CENTRAL UNIVERSITY

FAITH · INTEGRITY · EXCELLENCE

FEBRUARY COHORT EXAMINATION
2018 ACADEMIC YEAR

DEPARTMENT OF ECONOMICS

ECON 105: INTRODUCTION TO STATISTICS 1

TIME ALLOWED: 3 HOURS

STUDENT ID No. . . . .

INSTRUCTIONS:

ANSWER ALL QUESTIONS IN SECTION A
AND
ANSWER QUESTION '1' AND ANY OTHER TWO QUESTIONS IN SECTION B.
ALL QUESTIONS ARE TO BE ANSWERED IN THE ANSWER BOOKLET.

DO NOT TURN OVER THIS PAGE UNTIL YOU HAVE BEEN TOLD TO DO SO BY
THE INVIGILATOR.

Lecturer: Mr Ernest Somua–Wiafe

SECTION A

(Answer all questions in this section)

1. Which sampling method is less likely to give a representative sample?

a) convenience sampling

b) simple random sampling

c) stratified sampling

d) systematic sampling

2. Which of the following is/are an example of an interval scale of measurement?

I – Time of the day   II - Sales volume of MP3 players

III - Soft drink preference   IV – Temperature in degree Celsius

V - Rating of a finance professor

a) I   b) I & II   c) I, II & III   d) I & IV

3. The standard deviation of a data set measures the .... of the data set.

a) variables   b) spread   c) size   d) center

4. The number of classes in a frequency distribution depends on the size of the data set.

In general, the ....

a) smaller the data set, the smaller the number of classes

b) smaller the data set, the larger the number of classes

c) number of classes should be equal to the number of values in the data set divided by any constant

d) larger the data set, the smaller the number of classes
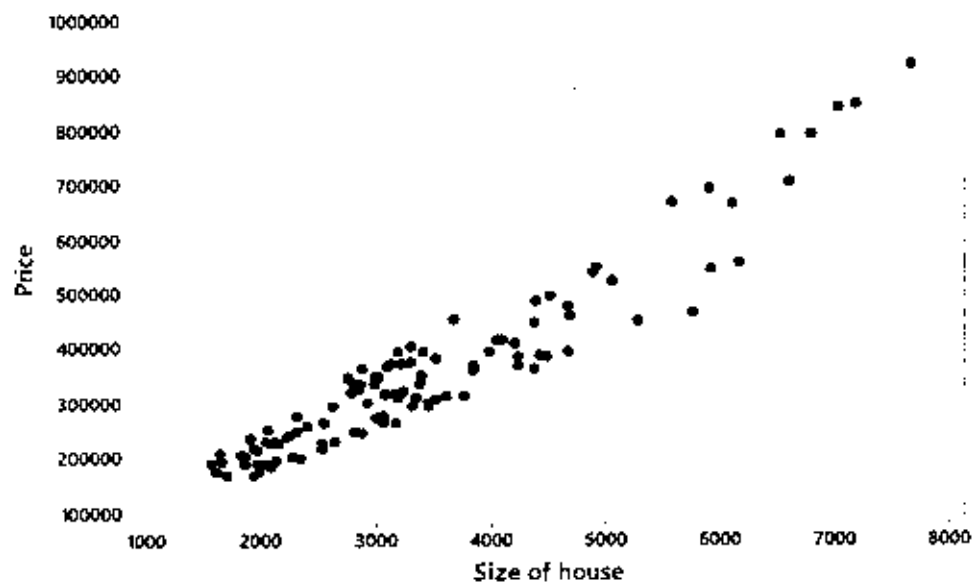
5. The distribution of the boxplot to the right is.............

a) skewed right

b) skewed left

c) symmetric

d) uniform
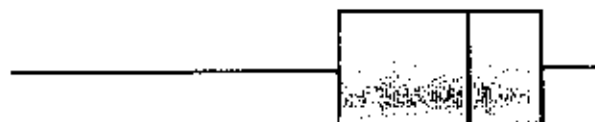
6. We would expect that the mean is.......

a) less than the median

b) more than the median

c) the same as the median

d) the same as the mode

Use the graph below to answer questions 7, 8 and 9

7. Which type of plot is displayed in the graph?

a) a dot plot of price versus size of house

b) a scatter plot of price versus price of house

c) a dot graph of size of house against price

d) a scatter graph of price against size of house

8. What is the relationship between the two variables displayed in the graph?

a) There exist strong positive relationship between price of house and the size of house

b) There exist weak positive relationship between price of house and the size of house

c) There exist strong negative relationship between price of house and the size of house

d) There exist weak negative relationship between price of house and the size of house

3

9. From the graph, we can clearly notice that ......

a) as the size of the house increases, the price of the house decreases

b) as the size of the house increases, the price of the house remains the same

c) as the size of the house decreases, the price of the house decreases

d) as the size of the house remains constant, the price of the house also remains constant

10. Assuming you have a data set given in two versions, one is the original ungrouped data, and the other is a set where the data values have been grouped. If you calculate the mean and standard deviation for the two versions of the data sets, you would expect to get

a) exactly the same mean and variation no matter which set you used for your calculations

b) exactly the same mean but a slightly different variation

c) slightly different mean but exactly the same variation

d) slightly different mean and a slightly different variation

11. In case where data is quantitative discrete data with a large sample size yet few distinct values, which type of data grouping methods would be appropriate to employ .....

a) single value grouping

b) decimal place grouping

c) cut point grouping

d) class limit grouping

12. A statistics lecturer wants to know how his students feel about an introduction to statistics course. He decides to administer a survey to a random sample of students who have taken the course. If from the various levels (level 100, level 200, level 300 and level 400) he surveyed some randomly selected students from each of the levels, what sampling method is he applying?

a) simple random sampling

b) stratified random sampling

c) cluster sampling

d) systematic sampling

Use the following information to answer questions 13, 14 and 15

A marketing company is interested in the percentage of people in Ghana who will buy a particular product. The company then randomly selects 10,000 people from the 10 regions in Ghana and determines the number of people from the sample who bought the product after advertisement.

13. From the information given, what is the type of the variable of interest?

a) Qualitative variable

b) Quantitative discrete variable

c) Quantitative continuous variable

d) None of the above

14. The group of people selected from the various 10 regions in Ghana is be termed as …

a) the population

b) an observation

c) simple random sampling

d) sample

15. The proportion of people from the selected 10,000 who intend to buy the product is referred as …

a) a parameter

b) a statistic

c) a variable

d) sample data

16. The prices of all Central University textbooks follow a bell-shaped distribution with a mean of $¢180$ and a standard deviation of $¢30$. According to the empirical rule, the percentage of books will fall within the price of $¢120$ and $¢240$?

a) 68%   b) 75%   c) 95%   d) 99.7%

17. The mean monthly electricity bills paid by all home owners in a town is GH$¢2,365$ with a standard deviation of GH$¢$ 340. Using Chebyshev's theorem, at least 75% of all home owners in this town pay a monthly electricity bills of about ….

a) $¢2,025$ to $¢2,705$ b) $¢1,685$ to $¢3,045$ c) $¢1,345$ to $¢3,385$ d) $¢1,005$ to $¢3,725$

18. All the following describes an outlier except ...

a) An outlier may be a data value that has been incorrectly recorded.

b) An outlier may be a data value that has its value similar to other data points.

c) An outlier may be from an observation that was incorrectly included in the data set.

d) An outlier may be an unusual data value that has been recorded correctly and belongs in the data set.

19. Any measure that is not likely to be influenced by few extreme values is termed as ...

a) a constant term

b) mean deviation

c) a trimmed mean

d) a resistance measure

20. If the event of interest is A, then which of the following statements is not correct?

a) the probability that A will not occur is $1 - P(A)$

b) the probability that A will occur is zero if event A is impossible

c) the probability A will occur is one if event A is certain

d) none of the above statements

21. Which of the following definitions under probability is correct?

a) an activity, which results in an outcome, is called an event

b) the probability of an event A is expressed in decimal form ranging from $0 < P(A) < 1$.

c) the sample space refers to all possible outcomes of an experiment

d) an event is the combination of all possible outcomes

22. Two events, A and B, are said to be mutually exclusive if:

a) $P(A \mid B) = 1$ b) $P(B \mid A) = 0.5$ c) $P(A \text{ and } B) = 0$ d) none of the above

23. Two events, A and B, are said to be independent if:

a) $P(A \text{ and } B) = P(A)P(B)$ c) $P(A \text{ or } B) = P(A) + P(B) - P(A \text{ and } B)$

b) $P(A \mid B) = P(B)$ d) $P(B \mid A) = P(A)$

Use the following information to answer questions 24 and 25.
The table below shows the results of a study on 175 people in which researchers examined the relationship between the presence of a certain mutated gene and colon cancer.

	Gene present	Gene absent	Total
Patient has cancer	70	20	90
Patient does not have cancer	25	60	85
Total	95	80	175

24. The probability that a randomly selected person is cancer free but has the mutated gene is closest to

a) 0.49   b) 0.71   c) 0.50   d) 0.14

25. The probability that a randomly selected person either has the mutated gene or has cancer is closest to

a) 0.66   b) 1.06   c) 0.54   d) 0.51

26. A staff member is told that his salary was in the 75th percentile. This means that

a) about 25 staff members earn more than him

b) about 75 staff members earn less than him

c) about 75% staff members earn less than him

d) about 75% staff members earn more than him

Use the following sample data to answer questions 27, 28, 29 and 30.

5   2   0   1   -2   6

27. Find the mean of the data.

a) 2.4

b) 2

c) 4.8

d) 3

28. The sample standard deviation of the data set is closest to

a) 4

b) 1.852

c) 2.170

d) 3.033

29. Find the median of the data.

a) 1

b) 1.5

c) 2

d) 2.5

30. Find the third quartile of the data.

a) 6

b) 5.75

c) 5.25

d) 5.5

SECTION B

### (Answer question one and any other two questions)

1a) A study on Central University student hall residents who utilize the university library was conducted. The first student was selected randomly, and then every eleventh student resident was interviewed as part of the sample. Data was collected on gender, duration (in hours) the individual spends at the library, and number of times the individual visits the library in the week. For purposes of making inferences, the proportion of female residents who use the library was obtained as 0.64.

    i)   What is the population for this study?

    ii)   Identify any statistic that is of interest to the research.

    iii) What is the random sampling method used?

    iv) For each of the variables sampled, identify the type of variable as either qualitative or quantitative discrete/continuous and state the level of measurement.

b) Indicate whether each of the following constitutes data collected from a population or a sample.

    i)   A group of 25 patients selected to test a new drug

    ii)   Number of items produced on a machine for each year from 1995 to 2017

    iii) Salaries of CEOs of all companies in Accra

    iv) Five hundred houses selected from Tema

    v)   Annual incomes of all 33 employees of a restaurant

c) Wellstone Incorporation produces markets replacement covers for cell phones in five different colors: bright white, metallic black, magnetic lime, tangerine orange, and fusion red. To estimate the demand for each color, the company set up a stand in the Accra Mall for several hours and asked randomly selected people which cover color their favorite was. The pie chart below displays the results:
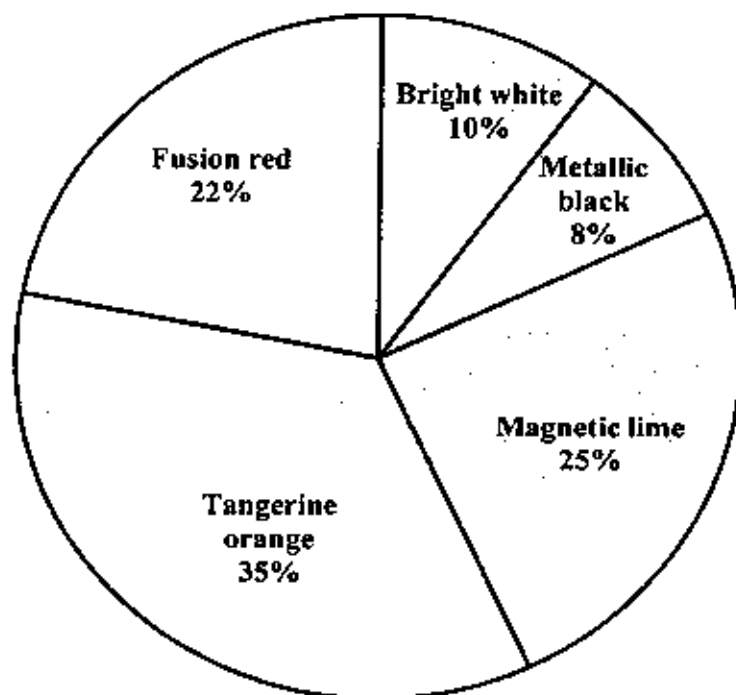
i. Construct a pareto chart for the different colors of cell phone covers. (*With the aid of a measuring ruler, draw the chart to reasonable scale*)

ii. Assume that Wellstone Inc. plans to produce 1 million cell phone covers, how many of each color would you recommend it produce?

2. The table below shows the number of times an issue was downloaded over the last 33 months. Suppose that you wish to summarize the number of downloads with a frequency distribution.

31	259	275	499	525	550	563	567	568	596	605
614	616	623	623	636	649	657	662	668	670	687
691	697	708	708	709	719	725	727	762	776	782

a. Which **type of grouping** would you recommend? Explain.

b. Create a grouped **frequency distribution table** for the data provided on the number of times an issue was downloaded over the last 33 months.

c. Compute the 90th percentile for the number of times the issue was downloaded.

d. Find the ungrouped mean of the data.

e. Find the grouped standard deviation and compare with on your answer in (d).

3. The National Association of Realtors reported the median home price in Ghana and the increase in median home price over a five-year period. Use the sample home prices (in ₵1000s) shown here to answer the following questions.

49	92	111	175	209	212	212
219	264	298	628	958	995	2325

a. What is the sample median home price?

b. In 2013, the National Association of Realtors reported a median home price of ₵139,300 in Ghana. What was the percentage increase in the median home price over the five-year period?

c. What are the first quartile and the third quartile for the sample data?

d. Provide a five-number summary for the home prices.

e. Obtain the limits and identify any outlier(s) in the data?

f. Considering the variations in the first, second, third and fourth quarters and comment on the shape of the distribution.

g. Why does the National Association of Realtors prefer to use the median home price in its reports over the mean home price?

10

4. The Internal Revenue Service (IRS) publishes data on adjusted gross incomes in Statistics of Income, Individual Income Tax Returns. The following frequency histogram shows one year's individual income tax returns for adjusted gross incomes of less than ₵50,000.

Histogram of Adjusted gross income (₵1000s)
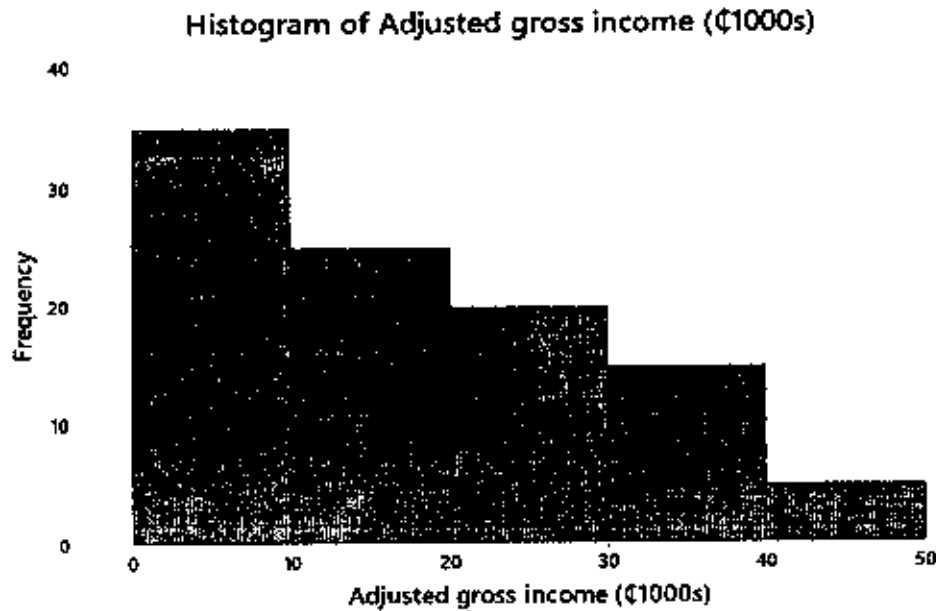
Use the histogram and the fact that adjusted gross incomes are expressed to the nearest whole cedis to answer each of the following questions.

a. Comment on the distribution of the adjusted gross income taking into consideration modality and skewness.

b. Create a grouped frequency table for the distribution and obtain the relative frequency and the cumulative relative frequency.

c. Approximately what percentage had an adjusted gross income of less than ₵30,000?

d. Obtain the sample mean and the sample standard deviation of the grouped data.

e. The IRS reported that 89,928,000 individual income tax returns had an adjusted gross income of less than ₵50,000. Approximately how many had an adjusted gross income between ₵30,000 and ₵49,999, inclusive?

5. A joint frequency distribution for number of farms, by acres and tenure of operator, is provided in the following contingency table. Frequencies are in thousands.

		Tenure of operator			
		T_1 Full Owner	T_2 Part Owner	T_3 Tenant	Total
Acres	A_1 Under 50		70	74	
	A_2 50–under 180	492	130	38	660
	A_3 180–under 500	198			368
	A_4 500–under 1000	51	84	14	149
	A_5 1000& over	41	114	17	172
	Total	1521	541		

a) Fill in the six missing entries.

b) How many cells does this contingency table have?

c) How many farms are operated by part owners and have between 500 and 1000 acres?

d) How many farms are not full-owner operated?

e) How many tenant-operated farms have 180 acres or more?

f) Construct a *joint probability distribution* corresponding to the table above

g) Using the general addition rule, compute $P(T_2$ OR $A_1)$ and interpret your answer in terms of percentages.

h) Find $P(T_2 \mid A_1)$ and interpret the probability you obtain in terms of conditional probability or percentages.

12

FORMULA SHEET

Location of a percentile: $L_p = \frac{p}{100}(n + 1)$

Percentile of a specific value y in a data set $p = \frac{x + 0.5y}{n} \times 100$

where x = number of data values less than the number you want to find the percentile for

where y = number of data values equal to the data value for which you want to find the percentile

IQR = Interquartile Range = $Q_3 - Q_1$

Lower limit = $Q_1 - 1.5 \times$ IQR

Upper limit = $Q_3 + 1.5 \times$ IQR

Pearson's Coefficient of Skewness: $sk = \frac{3(\bar{x} - Q_2)}{s}$ where Q_2 is the second quartile

Empirical Rule:

- About 68% of all values fall within 1 standard deviation of the mean
- About 95% of all values fall within 2 standard deviations of the mean
- About 99.7% (ie almost all the observations) fall within 3 standard deviations of the mean

Chebyshev's Theorem: At least $\left(1 - \frac{1}{k^2}\right)$ of the data values lie within k standard deviations of the mean

Standardization: $z = \frac{X - \mu}{\sigma}$

Grouping quantitative data

$2^k > n$ where k is the expected number of classes and n is the number of observations.

$i \geq \frac{H - L}{k}$ where i is the class interval, H is the maximum observed value, L is the minimum observed value, and k is the number of classes. Round up i to some convenient number, such as a multiple of 10 or 100

Ungrouped data

Population mean: $\mu = \dfrac{\sum x}{N}$

Sample mean: $\bar{X} = \dfrac{\sum x}{n}$

Geometric mean: $GM = \sqrt[n]{(X_1)(X_2)\dots(X_n)}$

Rate of increase over time: $GM = \sqrt[n]{\dfrac{Value\ at\ end\ of\ period}{Value\ at\ start\ of\ period}} - 1$

Mean Deviation: $MD = \dfrac{\sum |x - \bar{x}|}{n}$

Population Variance: $\sigma^2 = \dfrac{\sum (X - \mu)^2}{N}$

Sample Variance: $s^2 = \dfrac{\sum (X - \bar{X})^2}{n-1}$

Sample Standard Deviation: $s = \sqrt{s^2}$

Population Standard Deviation: $\sigma = \sqrt{\sigma^2}$

Grouped data

Population mean: $\mu = \dfrac{\sum fM}{N}$

Sample mean: $\bar{X} = \dfrac{\sum fM}{n}$

Population Variance: $\sigma^2 = \dfrac{\sum f(m-\mu)^2}{N}$

Sample Variance: $s^2 = \dfrac{\sum f(m-\bar{x})^2}{n-1}$

Probability

Complement rule: $P(A) = 1 - P(not\ A)$

General Addition Rule: Given that A and B are two events, $P(A\ or\ B) = P(A) + P(B) - P(A\ and\ B)$

General Multiplication Rule: Given that A and B are two events, $P(A\ and\ B) = P(A)P(B|A)$

Conditional Probability: $P(B|A) = \dfrac{P(A\ and\ B)}{P(A)}$

Bayes's Theorem: $P(A_j|B) = \dfrac{P(A_j)P(B|A_j)}{P(A_1)P(B|A_1) + P(A_2)P(B|A_2)}$ where $j = 1, 2$