. Median is one of the elements of this data set. ✓
C. Mean is less than median. D. Median cannot be obtained. E. None
8. _____ One of the following data presentations is different from the other.
A. Bar Chart ☒ B. Ogives C. Pie Chart D. A and B.
9. _____ A study was conducted to investigate the effect of a detergent plant upon the water quality of a river. As part of an environmental impact study, some fish were captured, tagged, and released. The following information was recorded for each fish:

Sex (0 = female, 1 = male), Length (cm),
Maturation (0 = young, 1 = adult), Weight (gram).

Which of the following is TRUE?

- A. The variables are nominal, ratio, nominal, and ratio respectively.
B. Sex is a continuous variable.
C. Length and Weight are discrete variable with ratio scale of measurement.
D. Maturation is ordinal scale.
E. All are wrong
10. _____ Which of the following measures of central tendency is not rigidly defined.
A. Mode ☒ B. Median C. Mean D. All E. None

Part II: Short Answer and Workout Part (20 Pts).

1. Identify the following as qualitative, discrete or continuous variable.

- discrete Number of Geomatic engineers on a project site (1 Pt).
- qualitative Land use land cover category (1 Pt).
- Continuous Elevation above sea level (1 Pt).

2. The frequency distribution of the net profits in million birr of 50 companies is shown in Table 2.

Table 2: Frequency Distribution

Class Boundary	<20	20-30	30-40	40-50	50-60	>60
Frequency	3	8	f_3	9	f_5	5

- If 30% of data are in the class 30-40, obtain the missing frequencies f_3 and f_5 (3pts)

$$\frac{f_3}{25 + f_3 + f_5} = 30\%$$

$$\frac{f_3}{25 + f_3 + f_5} \times 100 = 30 = \frac{f_3}{50} \times 100 = 30$$

~~$$100f_3 = 30(25 + f_3 + f_5)$$~~

$$100f_3 = \frac{30 \times 50}{100}$$

$$f_3 = 15$$

$$f_5 = 40 + f_5 = 50$$

$$f_5 = 50 - 40$$

$$f_5 = 10$$

$$-12-$$

5) = 25.5
e Median

$$= \frac{51}{2} = 25.5$$

the median class (3pts): 20-30

$$Md: ELCB \left(-\frac{N}{2} + CF \right) \frac{w}{f}$$

$$= 20 + \frac{(25 - 3) \cdot 10}{8}$$

$$= 20 + \frac{220}{8}$$

$$Md: \underline{\underline{47.5}}$$

$$Mo: \left(\frac{\Delta_1}{\Delta_1 + \Delta_2} \right)$$

$$= 20 + \left(\frac{8 - 3}{(8 - 3) + (8 - 15)} \right) 10$$

$$20 + \left(\frac{5}{-2} \right) 10$$

$$20 + \left(-\frac{50}{2} \right)$$

$$Mo: \underline{\underline{-5}}$$

c. What percentage of companies get at least a net profit of 40 million?
(3pts)

at least means greater than 40

$$So, \frac{34}{50} \times 100 = \frac{3400}{50} = \underline{\underline{68\%}}$$

3. A GIS expert would like to detect land use land cover change in Oromiya Regional State. Accordingly, the expert took samples of land use land cover at some districts in the Regional State. He coded the outcome of his analysis as follows: Residential Area as 1, Agricultural Area as 2, Forest Area as 3, Grass land Area as 4 and Bare land Area as 5. Therefore, the expert obtained the following data: 1, 1, 2, 3, 4, 3, 2, 2, 1, 2, 1, 2, 3, 1, 2, 5, 2, 1, 3, 3, 2, 2, 2, 1, 2, 2, 2, 1, 4, 5, 1, 2, 5

- a). The variable of interest _____ (1 Pt).
- b). _____ Is the variable ordinal or Nominal (1 Pt)?
- c). _____ What is the size of the sample (1 Pt)?
- d). Summarize the data using a measure of central tendency _____ (2Pts).
- e). Present the data using Pie-chart (3Pts).

1. Define the following terms in your words.

a) Statistics: _____

b) Sampling: _____

c) Population: _____

2. Classify the following as belonging to descriptive statistics [A] or inferential statistics [B]

a) According to insurance company figures, the chance of your living to age 63 is 62.8%. _____

b) 5% of all the deaths in St. Paul's hospital in 2011 were due to traffic accidents. _____

3. Fill the following words in the blank spaces: populations, statistics, and samples. "Statistical methods may be described as methods for drawing conclusions about _____ based on _____ computed from the _____."

4. A medical researcher is interested to estimate the mean serum cholesterol level of healthy men living in Addis Ababa. For a sample of 250 healthy men, he measured their serum cholesterol and the mean value is 220mg/ml with standard deviation 40mg/ml. Based on this information, what is the

a) Variable: is a certain characteristic which is chosen object to be

b) Population: a complete observation

c) Sample: a part or subset of population under study

d) Parameter: the statistical value obtained from a population

5. A study was conducted to measure the effect of smoking upon the birth weight of a baby. The following variables were measured for each woman in the study: Smoking status (yes, no), baby birth weight (g), rate of smoking (A, B, C, D), birth date. Determine the scales of measurements of these variables?

a) Smoking status: _____

b) Baby birth weight: _____

c) Rate of smoking: _____

d) Birth date: _____

6. A number calculated with complete population data and quantifies a characteristic of the population is called _____
7. Consider the following data: 1, 7, 3, 3, 6, 4 The mean and median for this data are:
 a) Mean: 4
 b) Median: 3
8. Let a and b be constants and let $y_i = ax_i + b$ for $i = 1, 2, \dots, n$. What are the relationships
 a) between $\bar{x}$ and $\bar{y}$: _____
 b) between s_x^2 and s_y^2 : _____
9. State at least two important characteristics of a good average.

Part II: Attempt the following questions by showing the necessary steps.

- 1) The frequency distribution for 100 observations is given below.
- Complete the class mark and less than cumulative frequency columns. [2 pts]
 - Determine the mean, mode and median? [6 pts]
 - What can you say about the distribution of the series? [1 pt]

Class Boundaries	Freq. (f)	Class Mark (X)	LSCF
59.0 - 60.5	10		
60.5 - 62.0	14		
62.0 - 63.5	21		
63.5 - 65.0	26		
65.0 - 66.5	15		
66.5 - 68.0	10		
68.0 - 69.5	4		

- 2) For a distribution the following sample results were obtained:

$$\sum_{i=1}^{20} x = 1000 \quad \sum_{i=1}^{20} x^2 = 50,320$$

Find the coefficient of variation and interpret the result.

- 3) The mean and variance of wage of employees in two hospitals, A and B, are given below: [3 pts]

	Hospital A	Hospital B
Mean Wage	120	85
Variance	25	9

- In which hospital is the distribution of wages more uniform?
- Suppose the wages of two employees in hospital A and B respectively are 130 and 95. Whose wage is better relative to the respective hospital?

[4 pts]

Page 2 of 2

-16-

Statistics - Assignment & mid-exam

Adama Science and Technology University

Probability and Statistics - Exercise

- 1) Classify the following sentences as belonging to the area of descriptive or inferential statistics.
 - a. As a result of recent cutbacks by oil producing nations, we can expect the price of gasoline to double in the next year.
 - b. At least 5% of all killings reported last year in city X were due to tourists.
 - c. Of all patients who received this particular type of drug at a clinic, 75% later developed significant side effect.
 - d. Of all U.S. kindergarten teachers, 32% say that knowing the alphabet is an essential skill.
 - e. Of the 800 U.S. kindergarten teachers polled, 32% say that knowing the alphabet is an essential skill.
- 2) For a hypothetical research situations below, identify the population and the sample.
 - a. A private university desires to estimate the percentage of its students who have full time jobs. You interview 50 of its students and ask whether they hold full time jobs.
 - b. All persons in a Kebele are called by telephone and asked to name the candidates that they favor for election to the Kebele leading committee.
- 3) Of the U.S. adult population, 36% has an allergy. A sample of 1200 randomly selected adults resulted in 33.2% reporting an allergy.
 - a. Describe the population.
 - b. What is the sample?
 - c. Describe the variable.
 - d. Identify the statistic and give its value.
 - e. Identify the parameter and give its value.
- 4) A new department store is getting ready to open. The store owner wants to know two qualitative and two quantitative variables about its customers that he might find helpful. Identify these variables.
- 5) The admissions office wants to estimate the cost of textbooks at ASTU. Let the variable x be the total cost of all textbooks purchased by a student this semester. The plan is to randomly identify 100 students and obtain their total textbook costs. The average cost for the 100 students will be used to estimate the average cost for all students.
 - a. Describe the parameter
 - b. Describe the population.
 - c. Describe the variable involved.
 - d. Describe the sample.
 - e. Describe the statistic and how you would use the 100 data values collected to calculate the statistic.

c) In your own words, explain why the parameter is fixed and the statistic varies.

7) Classify each of the following first as qualitative or quantitative and second as nominal, ordinal, interval or ratio scale. If quantitative, indicate whether the data are discrete or continuous.

- Time for swimmers to complete a 50m race.
- Months of the year Meskerm, Tikimit...
- Number of complaint letters received by a firm's administrator.
- Socioeconomic status of a family when classified as low, middle and upper classes.
- Blood type of individuals, A, B, AB and O.
- Quintals of wheat produced in Ethiopia last year.
- Miles driven by city-bus drivers each day.
- For 16 persons arrested for driving while intoxicated you record whether they live in urban, suburban, or rural areas.
- You observe 20 overweight persons and classify them as high, medium or low physical activity levels.
- Horsepower of motorcycle engines
- The hair color of children
- TV Show ratings (poor, fair, good, excellent)
- Ages of students enrolled in a martial arts course
- Weights of newborn children at Black Lion hospital.
- The breaking strength of a given type of string
- The number of stop signs in towns

8) Which of the diagrams is most appropriate for the following data set? Draw the diagram.

- Number of Students enrolled to a certain department from year 1 to 3

Year	Male	Female
1	50	20
2	45	15
3	40	10

- Net profit (in million birr) earned by different companies in 2010

Company	Net Profit
A	20
B	-15
C	30

9) Discuss the following

- The importance of averaging
- The characteristics of a good average
- The merits and demerits of mean, mode and median

arithmetic mean and geometric mean of two observations are 6.5 and 6 respectively, find the values of the observations.

The following frequency distribution is given.

- Sketch the histogram of the data, the frequency polygon and the less than ogive curve.
- Determine the arithmetic mean, mode, median and standard deviation of the distribution.

CLASSES	FREQUENCIES
59.0 - 60.4	10
60.5 - 61.9	14
62.0 - 63.4	21
63.5 - 64.9	26
65.0 - 66.4	15
66.5 - 67.9	10
68.0 - 69.4	4

12) Find the mean, mode, median and standard deviation of the following series.

X:	10	11	12	13	14
F:	3	12	18	10	3

13) The weekly salaries of six employees at a company are \$140, \$220, \$90, \$180, \$140, \$200. For these six salaries, find: (a) the mean (b) the median (c) the mode

14) Kebede has grades of 84, 65, and 76 on three math tests. What grade must he obtain on the next test to have an average of exactly 80 for the four tests?

15) Discuss the following

- Explain the meaning of dispersion
- What purpose does a measure of dispersion serve?
- Why are measures of central tendency not sufficient to study a frequency distribution?
- Distinguish between absolute and relative measures of dispersion.

16) The first four deviations from the mean in a sample of $n = 5$ reaction times were .3, .9, 1.0, and 1.3. What is the fifth deviation from the mean? Give a sample for which these are the five deviations from the mean.

-19-

Gypsum
94p
41