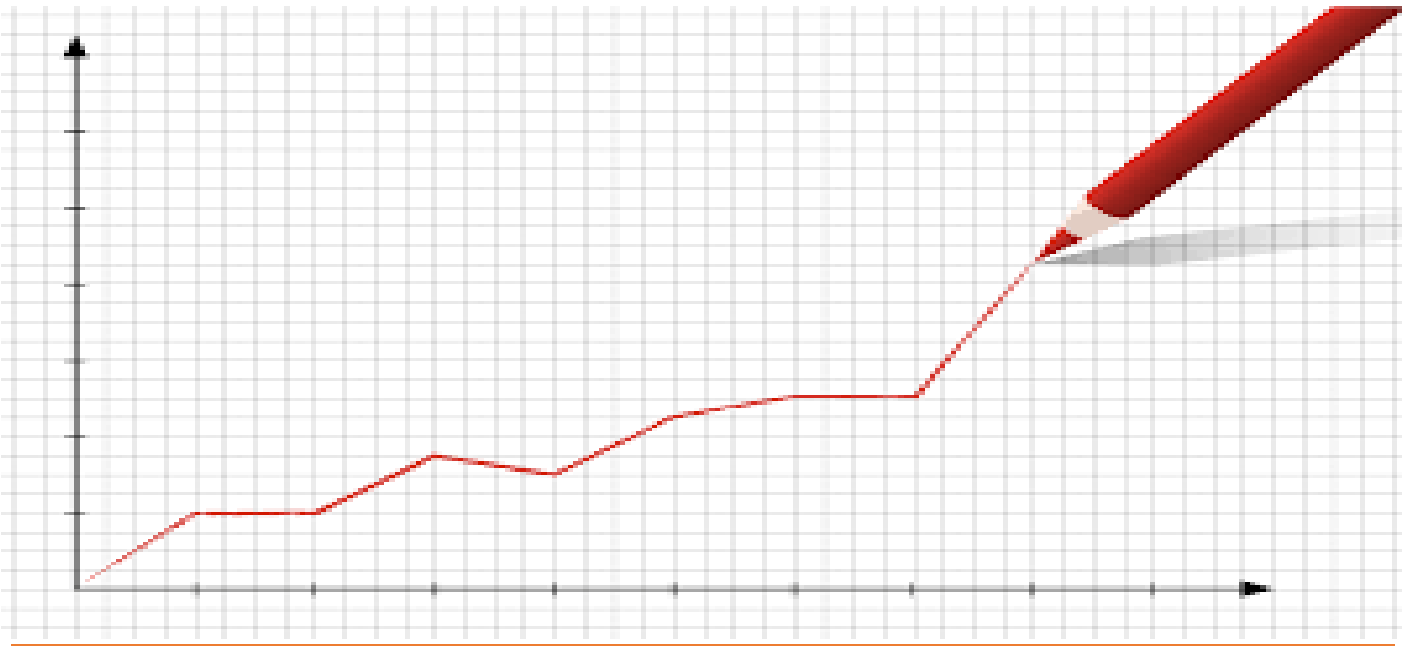
GEOGRAPHY NOTES FORM 1

STATISTICAL METHODS



Specific objectives

By the end of the topic, the learner should be able to:

a) define the term statistics

b) identify types and sources of statistical data

c) identify and describe methods of collecting and recording data

d) analyze and present statistical data using appropriate graphical methods

e) state the advantages and disadvantages of each method of data presentation.

STATISTICS

Statistics-numerical figures collected systematically and arranged for a particular purpose.

Statistical data-information presented inform of numbers e.g.

1) No. of students in a school
2) Mean daily temperature of a place
3) Amount of milk produced daily from a farm
4) Amount of money earned from exports annually.

Statistical methods-techniques of collecting, recording, analyzing, presenting and interpreting statistical data.

Significance of Statistics

i) Illustrates relationship between 2 or more varying quantities e.g. beans production and acreage under cultivation.
ii) Summarizes geographical information which saves time and space.
iii) Makes comparison between components e.g. province with the highest number of people.
iv) Prediction of future trends of weather and climate.
v) Prediction of natural disasters e.g. droughts and floods.
vi) Planning for provision of social amenities e.g. hospitals and schools.

Types of Statistical Data

Primary Data

First hand or original information from the field e.g.
Mean daily temperature from a weather station
Enumeration/census

Learner's Short Notes

Secondary/Derived Data

2nd hand information available in stored sources compiled by other researchers .e.g.

a. Textbooks
b. Reference books
c. Maps
d. Video/audio tapes
e. Textbooks
f. Newspapers
g. Magazines
h. Census reports
i. Slides
j. Census reports

Nature of Statistical Data

1. Discrete Data

Which is given in whole numbers e.g.

16 elephants
1093 tonnes of wheat .

2. Continuous Data

Facts and figures which can take any value e.g.

a. Fractions e.g.23 ¼
b. Decimals e.g. 6.20 mm
c. Values within range e.g. 0-30◦c

3. Grouped Data

Which is non precise/exact but values range in groups e.g.

Age group	Number of boys
15-19	32
20-24	8

Learner's Short Notes

Sources of Statistical Data

1. Primary Sources

People or places which have 1st hand or original information.

The information can be collected by observation, measuring, counting, photographing etc.

Advantages

i) Give first hand information
ii) The information cant be got from other sources

2. Secondary sources

Materials in which information collected by others was stored e.g. text books, reference books, etc.

Methods of Collecting Data (statistical Techniques)

1. Observation

Use of eyes to observe features or weather then information is recorded immediately e.g. cloud cover, rocks, soil, land forms, vegetation, etc.

Advantages

i) Gives 1st hand information which is reliable.
ii) Relevant material to the study is collected.
iii) Time saving since one doesn't have to look for data in many places.

Disadvantages

i) Data on past activities isn't available.
ii) May be hindered by weather conditions e.g. mist and dust storms.
iii) Ineffective for people with visual disabilities.
iv) Tiresome and expensive as it involves a lot of travelling because physical presence is required.

2. Interviewing

Gathering information from people by direct discussions then answers are recorded. It may be face to face or on a telephone. A questionnaire prepared in advance is used.

Learner's Short Notes

Guidelines

- One should be polite
- Warm and friendly
- Respondents/ interviewees should be assured information is confidential.
- Respondent should not be interrupted when answering questions.
- They should not be given clues but answers should come from them.

Advantages

i) Reliable first hand information is collected.
ii) Interviewer can seek clarification incase of ambiguity of answers.
iii) Can be used on illiterate.
iv) Interviewer can gauge the accuracy of responses.

Disadvantages

i) Time consuming since one person can be handled at a time.
ii) Expensive and tiresome as extensive travelling is required to meet the respondents.
iii) May encounter language barrier if the respondent doesn't speak the same language as the interviewer.
iv) A respondent may lie, exaggerate or distort facts leading to collection of wrong information.

3. Administering questionnaires

Set of systematically structured questions printed on paper used on interviews or sent to respondents to fill answers.

Types

Open-ended questionnaire-in which respondent is given a chance to express his views. The disadvantage is that different answers are given which are difficult to analyze.

Closed-ended (rigid) questionnaire-in which respondents are given answers to choose from.

Characteristics of a good questionnaire

i) Short
ii) Uses simple language
iii) Systematically arranged from simple to difficult
iv) Clear questions
v) Doesn't touch on respondent's privacy

Learner's Short Notes

Advantages

i) Comparisons can be made since questions are similar.
ii) First hand information which is relevant to current trends and situation is collected.
iii) Saves money on travelling as physical presence isn't required.
iv) Saves time as all respondents are handled at the same time.
v) A lot of information can be collected.

Disadvantages

i) Difficult analysis due to different answers.
ii) Some questionnaires may be sent back while blank by lazy respondents.
iii) Can't be used on illiterate respondents.
iv) Some respondents may write wrong information

4. Content analysis

Technique of collecting data from secondary sources.

This is by reading, watching films, viewing photographs and listening to get what is relevant.

Advantages

i) Easy to get data if analyzed.
ii) Cheap as there isn't extensive travelling
iii) Saves time as all information is in one place.
iv) Possible to get old data

Disadvantages

a) Difficult to verify accuracy of data
b) Data may be irrelevant to current trends
c) Up to date data may not be readily available

5. Measuring

Determining distances, areas, height or depth using instruments and recording.

Distance can be estimated by pacing or taking steps of equal and unknown length.

Learner's Short Notes

6. Collecting Samples

Getting a small part e.g. of soil, rock or vegetation to represent the whole to be used to carry out tests in the laboratory.

7. Counting/census taking

Arithmetical counting and recording.

8. Photographing

Capturing on film or video and still photographs.

9. Digging

Using tools such as hoe pick axe, spade or soil auger to get samples of soil and rocks.

10. Feeling and touching

Using fingers to feel the surfaces of soils and rocks to get their textures

11. sampling

Examining by taking a sample -a part representing the whole (population).

Types of Sampling

1. Random Sampling

Selection of members of a group haphazardly where every item has an equal chance of being selected e.g. to select 5 students to go for a tour from a class:

a. Class members write their names on pieces of paper
b. They are folded and put in a basket
c. The basket is shaken and fives papers are taken out

2. Systematic Sampling

Selection of members of a sample from an evenly distributed phenomena at regular intervals e.g. after every 10 items/members.

Learner's Short Notes

3. Stratified sampling

Selection of members of a sample by breaking the population into homogenous groups e.g. to select 6 students to go for a tour:

a) Break the class into boys and girls
b) Select 3 student from each group by random or systematic sampling
c) Combine units from each group to form the required sample.

4. Cluster Sampling

Selection of sample by dividing the sample into clusters with similar characteristics then a sample is taken from each cluster and representative choices from each cluster are combined to form a sample e.g. to sample the housing cost an estate is chosen to represent each group and representative choices are chosen from each estate and combined to form a sample.

Advantages

a) It's less expensive
b) It saves time
c) It avoids bias

Disadvantages

i) A poor selected sample can lead to misleading information
ii) Systematic sampling to an evenly distributed population

5. Experimentation

Conducting a test or investigation to provide evidence for or against a theory e.g. to determine the chemical composition of rocks and soils.

Advantages

i) First hand data is obtained
ii) Gives accurate results if properly conducted.
iii) It can lead to further discoveries

Learner's Short Notes

Disadvantages

i) May be expensive as it involves use of expensive equipment.
ii) May be time consuming
iii) Use of defective instruments may lead to inaccurate results
iv) Improper handling of equipment and chemicals may lead to accidents

Methods of Recording Data

Methods off storing information to avoid losing it.

1. Note Taking

Writing in a note book what is being observed, answers during interviews and then notes are compiled in school or office when writing report.

2. Filling In Questionnaires

Filling answers in questionnaires which are responses from a respondent by an interviewer or respondent himself which he/she then sends back.

3. Tallying

Making 4 vertical or slanting strokes and the 5th across the 4 to record data obtained by counting or measuring similar items.

Item	Frequency of occurrence	total
A	𝍸 |||	8
B	||||	3

4. Tabulation .

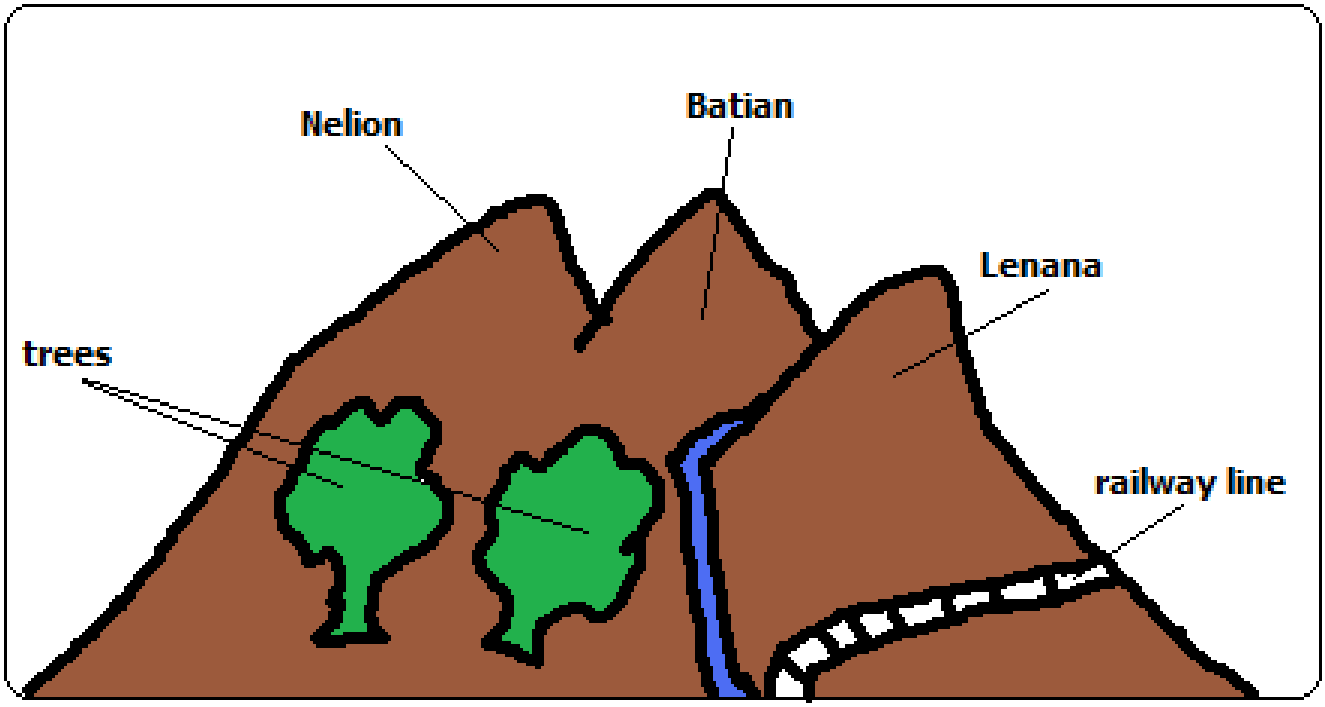
Drawing of tables and filling in data systematically e.g. weather recording sheets.

Month	J	F	M	A	M	J	J	A	S	O	N	D
Temp (◦c)	24	24	23	22	19	17	17	18	19	20	22	23
Rainfall (mm)	109	122	130	76	52	34	28	38	70	108	121	120

Learner's Short Notes

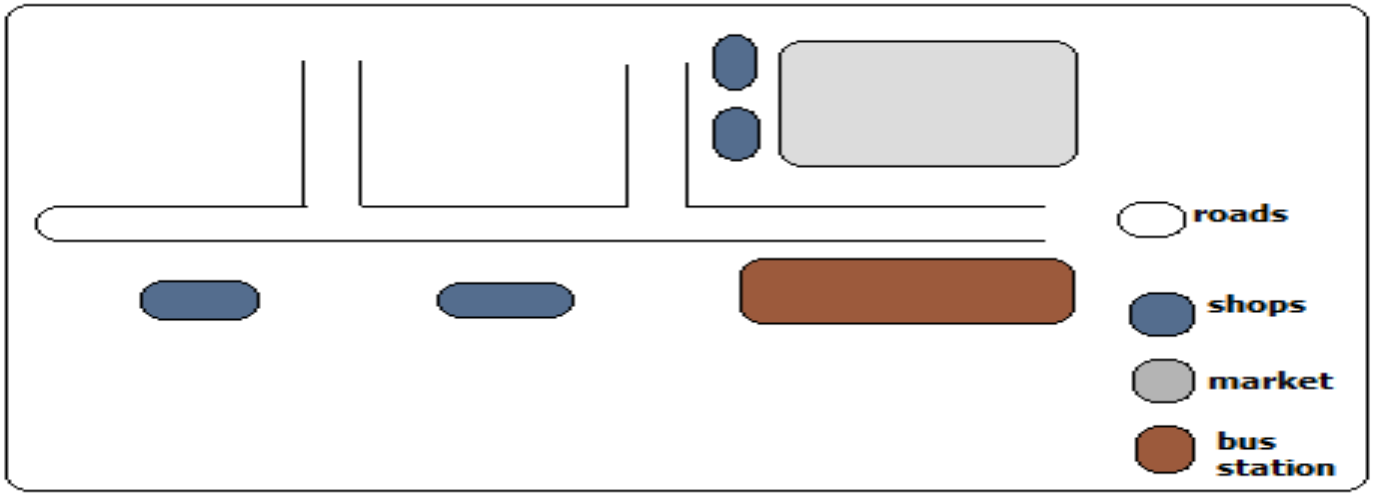
5. Field Sketching

Summarizing information observed in the field by making a rough drawing of landscape and labeling the essential information



6. Mapping/Drawing Maps

Drawing of a rough map of an area of study and labeling in words or symbols accompanied by key.



Learner's Short Notes

7. Tape Recording

Recording image of an object or landscape on a film which is processed to get a photograph then the photographs are labelled to avoid mix up during storage.

8. Labeling samples

Recording conversations during interviews on audio tapes using a tape recorder.

Permission should be got from the respondent to record his/her responses.

Advantages

i) It's used if responses are too many to be recorded on a note book.
ii) It allows smooth flow of discussion as asking respondents to repeat answers would irritate them.

Analysis of Data

Examining the numerical figures in detail.
Techniques of analyzing Data

1. Calculation of Percentages

-If in the study of a farm 10 hectares are devoted to coffee, what is the % of the area under coffee?

10/100×10%

The table below shows the number of tourists who visited Kenya from various parts of the world in 2006.

Place of Origin	No. of tourists per year	
	2005	2006
Europe	942000	965000
Africa	120000	154000
Asia	97000	128000
Total	1159000	1247000

a) Calculate percentage increase of tourists from Africa between 2005 and 2006.

Learner's Short Notes

2. Measures of Central Tendency

Outstanding general characteristics of the data.

a) **Arithmetic Mean**

The average

$$\text{Where } \bar{X} = \text{mean}$$
$$\Sigma X = \text{sum of values}$$
$$N = \text{number of values}$$

Advantages

i) Easy to calculate for a small data
ii) Summarizes data using a single digit
iii) Easy to understand and interpret

Disadvantages

i) Difficult to calculate for grouped data
ii) Affected by extreme values

b) Median

The middle value in a set of data arranged in order. M= (N+1)/2

(I) 20, 50, 90, 100, 150, 180, 200, 220, 240, 300, 360.
(II) 20, 50, 90, 100, 150, 180, 200, 220, 240, 300.

Advantages

i) Easy to calculate in a small data set
ii) Easy to understand as it's the value at the middle

Disadvantages

i) Difficult to calculate in a large data set
ii) Doesn't show data distribution

Learner's Short Notes

c) Calculation of Ranges

Difference between the largest and smallest values. Calculate the range of for the data above.

d) Mode

Most frequently recurring value in a set of data.

10, 2, 5, 9, 10, 11, 20, 15, 18, 10.

The mode is 10.

Advantages

i) Easy to find as no calculation is involved
ii) Easy to understand

Disadvantage

i) Rarely used as a measure of central tendency

Graphs: Statistical Presentation of Data

Two dimensional drawings which show relationships between two types of data representing two items also called variables. These are dependent variable Which is affected by the other e.g. temperature (on y-axis) and independent variable whose change is not affected by the other e.g. altitude(on x-axis)

Steps

i) Draw x and y axis.
ii) Choose suitable scale to accommodate the highest and lowest value.
iii) Plot the values accurately using faint dots.
iv) Join the dots using curved line. If it's a bar graph the dots should be at the middle of the top line. Years should also be at the middle. You should have also decided on the width of the bars.
v) In data without continuity e.g. crop production here should be gaps between bars and for one with continuity e.g. rainfall bars should not have gaps.
vi) Draw vertical lines on either side of the dot then draw horizontal line to join them with the dot.
vii) Shade uniformly if they are representing only one type of data and differently if representing one type of data.
viii) In combined line and bar graph temperature figures are plotted on the right hand side of y-axis while rainfall on the left
ix) Don't start exactly at zero.
x) Include temperature and rainfall scales.
xi) Start where the longest bar ends.

Learner's Short Notes

What a Well Drawn Graph Should Have

a. Title
b. Scale/scales
c. Labelled and marked x and y axis starting at zero.
d. Key if required e.g. in comparative bar graph.
e. Accurately plotted and lines, curves or bars properly drawn

1. Simple Line graph

The simplest of line graphs is the *single line graph*, so called because it displays information concerning one variable only, in terms of its frequencies.

Advantages

i) Easy to construct
ii) Easy to interpret
iii) Easy to read/estimate exact values.
iv) Shows trend or movement overtime.

Disadvantages

i) Doesn't give a clear impression on the quantity of data.
ii) May give false impression on the quantity especially when there was no production.
iii) Poor choice of vertical scale may exaggerate fluctuations in values.
iv) Difficult to find exact values by interpolation.

2. Simple Bar Graph/histogram

Advantages

i) Easy to construct.
ii) Easy to interpret.
iii) Easy to read.
iv) Gives a clear visual impression on the quantity of data.

Disadvantages

i) Poor choice of vertical scale may cause exaggeration of bars.
ii) Doesn't show continuity/variation of data overtime.
iii) Unsuitable technique when values exist in continuity.
iv) Not possible to obtain intermediate values from the graph

Learner's Short Notes

3. Combined Line and bar Graph

Advantages

i) Easy to construct.
ii) Easy to read.
iii) It shows relationship between two sets of data.

Disadvantages

i) Difficult to choose suitable scale when values of variables differ by great magnitude.
ii) Considerable variation of data represented by the line may cause the line the bars thus obscuring the relationship.
iii) Doesn't show relationship between the same sets of data of more than one place.

4. Wind rose

This is a chart on which data on wind direction for a given period of time (e.g. a month) is recorded.

Advantages

i) Easy to construct and read
ii) Easy to interpret and to understand
iii) Easy to compare the frequency of wind blowing in different directions
iv) Gives a good visual impression of the direction of wind.

Disadvantages

i) Difficult to know the exact numbers of days or time the wind blew in a certain direction.
ii) Values are difficult to extract since it involves measuring and calculating using the give scale.
iii) The wind blowing pattern over a given perod is difficult to discern from the diagram.

Learner's Short Notes

STATISTICAL METHODS PAST KCSE QUESTIONS

1. The table below shows petroleum production in thousand barrels per day for countries in the Middle East in April 2006. Use it to answer question (a)

Country	Production in '000"
Iran	3800
Kuwait	2550
Qatar	800
Saudi Arabia	9600
United Arab Emirates	2500
Iraq	1900

a)
   i) What is the difference in production between the highest and the lowest producer (1mk)
   ii) What is the total amount of petroleum produced in April 2006 in the region? (1mk)

b) State three conditions that are necessary for the formation of petroleum (3mks)

2. The graph below shows percentage value of some export commodities from Kenya between 1999 and 2003.

Use it to answer questions (a) and (b)

(a)

   i) What was the percentage value of the tea exported in the year 2000? (2mks)
   ii) What was the difference in the percentage values of the horticultural products and coffee exports in 1999? (2mks)
   iii) Describe the trend of the value of coffee exports from 1999 to 2003 (3mks)
   iv) Explain three factors which may have led to the increased export earnings from horticultural produce in Kenya between years 1999 and 2003 (6mks)
   v) Give three advantages of using simple line graphs to represent data. (3mks)

(b) State four reasons why Kenya's agricultural export earnings are generally low (4mks)

(c) State five reasons why the common market for Eastern and southern Africa (5mks)

3.

(a) Define the following terms

   i) Statistics

   ii) Statistical data

   iii) Statistical methods (6mks)

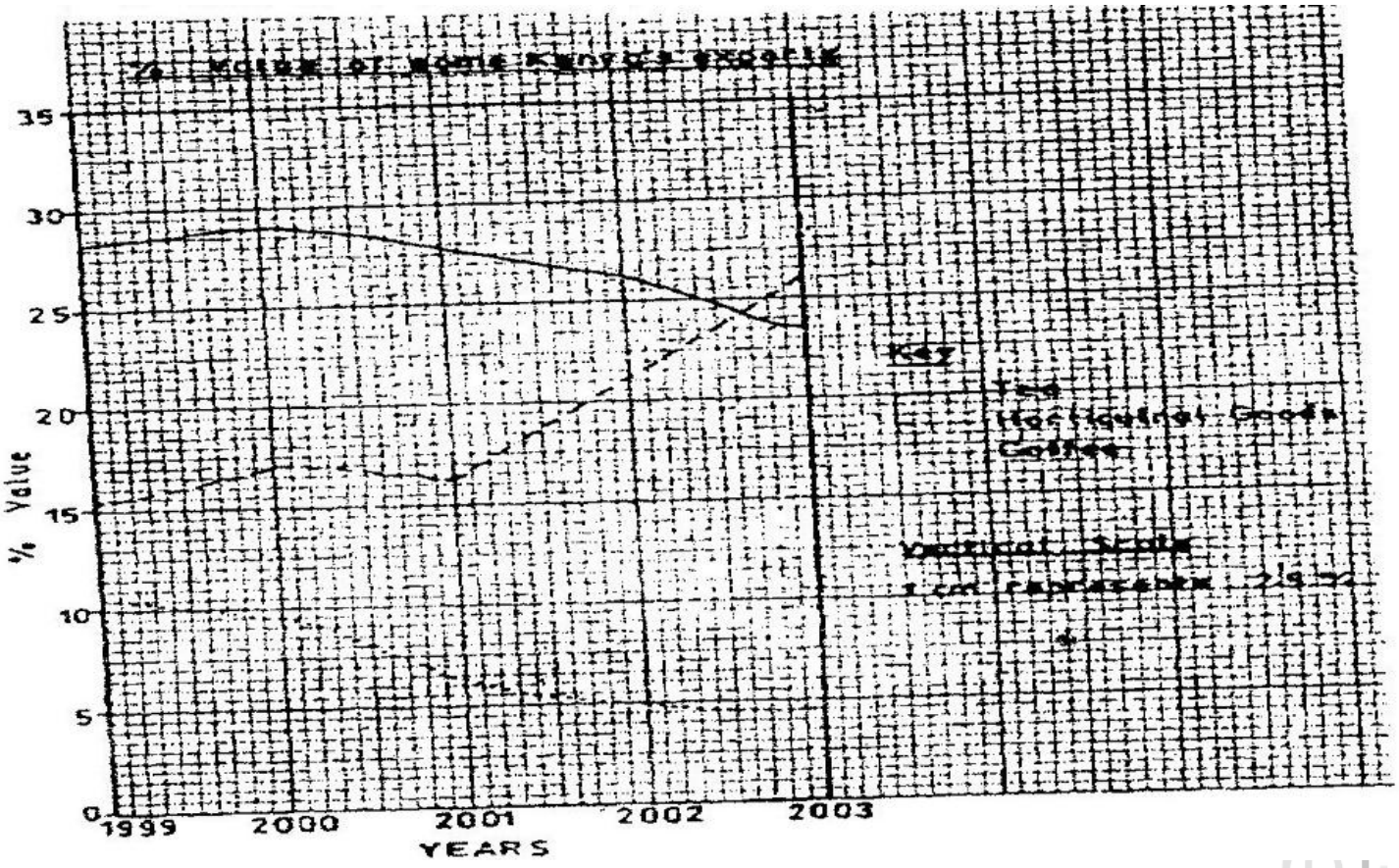
(b) State two types of statistical data. (2mks)

(c) Write down two types of questionnaires. (2mks)

4.

(a) What factors must be considered in selecting methods of data collection. (3mks)

(b) Differentiate between discrete data and continuous data giving relevant examples. (4mks)



5. (a) What is sampling (1mk)

   (b) State 3 types of sampling. (3mks)

6. (a) Name two main methods used in analyzing statistical data. (2mks)

   (b) What is the significance of statistics in geography? (5mks)

7. (i) Name two types of graphs that you have learnt about. (2mks)

   (ii) What are the advantages of using graphs named above in representing statistical data? Give advantages. (4mks)

8. (i) What is a questionnaire?

   (ii) State four advantages of using questionnaires in collection of' statistical data. (4mks)

   (iii) Explain oral interview method. (2mks)

9. Explain the following methods of data recording.

   a. Tabulation

   b. Photographing

   c. Tape recording

   d. Tallying

10. What is data? (2mks)

11. Marks 72, 60, 65, 70, 65, 80, 65, 70, 80, 84, 63, 75, 63, 71, 74

    Use the data above to find out mean and mode. (4mks)

12. With the help of data above explain how median is obtained. (3mks)

MARKING SCHEME

1. (a)

i) 9600-800 = 8,800,000 Barrels

ii) 21,150,000 Barrels.

iii) 21,150,000\30 = 705,000 Barrels

2. Graph

i) 29.3% (29 - 29.5%) 29%

ii) 4.75%/4.8%/4.9%
(4.75% - 4.9%)

(iii) Describe the trend of the value of coffee exports from years 1999 to 2003.

a. The value was generally declining over the five year period.
b. The value was highest in 1999.
c. The decline between 1999 and 2000 was minimal/ gradual.
d. The highest drop was between 2000 and 2001
e. There was a minimal drop between 2002 and 2003.
f. The decline between 2001 and 2002 was minimal/ gradual.
The value was lowest in 2003.

(iv) Explain three factors which may have led to the increased export earnings from horticultural produce in Kenya between years 1999 and 2003.

a. Improved technology which leads to advanced crop husbandry/increase the volume of fresh horticultural products.
β. Aggressive promotion of trade abroad leading to a wide/ready market in foreign countries.
c. Improved ways of packaging have made the produce more competitive/ attractive,
d. Improved infrastructure/air/road transport have helped in the quick means of transportation of fresh produce to the market.
e. The declining benefits from traditional agricultural exports leading to the expansion of the areas under horticultural crops.
f. The government has encouraged the formation of organizations that are assisting horticultural farmers.

v) Give three advantages of using simple line graphs to represent data.

a. Give clear visual impression.
b. Easy to construct.
c. Easy to interpret.
d. Can be used to represent a wide variety of variables.
e. Appropriate for comparison

(b) Reasons why Kenya's agricultural export earning generally are low

i) Kenya sells most of her agricultural products in their raw form and they are priced lowly.
ii) International prices keep fluctuating from year to year.
iii) Prices of some commodities are externally determined.
iv) There is competition from other producing countries/from other similar products.
v) Some products are inferior in quality.
vi) There are fixed quarters for some agricultural products.
vii) Decline in quantities of some agricultural exports.

3. (a)

i) Statistics It refers to the art or science that is concerned with the interpretation of numeric information.
ii) Statistical data Refers to the information collected and arranged in a systematic manner.
iii) Statistical methods Refers to the techniques used in collecting, recording, analyzing and presenting data.

(b) Primary data and Secondary data

(c) Closed-ended (rigid) - Open-ended

4. (a)

i) The method should be inexpensive.
ii) Should be time saving
iii) Should give accurate data
iv) Most applicable method

(b) Discreet data refers to the non-continuous data over time given in whole numbers only e.g.

i) Total population in a nation.
ii) Monthly rainfall totals.
iii) No. Of livestock per district
iv) Continuous data can be given in any value including decimals e.g. 1.8km.

5. Sampling refers to the process by which a representative portion of the whole phenomena under study is analyzed and generalized/ generalization is made.

Types of sampling

i) Systematic sampling
ii) Stratified sampling
iii) Random sampling

6. (a)
i) Calculation of percentages
ii) Measuring of Central tendency (mean, median and mode)
iii) Frequency distribution

(b)
i. Predicting for future trends.
ii. Showing changes through time
iii. Establishing Geographical relationships
iv. For economic planning
v. For explaining geographical phenomena.
vi. Useful for making comparisons.

7. (i)
a. Simple line graph
b. A combined line and bar graph
c. Simple bar graph

(ii)
a) The simple bar graph
b) Prominent values stick out well
c) Bars are appealing to the eye
d) Easy to draw, read and interpret the data represented

8.
(i) A set of pre-questions which are related to the topic of study.

(ii)
a. Its a source of first hand information
b. The researcher can ask for clarification from the respondent.
c. Similar questions are used for all respondents and comparison can easily be made.
d. When posted, rigid questionnaires reduce fieldwork expenses.

(iii) Interview involves collection of information by asking questions directly and recording the answers given. In this method the researcher established contact with the respondent and agree on time for face to face interview with the respondent. Interviews can also be carried out on telephone with the interviewee.

9. Methods of data recording
i) **Tabulation** This is recording of data by arranging facts of figures in form of table or list.
ii) **Photographing** This is done by use of a camera to record geographical information.
iii) **Tape recording** This can be done when one is collecting data through an oral interview where one uses tape recording device to record conversation.

iv) Tallying Used when the data is collected through counting. One counts and puts a vertical strike for every item counted, on the fifth count one puts a diagonal crossing the four strokes.

10. Data refers to facts and figures collected from the field.

11. The mean is

72 + 60 + 65 + 70 + 65 + 80 + 65 + 70 + 80 + 84 + 63 + 75 + 63 + 71 + 74

= 1057/12 = 70.47

12. This is got by arranging the data in an ascending order as follows: 60, 63, 63, 65, 65, 65, 70, 70, 71, 72, 74, 75, 80, 80, 8.4. The middle number is the median: 70