(7) Intraocular pressure (in mmHg) for a sample of elderly patients were recorded as follows:

1	2	3	4	5	6	4	3	2
15.2	16.1	17.3	14.5	16.3	13.6	12.7	13.4	14.8
(6)	(7)	(10)	(4)	(8)	(3)	(1)	(2)	(5)

a. Calculate the median
b. Calculate the mean
c. What is the range?

...disease of your choice.

Section III: LONG ANSWER QUESTIONS. (Answer BOTH Questions in this section)

1. Results from a survey involving the nutritional status of the youngest child and mother's occupation were summarised as below:

		Nutritional status – youngest child		
		Undernourished	Fair	Obese
Occupation of mother	Skilled	2	25	32
	Non-skilled	30	20	15

Use a test of hypothesis to determine whether there is any association between the nutrition status and occupation of the mother

Note $X^2_{.95}$ (z) = 5.991

c) What information do you require in order to work out the budget?

Done

2. Following an intervention to improve healthier life styles, the following diastolic blood pressure levels were recorded in a sample from the community under intervention:

Serial no.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
DBP level (mmHg)	98	87	79	76	92	87	90	79	85	88	77	78	84	79	90	76	86	84	77	89

a) What is the median blood pressure level?

b) Calculate the mean, variance and standard deviation (use the individual values)

c) Calculate the standard error of the mean

d) Construct a 95% confidence interval for the mean sugar level in the population from which this random sample was drawn.

Note: $t_{.975\,(20)} = 2.0860$ $t_{.975\,(19)} = 2.0930$ $t_{.975\,(18)} = 2.1009$ $t_{.975\,(17)} = 2.1098$

$t_{.95\,(20)} = 1.7247$ $t_{.95\,(19)} = 1.7281$ $t_{.95\,(18)} = 1.7341$ $t_{.95\,(17)} = 1.7396$

Section I SHORT ANSWER QUESTIONS. (question 1 and 2 are compulsory)

1. In a hospital-based survey on road safety, out of a random sample 225 incidents of road traffic crashes, over-speeding was found to have contributed to 45 of the crashes.

   a. What is the proportion of crashes with over-speeding as a contributing factor?

   b. Write down a 95% confidence interval for the proportion of road traffic crashes with households with over-speeding as a contributing factor in the population from which this sample was drawn.

   Note: the sampling variance of a proportion is given by: $\frac{p(1-p)}{n}$

2. Briefly outline the role and

6) A community diagnosis in Kiambu involving a cluster random sample of 500 households found 450 of the households had VIP latrines or toilets.

a. What is the proportion of households with no or poor quality toilet facilities?

b. Write down a 95% confidence interval for the proportion of households with no toilet facilities in this population :-

7) Briefly describe five (5) f

[illegible] ANSWER QUESTIONS. (question 1 and 2 are compulsory)

1. Out of a sample of 1160 women attending ANC who were tested for HIV in County 'X' 45 of them were found positive.
   a. What is the prevalence of HIV in the sample?
   b. Write down a 99% confidence interval for the prevalence of HIV in county 'X'. (Note $Z_{.995} = 2.576$)

QUESTIONS. (Answer BOTH questions in this section)

1. In a survey on a community living around a factory suspected to be causing air pollution, the number of hospital visits in the preceding year for the youngest child under 5 years is recorded for a sample of households as shown below:

USE CODED DATA METHOD

HH No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
No. of visits	4	7	6	6	3	7	4	9	5	8	7	8	9	7	9	7	6	4	7	3

a) What is the median number of visits?

b) Calculate the mean, variance and standard deviation (use the individual values)

c) Calculate the standard error of the mean

d) Construct a 95% confidence interval for the mean number of hospital visits for the youngest child less than 5 years in the population from which this random sample was drawn.

Note: $t_{.99}(19) = 2.539$; $t_{.995}(19) = 2.8609$; $t_{.975}(19) = 2.093$; $t_{.995}(20) = 2.845$

12. In a clinical trial involving 4,396 patients aged 65-74 whose systolic pressure was between 160 and 209 mmHg and whose diastolic pressure was less than 115 mmHg, patients were randomly assigned to initial therapy with a diuretic or a beta-blocker or a matched placebo, and then followed up for an average of 5.8 years. The main objective was to see if a policy of anti-hypertensive treatment reduces the risk of stroke, coronary heart disease and death. The main results are as follows:

	Active treatment		
	Diuretic	Beta-blocker	Placebo
No. of patients	1081	1102	2213
Strokes	45	56	134
Coronary events	48	80	159
Deaths	134	167	315

Using confidence intervals in both cases, is there significant evidence for there being differences in the risk of:

(a) Death between those taking diuretic and those taking placebo? (10 marks)

(b) Stroke between those taking beta-blocker and those taking placebo? (10 marks)

...tion Number on EACH answer paper

ANSWER EACH QUESTION ON A SEPARATE SHEET OF PAPER

Section I: SHORT ANSWER QUESTIONS (Questions 1 and 2 are compulsory)

1. In a recent community diagnosis in greater Kiambu involving random sample of 800 women of reproductive age, 550 were found to be currently using a family planning method.

a) What is the proportion of women of reproductive age who were currently using a method?

b) Write down a 99% confidence interval for the proportion women of reproductive age who were currently using a method in the population from which this sample was drawn.

Note: the sampling variance of a proportion is given by:

$$\frac{p(1-p)}{n}$$

2. Describe the mechanism of Global warming

Section II: SHORT ANSWER QUESTIONS (Answer any Eight [8] questions)

1. The following are distances from the household to the nearest public health facility as recorded in a recent community diagnosis:

HH No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Distance from facility	6	5	8	6	6	7	8	9	5	6	7	8	4	7	9	7

a) What is the median distance

b) Calculate the mean, variance and standard deviation (use the individual values)

c) Calculate the standard error of the mean

d) Construct a 99% confidence interval for the mean distance from household to health facility in the population from which this random sample was drawn.

Note: $t_{.995}(15) = 2.9467$

4. Given the following observations on systolic blood pressure in mmHg

Class limits	frequency
90-100	205
100-110	326
110-120	458
120-130	514
130-140	420
140-150	362
150-160	258

a) Calculate the median

b) Use the coded data method to calculate the arithmetic mean, variance and standard deviation

. Below are given some descriptive statistics for two variables A and B. Using these statistics and making calculations when and if necessary, answer the questions below:-

Reg. No:.....

	Variable A	Variable B
Minimum Value	0.05	-0.20
First quartile	3.07	13.38
Median	6.27	18.29
Mean	8.97	16.61
Third quartile	11.59	25.14
Maximum value	43.24	33.37
Variance	68.94	92.97

(a) The middle fifty percent of the ordered values of variable. A fall between which two values? Interquartile range = Third quartile ↔ first quartile

A = 11.59 − 3.07 = 8.52

B = 25.14 − 13.38

B = 13.38 — 25.14

(b) Which variable is left skewed? Give reason

→ B → Mean is less than median.

(c) Which of the variables, A and B is more variable?

B is more variable → has a greater variance and a larger interquartile range.

8.5
11.71

In an anthropometric study to assess the nutritional status of school-going children in a rural community, the average weight of children at 10-years of age was found to be 25.7 kg with a standard deviation of 4.5. Accept these values as population parameters and that weight is normally distributed.

(a) What is the probability that a given 10-year old child in the community will have a weight greater than 24 kg?

$Z = \frac{24-\mu}{\sigma} = \frac{24-25.7}{4.5} = \frac{-1.7}{4.5} = 0.37777 = 0.3520$

$1 - 0.3520 = 0.6480$

(b) What is the probability that the average weight of 48 10-year olds will lie between 25 and 27 kg?

$Z = \frac{25-25.7}{4.5} = 0.16$

$Z = \frac{27-25.7}{4.5} = 0.29 =$

(c) What minimum weight do 10 year-olds in the community need to have to be in the top 10% of the children by weight?

[7 marks]

Z_{score} of $0.38 = 0.3520$;
Z_{score} of $3.29 = 0.0005$;
Z_{score} of $1.08 = 0.1401$;
Z_{score} of $1.285 = 0.10$

Z_{score} of $1.96 = 0.025$;
Z_{score} of $1.645 = 0.05$
Z_{score} of $1.645 = 0.05$;

SCHOOL OF PUBLIC HEALTH

Level V T3 2018

BIOSTATISTICS

INSTRUCTION: IN-CLASS ASSIGNMENT. WORK IN PAIRS (8.8.2018 at 11.30 a.m)

(To be discussed on Monday, August, 13, 2018)

1. What is Biostatistics? What are the Uses of Biostatistics?

2. With examples differentiate between **Descriptive** and **Analytic** Biostatistics.

3. Fill in the space provided with the appropriate **scale of measurement** and **Type of Variable** for the following data.

	Type of Variable	Scale of Measurement
Taking the height of an expectant mother who presents at the ANC clinic.		
Identifying the stage of disease for breast cancer patient.		
Recording the time of the day the patient is discharged from hospital.		
Recording the blood group of a subject participating in a study.		
Recording area of residence of a patient presenting at a health facility.		

4. Most of the biological data conform to the Normal distribution. What are the characteristics of such distribution?

5. Given the standard normal distribution, find:
   a. $P(-2.87 \leq Z \leq 2.64)$

   b. $P(Z \leq 2.33)$

   c. $P(Z \leq -2.33)$

   d. $P(0.84 \leq Z \leq 2.45)$

   e. $P(Z \geq 2.71)$

6. The following are the ages (in years) of 9 men residing in one of the Home for the elderly n a certain County.

   63 64 64 70 72 76 77 81 81

   For the above data, determine the following:

   i. Median

   ii. Mean

   iii. Standard deviation

7. What types of graphs/diagrams can be used to display nominal or ordinal observations? Discrete or continuous observations?

8. Among females in a certain County between 18 and 49 years of age, diastolic blood pressure is normally distributed with mean $\mu = 77$ mm. Hg and standard deviation $\sigma = 11.6$ mm. Hg. What is the probability that a randomly selected woman in this population will

   have her diastolic blood pressure:

   i. less than 60 mm. Hg.?

   ii. greater than 90 mm. Hg.?

   iii. between 60 and 90 mm. Hg.?

9. Displayed below is a frequency distribution of the resting systolic blood pressures for a sample of 35 patients with ischematic heart disease.

Blood Pressure (mm Hg.)	Number of patients
115 - 124	4
125 – 134	5
135 – 144	5

145 -154	7
155 – 164	5
165 - 174	4
175 - 184	5
Total	35

a. Construct the cumulative frequency distribution and a histogram for the above data.

b. What is the probability that a randomly selected patients will have his/her systolic blood
Pressure between 125 and 154 mm Hg.?

c. Determine the following for the above data:

i. Modal class.

ii. Median.

iii. Mean.

iv. Standard deviation.

v. Semi Interquartile range.

(vi) How much of child could be prevented ...

AF = (% unexp − % exp) / % unexp

(12) In a community survey on women of reproductive age who were currently using family planning, households were classified as rural, peri-urban or urban and the results summarized as follows:

		Residence			
		Rural	Peri-urban	Urban	
Currently using family planning	Yes	75	80	85	240
	No	25	20	15	60
		100	100	100	300

a) State a null hypothesis for the issue under study
b) Carry out the appropriate analysis
c) State your conclusion, using a significance level of 0.05
d) Name the test you would use if in the analysis above the expected values were too small even after collapsing the table

Note:
$X^2_{95}(6) = 10.645$ $X^2_{95}(3) = 7.815$ $X^2_{95}(2) = 5.991$
$X^2_{975}(6) = 14.449$ $X^2_{975}(3) = 9.348$ $X^2_{975}(2) = 7.378$

monitoring of a health activity provides a means of improving service.

(4) Given the following observations on systolic blood pressure in mmHG

Class limits	Frequency
55-60	14
60-65	28
65-70	40
70-75	56
75-80	38
80-85	26
85-90	18

Calculate the mean and median

Answer the first three questions and any other four questions in this section.

1. In a family planning study where 320 women of reproductive age were interviewed, it was found that 250 of them had ever used at least one method of family planning.
   a) What is the proportion of women who had ever used at least one method?
   b) Write down a 95% confidence interval for the proportion women who had ever used at least one method in the population from which this sample was drawn.

Answer All three questions

11. The table below presents the results of a study conducted by Westergren et.al. (2001) to determine the association between amount of food eaten and whether feeding is dependent or independent.

Feeding		Dependent	Independent	Total
Eats ≤ ¾ of served food	Yes	59	33	92
	No	17	44	61
	Total	76	77	153

Calculate:

i) The probability of dependent feeding in those who eat ≤ ¾ of served food.

ii) The probability of independent feeding in those who eat ≤ ¾ of served food.

iii) The odds ration, a measure of association, for the data presented in the above table. Interpret your calculated odds ratio.

Point of Exit.

x = Midpoint

Section I: SHORT ANSWER QUESTIONS

1. The data below describes the distribution of plasma volumes for patients seen at a clinic in Nairobi.

2.75, 2.86, 3.37, 2.76, 2.62, 3.49, 3.05, 3.12

a) Calculate the mean and standard deviation (SD) of the data

b) Calculate the coefficient of variation

$SD = \sqrt{Variance}$

$Variance = \frac{\Sigma fx^2 - \Sigma fx}{n-1}$

1. Explain what you understand by the following:

a) Population : collection of all the items of interest

b) Sample : a subset of the pop. chosen to represent the pop

c) Variable : a thing that varies from subj to subj or time to time

d) Measure of central tendency : A single value that describe the way in which a group of data cluster around a central value, m, m, m.

e) Measure of location : ~~Quantities that rep the same atypical value of a random~~ Describes where an observation/random var. lies on the scale of obs.

2. Describe four indicators of community participation in health care (ttt)

DS/04/08

LEVEL IV REVISION EXERCISES IN BIOSTATISTICS

I) INTRODUCTION OF BIOSTATISTICS

(A) Define 'statistics' as a discipline. – science of collection, org., summarizing of data, interpret. & drawing
(B) Define Biostatistics and list its uses in public health. – Applic. of statistical methods in biology.
(C) Give different types of variables with examples. – Qualitative – sex, religion; Quantitative – weight, height, B.P., family size
(D) Give different types of scales of measurement with examples. – Nominal scale – ... Ordinal – ... Interval – temp. Ratio – Age
(E) What is the importance of knowledge of types of data one collects?
(F) Which are the two main areas of Statistics/Biostatistics? What is/are the purpose(s) for each area? Descriptive – Organization & summarizing; Inferential – drawing inference

II) PRESENTATION OF DATA

1. List the different forms of presenting data for the following type of variables:
   (A) Qualitative data: ...Bar graph... Pie chart...
   (B) Quantitative data: .....Histogram... frequency... Polygon... Ogive graph

2. The following are the age (in years) of 30 patients seen at a certain health facility during the year 2010.

11 25 46 7 31 13 35 49 4 28 18 30 20 44 50 2 33 6 53 9 27 36 15 40 18 30 2 52 53 8

a. Construct i) Frequency distribution table.
ii) A histogram.
iii) Frequency polygon.

	F	cf
0 – 5	3	3
6 – 11	5	8
12 – 17	2	10
18 – 23	5	15
24 – 29	3	18
30 – 35	7	25
36 – 41	1	26
42 – 47	2	28
48 – 53	2	30

3. The following data presents the religious affiliation of 44 subjects enrolled in a certain study.

C P LM P P C P C LM M M C P C P C P M C P P M
P C LM C P M C C P M M P P C C C M P P M C C

Construct a: (a) Bar diagram (b) Pie chart for the given data.
Comment on the religious affiliation of the enrolled subjects.

C – 16
P – 16
LM – 3

frequency distribution in A (2 a) above.

IV) THE NORMAL DISTRIBUTION

(a) When do we say that the data conform to the 'Normal distribution'? Present 95%.

(b) List the properties of the 'Normal distribution'. i) Symmetrical / bell shaped ii) Have 3 measures made & median.

(c) What is the 'Standard normal distribution'? What is its importance? iii) 68% of pop. 1 bet. 1SD away

→ To compare diff. pop

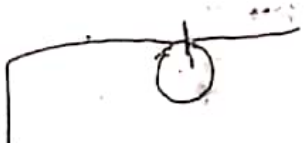
(d) Use the standard normal tables to determine the following probabilities:

a. $P(Z \geq 2.58) = 0.01$

b. $P(Z \leq 2.58)$

c. $P(-1.5 \leq Z \leq +1)$

d. $P(-1.96 \leq Z \leq -1.96)$



many factors eg occupation & mortality, marriage & morbidity

Section III: LONG ANSWER QUESTIONS. (Answer BOTH questions in this section)

1. The following are distances from the household to the nearest public health facility as recorded in a recent community diagnosis:

HH No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Distance from facility	6	5	8	6	6	7	8	9	5	6	7	8	4	7	8	9

a) What is the median distance $L_m + \left(\frac{\frac{n}{2} - cf}{f}\right)$:

b) Calculate the mean, variance and standard deviation (use the individual values)

c) Calculate the standard error of the mean $SEM = \frac{SD}{\sqrt{n}}$ SEM = cd/fn .

d) Construct a 99% confidence interval for the mean distance from household to health facility in the population from which this random sample was drawn.

$\bar{X} \pm t_{16\alpha/2} \times SEM$:

Note: t_{95} (15) = 2.9467

(2) An article on the relationship between diabetes and prolonged healing of wounds (Journal of the American Medical assoc. Nov 2, 1940, pp. 1518 – 152) gives the following data:

Null hyp. : diabetes does not cause prolong healing of wounds.

PATIENT	HEALING TIME		TOTAL
	Normal	Prolonged	
Diabetic	228	100	328
Non-diabetic	118	40	158
Total	346	140	486

a) Formulate the appropriate hypothesis to test the association

b) Carry out the test to make conclusions.(give the p-value

train to detect ear[illegible]

Parenteral supp. [illegible]

1. Growth monitoring + A[illegible]
2. integrated health + interventions
3. Prevention of specific micronutrient deficiency
4. Prev + [illegible] of infectious ds

Sensitivity

specificity

(c) What can you say about the performance of this test for breast cancer?

2. In a community survey on infant immunization in Kiambu County, the following data are obtained

chi square or Z test.

		Mother's Education Level		
		Primary	Secondary	Tertiary
Immunization Status	Fully Immunized	40	32	16
	Not fully Immunized	20	12	5

(a) State a null hypothesis for the problem under investigation.

(b) Carry out the appropriate analysis.

(c) Give your conclusion.

Define a communicable disease and the

are arising
- timely decision making
- project accountability
- provides foundation for evaluation & learning
- Provides info enabling mgt staff & assess implementation progress.

Given the following observations on systolic blood pressure in mmHG.

Class limits	Frequency		X	fx
55-60	14	14	57.5	805
60-65	28	42	62.5	1750
65-70	40	82	67.5	2700
70-75	56	138	72.5	4060
75-80	38	176	77.5	2945
80-85	26	202	82.5	2145
85-90	18	220	87.5	1575
	220			15980

Calculate the mean and median

$$\text{Mean} = \frac{\Sigma fx}{\Sigma f} = \frac{15980}{220} = 72.64$$

$$\text{Median} = L.L + \left(\frac{n}{2} - \text{Cf above}\right) \times$$

$\bar{x} = \frac{15980+45}{220} = 72.64$

(2) ... if no one smoked? [2 marks]

In a community survey on women of reproductive age who were currently using family planning, households were classified as rural, peri-urban or urban and the results summarized as follows:

		Residence		
		Rural	Peri-urban	Urban
Currently using family planning	Yes	75	80	85
	No	25	20	15

a) State a null hypothesis for the issue under study
b) Carry out the appropriate analysis
c) State your conclusion, using a significance level of 0.05
d) Name the test you would use if in the analysis above the expected values were too small even after collapsing the table

Note:
$\chi^2_{.95}(6) = 10.645$ $\chi^2_{.95}(3) = 7.815$ $\chi^2_{.95}(2) = 5.991$
$\chi^2_{.975}(6) = 14.449$ $\chi^2_{.975}(3) = 9.348$ $\chi^2_{.975}(2) = 7.378$

KINARCA H.D REVISION

BIOSTATISTICS

- In a survey carried out at border towns, migrant workers were asked to state their most important health problem with the following results:
  a) Calculate the relative frequencies
  b) Use an appropriate method of presentation for these data

	Health Problem	Frequency
1	Malaria	70
2	STI/HIV/ AIDS	55
3	Gastro-intestinal	40
4	Mental	20
5	Malnutrition	10
6	Others	5

KINARCA H.D REVISION

Edit title

- In a study of the television viewing habits of young people, a developmental psychologist selected a random sample of 300 first year university students- 100 boys and 200 girls. Each student was asked which of the following tv programs they like best: Sports, Movies, or Comedies.
  1. State the 'null' and 'alternative' hypotheses
  2. Do the boys' preferences for those tv programs differ significantly from the girls preference? Use 0.05 level of significance.

Viewing preferences

	Sports	Movies	Comedies	Total
Boys	50	30	20	100
Girls	50	80	70	200
TOTAL	100	110	90	300

Edit title

- State the 'null' and 'alternative' hypotheses
- Do the boys' preferences for those tv programs differ significantly from the girls preference? Use 0.05 level of significance.

NOTES

Edit title

- Explain the following terminologies as applies to Biostatistics giving one example of each to illustrate your definition.
  a) Observation
  b) Variable
  c) Binary variable
  d) Quantitative variable
  e) Frequency

KINARCA H.D REVISION

1. The following are distances from the household to the nearest public health facility as recorded in a recent community diagnosis:

HH No.	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
Distance from facility	6	5	9	6	6	7	8	9	5	6	7	8	4	7	9	7

5 5 6 6 6 6 7 7 7 7 8 8 9 9 9

a) What is the median distance

b) Calculate the mean, variance and standard deviation (use the individual values)

6.8125

c) Calculate the standard error of the mean 0.378

d) Construct a 99% confidence interval for the mean distance from household to health facility in the population from which this random sample was drawn.

Note: $t_{.995}(15) = 2.9467$

KINARA H.D REVISION

Edit title

6. A community diagnosis in Kiambu involving a cluster random sample of 500 households found 450 of the households had VIP latrines or toilets.
   a. What is the proportion of households with no or poor quality toilet facilities?
   b. Write down a 95% confidence interval for the proportion of households with no toilet facilities in this population :-

KINARCA H.D REVISION

Edit title

...and any other four questions in this section.

1. In a family planning study where 320 women of reproductive age were interviewed, it was found that 250 of them had ever used at least one method of family planning.
   a) What is the proportion of women who had ever used at least one method?
   b) Write down a 95% confidence interval for the proportion women who had ever used at least one method in the population from which this sample was drawn.

stats

KINARA H.D REVISION

NOTES

Edit title

- List three properties of a population which follows the normal distribution

NOTES