DEPARTMENT OF STATISTICS & ACTUARIAL SCIENCE
SCHOOL OF PHYSICAL & MATHEMATICAL SCIENCES
FIRST SEMESTER, 2023/2024 ACADEMIC YEAR
B.A/B.Sc. STATISTICS
STAT 111: INTRODUCTION TO STATISTICS AND PROBABILITY 1

EXERCISE 2 ON LECTURES 3 AND 4

1. A professor has decided to use weighted average for determining the final grades for his students. The home assignment will count for 20% of a student's grade, quizzes 10%, the term paper 10%, the mid-term 25% and the final 35%. From the following data, compute the final average for each student.

Student	Homework	Quizzes	Term Paper	Mid Term	Final
Akosua	85	89	94	87	90
Brenya	78	84	88	91	92
Crentsil	94	88	93	86	89

2. The ages (in years) of 30 employees of a certain company are given below

55	25	34	66	28
32	28	26	56	48
47	61	25	24	31
24	28	44	37	51
47	35	32	36	50
37	36	27	27	44

(a) Calculate the mean, median and mode for the raw data.

(b) Determine the interquartile range (IQR).

(c) Find the variance and hence the standard deviation of the raw scores.

(d) Construct a frequency distribution table for the above data and hence estimate the following quantities:

(i) mean, median and mode

(ii) variance and standard deviation

(iii) the first and third quartile values

(iv) 85^{th} Percentile value

(e) The following are formulas we can employ to determine the coefficient of skewness:

$$S_k = \frac{Mean - Mode}{StdDeviation} \quad (1)$$

$$S_k = \frac{3(Mean - Median)}{StdDeviation} \quad (2)$$

$$S_k = \frac{Q_1 - 2Q_2 + Q_3}{IQR} \quad (3)$$

$$S_k = \frac{n}{(n-1)(n-2)} \sum_{i=1}^{n} \left(\frac{x_i - \bar{x}}{s} \right)^3 \quad (4)$$

Using your frequency table, compute estimate of the coefficient of skewness using each formula and compare your estimates.

3. A random sample of 20 STAT 111 interim assessment scores from 2022/2023 academic year is listed below:

29, 26, 13, 23, 23, 25, 17, 22, 17, 19, 12, 26, 30, 30, 18, 14, 12, 26, 17, 18

(a) Calculate the sample median and the interquartile range (IQR).
(b) Obtain a box-and-whisker plot for the data
(c) Are there outliers in the dataset? Explain your answer.

4. (a) A sample of 18 data values gave a mean of 6.5 and standard deviation 2.4. Another sample of 12 data values gave a mean of 8 and standard deviation of 2.0. Find the mean and standard deviation of the pooled sample.

(b) The mean of a team of 18 players is 86.5 kg. A new player joined the team and the new mean became 86 kg. What is the weight of the new player?

(c) A random variable Y has the following summary statistics:

$$n = 12 \ , \ \sum_{i=1}^{n} Y_i = 24 \ , \ \sum_{i=1}^{n} Y_i^2 = 330 .$$

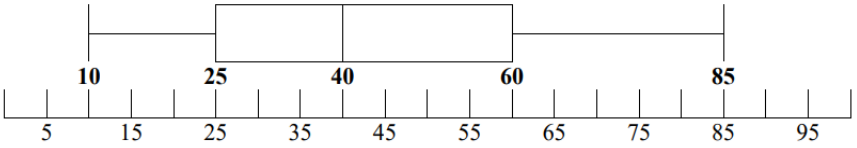
If $Y = T - 100$, find the mean and standard deviation of T.

5. (a) The geometric mean of five numbers is 25. Two numbers 26 and 30 were added to the data set. What will be the geometric mean of the new data set?

(b) Given the data; 18 , 22 , 23 , 23 , 25 , 28 , 34 and 35; find the harmonic mean.

(c) If the data in (b) is such that 18 appears 3 times, 22 appears 4 times, 35 appears 2 times with the remaining data being the same, compute the geomatic mean and the harmonic mean for the given data.

6. Consider the following box and whisker plot:



From the above box and whisker plot, ascertain the following:

(i) Median.

(ii) Smallest and the largest values.

(iii) First and third quartiles.

(iv) Whether the distribution is symmetrical or skewed, along with justification.

7. A research organization selected a sample of 30 visitors to a prestigious shopping mall. The data about the ages of the selected persons have been organized into the following table:

Age (in years)	Number of visitors
$18 \leq x < 23$	2
$23 \leq x < 28$	7
$28 \leq x < 33$	12
$33 \leq x < 38$	6
$38 \leq x < 43$	3

You are required to calculate the following:

(i) Mean

(ii) Sample variance and sample standard deviation.

(iii) Coefficient of variation.

(iv) Coefficient of Skewness and Coefficient of Kurtosis

(v) Comment briefly on the shape of the distribution.

8. The data in the following table shows the monthly maintenance cost and the ages of nine similar machines operating in a factory:

Machine	1	2	3	4	5	6	7
Age (in months) X	5	10	15	20	30	30	30
Cost Ghc 000) Y	19	24	25	30	31	32	30

(i) Describe the apparent relationship between maintenance cost and age of machine.

(ii) Find the Spearman's rank correlation coefficient of maintenance cost and age of machine.

(iii) Find the Pearson's product moment correlation coefficient of correlation and interpret your result.

(iv) Find the Kendal's rank correlation coefficient.

9. Fifteen students (named A-O) took classes in statistics and biology. The marks earned by each of the students are shown below.

Student	**A**	**B**	**C**	**D**	**E**	**F**	**G**	**H**	**I**	**J**	**K**	**L**	**M**	**N**	**O**
Statistics	74	53	67	63	77	57	60	47	76	54	80	92	53	52	80
Biology	60	68	64	66	71	66	55	71	82	73	84	59	63	55	79

(a) Use the information to obtain a scatter plot.

(b) Compute the product moment correlation coefficient and comment on its value.

(c) Compute the Spearman's rank correlation coefficient and comment on its value.

(d) Compute the Kendall's Tau correlation coefficient and comment on its value.

10. A survey was conducted to evaluate the effectiveness of the COVID-19 flu vaccine that had been administered in a small community in Accra. The survey of 1,000 inhabitants provided the information shown in the table below:

Status	Vaccine administered			Total
	No Vaccine	One Shot	Two Shots	
Flu	24	9	13	46
No Flu	289	100	565	954
Total	313	109	578	1,000

(a) Compute the expected frequency for each cell in the contingency table.

(b) Calculate the Chi-Square χ^2 statistic for the given data.

(c) Calculate the Phi ϕ coefficient for the given data and comment on its value.

(d) Calculate the Cramer's V statistic and comment on its value.

11. The table below displays the population (in millions) and the number of violent crimes (in millions) in US from 1982 to 1989.

Year	**1982**	**1983**	**1984**	**1985**	**1986**	**1987**	**1988**	**1989**
US Population	231	234	239	241	243	246	248	249
Violent Crimes	1.32	1.26	1.33	1.49	1.48	1.57	1.65	1.82

(i) Find the Spearman's rank correlation coefficient and interpret your result.
(ii) Find the Pearson's product moment correlation coefficient and interpret your result.
(iii) Find the Kendal's rank correlation coefficient and interpret your result.

12. A Psychological Society is interested in the relationship between anxiety level and the need to succeed in school. A random sample of 400 students took a test that measured anxiety level and need to succeed in school. The Table below shows the results:

Need to succeed in school	*Anxiety Level*					Total
	High	Med-High	Medium	Med-Low	Low	
High	35	42	53	15	10	155
Medium	18	48	63	33	31	193
Low	4	5	11	15	17	52
Total	57	95	127	63	58	400

(a) Compute the expected frequency for each cell in the contingency table.

(b) Calculate the Chi-Square χ^2 statistic for the given data.

(c) Calculate the Phi ϕ coefficient for the given data and comment on its value.

(d) Calculate the Cramer's V statistic and comment on its value.

13. The data below shows the distribution of living arrangements for university students by gender:

Gender	Living Arrangement			
	Dormitory	Apartment	With Parents	Other
Male	72	84	49	45
Female	91	86	88	35

(a) Compute the expected frequency for each cell in the contingency table.

(b) Calculate the Chi-Square χ^2 statistic for the given data.

(c) Calculate the Phi ϕ coefficient for the given data and comment on its value.

(d) Calculate the Cramer's V statistic and comment on its value.

14. Car manufacturers are interested in whether there is a relationship between the size of car an individual drives and the number of people in the driver's family (that is, whether car size and family size are independent). Suppose a sample of 800 car owners were randomly taken and the following Table contains the results:

Family Size	Car SIze			
	Sub & Compact	Mid-size	Full size	Van & Truck
1	20	35	40	35
2	20	50	70	80
3 – 4	20	50	100	90
5 +	20	30	70	70

(a) Compute the expected frequency for each cell in the contingency table.

(b) Calculate the Chi-Square χ^2 statistic for the given data.

(c) Calculate the Phi ϕ coefficient for the given data and comment on its value.

(d) Calculate the Cramer's V statistic and comment on its value.