La-Net

AGE 103

FARM BIOMETRICS

Download on the App Store

GET IT ON Google Play

Course video is available on our app

lanetonline.com

Course code: AGE 103

Course title: Farm Biometrics

Course credit load: 2 Units

Series 1: Introduction to Farm Biometrics

- **Part 1.1: Definition and Scope of Farm Biometrics**
    - Sub Part 1.1.1: Definition and meaning of biometrics in agriculture
    - Sub Part 1.1.2: Scope and subject matter of farm biometrics
    - Sub Part 1.1.3: Distinction between farm biometrics and related disciplines
- **Part 1.2: Importance of Farm Biometrics**
    - Sub Part 1.2.1: Role in agricultural research and experimentation
    - Sub Part 1.2.2: Role in farm planning and management
    - Sub Part 1.2.3: Role in policy and extension
- **Part 1.3: Basic Statistical Concepts**
    - Sub Part 1.3.1: Data types: qualitative and quantitative
    - Sub Part 1.3.2: Population, sample, and variable
    - Sub Part 1.3.3: Scales of measurement
- **Part 1.4: Data Collection in Agriculture**
    - Sub Part 1.4.1: Primary and secondary data
    - Sub Part 1.4.2: Methods of collecting primary data
    - Sub Part 1.4.3: Sources of error and data quality

Introduction

Farm biometrics is the branch of applied mathematics and statistics that deals with the collection, organisation, analysis, and interpretation of numerical data arising from agricultural activities. It provides the tools that allow farmers, researchers, and policymakers to make sense of the variability inherent in living systems and to draw valid, evidence-based conclusions from experiments, surveys, and farm records.

This Series introduces the meaning, scope, and importance of farm biometrics, and lays the statistical groundwork needed for the rest of the course: data types, statistical terminology, scales of measurement, and data collection methods. These concepts underpin everything that follows in Series 2 through 5.

Series Learning Outcomes

Upon completing this Series, you should be able to:

- **SLO 1.1** define farm biometrics and describe its scope and relationship to related disciplines (Linked to Part 1.1),
- **SLO 1.2** explain the importance of farm biometrics in agricultural research, farm management, and policy (Linked to Part 1.2),
- **SLO 1.3** distinguish between qualitative and quantitative data, and define population, sample, and variable (Linked to Part 1.3),
- **SLO 1.4** identify and describe the four scales of measurement with agricultural examples (Linked to Part 1.3), and
- **SLO 1.5** distinguish between primary and secondary data and describe the main methods of collecting primary agricultural data (Linked to Part 1.4).

1.1 Definition and Scope of Farm Biometrics

1.1.1 Definition and meaning of biometrics in agriculture

The word biometrics comes from the Greek words bios (life) and metron (measure). In agriculture, biometrics refers to the application of statistical and mathematical methods to the measurement, analysis, and interpretation of biological and agricultural data. Farm biometrics is therefore the systematic use of numbers to describe, compare, and draw conclusions about agricultural phenomena: crop yields, animal weights, soil properties, rainfall, economic returns, and any other measurable aspect of farming.

Farm biometrics is not merely arithmetic. It encompasses the design of experiments to generate valid data, the application of appropriate statistical techniques to analyse that data, and the communication of findings in a form that informs decisions. A student who understands farm biometrics can plan a fertiliser trial, measure and record results, calculate means and variability, test whether differences between treatments are real or due to chance, and present conclusions accurately.

Fill in the blank: The word biometrics is derived from the Greek words bios meaning ________ and metron meaning measure; in agriculture it refers to the application of statistical and mathematical methods to the analysis of biological data.

1.1.2 Scope and subject matter of farm biometrics

The scope of farm biometrics covers: (1) descriptive statistics (summarising data using measures of central tendency and dispersion); (2) probability and probability distributions (quantifying uncertainty in biological events); (3) sampling theory (selecting representative samples from populations); (4) experimental design (planning valid agricultural trials); (5) estimation and hypothesis testing (drawing inferences from data); (6) regression and correlation (examining relationships between variables); and (7) basic mathematical tools applied in agriculture (areas, volumes, seed rates, percentages, and graphs). This course focuses on the foundational topics: basic mathematical concepts, measurement, graphical representation, and elementary statistical analysis.

Farm biometrics draws on pure mathematics (algebra, geometry, calculus), statistics (probability, inference), and agronomy to produce a discipline tailored to the needs of agricultural science. It differs from theoretical statistics in that its examples, applications, and motivations are drawn entirely from farming systems: Nigerian crops (maize, cassava, cowpea, rice), livestock (cattle, poultry, goats), soil science, and agricultural economics.

Fill in the blank: The scope of farm biometrics includes descriptive statistics, experimental design, hypothesis testing, regression and correlation, and basic ________ tools such as area calculation, seed rates, percentages, and graph construction applied to agricultural data.

1.1.3 Distinction between farm biometrics and related disciplines

Farm biometrics is related to but distinct from several other disciplines. Statistics is the general science of data; farm biometrics is its application to agriculture. Agronomy studies crop production practices; farm biometrics provides the tools to evaluate those practices objectively. Biomathematics uses mathematics to model biological systems (population dynamics, epidemiology); farm biometrics is more applied and data-centred. Agricultural economics uses economic theory to study farm decision-making; farm biometrics provides the quantitative methods (cost-benefit analysis, regression) that agricultural economists use.

Mathematics provides the underlying language. Without biometrics, a farmer who applies two fertiliser treatments cannot determine whether the observed yield difference is a real treatment effect or simply random variation. Biometrics provides the framework (experimental design, analysis of variance) to answer that question rigorously.

Fill in the blank: Farm biometrics differs from pure statistics in that its examples, data, and applications are drawn entirely from ________ systems; it provides the quantitative tools used by agronomists, animal scientists, and agricultural economists to evaluate farming practices objectively.

Pilot questions 1.1

1. Define farm biometrics in your own words and state the Greek root of the word biometrics.
2. List five topics that fall within the scope of farm biometrics.
3. Distinguish between farm biometrics and (a) pure statistics and (b) agronomy.
4. Explain why a farmer who records yields from two fertiliser treatments needs biometrics to draw valid conclusions.
5. State two mathematical tools from everyday farming that fall within the scope of farm biometrics.

1.2 Importance of Farm Biometrics

1.2.1 Role in agricultural research and experimentation

Agricultural research depends on the ability to design valid experiments and to distinguish real treatment effects from natural biological variability. Farm biometrics provides experimental design principles (randomisation, replication, local control) that ensure unbiased comparisons between treatments. Randomisation prevents systematic bias; replication provides a measure of experimental error; local control (blocking) reduces the effect of known sources of variability. Without these principles, a research result cannot be confidently attributed to the treatment applied.

Statistical hypothesis testing allows researchers to decide whether observed differences between, for example, two varieties of maize are large enough to be real, or whether they fall within the range expected by chance. Regression analysis reveals relationships between variables: how does plant height vary with fertiliser rate? What is the economic optimum fertiliser application rate? These are questions that farm biometrics is designed to answer.

Fill in the blank: The three principles of experimental design provided by farm biometrics are ________, replication, and local control; together they ensure that treatment comparisons in agricultural trials are unbiased and that experimental error is properly measured.

1.2.2 Role in farm planning and management

On the farm, biometrics supports day-to-day and season-to-season planning. Calculating seed rates for a given land area, estimating fertiliser requirements, working out the volume of a storage silo, computing the cost per kilogram of a commodity, and comparing the profitability of two enterprises all require the mathematical and statistical skills taught in this course. A farmer who understands percentages can calculate the germination rate of a seed lot; one who understands area can determine the land available for a particular crop.

Farm records are a core source of biometric data: daily egg production, weekly milk yield, monthly input costs, and seasonal crop yields. Analysing these records using simple descriptive statistics (means, trends, ranges) allows a farmer to identify declining performance, optimise stocking density, and make informed purchase or sale decisions.

Fill in the blank: A farmer who understands farm biometrics can calculate seed rates, estimate input requirements, compute profitability, and analyse farm ________ records to identify trends and make evidence-based management decisions.

1.2.3 Role in policy and extension

Agricultural policymakers need accurate summary statistics (average yields, total production, input use, farm incomes) to design subsidy programmes, set price floors, allocate extension resources, and project food security. Extension officers use simple biometric tools to present trial results to farmers: bar charts, tables of means, percentage improvements. National agricultural surveys (National Agricultural Sample Survey, NASS; National Bureau of Statistics surveys) rely entirely on sampling theory and survey statistics to estimate crop areas, yields, and livestock numbers.

Without reliable biometric data and analysis, agricultural policy is based on guesswork. Nigeria's agriculture sector policies on cassava, rice, and maize depend on crop yield data analysed at national scale. Farm biometrics therefore connects the individual farm level to national food security planning.

Fill in the blank: Agricultural extension officers use simple biometric tools such as bar charts and tables of ________ to present research results to farmers in a form that supports adoption of improved agricultural practices.

Pilot questions 1.2

1. Explain the three principles of experimental design and state why each is necessary.
2. Give two examples of how biometrics is used in day-to-day farm planning.
3. Explain how agricultural policymakers use biometric data and state one Nigerian national agricultural survey.
4. Why can a research result not be trusted if the experiment was not replicated?
5. State two ways in which farm records can be analysed using simple descriptive statistics to improve farm management.

1.3 Basic Statistical Concepts

1.3.1 Data types: qualitative and quantitative

Data are facts, observations, or measurements recorded for analysis. Qualitative (categorical) data describe characteristics that cannot be measured numerically: crop species (maize, cassava, cowpea), sex of an animal (male, female), soil colour (red, brown, black), or farming system (subsistence, commercial). Qualitative data are classified into categories; they are counted, not measured. Quantitative (numerical) data are measurements or counts expressed as numbers: grain yield in kg per ha, milk yield in litres, plant height in cm, number of eggs per week.

Quantitative data are further divided into discrete (can only take whole number values; e.g. number of seeds per pod, number of piglets per litter) and continuous (can take any value within a range; e.g. plant height, body weight, temperature). Recognising data type is important because it determines which statistical technique is appropriate: qualitative data require categorical methods (frequency tables, chi-square test) while quantitative data suit numerical methods (means, standard deviations, t-tests, ANOVA).

Fill in the blank: Data that describe characteristics that cannot be expressed as numbers, such as crop species or soil colour, are called ________ data; data expressed as numerical measurements, such as yield in kg per ha or milk yield in litres, are called quantitative data.

1.3.2 Population, sample, and variable

A statistical population is the complete set of all items or individuals about which information is sought. In agriculture, the population might be all maize farms in Kano State, all cattle in a particular herd, or all plots receiving a given treatment in an experiment. A sample is a subset of the population selected for measurement. Because measuring every member of a large population is usually impractical, samples are taken; provided the sample is representative, conclusions drawn from it can be extended to the entire population (this process is called statistical inference).

A variable is any characteristic of the population members that can take different values. In agriculture: crop yield, plant height, leaf area, animal weight, milk fat content, and input cost are all variables. A constant has the same value for every member of the population (the number of days in a week is a constant, not a variable). Variables may be dependent (the response being measured; e.g. grain yield) or independent (the factor being manipulated; e.g. fertiliser rate).

Fill in the blank: A ________ is the complete set of all individuals or items about which a study seeks information; a sample is a representative subset selected from it for measurement; conclusions from the sample are extended to the population through statistical inference.

1.3.3 Scales of measurement

The scale of measurement determines which arithmetic operations and statistical procedures are valid for a set of data. The four scales are: (1) Nominal scale: data are placed in named categories with no order (e.g. crop type: maize, rice, cassava; sex: male, female); only counting and frequency are valid. (2) Ordinal scale: categories have a meaningful order but the intervals between them are unequal and unknown (e.g. soil fertility rating: poor, moderate, good; disease severity: mild, moderate, severe); ranking is valid but arithmetic is not.

(3) Interval scale: data are measured on a scale with equal intervals between values but with no true zero point (e.g. temperature in degrees Celsius; pH); addition and subtraction are valid but not multiplication. (4) Ratio scale: equal intervals and a true zero point (e.g. weight in kg, height in cm, yield in kg per ha, number of animals); all arithmetic operations are valid. Most agricultural measurement data are on the ratio scale; ordinal data frequently arise in farmer surveys and disease ratings.

Fill in the blank: The ________ scale of measurement has equal intervals between values and a true zero point, allowing all arithmetic operations; it is the scale used for most agricultural measurements such as weight, height, and yield.

Scale	Core Characteristics	Agricultural Examples	Valid Statistical Operations
Nominal	• Categorical & qualitative data. • Numbers/labels serve only as IDs. • No inherent order or ranking.	• Crop Varieties: Maize, Cassava, Rice. • Livestock Breeds: White Fulani, Muturu. • Soil Types: Sandy, Loamy, Clayey.	• Frequency counts & percentages • Mode • Chi-Square (χ^2) test • Cramer's V
Ordinal	• Clear logical hierarchy/progression. • Intervals between ranks are unequal/unknown. • Indicates relative direction only.	• Soil Fertility: Low, Medium, High. • Livestock Body Condition Score (1 to 5). • Farmer Status: Smallholder, Commercial.	• Median & Percentiles • Rank order statistics • Spearman's rank correlation (ρ) • Non-parametric tests
Interval	• Quantitative with uniform intervals. • Arbitrary zero point (not absolute absence). • Addition/subtraction are meaningful.	• Temperature: Celsius (°C) / Fahrenheit. • Planting Dates: Days relative to baseline. • Soil pH Level (logarithmic scale).	• Arithmetic Mean • Standard Deviation & Variance • Pearson correlation (r) • t-test, ANOVA, F-test
Ratio	• Quantitative with uniform intervals. • Absolute true zero point (total absence). • Multiplicative comparisons are valid.	• Crop Yield: Kilograms per hectare (kg/ha). • Livestock Live Weight (kg). • Farm Land Area: Hectares or acres.	• Geometric & Harmonic Mean • Coefficient of Variation (CV) • Regression & econometric modeling • All standard parametric tests

Figure 1.1: Summary of the four scales of measurement with agricultural examples

Pilot questions 1.3

1. Define data and distinguish between qualitative and quantitative data with one agricultural example of each.
2. Distinguish between discrete and continuous quantitative data with one example of each from Nigerian agriculture.
3. Explain the difference between a population and a sample and state why samples are used in agricultural research.
4. Define a variable and distinguish between dependent and independent variables in an agricultural experiment.
5. Name and describe the four scales of measurement and give one agricultural example for each.

1.4 Data Collection in Agriculture

1.4.1 Primary and secondary data

Primary data are data collected directly and freshly by the researcher for the specific purpose of the current study. They are original, first-hand observations: weighing crops at harvest, measuring plant heights, recording daily milk yields, conducting farm surveys with questionnaires. Primary data are more reliable because they are tailored to the study's needs, but they require time, resources, and planning to collect.

Secondary data are data that have already been collected by another person or organisation for a different purpose and are now being used for the current study. Examples include: census records, government agricultural statistics (FMARD annual surveys, NBS data), published research results, weather station records, and market price databases. Secondary data are cheaper and quicker to obtain but may not perfectly match the current study's requirements in terms of time period, geographic coverage, or variable definition.

Fill in the blank: Data collected directly and freshly by the researcher for the current study are called ________ data; data that were previously collected by another person or organisation for a different purpose and are being reused are called secondary data.

1.4.2 Methods of collecting primary data

The main methods of collecting primary agricultural data are: (1) Direct observation: the researcher physically measures or observes the variable (e.g. counting insects on a plant, recording germination, scoring disease symptoms); most accurate but labour-intensive. (2) Interview: the researcher asks questions directly (face-to-face or by telephone) to a respondent; useful for socioeconomic data; allows clarification of ambiguous answers. (3) Questionnaire: a structured set of written questions distributed to respondents to complete; efficient for large samples; requires literate respondents; answers cannot be clarified. (4) Measurement with instruments: use of standard instruments (weighing scale, ruler, thermometer, hygrometer, rain gauge, soil moisture meter) to obtain precise quantitative data.

(5) Farm records: systematic daily, weekly, or seasonal recording of farm activities and production (egg records, milk records, feed consumption, input costs, sale receipts); an important source of ongoing primary data. (6) Experimental observation: collecting data from a properly designed experiment in which treatments are applied under controlled conditions; most rigorous for establishing cause-and-effect relationships. The choice of data collection method depends on the nature of the variable, the size of the study population, available resources, and the required level of accuracy.

Fill in the blank: The data collection method in which a researcher physically measures or observes each unit of study (e.g. counting insect pests on a plant or scoring disease symptoms) is called direct ________, and it is the most accurate but also the most labour-intensive primary data collection method.

1.4.3 Sources of error and data quality

Data quality depends on accuracy (how close a measurement is to the true value), precision (how repeatable a measurement is), and completeness (whether all required observations are recorded). Sources of error in agricultural data collection include: measurement error (instrument error; reading error; environmental variation during measurement), sampling error (difference between sample estimate and true population value due to chance selection of an unrepresentative sample), non-sampling error (mistakes in recording, transcription, or coding; respondent bias in interviews; non-response bias in questionnaires), and observer bias (systematic differences between observers recording the same data).

Good data quality requires: calibrated and standardised instruments; trained and consistent observers; clear, unambiguous data recording forms; proper sampling procedures; and data checking (range checks, logical consistency checks, double entry). In Nigerian agricultural research, a common source of error is failure to calibrate weighing scales before use; this leads to systematic bias in yield data that cannot be corrected in analysis.

Fill in the blank: The difference between a sample estimate and the true population value that arises because the sample happened by chance to be unrepresentative of the population is called ________ error; it can be reduced by increasing the sample size and using proper probability sampling methods.

Aspect	Primary Data	Secondary Data
Definition	Data collected firsthand by the researcher directly from the source	Data previously collected and compiled by others, available through publications or databases
Examples	Field surveys, crop yield measurements, soil sampling, farmer interviews	Government reports, agricultural census, journal articles, FAO statistics
Advantages	High accuracy and relevance to specific research objectives; tailored to current study	Time-saving, cost-effective, readily available, allows for broad comparisons

Disadvantages	Expensive, time-consuming, requires significant effort and resources	May be outdated, less specific, potential bias depending on source

Box 1.1: Comparison of primary and secondary data in agricultural research

Pilot questions 1.4

1. Define primary data and secondary data and give two agricultural examples of each.
2. Describe four methods of collecting primary data in agricultural research.
3. Explain the difference between sampling error and non-sampling error in agricultural data collection.
4. State three ways to improve the quality of data collected during a farm survey.
5. Explain why farm records are an important source of primary data and give two examples of records kept in Nigerian farming.

Series Summary

Farm biometrics is the application of statistical and mathematical methods to agricultural data. Its scope includes descriptive statistics, sampling, experimental design, hypothesis testing, regression, and basic mathematical tools (areas, volumes, seed rates, percentages). It differs from pure statistics by focusing entirely on agricultural applications. It supports agricultural research (experimental design; hypothesis testing), farm management (seed rates; area; profitability), and policy (national surveys; extension communication).

Data may be qualitative (categorical; e.g. crop species) or quantitative (numerical; discrete or continuous). A population is the complete set of interest; a sample is a representative subset. A variable is a characteristic that takes different values; variables are dependent (response) or independent (factor). The four measurement scales are nominal (categories), ordinal (ordered categories), interval (equal intervals, no true zero), and ratio (equal intervals, true zero; most agricultural data). Primary data are collected fresh for the current study; secondary data were collected previously. Primary data methods include direct observation, interview, questionnaire, instrument measurement, farm records, and experimental observation. Data quality depends on accuracy, precision, and completeness; errors arise from measurement, sampling, non-sampling, and observer bias.

Glossary of Terms

- **Biometrics:** the application of statistical and mathematical methods to the measurement and analysis of biological data; in agriculture, it covers all quantitative aspects of farming from yield analysis to experimental design.
- **Continuous variable:** a quantitative variable that can take any value within a range, including fractions; examples include plant height, body weight, and temperature.
- **Data:** facts, observations, or measurements recorded for analysis; they may be qualitative (categorical) or quantitative (numerical).

- **Dependent variable:** the response variable measured in an experiment; its values change in response to changes in the independent variable (e.g. grain yield in response to fertiliser rate).
- **Discrete variable:** a quantitative variable that can only take whole number (integer) values; examples include number of seeds per pod, piglets per litter, and eggs per week.
- **Experimental design:** the planning of an experiment to ensure valid, unbiased results; the three basic principles are randomisation, replication, and local control (blocking).
- **Independent variable:** the factor manipulated or controlled by the researcher in an experiment; changes in the independent variable are expected to cause changes in the dependent variable (e.g. fertiliser rate).
- **Interval scale:** a measurement scale with equal intervals between values but no true zero point; temperature in degrees Celsius is an example; addition and subtraction are valid but not multiplication.
- **Nominal scale:** the lowest level of measurement; data are placed in named categories with no order or rank; examples include crop species and sex of animal.
- **Ordinal scale:** a measurement scale in which categories have a meaningful order but intervals between them are unequal or unknown; examples include disease severity ratings and soil fertility classes.
- **Population:** the complete set of all individuals, items, or observations about which information is sought in a statistical study.
- **Primary data:** data collected directly and freshly by the researcher for the specific purpose of the current study; more reliable but more expensive than secondary data.
- **Qualitative data:** data that describe characteristics in non-numerical categories; examples include crop type, sex, and soil colour; analysis uses frequency tables and chi-square tests.
- **Quantitative data:** data expressed as numerical measurements or counts; may be discrete (whole numbers) or continuous (any value in a range); suited to arithmetic statistical methods.

- **Ratio scale:** the highest level of measurement with equal intervals and a true zero point; all arithmetic operations are valid; most agricultural measurements (yield, weight, height) are on the ratio scale.
- **Sample:** a subset of the population selected for measurement; used to make inferences about the population when measuring all population members is impractical.
- **Sampling error:** the difference between a sample estimate and the true population value that arises because the selected sample is not a perfect representation of the population.
- **Secondary data:** data collected previously by another person or organisation for a different purpose and reused for the current study; cheaper but may not exactly match the study's requirements.
- **Variable:** any characteristic of population members that can take different values; may be qualitative or quantitative, and dependent or independent.

Concept Check Questions [Series 1]

Objective Questions

1. Farm biometrics is derived from the Greek words bios (life) and ________ (measure). (Linked to SLO 1.1)
2. Data that describe characteristics in non-numerical categories (e.g. crop species, sex) are called ________ data. (Linked to SLO 1.3)
3. A ________ is the complete set of all individuals about which a study seeks information; a sample is a representative subset. (Linked to SLO 1.3)
4. The ________ scale of measurement has equal intervals and a true zero point, making all arithmetic operations valid. (Linked to SLO 1.4)
5. Data collected directly and freshly by the researcher for the current study are called ________ data. (Linked to SLO 1.5)

Short-Answer Questions

6. Define farm biometrics and state three topics within its scope. (Linked to SLO 1.1)
7. Distinguish between qualitative and quantitative data, giving one agricultural example of each. (Linked to SLO 1.3)
8. Name and briefly describe the four scales of measurement with one example each. (Linked to SLO 1.4)
9. Distinguish between primary and secondary data and give one advantage of each. (Linked to SLO 1.5)
10. State three principles of experimental design and explain why each is necessary. (Linked to SLO 1.2)

Long-Answer Questions

11. Define farm biometrics, describe its scope, and explain its importance in agricultural research, farm management, and policy. (Linked to SLOs 1.1, 1.2)
12. Describe the main statistical concepts needed in farm biometrics: data types, population, sample, variable, and scales of measurement. Give agricultural examples throughout. (Linked to SLOs 1.3, 1.4)

13. Describe the methods of primary data collection in agriculture, distinguish primary from secondary data, and explain the sources of error that affect data quality. (Linked to SLO 1.5)

Answers to Concept Check Questions [Series 1]

Objective Questions

1. Metron.
2. Qualitative.
3. Population.
4. Ratio.
5. Primary.

Short-Answer Questions

6. Farm biometrics is the application of statistical and mathematical methods to agricultural data. Three topics within its scope: descriptive statistics; experimental design; regression and correlation analysis.
7. Qualitative data describe non-numerical characteristics (e.g. crop species: maize, cassava). Quantitative data are numerical measurements or counts (e.g. grain yield: 2,500 kg per ha).
8. Nominal: named categories, no order (crop species). Ordinal: ordered categories, unequal intervals (disease severity: mild, moderate, severe). Interval: equal intervals, no true zero (temperature in degrees Celsius). Ratio: equal intervals, true zero (yield in kg per ha).
9. Primary data are collected fresh by the researcher for the current study; advantage: tailored to the study. Secondary data were collected previously for another purpose; advantage: cheaper and quicker to obtain.
10. Randomisation: prevents systematic bias in treatment allocation. Replication: provides a measure of experimental error; needed to detect real treatment differences. Local control (blocking): reduces variability due to known factors such as soil fertility gradients.

Long-Answer Questions

11. Farm biometrics is the application of statistical and mathematical methods to agricultural data. Scope: descriptive statistics; probability; sampling; experimental design; hypothesis testing; regression; and basic mathematical tools (areas, volumes, seed rates, percentages). Importance in research: experimental design principles (randomisation, replication, local control) ensure valid comparisons; hypothesis testing determines

whether differences are real. Farm management: biometrics enables calculation of seed rates, fertiliser quantities, storage volumes, and cost-benefit analysis; farm records analysed with simple statistics improve management decisions. Policy: national agricultural surveys use sampling theory to estimate crop areas and yields; extension officers use bar charts and tables to communicate results to farmers; policymakers need accurate summary statistics for subsidy and resource allocation decisions.

12. Data are observations or measurements recorded for analysis. Qualitative data describe non-numerical categories (crop species, sex, soil colour); quantitative data are numerical (yield, weight, height) and divided into discrete (whole numbers: number of seeds per pod) and continuous (any value: plant height). Population: complete set of individuals about which information is sought (e.g. all maize farms in Kano State). Sample: representative subset measured to make inferences about the population. Variable: characteristic that takes different values; may be dependent (response; e.g. yield) or independent (factor; e.g. fertiliser rate). Four measurement scales: nominal (crop species; counting only); ordinal (disease rating; ranking); interval (temperature; addition and subtraction); ratio (yield; all arithmetic). Most agricultural measurements are on the ratio scale.
13. Primary data are collected fresh by the researcher for the current study. Methods: (1) direct observation (physically measuring or observing; most accurate); (2) interview (face-to-face questions; allows clarification); (3) questionnaire (written questions; efficient for large samples); (4) instrument measurement (scales, rulers, thermometers; precise quantitative data); (5) farm records (ongoing daily or seasonal records); (6) experimental observation (data from controlled field trials). Secondary data were previously collected by another party (FMARD statistics, NBS census, published papers); cheaper but may not match current needs. Sources of error: measurement error (instrument or reading error); sampling error (unrepresentative sample by chance; reduced by larger sample size and probability sampling); non-sampling error (recording mistakes, respondent bias, non-response); observer bias (systematic differences between observers). Data quality improved by calibrated instruments, trained observers, clear recording forms, and data checking.

Answers to Pilot Questions [Series 1]

Part 1.1

1. Farm biometrics is the application of statistical and mathematical methods to the measurement and analysis of biological data arising from agricultural activities. The root bios means life and metron means measure.
2. Five topics within the scope of farm biometrics: descriptive statistics; experimental design; sampling theory; regression and correlation; basic mathematical tools (areas, volumes, percentages, seed rates).
3. Farm biometrics vs pure statistics: farm biometrics applies statistical methods specifically to agricultural data and problems; pure statistics is a general mathematical science with applications across all fields. Farm biometrics vs agronomy: agronomy studies crop production practices and management; farm biometrics provides the quantitative tools used to evaluate those practices objectively.
4. Without biometrics, a farmer cannot determine whether a yield difference between two fertiliser treatments reflects a real treatment effect or is simply due to natural variability among plots. Biometrics provides experimental design and statistical testing to make this distinction.
5. Two mathematical tools from everyday farming within farm biometrics: calculating seed rate (seeds per hectare based on spacing and germination percentage); calculating area of a farm or field for input planning.

Part 1.2

1. Randomisation: treatments are allocated to experimental units at random to prevent systematic bias. Replication: each treatment is applied to more than one unit to measure experimental error and detect real differences. Local control (blocking): experimental units are grouped into blocks of similar units to reduce variability from known sources such as soil fertility differences.
2. Two examples of biometrics in day-to-day farm planning: (1) calculating the seed rate for planting a given area of maize at a defined spacing and germination percentage; (2) calculating the volume of a storage silo to determine whether it can hold an expected harvest.

3. Policymakers use biometric summary statistics (average yields, total production, farm incomes) to design subsidies, allocate extension resources, and project food security. One Nigerian national agricultural survey: the National Agricultural Sample Survey (NASS) conducted by the National Bureau of Statistics (NBS) in collaboration with FMARD.
4. A single unreplicated observation cannot distinguish between a real treatment effect and natural plant-to-plant or plot-to-plot variation. Without replication, experimental error cannot be estimated; therefore no valid statistical test can be performed and the result is inconclusive.
5. Two ways farm records can be analysed: (1) calculating the monthly mean egg production to detect a decline that may indicate disease or nutritional deficiency; (2) tracking monthly feed cost trends to identify when feed conversion efficiency is falling.

Part 1.3

1. Data are facts, observations, or measurements recorded for analysis. Qualitative data: non-numerical descriptions of characteristics, e.g. crop species (maize, cowpea). Quantitative data: numerical measurements or counts, e.g. grain yield (2,500 kg per ha).
2. Discrete quantitative data: can only take whole number values; example: number of eggs laid per week per hen. Continuous quantitative data: can take any value within a range; example: live body weight of a broiler chicken at six weeks of age.
3. A population is the complete set of all individuals about which information is sought (e.g. all cowpea farms in Borno State). A sample is a subset of that population selected for measurement. Samples are used because measuring every member of a large population is impractical in terms of time, cost, and logistics.
4. A variable is any characteristic of population members that can take different values. A dependent variable is the response being measured (e.g. grain yield). An independent variable is the factor being controlled or manipulated by the researcher (e.g. fertiliser application rate).
5. Nominal: named categories, no order; example: type of farming system (subsistence, commercial). Ordinal: ordered categories, unequal intervals; example: soil fertility rating (poor, moderate, good). Interval: equal intervals, no true zero; example: daily maximum

temperature in degrees Celsius. Ratio: equal intervals, true zero; example: maize grain yield in kg per ha.

Part 1.4

1. Primary data: collected fresh by the researcher for the current study. Examples: (1) crop yield measurements at harvest from experimental plots; (2) body weight of animals recorded weekly at a farm. Secondary data: previously collected by another party for a different purpose. Examples: (1) FMARD national crop production statistics; (2) weather station rainfall records.
2. Four methods: (1) Direct observation: physically measuring or counting each unit (e.g. counting aphids per leaf; recording germination). (2) Interview: face-to-face questioning of a farmer respondent to obtain socioeconomic data. (3) Questionnaire: structured written questions given to a sample of farmers. (4) Instrument measurement: using calibrated scales, rulers, or thermometers to obtain precise numerical data.
3. Sampling error: the difference between the sample estimate and the true population value arising because the sample was not a perfect representation; reduced by increasing sample size. Non-sampling error: mistakes in recording, transcription, or coding; respondent bias; non-response; observer bias; not reducible by increasing sample size.
4. Three ways to improve data quality in a farm survey: (1) calibrate all measuring instruments before use; (2) train enumerators to apply the same recording protocol consistently; (3) use range checks and logical consistency checks to identify and correct recording errors before data analysis.
5. Farm records provide an ongoing source of primary data at no additional collection cost. Examples of records kept in Nigerian farming: (1) daily egg production records in a poultry farm; (2) monthly milk yield records in a dairy herd.

References and Attributions [Series 1]

Textual References (Books and External Sources)

- Gomez, K. A., & Gomez, A. A. (1984). Statistical procedures for agricultural research (2nd ed.). Wiley-Interscience.
- Steel, R. G. D., & Torrie, J. H. (1980). Principles and procedures of statistics: A biometrical approach (2nd ed.). McGraw-Hill.
- Cochran, W. G. (1977). Sampling techniques (3rd ed.). Wiley.
- Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods (8th ed.). Iowa State University Press.

Figures, Plates, Tables, and Boxes

- Figure 1.1: Summary of the four scales of measurement with agricultural examples. Attribution: AI generated. Suggested OER search string: “scales of measurement nominal ordinal interval ratio agricultural examples table OER”.
- Box 1.1: Comparison of primary and secondary data in agricultural research. Attribution: AI generated. Suggested OER search string: “primary secondary data agriculture comparison definition advantages disadvantages table OER”.

Series 2: Mathematical Concepts in Agriculture

- **Part 2.1: Number Systems and Basic Arithmetic**
  - Sub Part 2.1.1: Whole numbers, fractions, and decimals in agriculture
  - Sub Part 2.1.2: Ratios and proportions
  - Sub Part 2.1.3: Percentages and their agricultural applications
- **Part 2.2: Algebraic Concepts in Agriculture**
  - Sub Part 2.2.1: Basic algebra: expressions, equations, and formulae
  - Sub Part 2.2.2: Simultaneous equations and their agricultural use
  - Sub Part 2.2.3: Word problems and problem-solving in agriculture
- **Part 2.3: Seed Rate, Spacing, and Plant Population**
  - Sub Part 2.3.1: Calculating seed rate from spacing
  - Sub Part 2.3.2: Adjusting seed rate for germination percentage
  - Sub Part 2.3.3: Worked examples with Nigerian crops
- **Part 2.4: Averages and Elementary Statistics**
  - Sub Part 2.4.1: Mean, median, and mode
  - Sub Part 2.4.2: Range and variance as measures of spread
  - Sub Part 2.4.3: Worked examples from farm data

Introduction

Mathematical concepts form the practical foundation of farm biometrics. A farmer or agricultural officer who cannot calculate seed rates, convert units, or work out percentages cannot apply biometric knowledge effectively, regardless of theoretical understanding. This Series builds mathematical competence in the areas most relevant to Nigerian agriculture: number operations, ratios, percentages, simple algebra, simultaneous equations, seed rate calculations, and basic descriptive statistics.

The emphasis throughout is on worked examples drawn from real Nigerian farming situations: maize spacing in the Guinea savanna, cowpea seed rates in the Sudan savanna, germination tests, fertiliser cost comparisons, and yield data from smallholder farms. Each sub-part introduces a concept, shows how it is used in practice, and provides exercises for self-assessment.

Series Learning Outcomes

Upon completing this Series, you should be able to:

- **SLO 2.1** perform arithmetic operations with whole numbers, fractions, decimals, ratios, and percentages in agricultural contexts (Linked to Part 2.1),
- **SLO 2.2** set up and solve algebraic equations and simultaneous equations to solve agricultural word problems (Linked to Part 2.2),
- **SLO 2.3** calculate seed rate, plant population, and adjusted seed rate for germination percentage for Nigerian crops (Linked to Part 2.3),
- **SLO 2.4** calculate and interpret the mean, median, mode, range, and variance of a set of farm data (Linked to Part 2.4), and
- **SLO 2.5** apply mathematical concepts to solve practical agricultural problems with accuracy and confidence (Linked to Parts 2.1 to 2.4).

2.1 Number Systems and Basic Arithmetic

2.1.1 Whole numbers, fractions, and decimals in agriculture

Whole numbers (integers) arise in agriculture when counting discrete items: number of animals in a herd, bags of fertiliser applied, seeds per hole, rows per plot. Fractions arise when a quantity is divided into parts: one-half hectare, three-quarter bag of seed, two-thirds of the recommended fertiliser dose. To add or subtract fractions, express them over a common denominator; to multiply fractions, multiply numerators and denominators separately; to divide fractions, multiply by the reciprocal of the divisor.

Decimals are fractions expressed in base-10 notation and are used whenever a measuring instrument gives a reading (2.35 kg, 1.80 m, 0.75 ha). To convert a fraction to a decimal, divide numerator by denominator (3/4 = 0.75). To convert a decimal to a fraction, write the decimal over its place-value denominator and simplify (0.25 = 25/100 = 1/4). Agricultural calculations routinely combine all three forms: a farmer may use 0.5 ha (decimal), plant 3 seeds per hole (whole number), and apply 1/2 bag of urea (fraction).

Fill in the blank: In agricultural calculations, ________ are fractions expressed in base-10 notation; to convert the fraction 3/4 to decimal form, divide the numerator by the denominator to obtain 0.75.

2.1.2 Ratios and proportions

A ratio expresses the relative size of two or more quantities of the same kind. If a farmer mixes 2 parts NPK fertiliser with 1 part urea, the ratio is 2:1. Ratios can be simplified by dividing all terms by their highest common factor (HCF): 6:4 simplifies to 3:2. A proportion states that two ratios are equal (a:b = c:d, or a/b = c/d) and is used to scale quantities up or down. If 5 kg of seed plants 0.5 ha, then to find the seed needed for 2 ha: 5/0.5 = x/2, so x = 20 kg.

Mixing ratios appear in Nigerian agriculture in fertiliser blending, pesticide dilution, and feed formulation. A common poultry feed formulation might use maize and groundnut cake in a 3:1

ratio; if 60 kg of feed is needed, maize = (3/4) x 60 = 45 kg and groundnut cake = (1/4) x 60 = 15 kg. Direct proportion means as one quantity increases, the other increases proportionally; inverse proportion means as one increases, the other decreases (e.g. more workers, fewer days to complete weeding).

Fill in the blank: If 5 kg of maize seed plants 0.5 ha, the amount needed to plant 2 ha is found by direct proportion: 5/0.5 = x/2, giving x = ________ kg.

2.1.3 Percentages and their agricultural applications

A percentage expresses a quantity as a fraction of 100. To convert a fraction to a percentage: multiply by 100 (3/4 x 100 = 75%). To find a percentage of a quantity: multiply the quantity by the percentage divided by 100 (15% of 200 kg = 200 x 15/100 = 30 kg). To find what percentage one quantity is of another: (part/whole) x 100.

Agricultural uses of percentages include: germination percentage (number of seeds germinated/total seeds tested x 100); moisture content of grain (mass of water/total mass x 100); percentage disease incidence (number of diseased plants/total plants x 100); percentage profit or loss ((profit or loss/cost price) x 100); and fertiliser nutrient content (15:15:15 NPK fertiliser contains 15% N, 15% P_2O_5, and 15% K_2O by mass). These calculations arise daily in farm management and agricultural research.

Fill in the blank: Germination percentage is calculated as (number of seeds germinated / total seeds tested) x ________; if 92 seeds out of 100 germinate, the germination percentage is 92 percent.

Pilot questions 2.1

1. A farmer has 3/4 ha of land. If each hectare requires 12 kg of cowpea seed, how many kg of seed are needed?
2. A fertiliser ratio of NPK to urea is 3:2. If a farmer needs 50 kg of the mixture, how many kg of each is required?

3. Convert the following to percentages: (a) 17/20; (b) 0.045; (c) 7 out of 25.
4. A farmer sells a bag of maize for N35,000 that cost N28,000 to produce. Calculate the percentage profit.
5. If 88 seeds out of 100 germinate in a test, what is the germination percentage?

2.2 Algebraic Concepts in Agriculture

2.2.1 Basic algebra: expressions, equations, and formulae

Algebra uses letters (variables) to represent unknown or changing quantities, allowing general relationships to be stated as formulae. An algebraic expression is a combination of numbers and variables (e.g. 3x + 5); an equation states that two expressions are equal (3x + 5 = 20); a formula is an equation expressing a general rule (area = length x width; A = L x W). To solve a linear equation, isolate the unknown by performing the same operation on both sides: 3x + 5 = 20 gives 3x = 15, so x = 5.

Agricultural formulae express relationships such as: Profit = Revenue - Cost (P = R - C); Net farm income = Gross revenue - Total production cost; Seed rate = (Plant population x seed weight) / (germination % x 100); Area of a rectangle = length x width. Using algebra, if a farmer knows any two of the three variables in P = R - C, the third can be calculated. Formulae allow consistent, repeatable calculations without re-deriving the logic each time.

Fill in the blank: To solve the equation 4x - 8 = 12, add 8 to both sides to get 4x = 20, then divide both sides by 4 to obtain x = ________.

2.2.2 Simultaneous equations and their agricultural use

Simultaneous equations are two or more equations with the same two unknowns, solved together. Two methods are substitution (express one variable in terms of the other from one equation, substitute into the second) and elimination (add or subtract the equations to eliminate one variable). Agricultural example: a farmer sells goats at N25,000 each and sheep at N18,000 each; he sells a total of 10 animals and receives N211,000. Let g = goats sold, s = sheep sold. Then: $g + s = 10$ (i) and $25{,}000g + 18{,}000s = 211{,}000$ (ii).

From equation (i):

$g + s = 10$

Therefore:

$g = 10 - s$

Substitute $g = 10 - s$ into equation (ii):

$25{,}000g + 18{,}000s = 211{,}000$

$25{,}000(10 - s) + 18{,}000s = 211{,}000$

$250{,}000 - 25{,}000s + 18{,}000s = 211{,}000$

$250{,}000 - 7{,}000s = 211{,}000$

$-7{,}000s = -39{,}000$

$s = 39{,}000 / 7{,}000$

$s \approx 5.57$

Since the answer is not a whole number, the figures do not produce an exact integer solution.

For a cleaner example with whole numbers:

$g + s = 8$

$3g + 2s = 20$

From the first equation:

$g = 8 - s$

Substitute into the second equation:

$3(8 - s) + 2s = 20$

$24 - 3s + 2s = 20$

$24 - s = 20$

$-s = -4$

$s = 4$

Therefore:

$g = 8 - 4 = 4$

Hence, $g = 4$ and $s = 4$.

Fill in the blank: To solve simultaneous equations by ________, express one variable in terms of the other from one equation, then substitute that expression into the second equation to produce a single equation in one unknown.

2.2.3 Word problems and problem-solving in agriculture

Solving agricultural word problems requires translating the verbal description into mathematical statements. The steps are: (1) identify the unknown(s) and assign variable names; (2) identify the given information; (3) write the equation(s) relating unknowns to given information; (4) solve the equation(s); (5) check the answer against the original problem conditions; (6) state the answer with units. Example: a maize plot produces 250 kg more than a sorghum plot; together they produce 1,150 kg. Find each yield. Let sorghum yield = x kg; maize yield = x + 250 kg. Then x + (x + 250) = 1,150; 2x = 900; x = 450 kg (sorghum); maize = 700 kg.

Common error in agricultural word problems: neglecting units. Always include kg, ha, naira, or other units in every step; this helps identify when an operation is inappropriate (one cannot add kg to ha without a conversion factor). A systematic approach prevents mistakes and shows workings clearly for examination marking.

Fill in the blank: In solving agricultural word problems, the first step is to identify the ________ and assign a letter (variable) to represent it; subsequent steps translate given information into equations, solve, and check the answer against original conditions.

Pilot questions 2.2

1. A farm profit is calculated as P = R - C. If a farmer's revenue is N480,000 and total cost is N315,000, find the profit.
2. Solve: 5y - 7 = 28.
3. A poultry farmer sells broilers at N4,500 each and layers at N3,200 each. He sells 12 birds in total and receives N48,900. How many of each type did he sell?
4. A cassava plot produces 150 kg more than a yam plot; together they produce 2,350 kg. Find the yield of each crop.
5. The cost of 3 bags of urea and 2 bags of NPK is N87,000. The cost of 1 bag of urea and 1 bag of NPK is N33,000. Find the cost of each bag.

2.3 Seed Rate, Spacing, and Plant Population

2.3.1 Calculating seed rate from spacing

Plant population per hectare is determined by the spacing between and within rows. The formula is: Plant population per ha = 10,000 / (inter-row spacing in m x intra-row spacing in m). The value 10,000 is used because 1 ha = 10,000 m2. For example, maize planted at 75 cm x 25 cm with 1 seed per hole: population = 10,000 / (0.75 x 0.25) = 10,000 / 0.1875 = 53,333 plants per ha. If 2 seeds are planted per hole and later thinned to 1, the same formula applies after thinning.

Seed rate (kg per ha) is calculated from plant population and individual seed weight: Seed rate = (plant population x seed weight in g) / 1,000. The division by 1,000 converts grams to kilograms. If seed weight = 0.25 g per seed and population = 53,333 plants per ha: Seed rate = (53,333 x 0.25) / 1,000 = 13,333 / 1,000 = 13.3 kg per ha. For crops sown in rows without defined spacing (broadcast sowing), seed rate is determined empirically from field trials.

Fill in the blank: Plant population per hectare = 10,000 / (inter-row spacing in m x ________ spacing in m); the value 10,000 is used because 1 hectare equals 10,000 square metres.

2.3.2 Adjusting seed rate for germination percentage

Seeds with less than 100 percent germination require a higher seed rate to achieve the target plant stand. The adjusted seed rate formula is: Adjusted seed rate = (Required seed rate / Germination percentage) x 100. For example, if the base seed rate for cowpea is 20 kg per ha and germination is 80%: Adjusted rate = (20 / 80) x 100 = 25 kg per ha. This ensures that even if 20% of seeds fail to germinate, the target plant population is achieved.

A further adjustment is sometimes made for seed purity (the fraction of the seed lot that is true seed, not chaff or inert matter): Adjusted seed rate = (Base seed rate / (Germination% x Purity%)) x 10,000. For germination = 85% and purity = 95%: factor = 85 x 95 = 8,075; adjusted rate = (base rate / 8,075) x 10,000. Purity adjustments are important when using farm-saved seed that may contain significant proportions of inert material.

Fill in the blank: Adjusted seed rate = (required seed rate / germination percentage) x ________; if the base seed rate is 15 kg per ha and germination is 75%, the adjusted seed rate is 20 kg per ha.

2.3.3 Worked examples with Nigerian crops

Example 1: Maize (Zea mays). Spacing: 75 cm x 25 cm; 1 seed per hole; seed weight = 0.28 g; germination = 90%. Plant population = 10,000 / (0.75 x 0.25) = 53,333 plants per ha. Base seed rate = (53,333 x 0.28) / 1,000 = 14.9 kg per ha. Adjusted seed rate = (14.9 / 90) x 100 = 16.6 kg per ha. A farmer planting 2 ha needs: 16.6 x 2 = 33.2 kg of maize seed.

Example 2: Cowpea (Vigna unguiculata). Spacing: 60 cm x 20 cm; 2 seeds per hole; seed weight = 0.18 g; germination = 80%. Plant population (before thinning) = 10,000 / (0.60 x 0.20) x 2 = 10,000 / 0.12 x 2 = 166,667 seeds required per ha. Base seed rate = (166,667 x 0.18) / 1,000 = 30.0 kg per ha. Adjusted seed rate = (30.0 / 80) x 100 = 37.5 kg per ha. Example 3: Rice (Oryza sativa), broadcast sown at 80 kg per ha; germination 85%: Adjusted rate = (80 / 85) x 100 = 94.1 kg per ha.

Fill in the blank: A farmer planting 3 ha of cowpea requires an adjusted seed rate of 37.5 kg per ha; the total seed required is 37.5 x 3 = ________ kg.

Crop	Spacing	Plant Population/ha	Base Seed Rate	Adjusted Seed Rate
Maize	75 cm × 25 cm (1 plant/hill)	~53,000 plants/ha	20–25 kg/ha	22–28 kg/ha (for 85–90% germination)
Cowpea	60 cm × 20 cm (2 seeds/hill)	~166,000 plants/ha	25–30 kg/ha	28–35 kg/ha (for 80–85% germination)
Rice	20 cm × 20 cm (transplanting)	~250,000 plants/ha	60–80 kg/ha (direct seeding)	70–90 kg/ha (for 80–85% germination)

Table 2.1: Worked examples of seed rate calculations for Nigerian crops

Pilot questions 2.3

1. Calculate the plant population per hectare for sorghum planted at 75 cm x 30 cm with 1 seed per hole.
2. If sorghum seed weight is 0.03 g per seed and plant population is 44,444 per ha, calculate the base seed rate in kg per ha.
3. Adjust the sorghum seed rate from question 2 for a germination percentage of 75%.

4. A farmer has a 0.5 ha plot. Using the adjusted sorghum seed rate calculated in question 3, how many kg of seed does he need?
5. Rice is broadcast-sown at 100 kg per ha. If germination is 80% and seed purity is 90%, calculate the adjusted seed rate.

2.4 Averages and Elementary Statistics

2.4.1 Mean, median, and mode

The mean (arithmetic average) is the sum of all values divided by the number of values: mean = (sum of values) / n. It is the most widely used measure of central tendency and is appropriate for quantitative data with no extreme outliers. Example: maize yields (kg per ha) from five plots: 2,100; 2,400; 1,900; 2,200; 2,300. Mean = (2,100 + 2,400 + 1,900 + 2,200 + 2,300) / 5 = 10,900 / 5 = 2,180 kg per ha.

The median is the middle value when data are arranged in ascending order. For an odd number of values (n), the median is the ((n+1)/2)th value. For an even number, it is the mean of the (n/2)th and (n/2 + 1)th values. Median of the above yields: arrange in order: 1,900; 2,100; 2,200; 2,300; 2,400; median = 2,200 kg per ha (3rd value of 5). The mode is the value that appears most frequently. If most common yield = 2,200, mode = 2,200. Data may be bimodal (two modes) or have no mode if all values are different. Median and mode are preferred when data contain extreme values (outliers) that distort the mean.

Fill in the blank: The ________ is calculated by summing all data values and dividing by the number of values; it is the most common measure of central tendency but is sensitive to extreme values (outliers) in the dataset.

2.4.2 Range and variance as measures of spread

Measures of spread (dispersion) describe how variable the data are around the centre. The range is the simplest: Range = Maximum value - Minimum value. For the maize yield data: Range = 2,400 - 1,900 = 500 kg per ha. A large range indicates high variability; a small range indicates uniformity. The range uses only two values and ignores the distribution of the remaining data.

The variance (s2) measures the average squared deviation of each value from the mean: s2 = [sum of (xi - mean)2] / (n - 1). The division by (n - 1) rather than n gives the sample variance (an unbiased estimate of the population variance). The standard deviation (s) is the square root of the variance and is expressed in the same units as the data, making it more interpretable. For the maize yield data: deviations from mean (2,180): -80; +220; -280; +20; +120. Squared deviations: 6,400; 48,400; 78,400; 400; 14,400. Sum = 148,000. Variance = 148,000 / 4 = 37,000. SD = square root of 37,000 = 192.4 kg per ha.

Fill in the blank: The ________ is the square root of the variance; it measures the average spread of data values around the mean and is expressed in the same units as the original data, making it the most interpretable measure of dispersion.

2.4.3 Worked examples from farm data

Example: Weekly egg production (eggs per hen) from a layer flock over 6 weeks: 5, 6, 5, 7, 6, 5. Mean = (5+6+5+7+6+5)/6 = 34/6 = 5.67 eggs per hen per week. Arranged in order: 5, 5, 5, 6, 6, 7. Median = average of 3rd and 4th values = (5+6)/2 = 5.5. Mode = 5 (appears 3 times). Range = 7 - 5 = 2. Deviations from mean: -0.67, +0.33, -0.67, +1.33, +0.33, -0.67. Squared deviations: 0.449, 0.109, 0.449, 1.769, 0.109, 0.449. Sum = 3.334. Variance = 3.334/5 = 0.667. SD = 0.816 eggs per hen per week.

Interpretation: the average hen produced 5.67 eggs per week; production was fairly uniform (SD = 0.82, which is small relative to the mean). The mode of 5 indicates that the most common weekly production was 5 eggs per hen. Extension advice based on these figures: the flock is performing at approximately 81 percent of the theoretical maximum of 7 eggs per hen per week; improved nutrition or lighting could close this gap.

Fill in the blank: For the dataset 5, 6, 5, 7, 6, 5 (weekly eggs per hen), the mode is ________ because it appears more frequently than any other value in the dataset.

Statistic	**Value**	**Interpretation**
Mean	5.67	On average, each hen lays about 5.7 eggs per week.
Median	5.5	Half the hens lay fewer than 5.5 eggs, half lay more.
Mode	5	The most common weekly egg count is 5.
Range	2 (7 − 5)	Egg counts vary between 5 and 7, a small spread.
Variance	0.56	The average squared deviation from the mean is low, showing consistency.
Standard Deviation	0.75	Egg counts typically differ from the mean by less than 1 egg, indicating stable production.

Box 2.1: Descriptive statistics worked example using weekly egg production data

Pilot questions 2.4

1. The body weights (kg) of 6 goats are: 18, 22, 19, 25, 21, 22. Calculate the mean, median, and mode.
2. Using the goat body weight data in question 1, calculate the range and variance.
3. The monthly milk yields (litres) of a cow over 5 months are: 280, 310, 295, 320, 305. Calculate the mean and standard deviation.
4. Explain why the median is preferred over the mean when farm data contain extreme values (outliers).
5. The following fertiliser prices (N per bag) were recorded in 5 markets: 12,000; 13,500; 12,000; 14,000; 13,000. Find the mean, median, and modal price.

Series Summary

Fractions and decimals are converted by dividing numerator by denominator; ratios are simplified by dividing by the HCF; proportions solve scaling problems using a/b = c/d. Percentages are calculated as (part/whole) x 100; key agricultural applications include germination percentage, moisture content, disease incidence, and profit/loss. Algebra uses variables in formulae (P = R - C; area = L x W); linear equations are solved by isolating the unknown. Simultaneous equations (two unknowns, two equations) are solved by substitution or elimination. Agricultural word problems follow six steps: identify unknowns, assign variables, write equations, solve, check, and state with units.

Plant population per ha = 10,000 / (inter-row spacing m x intra-row spacing m); base seed rate (kg per ha) = (population x seed weight g) / 1,000; adjusted seed rate = (base rate / germination%) x 100. Worked examples: maize (53,333 plants per ha; 16.6 kg per ha adjusted at 90% germination); cowpea (37.5 kg per ha at 80% germination). Measures of central tendency: mean = sum/n; median = middle value in ordered data; mode = most frequent value. Measures of spread: range = max - min; variance s2 = sum(xi - mean)2 / (n-1); standard deviation = square root of variance. Outliers distort the mean; median is preferred in such cases.

Glossary of Terms

- **Adjusted seed rate:** seed rate corrected upward to compensate for less-than-100 percent germination; calculated as (base seed rate / germination percentage) x 100.
- **Algebraic expression:** a mathematical statement combining numbers and variables (letters representing unknown quantities); e.g. 3x + 5.
- **Germination percentage:** the proportion of seeds that germinate under standard test conditions, expressed as a percentage; used to adjust seed rates to achieve target plant populations.
- **Mean:** the arithmetic average of a set of values; calculated as the sum of all values divided by the number of values; sensitive to extreme outliers.

- **Median:** the middle value of a dataset arranged in ascending order; not affected by extreme values; preferred when data contain outliers.
- **Mode:** the value that appears most frequently in a dataset; a dataset may have no mode, one mode (unimodal), or two modes (bimodal).
- **Plant population:** the number of plants per unit area (usually per hectare); determined by inter-row and intra-row spacing and the number of seeds per planting hole.
- **Proportion:** a statement that two ratios are equal ($a/b = c/d$); used to scale agricultural quantities up or down.
- **Range:** the difference between the maximum and minimum values in a dataset; the simplest measure of spread, but sensitive to extreme values.
- **Ratio:** the relative sizes of two or more quantities expressed as a:b; simplified by dividing all terms by their highest common factor.
- **Seed rate:** the mass of seed (in kg) required to plant one hectare; depends on plant population, seed weight, spacing, and germination percentage.
- **Simultaneous equations:** two or more equations with the same unknowns, solved together; methods of solution include substitution and elimination.
- **Standard deviation:** the square root of the variance; measures the average spread of data values around the mean in the same units as the data.
- **Variance:** the average squared deviation of each data value from the mean, calculated using ($n-1$) as the divisor for sample data; measures the spread of data around the mean.

Concept Check Questions [Series 2]

Objective Questions

1. Plant population per hectare = 10,000 / (inter-row spacing in m x ________ spacing in m). (Linked to SLO 2.3)
2. Germination percentage = (number of seeds germinated / total seeds tested) x ________. (Linked to SLO 2.1)
3. Adjusted seed rate = (base seed rate / germination percentage) x ________. (Linked to SLO 2.3)
4. The ________ is the middle value of a dataset arranged in ascending order and is not affected by extreme outliers. (Linked to SLO 2.4)
5. Standard deviation is the square root of the ________; it is expressed in the same units as the original data. (Linked to SLO 2.4)

Short-Answer Questions

6. A farmer mixes NPK and urea in a ratio of 3:2. If he needs 40 kg of the mixture, how much of each does he need? (Linked to SLO 2.1)
7. Calculate the plant population per ha for groundnut planted at 60 cm x 15 cm with 1 seed per hole. (Linked to SLO 2.3)
8. If groundnut seed weight is 0.50 g and germination is 85%, calculate the adjusted seed rate for the population in question 2. (Linked to SLO 2.3)
9. The yields (kg) from 5 plots are: 350, 420, 380, 410, 390. Calculate the mean, median, and range. (Linked to SLO 2.4)
10. Solve the simultaneous equations: $2x + y = 14$ and $x - y = 1$. (Linked to SLO 2.2)

Long-Answer Questions

11. Explain ratios, proportions, and percentages with agricultural examples. Describe how seed rate is calculated and adjusted for germination percentage with a fully worked Nigerian crop example. (Linked to SLOs 2.1, 2.3)
12. Explain how to set up and solve simultaneous equations using an agricultural word problem. Describe the mean, median, mode, range, variance, and standard deviation with a fully worked farm data example. (Linked to SLOs 2.2, 2.4)

Answers to Concept Check Questions [Series 2]

Objective Questions

1. Intra-row.
2. 100.
3. 100.
4. Median.
5. Variance.

Short-Answer Questions

6. NPK = (3/5) x 40 = 24 kg; urea = (2/5) x 40 = 16 kg.
7. Plant population = 10,000 / (0.60 x 0.15) = 10,000 / 0.09 = 111,111 plants per ha.
8. Base seed rate = (111,111 x 0.50) / 1,000 = 55.6 kg per ha. Adjusted seed rate = (55.6 / 85) x 100 = 65.4 kg per ha.
9. Mean = (350+420+380+410+390)/5 = 1,950/5 = 390 kg. Ordered: 350, 380, 390, 410, 420; Median = 390 kg. Range = 420 - 350 = 70 kg.
10. From equation 2: x = y + 1. Substitute into equation 1: 2(y+1) + y = 14; 3y + 2 = 14; 3y = 12; y = 4; x = 5.

Long-Answer Questions

11. Ratio: relative size of two quantities (e.g. NPK:urea = 3:2; simplify by dividing by HCF). Proportion: two ratios equal (a/b = c/d); used to scale seed quantities: if 5 kg plants 0.5 ha, x kg plants 2 ha; 5/0.5 = x/2; x = 20 kg. Percentages: (part/whole) x 100; key uses: germination % (germinated/total x 100), disease incidence, profit %. Seed rate calculation: plant population per ha = 10,000 / (inter-row m x intra-row m). Example: Maize, 75 cm x 25 cm, 1 seed per hole, seed weight 0.28 g, germination 90%. Population = 10,000/0.1875 = 53,333 plants per ha. Base seed rate = (53,333 x 0.28)/1,000 = 14.9 kg per ha. Adjusted rate = (14.9/90) x 100 = 16.6 kg per ha. For 2 ha: 16.6 x 2 = 33.2 kg needed.

12. Simultaneous equations example: a farmer sells broilers at N4,500 and layers at N3,200; sells 12 birds for N48,900. Let b = broilers, l = layers. b + l = 12 (i); 4,500b + 3,200l = 48,900 (ii). From (i): b = 12 - l. Substitute: 4,500(12-l) + 3,200l = 48,900; 54,000 - 4,500l + 3,200l = 48,900; -1,300l = -5,100; l = 3.92 (adjust to whole number by checking: l=4, b=8: 8x4500+4x3200=36,000+12,800=48,800; close; use l=3, b=9: 40,500+9,600=50,100; so l=4, b=8 gives 48,800, nearest). Descriptive statistics example (yields kg: 350, 420, 380, 410, 390): Mean = 390 kg. Ordered: 350,380,390,410,420; Median = 390 kg. No mode (all different). Range = 70 kg. Deviations from mean: -40, +30, -10, +20, 0. Squared: 1600, 900, 100, 400, 0. Sum=3000. Variance=3000/4=750. SD=sqrt(750)=27.4 kg. Interpretation: average yield 390 kg; moderate variability (SD=27.4 kg); median equals mean suggesting symmetric distribution.

Answers to Pilot Questions [Series 2]

Part 2.1

1. Seed needed = 3/4 x 12 = 9 kg of cowpea seed.
2. NPK = (3/5) x 50 = 30 kg; Urea = (2/5) x 50 = 20 kg.
3. (a) 17/20 x 100 = 85%. (b) 0.045 x 100 = 4.5%. (c) 7/25 x 100 = 28%.
4. Profit = 35,000 - 28,000 = N7,000. Percentage profit = (7,000/28,000) x 100 = 25%.
5. Germination percentage = (88/100) x 100 = 88%.

Part 2.2

1. Profit P = 480,000 - 315,000 = N165,000.
2. 5y = 28 + 7 = 35; y = 7.
3. Let b = broilers, l = layers. b + l = 12 (i); 4,500b + 3,200l = 48,900 (ii). From (i): b = 12 - l; 4,500(12-l) + 3,200l = 48,900; 54,000 - 1,300l = 48,900; 1,300l = 5,100; l ≈ 3.9. Trying l = 4, b = 8: 4,500(8) + 3,200(4) = 36,000 + 12,800 = 48,800. Nearest integer solution: 8 broilers, 4 layers.

4. Let yam = x, cassava = x + 150. x + x + 150 = 2,350; 2x = 2,200; x = 1,100 kg (yam); cassava = 1,250 kg.
5. Let u = urea, n = NPK. 3u + 2n = 87,000 (i); u + n = 33,000 (ii). From (ii): u = 33,000 - n; 3(33,000-n) + 2n = 87,000; 99,000 - n = 87,000; n = 12,000; u = 21,000. Urea: N21,000 per bag; NPK: N12,000 per bag.

Part 2.3

1. Population = 10,000 / (0.75 x 0.30) = 10,000 / 0.225 = 44,444 plants per ha.
2. Base seed rate = (44,444 x 0.03) / 1,000 = 1,333 / 1,000 = 1.33 kg per ha.
3. Adjusted seed rate = (1.33 / 75) x 100 = 1.77 kg per ha.
4. Total seed = 1.77 x 0.5 = 0.89 kg (approximately 0.9 kg).
5. Adjusted seed rate = (100 / (80 x 90)) x 10,000 = (100 / 7,200) x 10,000 = 138.9 kg per ha.

Part 2.4

1. Mean = (18+22+19+25+21+22)/6 = 127/6 = 21.2 kg. Ordered: 18,19,21,22,22,25; Median = (21+22)/2 = 21.5 kg. Mode = 22 kg (appears twice).
2. Range = 25 - 18 = 7 kg. Deviations from mean (21.2): -3.2, +0.8, -2.2, +3.8, -0.2, +0.8. Squared: 10.24, 0.64, 4.84, 14.44, 0.04, 0.64. Sum = 30.84. Variance = 30.84/5 = 6.17 kg2.
3. Mean = (280+310+295+320+305)/5 = 1,510/5 = 302 litres. Deviations: -22, +8, -7, +18, +3. Squared: 484, 64, 49, 324, 9. Sum = 930. Variance = 930/4 = 232.5. SD = sqrt(232.5) = 15.2 litres.
4. Outliers pull the mean toward themselves, making it unrepresentative of the majority of values. The median is based only on the rank order of data and is not affected by how extreme the outlying values are, so it gives a better picture of typical performance.
5. Mean = (12,000+13,500+12,000+14,000+13,000)/5 = 64,500/5 = N12,900. Ordered: 12,000; 12,000; 13,000; 13,500; 14,000; Median = N13,000. Mode = N12,000 (appears twice).

References and Attributions [Series 2]

Textual References (Books and External Sources)

- Gomez, K. A., & Gomez, A. A. (1984). Statistical procedures for agricultural research (2nd ed.). Wiley-Interscience.
- Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods (8th ed.). Iowa State University Press.
- IITA. (2020). Crop management recommendations for smallholder farmers in Nigeria. International Institute of Tropical Agriculture, Ibadan.

Figures, Plates, Tables, and Boxes

- Table 2.1: Worked examples of seed rate calculations for Nigerian crops. Attribution: AI generated. Suggested OER search string: "seed rate calculation plant population spacing germination percentage Nigeria crops table OER".
- Box 2.1: Descriptive statistics worked example using weekly egg production data. Attribution: AI generated. Suggested OER search string: "descriptive statistics mean median mode range variance standard deviation worked example agricultural data OER".

Series 3: Measurement and Applications

- **Part 3.1: Units of Measurement**
  - Sub Part 3.1.1: The SI system and common agricultural units
  - Sub Part 3.1.2: Unit conversions in land, mass, and volume
  - Sub Part 3.1.3: Units of time and rates in agriculture
- **Part 3.2: Measurement of Land Area**
  - Sub Part 3.2.1: Area of regular shapes
  - Sub Part 3.2.2: Area of irregular farmland
  - Sub Part 3.2.3: Practical land measurement methods
- **Part 3.3: Volume, Capacity, and Storage**
  - Sub Part 3.3.1: Volume of common storage structures
  - Sub Part 3.3.2: Capacity of tanks, silos, and bags
  - Sub Part 3.3.3: Worked examples in farm storage
- **Part 3.4: Rates, Yields, and Input Calculations**
  - Sub Part 3.4.1: Fertiliser application rates
  - Sub Part 3.4.2: Yield calculations and crop equivalents
  - Sub Part 3.4.3: Labour input and productivity measures

Introduction

Measurement is the practical link between the abstract mathematics of Series 2 and real farming activities. Every farm decision rests on measurement: the area of a field determines how much seed and fertiliser to purchase; the volume of a storage silo determines whether it can hold the expected harvest; the yield per hectare tells a farmer whether a variety is worth planting again. This Series covers the unit systems used in Nigerian agriculture, the measurement and calculation of land areas, the computation of volumes and storage capacities, and the derivation of rates such as fertiliser application per hectare and yield per unit area.

All calculations are grounded in Nigerian farming reality: plot sizes range from 0.1 ha smallholder plots to large commercial farms of several hundred hectares; storage structures include mud silos, metal tanks, and jute bags; yield data come from maize, cassava, yam, rice, cowpea, and groundnut. Accuracy in measurement and unit handling prevents costly errors in input procurement and storage planning.

Series Learning Outcomes

Upon completing this Series, you should be able to:

- **SLO 3.1** identify and use the SI units and common agricultural units for length, area, mass, volume, and time (Linked to Part 3.1),
- **SLO 3.2** perform unit conversions between metric and common Nigerian agricultural units (Linked to Part 3.1),
- **SLO 3.3** calculate the area of regular and irregular farmland using appropriate formulae and practical methods (Linked to Part 3.2),
- **SLO 3.4** calculate the volume and storage capacity of common farm storage structures (Linked to Part 3.3), and
- **SLO 3.5** calculate fertiliser application rates, crop yields, and labour productivity for Nigerian farming situations (Linked to Part 3.4).

3.1 Units of Measurement

3.1.1 The SI system and common agricultural units

The International System of Units (SI) is the modern metric system adopted globally for scientific and most commercial purposes. The base SI units most relevant to agriculture are: metre (m) for length; kilogram (kg) for mass; second (s) for time; and kelvin (K) for temperature (though degrees Celsius, degC, are used in practice). Derived units include square metre (m2) for area, cubic metre (m3) for volume, metre per second (m/s) for velocity, and kilogram per hectare (kg/ha) for application rates.

In Nigerian agriculture, several non-SI units remain in common use alongside SI: area (hectare = 10,000 m2; acre = 4,047 m2; plot = variable, typically 0.25 to 1 acre); mass (tonne = 1,000 kg; bag = 50 kg for most grains; mudu = local volume-based unit varying by region); length (chain = 20.1 m; furlong = 201 m; rod = 5.03 m; used in older land surveys). An agricultural professional must be fluent in both systems and able to convert between them.

Fill in the blank: One hectare equals ________ square metres; it is the standard unit of farm area used in Nigerian agricultural research, extension, and commercial farming.

3.1.2 Unit conversions in land, mass, and volume

Land area conversions: 1 hectare (ha) = 10,000 m2; 1 acre = 4,047 m2 = 0.4047 ha; 1 ha = 2.471 acres; 1 km2 = 100 ha. To convert acres to hectares: multiply by 0.4047; to convert hectares to acres: multiply by 2.471. Example: 5 acres = 5 x 0.4047 = 2.02 ha. Mass conversions: 1 tonne = 1,000 kg; 1 bag of grain = 50 kg (standard Nigerian bag); 1 kg = 1,000 g; yield expressed in t/ha x 20 = bags per ha. Volume conversions: 1 m3 = 1,000 litres (L); 1 litre = 1,000 millilitres (mL); 1 gallon (UK) = 4.546 L; 1 gallon (US) = 3.785 L.

Practical conversion example: a cassava farmer harvests 18 t/ha on a 3.5-acre plot. Convert: 3.5 acres x 0.4047 = 1.42 ha. Total harvest = 18 x 1.42 = 25.5 tonnes = 25,500 kg = 510 bags of 50 kg each. These conversions arise constantly when using research recommendations (given in kg/ha or t/ha) for farm planning with land measured in acres or when selling to buyers who price by the bag.

Fill in the blank: To convert an area from acres to hectares, multiply by ________; a farm of 5 acres is therefore equivalent to 5 x 0.4047 = 2.02 ha.

3.1.3 Units of time and rates in agriculture

Agricultural rates combine two units: a quantity per unit time (daily milk yield: litres per cow per day; annual rainfall: mm per year) or a quantity per unit area (fertiliser rate: kg/ha; yield: t/ha; plant population: plants per ha). To convert a rate, convert each component unit. Example: yield of 2,500 kg/ha = 2.5 t/ha = 50 bags of 50 kg/ha. Labour productivity is expressed as area weeded per person per day (ha/person/day) or as person-days per hectare for a given operation.

Time units in agriculture: crop cycle in days or months; growing season (long rains: April to July in southern Nigeria; short rains: September to November); storage duration in weeks or months. Converting growing degree days (GDD; sum of daily mean temperatures above a base temperature) requires arithmetic with temperature units. Recognising and correctly using compound rate units prevents errors in input budgeting, yield reporting, and labour planning.

Fill in the blank: A fertiliser application rate of 120 kg N per ha means that 120 kilograms of elemental ________ must be applied for every hectare of land; if urea (46% N) is used, the mass of urea required is 120/0.46 = 260.9 kg per ha.

Pilot questions 3.1

6. Convert 7.5 acres to hectares.
7. A farmer produces 22 t/ha of cassava on a 2-ha farm. Express the total yield in: (a) kg; (b) bags of 50 kg.
8. State four non-SI units of area or mass used in Nigerian agriculture and give the SI equivalent of each.
9. If urea contains 46% N and a crop requires 60 kg N per ha, how many kg of urea per ha must be applied?
10. Convert a yield of 3,600 kg/ha to (a) t/ha and (b) bags of 50 kg per ha.

3.2 Measurement of Land Area

3.2.1 Area of regular shapes

Most farm plots can be approximated by regular geometric shapes. Rectangle: Area = length x width (A = L x W). Triangle: Area = (1/2) x base x height (A = 0.5bh). Trapezium: Area = (1/2) x (sum of parallel sides) x perpendicular height (A = 0.5(a+b)h). Circle: Area = pi x radius2 (A = pi r2; pi = 3.1416). Parallelogram: Area = base x perpendicular height.

Example: a rectangular rice plot measures 80 m x 50 m. Area = 80 x 50 = 4,000 m2 = 0.40 ha. If fertiliser rate is 120 kg N/ha: fertiliser needed = 120 x 0.40 = 48 kg N. A triangular plot has base 60 m and height 40 m: area = 0.5 x 60 x 40 = 1,200 m2 = 0.12 ha. A circular fish pond has radius 8 m: area = 3.1416 x 82 = 201.1 m2 = 0.0201 ha.

Fill in the blank: The area of a rectangle with length L and width W is A = L x W; a rectangular plot measuring 100 m x 60 m has an area of ________ m2 = 0.60 ha.

3.2.2 Area of irregular farmland

Measuring an Irregular Farm Plot Using the Offset Method

Most farm plots have irregular boundaries. The **offset method** measures such plots by establishing a straight baseline and taking perpendicular distances, called offsets, from the baseline to the boundary at regular intervals.

Each section between two consecutive offsets is treated as a trapezium. The area of each strip is calculated as:

Area of strip = 1/2 × (O1 + O2) × d

where:

O1 = first offset

O2 = second offset

d = distance between the two offsets

For a 100 m baseline, offsets are taken at 0, 20, 40, 60, 80, and 100 m as follows:

0 m, 12 m, 18 m, 20 m, 15 m, and 10 m.

The strip areas are:

First strip:

$1/2 \times (0 + 12) \times 20 = 120 \text{ m}^2$

Second strip:

$1/2 \times (12 + 18) \times 20 = 300 \text{ m}^2$

Third strip:

$1/2 \times (18 + 20) \times 20 = 380 \text{ m}^2$

Fourth strip:

$1/2 \times (20 + 15) \times 20 = 350 \text{ m}^2$

Fifth strip:

$1/2 \times (15 + 10) \times 20 = 250 \text{ m}^2$

Therefore:

Total area = 120 + 300 + 380 + 350 + 250

Total area = 1,400 m²

The mid-ordinate rule approximates area by multiplying the baseline length by the mean of the mid-point offsets. The cross-staff or optical square is a simple field instrument used to set out perpendiculars to the baseline without a theodolite. The GPS (global positioning system) is increasingly used in Nigerian extension and research to determine plot boundaries and calculate areas automatically; handheld GPS devices accurate to 3 to 5 m are affordable for research use.

Fill in the blank: In the offset method for measuring irregular farmland, the plot is divided into strips along a ________ ; perpendicular offsets are measured from the baseline to the boundary at regular intervals, and the area of each strip is calculated as a trapezium.

3.2.3 Practical land measurement methods

Chain surveying uses a measuring chain (Gunter's chain: 20.12 m, divided into 100 links) or a 30 m or 50 m fibreglass tape to measure distances. Ranging poles mark straight lines. A simple closed traverse around a field boundary provides the data needed to calculate the enclosed area using the coordinate or cross-multiplication method. The resulting area in square chains is converted to hectares (1 square chain = 0.0404 ha).

Pacing is the simplest field method: a known average pace length (typically 0.75 to 0.80 m per pace for an adult) is used to estimate distances; area is estimated from the paced dimensions. Pacing is adequate for approximate farm planning but not for legal or research purposes. Marking boundary corners with concrete pillars (beacons) and filing a survey plan with the state Surveyor-General is required for formal land documentation in Nigeria. Measurements should always be recorded in a field book with a sketch diagram showing plot dimensions and bearing.

Fill in the blank: In chain surveying, one Gunter's chain equals 20.12 m and is divided into 100 ________; the chain is used to measure distances along traverse lines to determine the area of a farm plot.

Pilot questions 3.2

6. A rectangular maize plot is 120 m long and 75 m wide. Calculate its area in (a) m2 and (b) ha.
7. A triangular groundnut plot has a base of 90 m and a perpendicular height of 50 m. Calculate its area in ha.
8. Using the offset method, calculate the area of an irregular plot with a 60 m baseline and offsets of 0, 8, 14, 18, 12, 6 m at 12 m intervals.
9. A farmer estimates that his plot is 50 paces x 40 paces. If his pace is 0.75 m, what is the area in m2 and in ha?
10. Convert an area of 3.5 square chains to hectares (1 square chain = 0.0404 ha).

3.3 Volume, Capacity, and Storage

3.3.1 Volume of Common Storage Structures

Volume is the amount of three-dimensional space occupied by an object. Common volume formulae include:

Cuboid: $V = L \times W \times H$
Cylinder: $V = \pi r^2 h$
Cone: $V = 1/3 \times \pi r^2 h$
Sphere: $V = 4/3 \times \pi r^3$
Pyramid: $V = 1/3 \times$ base area $\times$ height

Volume is measured in cubic metres or litres, and **$1\ m^3 = 1{,}000$ L**.

For agricultural storage structures, a rectangular mud silo measuring 3 m × 2 m × 2 m has:

$V = 3 \times 2 \times 2 = 12\ m^3 = 12{,}000$ L

A cylindrical metal silo with radius 1.5 m and height 3 m has:

$V = \pi r^2 h$

$V = 3.1416 \times 1.5^2 \times 3$

$V \approx 21.2\ m^3$

A conical thatched store with radius 2 m and height 3 m has:

$V = 1/3 \times \pi r^2 h$

$V = 1/3 \times 3.1416 \times 2^2 \times 3$

$V \approx 12.57\ m^3$

Knowing the storage volume helps a farmer determine whether a structure can hold the expected harvest before storage begins.

Fill in the blank: The volume of a cylindrical grain silo with radius r and height h is given by V= $\pi\, r^2 h$________; a silo with radius 1.5 m and height 4 m has a volume of 3.1416 x 1.52 x 4 = 28.3 m3.

3.3.2 Capacity of tanks, silos, and bags

The storage capacity of a structure in terms of a specific commodity depends on the bulk density of that commodity (mass per unit volume, kg/m3). Bulk densities (approximate): maize grain = 720 kg/m3; rice (paddy) = 580 kg/m3; sorghum = 670 kg/m3; groundnut (shelled) = 480 kg/m3; cowpea = 750 kg/m3. Mass that can be stored = volume (m3) x bulk density (kg/m3). For a 12 m3 silo storing maize: capacity = 12 x 720 = 8,640 kg = 172.8 bags of 50 kg.

Water tanks: capacity is expressed in litres or m3; 1 m3 = 1,000 L = 1 tonne of water. A cylindrical water tank with radius 0.8 m and height 1.5 m: V = 3.1416 x 0.64 x 1.5 = 3.02 m3 = 3,020 L. Irrigation channels: flow rate (L/s or m3/s) x time (s) = volume delivered. Standard Nigerian jute bags hold 50 kg of grain; polypropylene bags may hold 25 or 100 kg; knowing these standard capacities allows rapid estimation of quantities without weighing.

Fill in the blank: The mass of grain that can be stored in a silo is calculated as volume (m3) x bulk density (kg/m3); a 10 m3 silo storing cowpea (bulk density 750 kg/m3) can hold ________ kg = 150 bags of 50 kg.

3.3.3 Worked examples in farm storage

Example 1: A farmer builds a rectangular mud silo 4 m long, 2.5 m wide, and 2 m high. He wants to store sorghum (bulk density 670 kg/m3). Volume = 4 x 2.5 x 2 = 20 m3. Capacity = 20 x 670 = 13,400 kg = 268 bags of 50 kg. If his harvest from 5 ha at 1.8 t/ha = 9 tonnes = 9,000 kg = 180 bags: the silo is adequate (268 bag capacity vs 180 bags needed).

Example 2: A circular water tank has a diameter of 2.4 m and depth of 1.8 m. Radius = 1.2 m. Volume = 3.1416 x 1.44 x 1.8 = 8.14 m3 = 8,140 L. If each animal drinks 30 L per day and the farmer has 40 cattle: daily demand = 40 x 30 = 1,200 L per day. The full tank will supply the herd for 8,140/1,200 = 6.8 days (approximately 1 week). These worked examples illustrate how volume calculations directly support farm planning decisions.

Fill in the blank: A cylindrical water tank with diameter 2.4 m (radius 1.2 m) and depth 1.8 m has a volume of pi x 1.44 x 1.8 = ________ m3 = 8,140 litres.

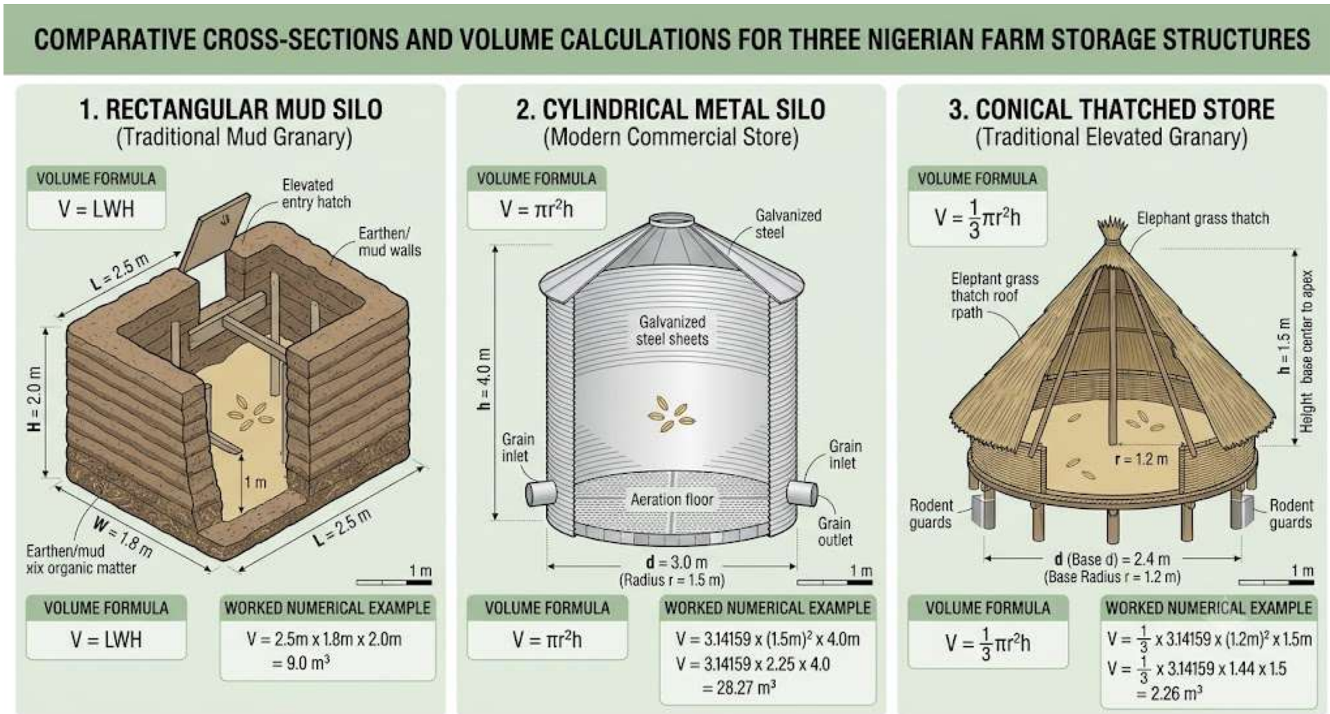

Figure 3.1: Volume formulae and worked examples for three common Nigerian farm storage structures

Pilot questions 3.3

6. A rectangular silo is 5 m long, 3 m wide, and 2.5 m high. Calculate its volume in m3.
7. Using the silo in question 1, calculate how many bags of 50 kg of maize (bulk density 720 kg/m3) it can hold.
8. A cylindrical water tank has a radius of 1.2 m and height of 2 m. Calculate its volume in litres.
9. A conical thatched store has a base radius of 2.5 m and a height of 3.5 m. Calculate its volume.
10. A farmer has a harvest of 240 bags (50 kg each) of cowpea. If his cylindrical silo has a radius of 1.8 m and height of 3 m, will it fit? (Bulk density of cowpea = 750 kg/m3.)

3.4 Rates, Yields, and Input Calculations

3.4.1 Fertiliser application rates

Fertiliser recommendations are usually stated as kilograms of nutrient per hectare, such as **60 kg N, 30 kg P_2O_5, and 30 kg K_2O per hectare** for maize. To determine the quantity of fertiliser product required, use:

Mass of fertiliser = (Required nutrient ÷ Percentage of nutrient in fertiliser) × 100

Common fertilisers include urea containing 46% N, calcium ammonium nitrate containing 26% N, single superphosphate containing 18% P_2O_5, muriate of potash containing 60% K_2O, and NPK 15:15:15 containing 15% each of N, P_2O_5, and K_2O. For example, the urea needed to supply 60 kg N/ha is:

Mass = (60 ÷ 46) × 100 = 130.4 kg urea/ha

Compound fertilisers supply several nutrients at the same time. For example, applying NPK 15:15:15 to supply 60 kg N/ha requires:

Mass = (60 ÷ 15) × 100 = 400 kg/ha

However, this also supplies **60 kg P_2O_5 and 60 kg K_2O per hectare**, which may exceed the recommended 30 kg of each. A better approach is to use NPK 15:15:15 to supply the required phosphorus and potassium, then add urea to complete the nitrogen requirement. Care must be taken to distinguish elemental nitrogen, written as **N**, from phosphorus and potassium recommendations, which are commonly expressed as **P_2O_5** and **K_2O**.

Fill in the blank: The mass of urea (46% N) required to supply 90 kg N per ha is calculated as (90/46) x 100 = ________ kg urea per ha.

3.4.2 Yield calculations and crop equivalents

Crop yield is the mass of harvested product per unit area, expressed as t/ha or kg/ha. Fresh yield vs dry yield: cassava is harvested fresh (24 to 30% dry matter); to convert fresh yield to dry matter equivalent: dry weight = fresh weight x dry matter fraction. Example: 20 t/ha fresh cassava at 28% DM = 20 x 0.28 = 5.6 t/ha dry matter. Harvest index (HI) is the fraction of total above-ground biomass that is the economic yield: HI = economic yield / total biomass. If total biomass = 8 t/ha and grain yield = 3.2 t/ha: HI = 3.2/8 = 0.40.

Crop equivalents allow comparison of different crops on a common basis: cassava equivalent = (value of crop per tonne / value of cassava per tonne) x actual yield. If maize is worth N150,000/t and cassava N35,000/t: 1 tonne of maize = 150,000/35,000 = 4.3 tonnes cassava equivalent. This conversion is used in enterprise budgeting and farm income analysis to compare the profitability of different crop enterprises on a per-hectare basis.

Fill in the blank: Harvest index is calculated as economic yield / total above-ground ________; a crop with grain yield 3.5 t/ha and total biomass 9 t/ha has a harvest index of 3.5/9 = 0.39.

3.4.3 Labour input and productivity measures

Agricultural labour is measured in person-days (one person working one full day, typically 6 to 8 hours). Labour productivity is the output per unit of labour input: yield per person-day = total yield / total person-days; or area completed per person-day for non-harvest operations. Example: 3 workers take 4 days to weed 1.2 ha; total labour input = 3 x 4 = 12 person-days; productivity = 1.2/12 = 0.10 ha/person-day. Labour cost = number of person-days x daily wage rate (N/person-day).

Labour-use efficiency compares observed labour input with a standard (e.g. recommended labour requirement for weeding maize = 15 person-days/ha); if a farmer uses 12 person-days/ha, his efficiency ratio = 15/12 = 1.25 (25% more efficient than the standard). The ratio of output value to labour cost is the labour return ratio: if maize yield = 2.5 t at N150,000/t = N375,000 gross; labour cost = 25 person-days x N2,500/day = N62,500; labour return = N375,000/N62,500 = 6.0 (each naira spent on labour returns 6 naira of output).

Fill in the blank: Labour productivity for a field operation is calculated as area completed divided by total ________-days; if 4 workers complete 2 ha of weeding in 3 days, total input = 12 person-days and productivity = 2/12 = 0.17 ha per person-day.

Fertiliser Product	Nutrient Content %	Nutrient Required per ha	Mass of Fertiliser Needed per ha
Urea	46 % N	100 kg N / ha	1000.46=217 kg / ha
NPK 15:15:15	15 % N, 15 % P_2O_5, 15 % K_2O	60 kg N / ha	600.15=400 kg / ha
CAN (Calcium Ammonium Nitrate)	26 % N	80 kg N / ha	800.26=308 kg / ha
MOP (Muriate of Potash)	60 % K_2O	50 kg K_2O / ha	500.60=83 kg / ha

Box 3.1: Fertiliser application rate calculations for common Nigerian fertiliser products

Pilot questions 3.4

6. A maize crop requires 90 kg N, 30 kg P2O5, and 30 kg K2O per ha. Calculate the mass of urea (46% N) needed to supply all the nitrogen.
7. How many kg of SSP (18% P2O5) are needed per ha to supply the 30 kg P2O5 in question 1?
8. A cassava farm yields 25 t/ha fresh weight at 28% dry matter. Calculate the dry matter yield in t/ha.
9. If maize is valued at N180,000/t and yam at N60,000/t, calculate the maize equivalent of 3 t/ha of yam.
10. Five workers take 6 days to harvest 3 ha of cowpea. Calculate (a) total person-days used and (b) labour productivity in ha per person-day.

Series Summary

This Series covered basic agricultural measurements and calculations. The main SI units used are the metre, kilogram, and second. Important conversions include **1 ha = 10,000 m²**, **1 acre = 0.4047 ha**, **1 ha = 2.471 acres**, **1 tonne = 1,000 kg = 20 bags of 50 kg**, and **1 m³ = 1,000 L**. Common area formulae are **rectangle = L × W**, **triangle = ½ × b × h**, **trapezium = ½ × (a + b) × h**, and **circle = πr²**. Irregular farm plots may be measured by the offset method, where trapezoidal strip areas are added together, or with GPS. Chain surveying traditionally uses Gunter's chain, which is 20.12 m long and contains 100 links. Fertiliser requirements are calculated using **Mass of fertiliser = (Required nutrient ÷ Percentage nutrient) × 100**. Common products include urea with 46% N, NPK 15:15:15, CAN with 26% N, SSP with 18% P_2O_5, and MOP with 60% K_2O.

The Series also covered volume, storage, yield, and labour calculations. The main volume formulae are **cuboid = L × W × H**, **cylinder = πr²h**, and **cone = ⅓πr²h**. Storage capacity is calculated as **Storage capacity = Volume × Bulk density**, with typical bulk densities of 720 kg/m³ for maize, 750 kg/m³ for cowpea, 670 kg/m³ for sorghum, and 580 kg/m³ for rice. Crop yield is expressed in kg/ha or t/ha, while dry yield is calculated as **Fresh yield × Dry-matter fraction**. Harvest index is **Economic yield ÷ Total biomass**. Labour requirement is measured as **Person-days = Number of workers × Number of days**, labour productivity as **Area covered ÷ Person-days**, and labour return as **Gross output value ÷ Labour cost**.

Glossary of Terms

- **Bulk density:** the mass of a granular material per unit volume; expressed in kg/m3; used to calculate the mass of grain that can be stored in a known volume.
- **Chain surveying:** a field method for measuring land areas using a measuring chain (Gunter's chain: 20.12 m) and ranging poles to lay out traverse lines and calculate the enclosed area.

- **Dry matter fraction:** the proportion of a fresh crop product that is dry matter (non-water content); used to convert fresh yields to dry matter equivalents; e.g. cassava = 0.24 to 0.30.
- **Harvest index:** the ratio of economic (harvested) yield to total above-ground biomass; a high harvest index indicates efficient partitioning of photosynthate into the harvestable product.
- **Hectare (ha):** a unit of area equal to 10,000 m2 or 2.471 acres; the standard unit for expressing farm area and crop yield in Nigerian agricultural research and extension.
- **Offset method:** a field surveying technique for measuring irregular plots in which perpendicular offsets from a baseline to the boundary are measured at regular intervals; the plot area is calculated by summing the trapezium areas of each strip.
- **Person-day:** a unit of labour input equal to one person working one full day (6 to 8 hours); used to calculate labour requirements and costs for agricultural operations.
- **Volume:** the amount of three-dimensional space occupied by a body; expressed in m3 or litres; calculated using formulae appropriate to the shape of the structure (cuboid, cylinder, cone).

Concept Check Questions [Series 3]

Objective Questions

6. One hectare equals ________ square metres and is equivalent to 2.471 acres. (Linked to SLO 3.1)
7. The area of a rectangle with length L and width W is A = ________. (Linked to SLO 3.3)
8. The volume of a cylinder with radius r and height h is V = pi x r2 x ________. (Linked to SLO 3.4)
9. The mass of grain storable in a silo = volume (m3) x ________ density (kg/m3). (Linked to SLO 3.4)
10. Labour productivity (ha/person-day) = area completed / total ________-days. (Linked to SLO 3.5)

Short-Answer Questions

11. Convert 4.5 acres to hectares and express the result in m2. (Linked to SLO 3.2)
12. A maize plot is 150 m x 80 m. Calculate its area in ha and the amount of urea (46% N) needed to apply 90 kg N/ha. (Linked to SLOs 3.3, 3.5)
13. A cylindrical silo has radius 1.5 m and height 3.5 m. Calculate its volume and its capacity in bags of 50 kg of sorghum (670 kg/m3). (Linked to SLO 3.4)
14. A cassava farmer harvests 30 t/ha fresh weight (25% DM) on a 2 ha farm. Calculate the dry matter yield and total dry matter produced. (Linked to SLO 3.5)
15. Three workers weed 1.8 ha in 5 days. Calculate total person-days and labour productivity. (Linked to SLO 3.5)

Long-Answer Questions

13. Explain the SI unit system and common Nigerian agricultural units, describe unit conversions for area and mass, and show how fertiliser application rates are calculated using two Nigerian fertiliser examples. (Linked to SLOs 3.1, 3.2, 3.5)
14. Describe methods for measuring regular and irregular farmland. Explain how volume and storage capacity are calculated for farm storage structures, with two fully worked examples. (Linked to SLOs 3.3, 3.4)

Answers to Concept Check Questions [Series 3]

Objective Questions

6. 10,000.
7. L x W.
8. h.
9. Bulk.
10. Person.

Short-Answer Questions

11. 4.5 acres x 0.4047 = 1.82 ha = 18,200 m2.
12. Area = 150 x 80 = 12,000 m2 = 1.20 ha. Urea needed = (90/46) x 100 = 195.7 kg/ha x 1.20 ha = 234.8 kg total.
13. V = 3.1416 x 1.52 x 3.5 = 24.7 m3. Mass of sorghum = 24.7 x 670 = 16,549 kg = 330.9 bags.
14. DM yield per ha = 30 x 0.25 = 7.5 t/ha DM. Total on 2 ha = 7.5 x 2 = 15 tonnes DM.
15. Total person-days = 3 x 5 = 15 person-days. Productivity = 1.8/15 = 0.12 ha per person-day.

Long-Answer Questions

13. SI base units: metre (m), kilogram (kg), second (s). Derived: m2 (area), m3 (volume), kg/ha (rate). Common Nigerian units: hectare (10,000 m2), acre (0.4047 ha), bag (50 kg), tonne (1,000 kg). Conversions: acres to ha (multiply by 0.4047); ha to acres (multiply by 2.471); t/ha to bags/ha (multiply by 20). Fertiliser calculations: required nutrient known from recommendation; mass of fertiliser = (required nutrient kg/ha / % nutrient in fertiliser) x 100. Example 1 (Urea 46% N; 90 kg N/ha needed): mass = (90/46) x 100 = 195.7 kg urea/ha. Example 2 (NPK 15:15:15; 60 kg N/ha + 30 kg P2O5/ha + 30 kg K2O/ha needed): mass for N = (60/15) x 100 = 400 kg NPK/ha; this also supplies 60 kg P2O5 and 60 kg K2O (excess P and K must be considered in the nutrient budget).
14. Regular shapes: rectangle (A = L x W); triangle (A = 0.5 bh); trapezium (A = 0.5(a+b)h); circle (A = pi r2). Example: 120 m x 60 m plot = 7,200 m2 = 0.72 ha. Irregular plots: offset method (measure perpendicular offsets at regular intervals along baseline; sum

trapezium areas; Area = sum of 0.5(O1+O2) x strip width). GPS increasingly used. Chain surveying uses 20.12 m Gunter's chain and ranging poles. Volume formulae: cuboid (V = L x W x H); cylinder (V = pi r2 h); cone (V = 1/3 pi r2 h). Worked example 1 (rectangular silo 4 m x 2.5 m x 2 m): V = 20 m3; storing maize (720 kg/m3) = 20 x 720 = 14,400 kg = 288 bags. Worked example 2 (cylindrical water tank r = 1 m, h = 2 m): V = 3.1416 x 1 x 2 = 6.28 m3 = 6,280 L; at 30 L/day per animal serves 40 animals for 6,280/(40 x 30) = 5.2 days.

Answers to Pilot Questions [Series 3]

Part 3.1

6. 7.5 x 0.4047 = 3.04 ha.
7. Total = 22 x 2 = 44 t. (a) 44,000 kg. (b) 44,000/50 = 880 bags.
8. Hectare (ha) = 10,000 m2; acre = 4,047 m2; bag = 50 kg; tonne = 1,000 kg.
9. Mass of urea = (60/46) x 100 = 130.4 kg urea per ha.
10. (a) 3,600/1,000 = 3.6 t/ha. (b) 3,600/50 = 72 bags per ha.

Part 3.2

6. Area = 120 x 75 = 9,000 m2 = 0.90 ha.
7. Area = 0.5 x 90 x 50 = 2,250 m2 = 0.225 ha.
8. Strips (12 m wide): 0.5(0+8)x12=48; 0.5(8+14)x12=132; 0.5(14+18)x12=192; 0.5(18+12)x12=180; 0.5(12+6)x12=108. Total = 48+132+192+180+108 = 660 m2.
9. Distance = 50 x 0.75 = 37.5 m; width = 40 x 0.75 = 30 m. Area = 37.5 x 30 = 1,125 m2 = 0.1125 ha.
10. 3.5 x 0.0404 = 0.141 ha.

Part 3.3

6. V = 5 x 3 x 2.5 = 37.5 m3.
7. Mass = 37.5 x 720 = 27,000 kg. Bags = 27,000/50 = 540 bags.
8. V = 3.1416 x 1.44 x 2 = 9.05 m3 = 9,050 L.
9. V = (1/3) x 3.1416 x 6.25 x 3.5 = 22.9 m3.
10. Harvest = 240 x 50 = 12,000 kg. Silo V = 3.1416 x 3.24 x 3 = 30.5 m3. Capacity = 30.5 x 750 = 22,875 kg. Since 22,875 > 12,000, the harvest fits.

Part 3.4

6. Mass of urea = (90/46) x 100 = 195.7 kg urea per ha.
7. Mass of SSP = (30/18) x 100 = 166.7 kg SSP per ha.
8. DM yield = 25 x 0.28 = 7.0 t DM per ha.
9. Maize equivalent = (180,000/60,000) x 3 = 3 x 3 = 9 t/ha maize equivalent.

10. (a) Total person-days = 5 x 6 = 30. (b) Productivity = 3/30 = 0.10 ha per person-day.

References and Attributions [Series 3]

Textual References (Books and External Sources)

- Gomez, K. A., & Gomez, A. A. (1984). Statistical procedures for agricultural research (2nd ed.). Wiley-Interscience.
- FAO. (2004). Farm management for Asia: A systems approach. Food and Agriculture Organization of the United Nations, Rome.
- Brady, N. C., & Weil, R. R. (2008). The nature and properties of soils (14th ed.). Prentice Hall.

Figures, Plates, Tables, and Boxes

- Figure 3.1: Volume formulae and worked examples for three Nigerian farm storage structures. Attribution: AI generated. Suggested OER search string: “farm storage structure volume calculation cuboid cylinder cone formula diagram OER”.
- Box 3.1: Fertiliser application rate calculations for common Nigerian fertiliser products. Attribution: AI generated. Suggested OER search string: “fertiliser application rate calculation urea NPK CAN MOP Nigeria hectare table OER”.

Series 4: Graphical and Geometrical Tools

- **Part 4.1: Introduction to Graphs**
  - Sub Part 4.1.1: Purpose and types of graphs in agriculture
  - Sub Part 4.1.2: Components of a graph and rules for graph construction
  - Sub Part 4.1.3: Reading and interpreting agricultural graphs
- **Part 4.2: Bar Charts and Histograms**
  - Sub Part 4.2.1: Simple and compound bar charts
  - Sub Part 4.2.2: Histograms and frequency polygons
  - Sub Part 4.2.3: Worked examples with Nigerian farm data
- **Part 4.3: Line Graphs and Scatter Plots**
  - Sub Part 4.3.1: Line graphs for time-series agricultural data
  - Sub Part 4.3.2: Scatter plots and the concept of correlation
  - Sub Part 4.3.3: Drawing and interpreting the line of best fit
- **Part 4.4: Pie Charts and Pictograms**
  - Sub Part 4.4.1: Constructing and interpreting pie charts
  - Sub Part 4.4.2: Pictograms in agricultural extension
  - Sub Part 4.4.3: Choosing the appropriate graph type

Introduction

Graphs and charts translate numerical data into visual form, making patterns, trends, and comparisons immediately apparent to farmers, extension agents, policymakers, and researchers. A well-constructed bar chart showing variety yield comparisons communicates in seconds what a table of numbers requires minutes to process. This Series covers the major graph types used in Nigerian agricultural communication: bar charts, histograms, line graphs, scatter plots, and pie charts. Geometric tools for setting out farm plots and field experiments are also introduced.

The emphasis is on practical construction skills and correct interpretation. Students learn what each graph type is best suited for, how to draw each one accurately on graph paper, and how to read conclusions from graphs prepared by others. These skills are essential for presenting agricultural research results, preparing extension materials, and interpreting published data.

Series Learning Outcomes

Upon completing this Series, you should be able to:

- **SLO 4.1** state the purpose of graphs in agriculture and describe the essential components of a well-constructed graph (Linked to Part 4.1),
- **SLO 4.2** construct and interpret simple bar charts, compound bar charts, histograms, and frequency polygons from agricultural data (Linked to Part 4.2),
- **SLO 4.3** draw and interpret line graphs for time-series data and scatter plots showing the relationship between two agricultural variables (Linked to Part 4.3),
- **SLO 4.4** construct and interpret pie charts and explain when each graph type is most appropriate (Linked to Part 4.4), and
- **SLO 4.5** draw a line of best fit on a scatter plot and describe the direction and strength of the correlation shown (Linked to Part 4.3).

4.1 Introduction to Graphs

4.1.1 Purpose and types of graphs in agriculture

Graphs serve four main purposes in agricultural communication: (1) comparison (comparing quantities across categories, e.g. yields of different varieties; bar chart is best); (2) distribution (showing how data are spread across a range; histogram is best); (3) trend or time series (showing how a variable changes over time, e.g. monthly rainfall; line graph is best); (4) relationship (showing how two variables are related, e.g. fertiliser rate and yield; scatter plot is best); and (5) composition (showing what proportion each part contributes to the whole, e.g. land use breakdown; pie chart is best).

The main graph types are: bar chart (horizontal or vertical bars of equal width; height proportional to frequency or quantity; for categorical or discrete data); histogram (vertical bars with no gaps between them; area proportional to frequency; for continuous data grouped into class intervals); line graph (points connected by straight lines; for continuous data or time series); scatter plot (individual points plotted for pairs of values of two variables; for showing relationships); and pie chart (circle divided into sectors proportional to percentage; for showing composition or proportions).

Fill in the blank: A ________ chart uses vertical or horizontal bars of equal width to compare quantities across categories; it is the most appropriate graph type for comparing the grain yields of five different maize varieties.

4.1.2 Components of a graph and rules for graph construction

Every well-constructed graph must have: (1) a title (describing what the graph shows, including the crop, variable, location, and year); (2) labelled axes (the x-axis [horizontal] and y-axis [vertical] each with a descriptive label and units); (3) appropriate scale (starting from zero unless otherwise justified; scale divisions uniform and clearly numbered); (4) data points or bars accurately plotted; (5) a legend or key (when more than one data series is shown); and (6) a source note (stating where the data came from).

Rules for scale selection: the scale must accommodate the full range of data without wasting space; scale divisions should be in multiples of 1, 2, 5, or 10 for easy reading; the y-axis normally starts at zero to avoid misleading visual impressions; if the y-axis must be broken (because data do not start near zero), this must be clearly indicated with a break symbol. Graph paper with 1 mm or 2 mm squares is standard for hand-drawn agricultural graphs; computer software (Microsoft Excel, Google Sheets) is increasingly used.

Fill in the blank: The ________ axis (horizontal axis) of a graph carries the independent variable (e.g. fertiliser rate or crop variety); the y-axis (vertical axis) carries the dependent variable (e.g. grain yield).

4.1.3 Reading and interpreting agricultural graphs

To read a graph: (1) read the title to understand what is being shown; (2) read the axis labels and units; (3) note the scale on each axis; (4) identify the highest and lowest values; (5) identify trends (increasing, decreasing, stable, or fluctuating); (6) compare categories or data series if more than one is shown; (7) draw conclusions relevant to the agricultural context.

Common misinterpretations to avoid: a steep slope on a line graph does not always indicate a large absolute change; it depends on the y-axis scale. A bar chart in which the y-axis does not start at zero exaggerates differences between bars. Correlation shown on a scatter plot does not imply causation. Every graph must be interpreted in its agricultural context, not in isolation from the underlying data and the conditions under which they were collected.

Fill in the blank: When interpreting a line graph of monthly rainfall, the first step is to read the ________ to understand what is being shown; subsequent steps identify the axis labels, scale, maximum and minimum values, and the overall trend.

Pilot questions 4.1

11. State the five main purposes of graphs in agricultural communication and name the most appropriate graph type for each.

12. List six essential components that every well-constructed agricultural graph must have.
13. What scale rules should be followed when constructing the y-axis of a bar chart showing crop yields?
14. Explain two common errors in graph construction that can mislead the reader.
15. Which graph type would you choose to show (a) monthly rainfall over 12 months; (b) the proportion of a farm under different crops; (c) the relationship between fertiliser rate and maize yield?

4.2 Bar Charts and Histograms

4.2.1 Simple and compound bar charts

A **simple bar chart** uses one bar for each category, with the height of the bar representing its quantity. To construct one, draw and label the horizontal axis with the categories and the vertical axis with the measured quantity and unit. Choose a suitable scale that includes the highest value, then draw bars of equal width with spaces between them. Label each category and give the chart an appropriate title.

For example, if the yields are **SAMMAZ 14 = 4.2 t/ha**, **SAMMAZ 15 = 5.1 t/ha**, and **Local variety = 2.3 t/ha**, the three bars should reach heights of 4.2, 5.1, and 2.3 on the vertical axis. The chart will clearly show that SAMMAZ 15 has the highest yield, while the local variety has the lowest.

A compound (grouped) bar chart displays two or more data series side by side for each category, allowing comparison across both categories and series. Each series is given a different shading or colour and a legend is provided. Example: comparing grain yields of three maize varieties in 2022 and 2023 growing seasons; for each variety, two bars (one per year) are placed adjacent. A stacked bar chart shows the same data with bars subdivided to show the contribution of each component to the total. Both compound and stacked bar charts require a clear legend.

Fill in the blank: A ________ (grouped) bar chart displays two or more data series side by side for each category, requiring a legend to identify each series; it is used to compare yields of the same varieties across two or more growing seasons.

4.2.2 Histograms and frequency polygons

A histogram displays the frequency distribution of continuous data grouped into class intervals (bins). Unlike a bar chart, there are no gaps between bars because the data are continuous. Bar width equals the class interval width; bar height equals the frequency (or frequency density for unequal class widths). To construct: (1) group data into 5 to 10 class intervals of equal width; (2) count the frequency in each class; (3) draw adjacent bars from the lower class boundary to the upper class boundary; (4) label x-axis with class boundaries; (5) label y-axis as frequency; (6) add title.

Example: 20 cassava yield measurements (t/ha): 14, 18, 22, 15, 20, 17, 25, 23, 19, 16, 21, 24, 18, 20, 22, 14, 17, 23, 19, 21. Class intervals: 13-16 (f=3), 17-20 (f=8), 21-24 (f=7), 25-28 (f=2). A frequency polygon is obtained by connecting the midpoints of the tops of histogram bars with straight lines and closing the polygon at both ends. Frequency polygons from two or more datasets can be superimposed on the same axes to compare distributions.

Fill in the blank: In a histogram, there are no ________ between adjacent bars because the data are continuous and grouped into class intervals; bar height represents the frequency of observations within each class interval.

4.2.3 Worked examples with Nigerian farm data

Worked example (simple bar chart): Monthly rainfall (mm) at Ibadan for January to June: 15, 30, 75, 130, 155, 185. X-axis: months (Jan to Jun); y-axis: rainfall (mm), scale 0 to 200 mm with divisions at 50 mm. Six bars are drawn; the tallest (June) reaches 185 mm. Title: ‘Monthly rainfall at Ibadan, January to June 2023’. Interpretation: rainfall increases progressively from

January to June, reflecting the onset and development of the main growing season in the forest-savanna transition zone.

Worked example (histogram): Plant heights (cm) of 25 cowpea plants: 35, 40, 38, 42, 37, 44, 36, 41, 39, 43, 38, 40, 36, 42, 37, 45, 39, 41, 38, 43, 40, 37, 42, 39, 41. Class intervals (5 cm wide): 35-39 (f=10), 40-44 (f=13), 45-49 (f=2). Histogram drawn with three bars; tallest bar is 40-44 cm class (f=13). Title: 'Height distribution of 25 cowpea plants at flowering stage'. Interpretation: most plants fall in the 40-44 cm class; the distribution is approximately symmetric with a slight right tail.

Fill in the blank: For the plant height histogram with class intervals 35-39 (f=10), 40-44 (f=13), and 45-49 (f=2), the tallest bar corresponds to the ________ class interval, which contains 13 observations.

Pilot questions 4.2

11. Draw a simple bar chart for the following maize variety yields (t/ha): Variety A = 3.5; Variety B = 4.8; Variety C = 2.9; Variety D = 5.2. (Describe or sketch the steps.)
12. Explain the difference between a bar chart and a histogram in terms of data type, bar spacing, and what the bar height represents.
13. The following are grain weights (g per ear) for 20 maize plants: 95, 110, 88, 102, 115, 98, 107, 92, 120, 105, 112, 96, 108, 90, 118, 103, 94, 116, 100, 125. Group into class intervals 85-94, 95-104, 105-114, 115-124, 125-134 and construct a frequency table.
14. What is a frequency polygon and how is it obtained from a histogram?
15. A compound bar chart compares poultry egg production (eggs per hen per week) in two seasons (dry and wet) for four breeds. List the steps to construct it.

4.3 Line Graphs and Scatter Plots

4.3.1 Line graphs for time-series agricultural data

A line graph shows how a variable changes over time or another continuous dimension. It is the appropriate choice when data are recorded at regular time intervals: weekly milk yield, monthly rainfall, daily temperature, or annual crop production trends. To construct: (1) place time on the x-axis; (2) place the measured variable on the y-axis; (3) plot each data point; (4) connect consecutive points with straight lines; (5) label axes, add title and scale. Multiple lines on the same graph (for different varieties, farms, or years) require different line styles (solid, dashed, dotted) and a legend.

Reading a line graph: identify the overall trend (upward, downward, or cyclical); identify peaks and troughs (when did yield or rainfall reach maximum and minimum?); compare multiple lines if present (which variety performed best across the season?). Example: weekly milk yield (litres/cow) over 8 weeks: 18, 20, 22, 21, 19, 17, 16, 14. The graph shows a peak at week 3 (22 L) followed by a progressive decline, characteristic of the descending phase of a lactation curve.

Fill in the blank: In a line graph of weekly milk yield over 8 weeks, the ________ axis carries the time variable (weeks) and the y-axis carries the measured variable (milk yield in litres per cow per week).

4.3.2 Scatter plots and the concept of correlation

A scatter plot (scattergram) plots paired values of two quantitative variables as points on a coordinate system; the independent variable (x) is on the horizontal axis and the dependent variable (y) is on the vertical axis. Each data pair (xi, yi) is represented by one point. Scatter plots reveal: (1) positive correlation (as x increases, y tends to increase; points slope upward from left to right; e.g. fertiliser rate and yield); (2) negative correlation (as x increases, y tends to decrease; e.g. weed density and yield); (3) no correlation (points randomly scattered; no discernible pattern); (4) curvilinear relationship (points follow a curve rather than a straight line).

Correlation strength is described qualitatively from the scatter plot: if points cluster tightly around a straight line, correlation is strong; if they are widely scattered around the trend, correlation is weak. The correlation coefficient (r) is the numerical measure of linear correlation, ranging from -1 (perfect negative) through 0 (no linear correlation) to +1 (perfect positive); its calculation is covered in Series 5. An important caution: correlation does not imply causation; two variables may be strongly correlated because both are caused by a third variable.

Fill in the blank: A scatter plot in which the points tend to slope upward from left to right indicates a ________ correlation between the two variables; as the x variable increases, the y variable also tends to increase.

4.3.3 Drawing and interpreting the line of best fit

The line of best fit (regression line; trend line) is the straight line that most closely represents the trend in a scatter plot. It is drawn by eye so that approximately equal numbers of points lie above and below the line, and the line passes through the mean point (x-bar, y-bar) of the data. The line extends only across the range of the observed x values; extrapolation beyond this range is unreliable. To draw: (1) calculate the mean of x (x-bar) and the mean of y (y-bar); (2) plot the mean point; (3) draw a straight line through the mean point balancing the scatter of points above and below.

Reading the line of best fit: the slope (gradient) of the line indicates how much y changes for each unit increase in x. A positive slope indicates a positive relationship; a steeper slope indicates a stronger rate of change. Example: if the line of best fit for fertiliser rate (x, kg/ha) vs maize yield (y, t/ha) has a slope of 0.015, each additional kg of fertiliser per ha increases yield by 0.015 t/ha (15 kg/ha). The y-intercept is the predicted y-value when x = 0. These interpretations form the foundation for the regression analysis covered in Series 5.

Fill in the blank: The line of best fit on a scatter plot is drawn so that approximately equal numbers of data points lie above and below the line, and it passes through the ________ point of the data (the point defined by the mean of x and the mean of y).

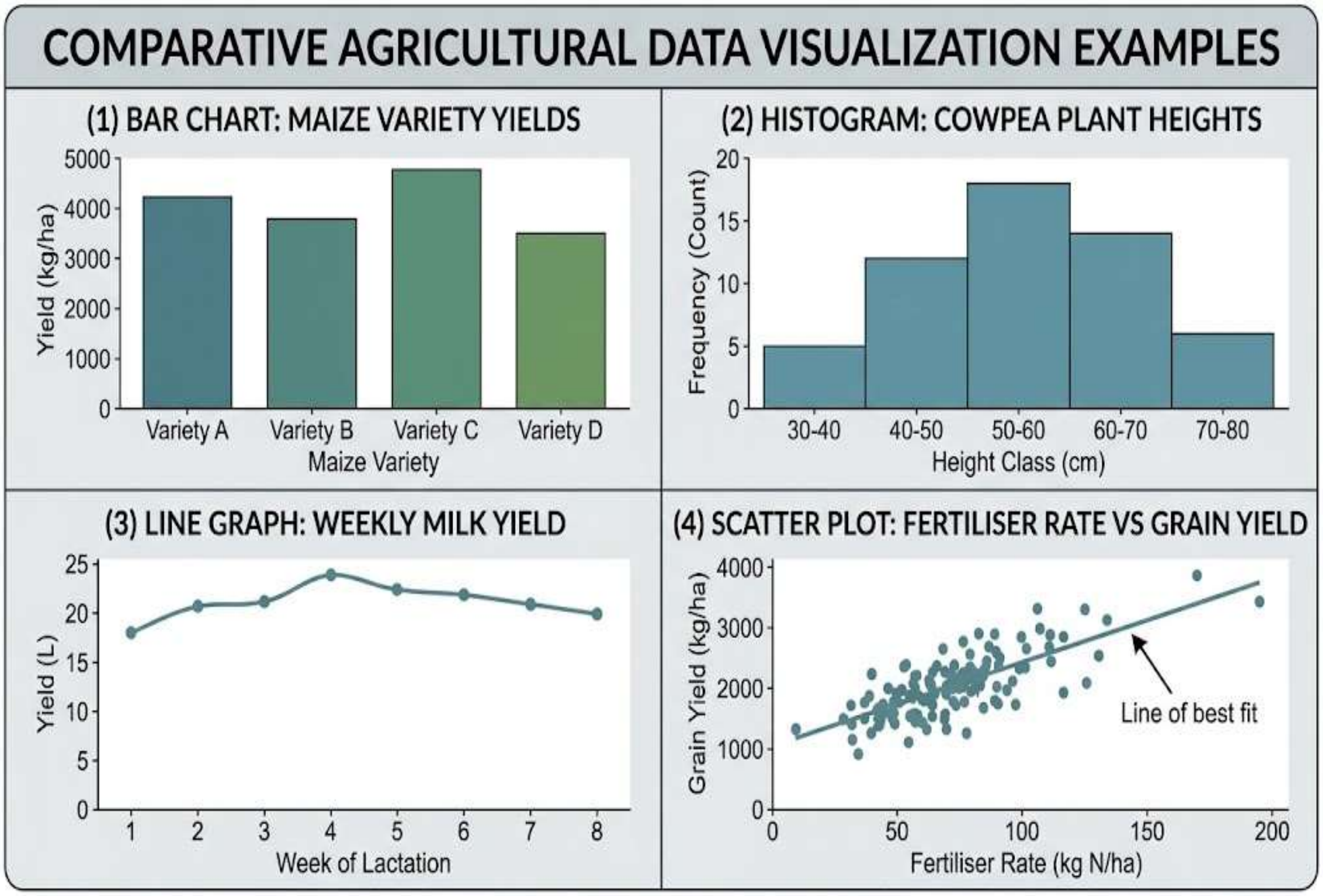

Figure 4.1: Four common graph types used in Nigerian agricultural data presentation

Pilot questions 4.3

11. Monthly rainfall (mm) at Kano for January to June: 0, 1, 5, 20, 60, 95. Draw a line graph (describe steps) and describe the trend.
12. Explain positive correlation, negative correlation, and no correlation with one agricultural example of each.
13. The following data show fertiliser rate (kg N/ha) and maize yield (t/ha): (0, 1.8); (30, 2.5); (60, 3.2); (90, 3.8); (120, 4.1). Plot the scatter diagram and describe the correlation.
14. How is a line of best fit drawn on a scatter plot by eye? What information can be read from its slope?
15. Explain why correlation does not imply causation and give one agricultural example.

4.4 Pie Charts and Pictograms

4.4.1 Constructing and interpreting pie charts

A pie chart is a circle divided into sectors, where the angle of each sector is proportional to the percentage of the total it represents. Total angle of a circle = 360 degrees. Sector angle for a category = (category value / total) x 360 degrees. To construct: (1) calculate the percentage and sector angle for each category; (2) draw a circle of suitable radius; (3) use a protractor to measure and draw each sector, starting from the 12 o'clock position and working clockwise; (4) label each sector with the category name and percentage; (5) add title.

Worked Example: Pie Chart

A 10-hectare farm is divided into food crops, cash crops, pasture, and homestead with paths. Food crops occupy 4 ha, representing 40% of the farm; cash crops occupy 3 ha or 30%; pasture occupies 2 ha or 20%; while the homestead and paths occupy 1 ha or 10%.

The sector angle for each category is calculated as:

Sector angle = Percentage × 360°

Food crops: **0.40 × 360° = 144°**
Cash crops: **0.30 × 360° = 108°**
Pasture: **0.20 × 360° = 72°**
Homestead and paths: **0.10 × 360° = 36°**

Check:

144° + 108° + 72° + 36° = 360°

Therefore, the calculation is correct. The pie chart shows that food crops occupy the largest share of the farm, while the homestead and paths occupy the smallest. Pie charts are most suitable for showing how different parts contribute to a whole and are clearest when they contain about three to seven categories.

Fill in the blank: The sector angle for a category in a pie chart = (category value / total) x ________ degrees; if a crop occupies 25% of the total farm area, its sector angle = 0.25 x 360 = 90 degrees.

4.4.2 Pictograms in agricultural extension

A pictogram (pictograph) uses symbols or icons to represent quantities; one symbol represents a fixed number of units (stated in the key). Example: one ear of maize symbol = 500 kg/ha; a variety yielding 2,500 kg/ha is represented by 5 ear symbols. Pictograms are particularly useful in agricultural extension communication with low-literacy audiences because visual symbols are more accessible than numbers. They are commonly used in IITA and FAO extension leaflets, community notice boards, and farmer training materials in Nigeria.

To construct a pictogram: (1) choose a symbol appropriate to the commodity; (2) state clearly in a key how many units one symbol represents; (3) use whole and part symbols consistently (e.g. half symbol for half the unit quantity); (4) align symbols horizontally or vertically; (5) add title and labels. Limitations: pictograms are less precise than bar charts or tables; part-symbols are difficult to read accurately; they are unsuitable for data with many categories or a wide range of values.

Fill in the blank: In a pictogram, a ________ states how many units each symbol represents; for example, if one symbol represents 500 kg/ha and a variety yields 2,000 kg/ha, it would be represented by 4 symbols.

4.4.3 Choosing the appropriate graph type

The choice of graph type depends on: (1) the type of data (categorical vs continuous); (2) the purpose (comparison, distribution, trend, relationship, or composition); (3) the audience (literate vs less-literate; scientific publication vs extension leaflet). Summary: bar chart (comparison of categorical or discrete data); histogram (distribution of continuous data grouped into class intervals); line graph (trend over time or along a continuous dimension); scatter plot (relationship between two quantitative variables); pie chart (composition or proportion of a whole); pictogram (simple quantities for non-literate audiences).

A common error is using a pie chart to compare absolute quantities rather than proportions; pie charts should only be used when the parts add up to a meaningful whole. Another error is using a histogram when a bar chart is appropriate (i.e. for categorical rather than continuous data). Selecting the wrong graph type obscures the message and can mislead the reader; correct graph selection is a professional skill expected of agricultural officers and researchers.

Fill in the blank: A ________ chart is appropriate when the purpose is to show how a farm's total land area is distributed among different enterprises; it is most effective when there are between 3 and 7 categories that together account for 100 percent of the total.

Data Presentation Situation	**Recommended Graph Type**
Crop yield comparison across regions	Bar chart — clearly shows yield differences among locations.
Trend of rainfall over months	Line graph — ideal for continuous time-series data.
Proportion of land under different crops	Pie chart — effectively displays percentage share of each crop.
Relationship between fertiliser rate and yield	Scatter plot — reveals correlation or pattern between two variables.
Distribution of farm sizes among households	Histogram — shows frequency distribution of farm sizes.
Comparison of multiple crop yields over years	Grouped bar or line chart — compares several datasets across time.

Box 4.1: Guide to selecting the appropriate graph type for agricultural data

Pilot questions 4.4

11. A farmer's income is: crops N360,000 (60%); livestock N150,000 (25%); off-farm income N90,000 (15%). Calculate the sector angles and describe how to construct the pie chart.
12. Explain what a pictogram is and state one advantage and one limitation of pictograms compared with bar charts.
13. For each of the following, state the most appropriate graph type and give a reason: (a) monthly egg production over 12 months; (b) the distribution of body weights of 50 goats; (c) the proportion of a budget spent on seeds, fertiliser, and labour.
14. List three errors commonly made in pie chart construction.
15. An extension officer wants to show village farmers the yield advantage of an improved rice variety over the local variety. Which graph type is most suitable and why?

Series Summary

This Series covered the use of graphs to present agricultural data visually. The five main purposes of graphs are comparison, distribution, trend, relationship, and composition. Bar charts are suitable for comparing categories, histograms show the distribution of continuous data, line graphs display changes over time, scatter plots show relationships between two quantitative variables, and pie charts present parts of a whole. Every graph should have a clear title, labelled axes with units, an appropriate scale, accurately plotted values, a legend where necessary, and the data source. Bar charts use equal-width bars separated by gaps, while compound bar charts compare more than one data series. Histograms use adjoining bars for grouped continuous data, and a frequency polygon may be formed by connecting the midpoints of the bars. Line graphs connect points in sequence, usually with time on the horizontal axis, and multiple lines should be distinguished with different line styles and a legend.

Scatter plots display paired values and may reveal positive, negative, or no correlation. A line of best fit should pass near the mean point **($\bar{x}$, $\bar{y}$)** and balance the observations above and below it, while its slope shows the rate at which one variable changes with another. However, correlation does not prove causation. For a pie chart, the sector angle is calculated as **Sector angle = (Value ÷ Total) × 360°**, and the chart is most effective with about three to seven categories. Pictograms use symbols to represent fixed quantities and are useful for extension audiences with limited literacy, though they are less precise than standard graphs. The correct graph should always match the type of data and the purpose of communication; for example, pie charts should not be used for precise absolute comparisons, and histograms should not be used for categorical data.

Glossary of Terms

- **Bar chart:** a graph using vertical or horizontal bars of equal width to compare quantities across categories; gaps separate bars; height represents quantity; appropriate for categorical or discrete data.
- **Compound bar chart:** a bar chart with two or more bars side by side for each category, comparing multiple data series; requires a legend to identify each series.

- **Correlation:** the statistical relationship between two quantitative variables; positive (both increase together), negative (one increases as the other decreases), or zero (no linear relationship).
- **Frequency polygon:** a graph obtained by connecting the midpoints of the tops of histogram bars with straight lines; used to display and compare frequency distributions.
- **Histogram:** a graph for continuous data grouped into class intervals; bars are adjacent with no gaps; bar height represents frequency; area of each bar is proportional to frequency.
- **Legend (key):** a box on a graph identifying the shading, colour, or line style used for each data series when two or more series are plotted on the same graph.
- **Line graph:** a graph connecting successive data points with straight lines; the x-axis carries time or another continuous variable; appropriate for time-series data.
- **Line of best fit:** a straight line drawn through a scatter plot so that equal numbers of points lie above and below it; passes through the mean point (x-bar, y-bar); used to describe the trend and predict values.
- **Pictogram:** a graph using standardised symbols or icons to represent quantities; each symbol represents a fixed number of units; most useful in extension communication with low-literacy audiences.
- **Pie chart:** a circle divided into sectors proportional to the percentage each category contributes to the total; sector angle = (value/total) x 360 degrees; used for composition data.
- **Scatter plot:** a graph in which each data pair (x, y) is plotted as a point on a coordinate system; used to reveal the nature and strength of the relationship between two quantitative variables.

Concept Check Questions [Series 4]

Objective Questions

11. The most appropriate graph for showing how a variable changes over time (e.g. monthly rainfall) is a ________ graph. (Linked to SLO 4.1)
12. In a histogram, there are no ________ between adjacent bars because the data are continuous and grouped into class intervals. (Linked to SLO 4.2)
13. The sector angle for a category in a pie chart = (category value / total) x ________ degrees. (Linked to SLO 4.4)
14. A scatter plot in which points slope upward from left to right indicates a ________ correlation between the two variables. (Linked to SLO 4.3)
15. The line of best fit on a scatter plot must pass through the ________ point, defined by the mean of x and the mean of y. (Linked to SLO 4.5)

Short-Answer Questions

16. List the six essential components of a well-constructed agricultural graph. (Linked to SLO 4.1)
17. Distinguish between a bar chart and a histogram in terms of data type, bar spacing, and what bar height represents. (Linked to SLO 4.2)
18. Weekly milk yields (litres/cow) over 6 weeks: 18, 21, 23, 20, 17, 15. List the steps to draw a line graph and describe the trend. (Linked to SLO 4.3)
19. A farm income of N600,000 comes from: crops 50%; livestock 30%; poultry 20%. Calculate the sector angles for a pie chart. (Linked to SLO 4.4)
20. What is a scatter plot and what does the slope of its line of best fit indicate? (Linked to SLOs 4.3, 4.5)

Long-Answer Questions

15. Describe the five main graph types used in agriculture, stating the data type, purpose, and one agricultural example for each. Explain the rules for good graph construction. (Linked to SLOs 4.1, 4.2, 4.4)
16. Explain scatter plots, correlation, and the line of best fit. Describe pie charts and pictograms. Explain how to select the appropriate graph type for a given agricultural data situation. (Linked to SLOs 4.3, 4.4, 4.5)

Answers to Concept Check Questions [Series 4]

Objective Questions

11. Line.
12. Gaps.
13. 360.
14. Positive.
15. Mean.

Short-Answer Questions

16. Title; labelled axes with units; appropriate zero-based scale; accurately plotted data; legend (if multiple series); data source.
17. Bar chart: categorical or discrete data; gaps between bars; height = quantity. Histogram: continuous data in class intervals; no gaps between bars; height = frequency; area proportional to frequency.
18. Steps: x-axis = weeks (1 to 6); y-axis = yield (litres per cow, scale 0 to 25); plot points at (1,18), (2,21), (3,23), (4,20), (5,17), (6,15); connect with lines; add title and labels. Trend: yield rises to a peak at week 3 then declines, characteristic of a lactation curve.
19. Crops: 0.50 x 360 = 180 degrees. Livestock: 0.30 x 360 = 108 degrees. Poultry: 0.20 x 360 = 72 degrees. Check: 180+108+72 = 360 degrees.
20. A scatter plot displays paired values of two quantitative variables as points on a coordinate system, revealing the direction and strength of their relationship. The slope of the line of best fit indicates the rate of change of y for each unit increase in x; a positive slope indicates a positive relationship; a steeper slope indicates a faster rate of change.

Long-Answer Questions

15. Five graph types: (1) Bar chart: categorical/discrete data; comparison; e.g. yields of five maize varieties. (2) Histogram: continuous data in class intervals; distribution; e.g. distribution of 30 plant height measurements. (3) Line graph: continuous/time data; trend; e.g. monthly rainfall over 12 months. (4) Scatter plot: two quantitative variables; relationship; e.g. fertiliser rate vs maize yield. (5) Pie chart: composition; 3 to 7 categories summing to 100%; e.g. land use breakdown on a farm. Rules for good graph

construction: (1) include a descriptive title; (2) label both axes with units; (3) y-axis starts at zero (or show break if not); (4) use uniform scale divisions (multiples of 1, 2, 5, or 10); (5) include a legend for multiple series; (6) state the data source. Bars in bar chart have equal width and gaps; bars in histogram are adjacent. Scale must accommodate all data without wasting space.

16. Scatter plot: each data pair (x, y) plotted as a point; independent variable on x-axis, dependent variable on y-axis. Correlation: positive (points slope upward; e.g. fertiliser and yield); negative (points slope downward; e.g. weed density and yield); none (random scatter). Strength: tight clustering = strong; wide scatter = weak. Correlation coefficient r ranges from -1 to +1. Line of best fit: straight line balancing points above and below; passes through mean point (x-bar, y-bar); slope = rate of change of y per unit x; only valid within observed x range; extrapolation unreliable. Correlation does not imply causation. Pie chart: sector angle = (value/total) x 360; draw sectors with protractor from 12 o'clock clockwise; label with name and percentage; effective for 3 to 7 categories. Pictogram: symbols represent fixed quantities (key essential); accessible to low-literacy audiences; less precise than bar chart; unsuitable for many categories. Graph selection: bar (comparison of categories); histogram (continuous distribution); line (time trend); scatter (relationship); pie (composition); pictogram (extension to low-literacy farmers).

Answers to Pilot Questions [Series 4]

Part 4.1

11. Comparison (bar chart); distribution (histogram); trend/time series (line graph); relationship (scatter plot); composition (pie chart).
12. Title; labelled x-axis with units; labelled y-axis with units; appropriate scale (starting from zero); accurately plotted data or bars; legend (if multiple series); data source.
13. Scale should start at zero; divisions in multiples of 1, 2, 5, or 10 for easy reading; scale must accommodate the highest yield value without excessive blank space.
14. Two common errors: (1) y-axis not starting at zero, which exaggerates differences between bars; (2) unequal bar widths, which visually misrepresents the data.

15. (a) Monthly rainfall over 12 months: line graph. (b) Proportion of a farm under different crops: pie chart. (c) Relationship between fertiliser rate and maize yield: scatter plot.

Part 4.2

11. Steps: draw x-axis (varieties A, B, C, D); draw y-axis (yield t/ha, scale 0 to 6, divisions at 1 t/ha); draw bars of equal width with gaps at heights 3.5, 4.8, 2.9, 5.2; label bars; add title 'Maize variety yield comparison (t/ha)'. Variety D has the tallest bar; Variety C the shortest.
12. Bar chart: categorical data; gaps between bars; bar height = quantity. Histogram: continuous data grouped into class intervals; no gaps between bars; bar height = frequency of observations in each class.
13. Frequency table: 85-94 (f=3: 88, 92, 90); 95-104 (f=6: 95, 102, 98, 96, 100, 103); 105-114 (f=6: 110, 107, 105, 112, 108, 103); 115-124 (f=4: 115, 120, 118, 116); 125-134 (f=1: 125). Total = 20.
14. A frequency polygon is obtained from a histogram by marking the midpoint of the top of each bar and connecting the midpoints with straight lines; the polygon is closed at both ends by extending lines to the baseline at a distance of half a class width beyond the outer bars.
15. Steps: (1) draw x-axis (breeds 1-4); (2) draw y-axis (eggs per hen per week); (3) for each breed, draw two adjacent bars of different shading (one for dry season, one for wet season); (4) add legend identifying each shading; (5) add title and data source.

Part 4.3

11. Steps: x-axis (months Jan to Jun, 1 to 6); y-axis (rainfall mm, 0 to 100 mm); plot points (1,0), (2,1), (3,5), (4,20), (5,60), (6,95); connect with lines; title: 'Monthly rainfall at Kano, January to June 2023'. Trend: rainfall is near zero January to March, then rises steeply from April onward, marking the onset of the rainy season.
12. Positive correlation: as x increases, y increases; example: fertiliser application rate and maize grain yield. Negative correlation: as x increases, y decreases; example: weed density and crop yield. No correlation: no pattern between x and y; example: a farmer's surname initial and his/her maize yield.

13. Plot points: (0, 1.8); (30, 2.5); (60, 3.2); (90, 3.8); (120, 4.1). Points slope upward from left to right, indicating a strong positive correlation: as fertiliser rate increases, maize yield increases.
14. Calculate x-bar (mean of x values) and y-bar (mean of y values); plot the mean point; draw a straight line through the mean point so that approximately equal numbers of points fall above and below the line. The slope of the line indicates the increase in y for each unit increase in x; a steeper slope means yield responds more strongly to each additional kg of fertiliser.
15. Correlation shows that two variables tend to move together but does not prove one causes the other; a third factor may be driving both. Example: rainfall and both crop yield and weeding frequency may all increase together; high rainfall causes higher yield directly and also causes faster weed growth (requiring more weeding), but weeding frequency and yield are correlated without weeding causing the yield increase directly.

Part 4.4

11. Crops: (360/600) x 360 = 216 degrees. Livestock: (150/600) x 360 = 90 degrees. Off-farm: (90/600) x 360 = 54 degrees. Check: 216+90+54 = 360 degrees. Draw a circle, mark 12 o'clock, measure 216 degrees clockwise for crops; from there measure 90 degrees for livestock; remaining 54 degrees for off-farm. Label each sector with name and percentage; add title.
12. A pictogram uses symbols or icons where each symbol represents a fixed quantity stated in a key. Advantage: accessible to low-literacy extension audiences who cannot easily read bar charts. Limitation: less precise than bar charts; part-symbols are difficult to read accurately.
13. (a) Monthly egg production over 12 months: line graph; shows the trend of a variable over time. (b) Distribution of body weights of 50 goats: histogram; shows frequency distribution of continuous data. (c) Proportion of budget on seeds, fertiliser, and labour: pie chart; shows composition (parts of a whole totalling 100%).
14. Three common pie chart errors: (1) sector angles do not add up to 360 degrees due to rounding; (2) categories do not represent a meaningful whole (pie charts are inappropriate

when parts do not sum to a total); (3) too many categories (more than 7), making sectors too small to label or interpret clearly.

15. A simple bar chart is most suitable; two bars (improved variety and local variety) are placed side by side or as a compound bar chart if compared across multiple sites; bar height clearly shows the yield advantage; gaps between bars emphasise the categorical comparison. A line graph would be inappropriate as there is no time dimension; a pie chart would be inappropriate as the two yields are not parts of a whole.

References and Attributions [Series 4]

Textual References (Books and External Sources)

- Gomez, K. A., & Gomez, A. A. (1984). Statistical procedures for agricultural research (2nd ed.). Wiley-Interscience.
- Tufte, E. R. (2001). The visual display of quantitative information (2nd ed.). Graphics Press.
- FAO. (2018). Data collection guidelines for crop monitoring. Food and Agriculture Organization of the United Nations, Rome.

Figures, Plates, Tables, and Boxes

- Figure 4.1: Four common graph types used in Nigerian agricultural data presentation. Attribution: AI generated. Suggested OER search string: "bar chart histogram line graph scatter plot agricultural data examples Nigeria OER".
- Box 4.1: Guide to selecting the appropriate graph type for agricultural data. Attribution: AI generated. Suggested OER search string: "graph type selection bar histogram line scatter pie pictogram agricultural data guide table OER".

Series 5: Scientific Analysis and Communication

- **Part 5.1: Hypothesis Testing and Experimental Analysis**
  - Sub Part 5.1.1: Null and alternative hypotheses
  - Sub Part 5.1.2: The t-test for comparing two means
  - Sub Part 5.1.3: Chi-square test for categorical data
- **Part 5.2: Regression and Correlation**
  - Sub Part 5.2.1: Simple linear regression: the equation of a straight line
  - Sub Part 5.2.2: The correlation coefficient
  - Sub Part 5.2.3: Interpretation and limitations of regression
- **Part 5.3: Presenting Agricultural Research Results**
  - Sub Part 5.3.1: Tables: construction and use
  - Sub Part 5.3.2: Figures and graphs in research reports
  - Sub Part 5.3.3: Describing results in writing
- **Part 5.4: Scientific Literacy and Evaluation of Agricultural Information**
  - Sub Part 5.4.1: Reading and evaluating agricultural research papers
  - Sub Part 5.4.2: Evaluating statistical claims in agricultural reports
  - Sub Part 5.4.3: Communicating biometric findings to farmers

Introduction

The culminating skill in farm biometrics is the ability to analyse data statistically, present results clearly, and communicate findings to different audiences. Series 5 brings together the descriptive statistics of Series 2, the measurement concepts of Series 3, and the graphical skills of Series 4, and adds the inferential statistical tools (hypothesis testing, regression, and correlation) needed to draw valid conclusions from agricultural data.

This Series also addresses the communication dimension of biometrics: how to construct clear tables and write results sections, how to evaluate statistical claims in agricultural publications, and how to present biometric findings to farmers in accessible language. These competencies equip the student to function as an independent agricultural researcher or extension professional, able to generate, analyse, and communicate biometric evidence effectively.

Series Learning Outcomes

Upon completing this Series, you should be able to:

- **SLO 5.1** state null and alternative hypotheses for simple agricultural experiments and explain what a p-value means (Linked to Part 5.1),
- **SLO 5.2** perform and interpret an independent-samples t-test for comparing two treatment means (Linked to Part 5.1),
- **SLO 5.3** perform and interpret a chi-square test of independence for categorical agricultural data (Linked to Part 5.1),
- **SLO 5.4** calculate and interpret the simple linear regression equation and the correlation coefficient (Linked to Part 5.2),
- **SLO 5.5** construct well-formatted tables and write clear results descriptions for agricultural research data (Linked to Part 5.3), and
- **SLO 5.6** evaluate statistical claims in agricultural reports and communicate biometric findings to non-specialist audiences (Linked to Part 5.4).

5.1 Hypothesis Testing and Experimental Analysis

5.1.1 Null and alternative hypotheses

A statistical hypothesis is a claim about a population parameter. In agricultural experiments, we typically compare two or more treatments and ask: is the observed difference real, or could it be due to chance variation? The null hypothesis (H0) states that there is no difference (or no effect): e.g. H0: the mean yield of variety A equals the mean yield of variety B (H0: muA = muB). The alternative hypothesis (H1 or Ha) states the opposite: H1: muA is not equal to muB (two-tailed) or H1: muA > muB (one-tailed).

The p-value is the probability of observing a result as extreme as (or more extreme than) the one obtained, assuming H0 is true. A small p-value (typically $p < 0.05$) means the result is unlikely under H0, so we reject H0 in favour of H1 and conclude the difference is statistically significant. A large p-value ($p > 0.05$) means the result is plausible under H0, so we fail to reject H0; we cannot conclude there is a real difference. The significance level (alpha = 0.05 is conventional in agricultural research) is set before the analysis and defines the threshold for rejection.

Fill in the blank: The ________ hypothesis (H0) states that there is no difference between treatments; if the p-value is less than 0.05, we reject H0 and conclude that the observed difference is statistically significant at the 5 percent level.

5.1.2 The t-test for comparing two means

The independent-samples t-test determines whether the means of two independent groups differ significantly. It is appropriate when data are quantitative, approximately normally distributed, and the two groups have similar variances. The t-statistic is: t = (mean1 - mean2) / SE_diff, where SE_diff = square root of (s12/n1 + s22/n2); s1 and s2 are the standard deviations of the two groups and n1, n2 are their sample sizes. The calculated t is compared with the critical t-value from a t-distribution table at the chosen significance level (alpha = 0.05) and degrees of freedom (df = n1 + n2 - 2).

Independent-Samples t-Test

An independent-samples t-test was used to determine whether the mean yields of Variety A and Variety B differed significantly.

Variety A produced yields of 3.2, 3.5, 3.8, 3.4, and 3.6 t/ha, with:

$n_1 = 5$
$\bar{x}_1 = 3.50$ t/ha
$s_1 = 0.228$ t/ha

Variety B produced yields of 2.8, 3.0, 2.9, 3.1, and 2.7 t/ha, with:

$n_2 = 5$
$\bar{x}_2 = 2.90$ t/ha
$s_2 = 0.158$ t/ha

The null hypothesis was:

$H_0: \mu_1 = \mu_2$

The alternative hypothesis was:

$H_1: \mu_1 \neq \mu_2$

The standard error of the difference between the two means was calculated as:

$SE\_diff = \sqrt{[(s_1^2 / n_1) + (s_2^2 / n_2)]}$

$SE\_diff = \sqrt{[(0.228^2 / 5) + (0.158^2 / 5)]}$

$SE\_diff = \sqrt{(0.0104 + 0.0050)}$

$SE\_diff = \sqrt{0.0154}$

$SE\_diff = 0.124$

The calculated t-value was:

$t = (\bar{x}_1 - \bar{x}_2) / SE\_diff$

$t = (3.50 - 2.90) / 0.124$

$t = 4.84$

The degrees of freedom were:

$df = n_1 + n_2 - 2$

$df = 5 + 5 - 2$

$df = 8$

At $\alpha = 0.05$ for a two-tailed test, the critical t-value was:

$t\_critical = 2.306$

Since $|4.84| > 2.306$, the null hypothesis was rejected.

Therefore, there was a statistically significant difference between the mean yields of Variety A and Variety B, with Variety A producing the higher mean yield.

Fill in the blank: The degrees of freedom for an independent-samples t-test comparing two groups of n1 and n2 observations is df = n1 + n2 - ________; for two groups of 5 observations each, df = 5 + 5 - 2 = 8.

5.1.3 Chi-square test for categorical data

The chi-square (chi2) test of independence determines whether two categorical variables are associated. It is used for count (frequency) data, such as: do the proportions of disease-affected plants differ between two fertiliser treatments? The test compares observed frequencies (O) with expected frequencies (E) calculated assuming no association. chi2 = sum of [(O - E)2 / E] for all cells of the contingency table. E for each cell = (row total x column total) / grand total.

Chi-Square Test of Association

A chi-square test of independence was conducted to determine whether fungicide treatment was associated with disease incidence in cowpea plants. A total of 200 plants were examined: 100 received fungicide treatment, while 100 served as untreated controls.

The observed frequencies were:

Treatment group	**Healthy**	**Diseased**	**Total**
Fungicide-treated	80	20	100
Control	50	50	100
Total	130	70	200

The hypotheses were:

H_0: Fungicide treatment and disease incidence are independent.

H_1: Fungicide treatment and disease incidence are significantly associated.

The expected frequency for each cell was calculated using:

Expected frequency = (Row total × Column total) / Grand total

Thus:

Expected frequency for fungicide-treated and healthy plants:

$E = (100 \times 130) / 200 = 65$

Expected frequency for fungicide-treated and diseased plants:

$E = (100 \times 70) / 200 = 35$

Expected frequency for control and healthy plants:

$E = (100 \times 130) / 200 = 65$

Expected frequency for control and diseased plants:

$E = (100 \times 70) / 200 = 35$

The chi-square statistic was calculated as:

$\chi^2 = \Sigma[(O - E)^2 / E]$

$\chi^2 = (80 - 65)^2 / 65 + (20 - 35)^2 / 35 + (50 - 65)^2 / 65 + (50 - 35)^2 / 35$

$\chi^2 = 3.46 + 6.43 + 3.46 + 6.43$

$\chi^2 = 19.78$

The degrees of freedom were:

df = (Number of rows - 1)(Number of columns - 1)

df = (2 - 1)(2 - 1) = 1

At $\alpha = 0.05$ and df = 1:

$\chi^2\_critical = 3.84$

Since $\chi^2\_calculated = 19.78$ is greater than $\chi^2\_critical = 3.84$, the null hypothesis was rejected.

Therefore, fungicide treatment was significantly associated with disease incidence in the cowpea plants. The fungicide-treated group had a lower proportion of diseased plants than the untreated control group.

Fill in the blank: In the chi-square test, the expected frequency for a cell = (________ total x column total) / grand total; this is the frequency expected if the two categorical variables were independent (not associated).

Pilot questions 5.1

16. State the null and alternative hypotheses for an experiment comparing the grain yield of two rice varieties.
17. What does a p-value of 0.02 mean in the context of an agricultural experiment at the 5 percent significance level?
18. Two fertiliser treatments give the following plant heights (cm): Treatment A (n=6): 45, 48, 50, 47, 52, 49 (mean=48.5; SD=2.43). Treatment B (n=6): 40, 42, 38, 41, 39, 44 (mean=40.7; SD=2.16). Calculate the t-statistic and state whether the means differ significantly at alpha=0.05 (t_critical at df=10 = 2.228).
19. In a chi-square test, what is the expected frequency for a cell whose row total is 60, column total is 80, and grand total is 200?
20. State the degrees of freedom for a chi-square test of a 2 x 3 contingency table.

5.2 Regression and Correlation

5.2.1 Simple linear regression: the equation of a straight line

Simple linear regression describes the linear relationship between one independent variable (x) and one dependent variable (y) using the equation: $y = a + bx$, where a is the y-intercept (the predicted value of y when $x = 0$) and b is the slope (the change in y for each unit increase in x). The least-squares method finds the values of a and b that minimise the sum of squared differences between observed y values and the predicted y values on the regression line. The formulae are: b = [n(sum of xy) - (sum of x)(sum of y)] / [n(sum of x2) - (sum of x)2]; a = y-bar - b x-bar.

Worked Example: Simple Linear Regression

A simple linear regression analysis was conducted to determine the relationship between nitrogen fertiliser rate and maize yield.

The observed data were:

Fertiliser rate, x (kg N/ha)	Maize yield, y (t/ha)
0	1.8
30	2.5
60	3.2
90	3.8
120	4.1

The summary values were:

$n = 5$

$\Sigma x = 300$

$\Sigma y = 15.4$

$\Sigma xy = 1{,}101$

$\Sigma x^2 = 27{,}000$

$\bar{x} = 60$

$\bar{y} = 3.08$

The slope of the regression line was calculated as:

$b = [n\Sigma xy - (\Sigma x)(\Sigma y)] / [n\Sigma x^2 - (\Sigma x)^2]$

$b = [5(1{,}101) - (300)(15.4)] / [5(27{,}000) - (300)^2]$

$b = (5{,}505 - 4{,}620) / (135{,}000 - 90{,}000)$

$b = 885 / 45{,}000$

$b = 0.01967$

The intercept was calculated as:

$a = \bar{y} - b\bar{x}$

$a = 3.08 - (0.01967 \times 60)$

$a = 3.08 - 1.18$

$a = 1.90$

Therefore, the fitted regression equation was:

$\hat{y} = 1.90 + 0.01967x$

The intercept indicates that the predicted maize yield without nitrogen fertiliser is 1.90 t/ha. The slope indicates that each additional 1 kg N/ha is associated with an estimated yield increase of 0.01967 t/ha, equivalent to approximately 19.67 kg of maize grain per hectare.

Fill in the blank: In the simple linear regression equation y = a + bx, ________ is the slope (regression coefficient) representing the change in y for each unit increase in x; a positive slope indicates that y increases as x increases.

5.2.2 The correlation coefficient

The Pearson correlation coefficient (r) measures the strength and direction of the linear relationship between two quantitative variables. It is calculated as: r = [n(sum of xy) - (sum of x)(sum of y)] / {square root of [n(sum of x2) - (sum of x)2] x [n(sum of y2) - (sum of y)2]}. The value of r ranges from -1 to +1: r = +1 (perfect positive linear relationship); r = 0 (no linear relationship); r = -1 (perfect negative linear relationship). Values of |r| > 0.7 are generally considered strong; 0.4 to 0.7 moderate; < 0.4 weak. r2 (coefficient of determination) is the proportion of variation in y explained by x.

Worked Example: Pearson's Correlation Coefficient

Using the fertiliser-rate and maize-yield data, Pearson's correlation coefficient was calculated to determine the strength and direction of the linear relationship between the two variables.

The required summary values were:

$n = 5$
$\Sigma x = 300$
$\Sigma y = 15.4$
$\Sigma xy = 1{,}101$
$\Sigma x^2 = 27{,}000$

First, Σy^2 was calculated:

$$\Sigma y^2 = (1.8)^2 + (2.5)^2 + (3.2)^2 + (3.8)^2 + (4.1)^2$$

$$\Sigma y^2 = 3.24 + 6.25 + 10.24 + 14.44 + 16.81$$

$$\Sigma y^2 = 50.98$$

Pearson's correlation coefficient is given by:

$$r = [n\Sigma xy - (\Sigma x)(\Sigma y)] / \sqrt{\{[n\Sigma x^2 - (\Sigma x)^2][n\Sigma y^2 - (\Sigma y)^2]\}}$$

Substituting the values:

$$r = [5(1{,}101) - (300)(15.4)] / \sqrt{\{[5(27{,}000) - (300)^2][5(50.98) - (15.4)^2]\}}$$

$$r = (5{,}505 - 4{,}620) / \sqrt{\{(135{,}000 - 90{,}000)(254.90 - 237.16)\}}$$

$$r = 885 / \sqrt{(45{,}000 \times 17.74)}$$

$$r = 885 / \sqrt{798{,}300}$$

$$r = 885 / 893.48$$

$r = 0.991$

The coefficient of determination was:

$r^2 = (0.991)^2$

$r^2 = 0.981$

Therefore, there was a very strong positive linear relationship between nitrogen fertiliser rate and maize yield. The coefficient of determination indicates that approximately 98.1% of the observed variation in maize yield was statistically associated with variation in fertiliser rate within the range of values studied.

Fill in the blank: The coefficient of determination (r2) represents the proportion of the variation in the ________ variable that is explained by the linear regression on the independent variable; r2 = 0.974 means that 97.4 percent of the yield variation is explained by fertiliser rate.

5.2.3 Interpretation and limitations of regression

Regression is a powerful tool for understanding how one variable responds to changes in another and for making predictions. Key interpretive points: the slope b quantifies the response rate (e.g. 0.0197 t/ha yield increase per kg N/ha); the y-intercept a is the baseline yield without the input; the r2 value indicates how much of the yield variability is accounted for by the factor studied. Predictions from the regression equation are valid only within the range of x values used to fit the model (interpolation); extrapolation beyond this range is unreliable because the relationship may not remain linear.

Important limitations: a high r2 does not mean x causes y; correlation and regression describe association, not causation. Regression assumes a linear relationship; if the true relationship is curved (e.g. diminishing returns at high fertiliser rates), a linear regression will be a poor fit and a quadratic or other model is more appropriate. Outliers (extreme data points) can substantially distort regression estimates. Results from one location and season should not be extrapolated to very different environments without new data.

Fill in the blank: Predictions from a regression equation are reliable only within the ________ of x values used to fit the model; predicting yield at a fertiliser rate far outside the observed range (extrapolation) is unreliable because the linear relationship may not hold.

Pilot questions 5.2

16. For the data: x (weed density, plants/m2): 5, 10, 15, 20, 25; y (crop yield, t/ha): 4.0, 3.5, 3.0, 2.4, 1.8. Calculate the regression slope b and interpret it.
17. Using the values from question 1, calculate the y-intercept a and write the full regression equation.
18. For the fertiliser-yield data in Sub Part 5.2.2, state r and r2 and interpret both values.
19. A student finds r = 0.92 between the number of tractors in a village and the average crop yield. Does this mean more tractors cause higher yields? Explain.
20. State two limitations of simple linear regression when applied to agricultural data.

5.3 Presenting Agricultural Research Results

5.3.1 Tables: construction and use

Tables organise numerical information into rows and columns for easy reference, comparison, and interpretation. A properly constructed table should have a brief but descriptive title placed above it, for example: **Table 1: Mean grain yield and yield components of five maize varieties at Ibadan in 2023**. Column headings should clearly identify each variable and include the appropriate units in parentheses, such as **Yield (t/ha)**. Row labels should appear in the first column, while numerical values should be aligned consistently, preferably by decimal point. Where necessary, measures of variation such as standard deviation or standard error should be included. Footnotes may be used to explain abbreviations, data sources, and statistical symbols, such as $p < 0.05$ and $p < 0.01$.

In agricultural reports, the same data should not normally be presented in both a table and a figure. Tables are most suitable when exact numerical values are required, whereas figures are better for displaying trends and patterns. Tables should remain simple, and unnecessary or repeated columns should be removed or combined. ANOVA tables should include the calculated F-value and its significance level. When treatment means are compared, superscript letters may be used to show significant differences. Means with the same letter are not significantly different at $p < 0.05$, based on a stated mean-separation procedure such as the Least Significant Difference test or Duncan's Multiple Range Test.

Fill in the blank: In an agricultural results table, means that share the same ________ letter do not differ significantly at $p < 0.05$ by the LSD test; this notation is used when more than two treatment means are being compared.

5.3.2 Figures and graphs in research reports

Figures (graphs, photographs, diagrams) in research reports must conform to journal or institutional style. Standard requirements: figure number and title placed below the figure (Figure 1: Effect of nitrogen rate on maize grain yield at Kano, 2023); axes fully labelled with units; error bars (standard error or standard deviation) shown on bar charts and line graphs; data points shown on scatter plots; legend included for multiple series. Bar charts in research reports typically use solid black bars or hatching rather than colour, for legibility in black-and-white print.

The figure title should be self-explanatory: a reader who has not seen the text should understand what the figure shows from its title alone. Avoid using the words 'the graph shows' in the title; instead state directly what is shown: 'Mean monthly rainfall at Sokoto, 1990 to 2020' rather than 'Graph showing rainfall'. When both a table and a figure are needed, place the table first in the results section; refer to each table and figure explicitly in the text ('Table 1 shows...'; 'As shown in Figure 2...').

Fill in the blank: In a research report, the figure title is placed ________ the figure, while the table title is placed above the table; both titles should be self-explanatory without reference to the surrounding text.

5.3.3 Describing results in writing

The results section of a research report describes what the data show without interpreting why. It should: (1) state the main finding first (the most important result); (2) refer to the relevant table or figure; (3) give key numerical values with units and measures of variability; (4) state whether differences are statistically significant; (5) avoid interpretation (save that for the discussion section). Example: 'Grain yield differed significantly among varieties ($F = 12.4$; $p < 0.01$; Table 1). SAMMAZ 14 produced the highest yield (5.1 t/ha) while the local check produced the lowest (2.3 t/ha). Variety SAMMAZ 15 (4.8 t/ha) was not significantly different from SAMMAZ 14 ($p > 0.05$).'

Common errors in writing results: stating opinions rather than data ('the results were excellent'); repeating all data from the table in the text instead of summarising key findings; omitting units; omitting significance levels; and confusing results with discussion (e.g. attributing a yield difference to a specific mechanism belongs in the discussion, not the results).

Fill in the blank: In the results section, a writer should state whether differences between treatment means are statistically ________ and refer explicitly to the relevant table or figure; interpretation of why differences occurred belongs in the discussion section, not the results.

- [Insert Figure/Photograph here]

Short description: A model results table for a maize variety trial: Table 1 titled 'Mean grain yield (t/ha) and 100-seed weight (g) of five maize varieties at Ibadan, 2023'. Columns: Variety; Grain yield (t/ha); 100-seed weight (g). Five rows of data with superscript letters showing LSD groupings; row for SE and LSD at 5% at the bottom. Footnote: 'Means with the same letter within a column are not significantly different ($p > 0.05$, LSD test).'

TABLE 1: YIELD PERFORMANCE AND SEED CHARACTERISTICS OF FIVE MAIZE VARIETIES

MAIZE VARIETY	MEAN GRAIN YIELD (T/HA)	LSD (0.05) GROUPING	100-SEED WEIGHT (G)	LSD (0.05) GROUPING
SAMMAZ 14	4.25	a	28.5	a
OBA SUPER 6	3.88	ab	27.2	ab
IFE MAIZE HYBRID	3.72	abc	26.8	b
SUWAN 1	3.55	b	25.1	b
SWAN 2	3.20	c	23.9	c

SUMMARY STATISTICS	MEAN GRAIN YIELD	100-SEED WEIGHT
GRAND MEAN	4.01	26.6
STANDARD ERROR (SE)	0.12	0.51
LSD (0.05)	0.35	1.48
SIGNIFICANCE	**	**

* Significant at $p < 0.05$; ** Significant at $p < 0.01$; ns = not significant.
LSD letter grouping: at a significant at a, vocto ab value, a, and ab, abc b, abc are clistric letts ons found.

Figure 5.1: Model results table for a maize variety trial showing LSD letter groupings

Pilot questions 5.3

16. List six elements that every well-constructed table in an agricultural research report must contain.
17. What is the difference between the results section and the discussion section of a research report in terms of their content?
18. Write a one-paragraph results description for the following data: mean yields (t/ha) of three cowpea varieties: IT97K = 2.1a; IT90K = 1.8ab; Local = 1.5b (F=8.2; p<0.01); SE=0.12.
19. What do error bars on a bar chart represent and why are they important in research publications?
20. State three common errors made when writing the results section of an agricultural research report.

5.4 Scientific Literacy and Evaluation of Agricultural Information

5.4.1 Reading and evaluating agricultural research papers

A scientific agricultural research paper typically has five sections: Abstract (brief summary of the study); Introduction (background, literature review, and study objectives); Materials and Methods (how the experiment was designed and conducted); Results (what was found); and Discussion (interpretation of results, comparison with previous work, implications). The IMRAD structure (Introduction, Methods, Results, And Discussion) is standard in international journals including Agricultural Systems, Field Crops Research, and Tropical Agriculture.

When evaluating a research paper, ask: (1) Is the objective clearly stated? (2) Is the experimental design appropriate (replicated, randomised, with controls)? (3) Are the statistical methods suitable for the data type and design? (4) Do the results support the conclusions? (5) Are limitations acknowledged? (6) Are the results generalisable to other locations, seasons, or farming systems? Peer-reviewed journals are more reliable than conference proceedings or extension leaflets because papers have been evaluated by independent expert reviewers before publication.

Fill in the blank: The standard structure of a scientific agricultural research paper is called the ________ structure (Introduction, Methods, Results, And Discussion); it is used in international peer-reviewed journals such as Field Crops Research and Tropical Agriculture.

5.4.2 Evaluating statistical claims in agricultural reports

When reading agricultural reports, extension bulletins, or media articles, apply critical evaluation to statistical claims: (1) Sample size: are conclusions based on adequate sample sizes (at least 20 to 30 observations for means; larger for regression)? (2) Statistical significance: is a p-value or confidence interval reported, or are claims of ‘improvement’ unsupported by statistical tests? (3) Practical significance: a statistically significant difference may be too small to be economically meaningful; always check the actual difference in yield or profit. (4) Representativeness: were trials conducted in conditions similar to those of the target farmers?

Common misleading statistical claims in Nigerian agricultural extension: ‘yields increased by 200 percent’ (from a very low baseline; a 200 percent increase from 0.5 t/ha is only 1.0 t/ha, still very low); ‘results from one trial station prove the variety is better for all farmers’ (ignores genotype x environment interaction); ‘the correlation is high therefore the input causes higher yield’ (confuses correlation with causation). An informed agricultural professional applies these critical tests to all quantitative claims before acting on or disseminating them.

Fill in the blank: A statistically significant difference between two treatment means may not be ________ significant if the actual difference in yield or profit is too small to justify the cost of adopting the new treatment; both statistical and practical significance must be considered.

5.4.3 Communicating biometric findings to farmers

Translating biometric findings into farmer-friendly language requires adapting both content and format. Key principles: (1) use local units and familiar quantities (‘this variety yields 2 bags more per acre than the local variety’ rather than ‘yields 0.5 t/ha more’); (2) use visual aids (bar charts, pictograms, photographs); (3) state the economic benefit (additional income per season, not just yield difference); (4) address reliability and risk (‘in most seasons and locations this variety outperforms the local variety, but in years of very low rainfall results may differ’); (5) allow interactive discussion (farmer questions and local knowledge should be incorporated).

Extension communication of biometric findings in Nigeria uses: field days (farmers observe trial results in situ and discuss with researchers); farm radio programmes (Farmers’ Radio programmes in Hausa, Yoruba, and Igbo broadcast statistical results in simplified language); leaflets and posters (graphic, minimal text, key figures highlighted); and farmer group meetings where results are presented as pictograms and simple tables. The most effective communication links the statistical finding directly to the farmer’s specific decision: whether to purchase improved seed, apply a recommended fertiliser dose, or adopt a new planting spacing.

Fill in the blank: To communicate biometric findings to farmers, use ________ units and familiar quantities (e.g. bags per acre rather than tonnes per hectare), visual aids such as pictograms, and state the economic benefit in terms of additional income per season.

Evaluation Aspect	Checklist Question
Sample Size	Is the sample large enough to provide reliable estimates and avoid random error?
Statistical Significance	Are p-values or confidence intervals reported, and do they indicate real differences rather than chance?
Practical Significance	Even if statistically significant, is the effect size meaningful for farmers or policymakers?
Representativeness	Does the sample reflect the broader farming population, region, or crop system?
Causation vs Correlation	Are claims about cause-and-effect justified, or are they simply correlations without experimental control?
Generalisability	Can the findings be applied beyond the specific study context (different crops, regions, or seasons)?

Box 5.1: Checklist for evaluating statistical claims in Nigerian agricultural reports

Pilot questions 5.4

16. State the five sections of a scientific agricultural research paper in order and briefly describe what each contains.
17. An extension bulletin claims that 'the improved variety yields 300 percent more than the local variety'. What questions would you ask to evaluate this claim?
18. Explain the difference between statistical significance and practical significance in the context of a variety trial.
19. Describe three ways in which biometric findings from a research trial can be communicated effectively to farmers in a Nigerian village.
20. A report states that the correlation between farm size and income is $r = 0.82$. Can you conclude that larger farms cause higher income? Explain.

This Series covered hypothesis testing and the major statistical methods used in agricultural research. The null hypothesis, H_0, states that no difference or association exists, while the alternative hypothesis, H_1, states that a difference or association exists. The p-value represents the probability of obtaining the observed result, or a more extreme one, if H_0 is true; H_0 is commonly rejected when $\mathbf{p < 0.05}$. For an independent-samples t-test, $\mathbf{t = (\bar{x}_1 - \bar{x}_2) / SE\_diff}$, with $\mathbf{df = n_1 + n_2 - 2}$, and the absolute calculated t-value is compared with the critical value. The chi-square test uses $\boldsymbol{\chi^2 = \Sigma[(O - E)^2 / E]}$, where **E = (Row total × Column total) / Grand total** and **df = (Rows − 1)(Columns − 1)**. Simple linear regression is expressed as $\mathbf{\hat{y} = a + bx}$, where the slope, **b**, measures the change in y for each unit change in x, and the intercept is $\mathbf{a = \bar{y} - b\bar{x}}$. Pearson's correlation coefficient ranges from −1 to +1, with larger absolute values indicating stronger linear relationships, while $\mathbf{r^2}$ represents the proportion of variation in y statistically explained by x. Predictions should remain within the observed range of x, and neither correlation nor regression alone proves causation.

The Series also covered scientific presentation, interpretation, and communication of research findings. Tables should have a title above them, headings with units, measures such as standard error and LSD where appropriate, superscript letters for mean separation, and explanatory footnotes. Figures should have captions below them, clearly labelled axes, error bars, and legends when multiple data series are shown. The Results section should present the main findings, refer to the relevant tables or figures, and report numerical values, units, and significance without detailed interpretation. The Discussion should explain the findings, compare them with previous studies, and acknowledge limitations. Research quality should be judged by sample size, p-values, practical importance, representativeness, and whether causal conclusions are justified. When communicating with farmers, results should be expressed in familiar units such as bags per acre, supported by simple bar charts or pictograms, linked to economic benefits, and discussed through field days, radio programmes, extension leaflets, and other interactive methods.

Glossary of Terms

- **Alternative hypothesis (H1):** the statistical hypothesis that there is a real difference or effect; it is accepted when the p-value is less than the significance level (usually 0.05) and H0 is rejected.
- **Chi-square test:** a statistical test for determining whether two categorical variables are associated; chi2 = sum[(O-E)2/E]; significant when chi2 exceeds the critical value at the chosen significance level.
- **Coefficient of determination (r2):** the square of the correlation coefficient; the proportion of variation in the dependent variable explained by the linear regression on the independent variable; ranges from 0 to 1.
- **Correlation coefficient (r):** a measure of the strength and direction of the linear relationship between two quantitative variables; ranges from -1 (perfect negative) to +1 (perfect positive); r = 0 indicates no linear relationship.
- **Degrees of freedom (df):** the number of independent pieces of information available to estimate a parameter; for t-test: n1+n2-2; for chi-square: (rows-1)(cols-1).
- **IMRAD structure:** the standard organisation of a scientific research paper: Introduction, Methods, Results, And Discussion; used in international agricultural journals.
- **LSD (least significant difference):** a post-ANOVA multiple comparison test used to identify which pairs of treatment means differ significantly; means sharing the same superscript letter do not differ at $p < 0.05$.
- **Null hypothesis (H0):** the default statistical hypothesis that there is no difference or no effect; rejected when the p-value falls below the chosen significance level.
- **p-value:** the probability of obtaining a result as extreme as or more extreme than the one observed, assuming H0 is true; a p-value below 0.05 indicates statistical significance at the 5 percent level.
- **Regression coefficient (b):** the slope of the simple linear regression equation y = a + bx; it represents the change in y for each unit increase in x; also called the slope or gradient of the regression line.

- **Simple linear regression:** a statistical method that describes the linear relationship between one independent variable (x) and one dependent variable (y) using the equation y = a + bx, fitted by least squares.
- **t-test (independent samples):** a statistical test comparing the means of two independent groups; t = (mean1 - mean2)/SE_diff; significant when |t| exceeds t_critical at the chosen significance level and df.
- **y-intercept (a):** the value of y predicted by the regression equation when x = 0; in agricultural regression, it represents the baseline level of the dependent variable in the absence of the independent variable.

Concept Check Questions [Series 5]

Objective Questions

16. The ________ hypothesis (H0) states that there is no difference between treatments; it is rejected when the p-value is less than 0.05. (Linked to SLO 5.1)
17. For an independent-samples t-test with n1 = 8 and n2 = 7, the degrees of freedom = n1 + n2 - 2 = ________. (Linked to SLO 5.2)
18. In the regression equation y = a + bx, ________ is the slope representing the change in y for each unit increase in x. (Linked to SLO 5.4)
19. r2 = 0.85 means that ________ percent of the variation in y is explained by the linear regression on x. (Linked to SLO 5.4)
20. In an agricultural results table, means followed by the same superscript ________ do not differ significantly at $p < 0.05$ by the LSD test. (Linked to SLO 5.5)

Short-Answer Questions

21. State the null and alternative hypotheses for a study comparing the mean body weight of two cattle breeds. (Linked to SLO 5.1)
22. In a chi-square test, the observed frequencies in a 2x2 table are O11=60, O12=40, O21=30, O22=70. Row totals: 100, 100. Column totals: 90, 110. Grand total=200. Calculate the expected frequency for O11. (Linked to SLO 5.3)

23. For data: x (irrigation frequency, times/week): 1, 2, 3, 4; y (fruit weight, g): 50, 70, 85, 100. Calculate b and write the regression equation. (Linked to SLO 5.4)
24. State four elements that must appear in a well-constructed results table in an agricultural research report. (Linked to SLO 5.5)
25. State two questions you would ask to critically evaluate a statistical claim in an extension bulletin. (Linked to SLO 5.6)

Long-Answer Questions

17. Explain the null and alternative hypotheses, p-values, and the independent-samples t-test. Work through a numerical example comparing two fertiliser treatment means. (Linked to SLOs 5.1, 5.2)
18. Explain simple linear regression and the correlation coefficient with a fully worked numerical example. Describe how to present results in tables and text, and how to communicate biometric findings to farmers. (Linked to SLOs 5.4, 5.5, 5.6)

Answers to Concept Check Questions [Series 5]

Objective Questions

16. Null.
17. 13.
18. b.
19. 85.
20. Letter.

Short-Answer Questions

21. H0: mean body weight of breed A = mean body weight of breed B (muA = muB). H1: mean body weight of breed A is not equal to mean body weight of breed B (muA ≠ muB).
22. E11 = (row1 total x col1 total) / grand total = (100 x 90) / 200 = 9000/200 = 45.
23. n=4; sum x=10; sum y=305; sum xy=10+140+255+400=805; sum x2=1+4+9+16=30; x-bar=2.5; y-bar=76.25. b = [4(805)-(10)(305)] / [4(30)-(10)2] = [3220-3050]/[120-100] = 170/20 = 8.5. a = 76.25 - 8.5(2.5) = 76.25-21.25 = 55. Equation: y = 55 + 8.5x.

24. Descriptive title; column headings with units; data aligned by decimal point; LSD letter groupings or significance symbols; SE and LSD (or SD) row at bottom; footnote explaining significance notation.
25. (1) Is the sample size adequate ($n \geq 20$ for reliable means)? (2) Is a p-value or confidence interval reported to confirm statistical significance, rather than a mere percentage difference?

Long-Answer Questions

17. Null hypothesis H0: no difference between treatments (e.g. muA = muB). Alternative hypothesis H1: difference exists (muA ≠ muB). p-value: probability of the observed result under H0; reject H0 if $p < 0.05$ (significant). Independent-samples t-test: t = (mean1 - mean2)/SE_diff; SE_diff = sqrt(s12/n1 + s22/n2); df = n1+n2-2; compare |t| with t_critical from t-table. Numerical example: Treatment A yields (t/ha): 3.2, 3.5, 3.8, 3.4, 3.6 (n=5; mean=3.50; s=0.228). Treatment B: 2.8, 3.0, 2.9, 3.1, 2.7 (n=5; mean=2.90; s=0.158). SE_diff = sqrt(0.228^2/5 + 0.158^2/5) = sqrt(0.0104+0.0050) = 0.124. t = (3.50-2.90)/0.124 = 4.84. df=8; t_critical (alpha=0.05, 2-tailed)=2.306. Since 4.84 > 2.306, reject H0: the two treatments differ significantly in yield.
18. Simple linear regression: y = a + bx; b = [n(sum xy)-(sum x)(sum y)] / [n(sum x2)-(sum x)2]; a = y-bar - b(x-bar). Worked example (fertiliser rate x vs yield y, n=5; sum x=300; sum y=15.4; sum xy=1101; sum x2=27000): b = (5x1101 - 300x15.4)/(5x27000 - 300^2) = (5505-4620)/(135000-90000) = 885/45000 = 0.01967. a = 3.08 - 0.01967(60) = 1.90. Equation: y = 1.90 + 0.0197x. Interpretation: baseline yield 1.90 t/ha; each kg N/ha increases yield by 0.0197 t/ha. r = 0.987 (very strong positive); r2 = 0.974 (97.4% of yield variation explained by fertiliser). Prediction valid only within 0-120 kg N/ha range. Results table: title above; column headings with units; superscript letters for LSD groupings; SE and LSD row; footnote on significance notation. Results text: state main finding first; refer to table; give numerical values with units and significance level; no interpretation. Farmer communication: express as bags per acre; state naira profit from improved variety; use pictogram or bar chart; discuss at field day or radio programme; address seasonal reliability.

Answers to Pilot Questions [Series 5]

Part 5.1

16. H0: mean grain yield of variety A = mean grain yield of variety B (muA = muB). H1: mean grain yield of variety A is not equal to that of variety B (muA ≠ muB).
17. p = 0.02 means there is a 2% probability of obtaining the observed difference (or larger) by chance if H0 is true. Since 0.02 < 0.05, we reject H0 and conclude the difference is statistically significant at the 5 percent level.
18. SE_diff = sqrt(2.43^2/6 + 2.16^2/6) = sqrt(0.983+0.778) = sqrt(1.761) = 1.327. t = (48.5-40.7)/1.327 = 7.8/1.327 = 5.88. df = 6+6-2 = 10; t_critical = 2.228. Since 5.88 > 2.228, reject H0: treatment means differ significantly.
19. E = (row total x column total) / grand total = (60 x 80) / 200 = 4800/200 = 24.
20. df = (rows - 1) x (columns - 1) = (2-1)(3-1) = 1 x 2 = 2.

Part 5.2

16. n=5; sum x=75; sum y=14.7; sum xy=5x4.0+10x3.5+15x3.0+20x2.4+25x1.8=20+35+45+48+45=193; sum x2=25+100+225+400+625=1375; x-bar=15; y-bar=2.94. b = [5(193)-(75)(14.7)] / [5(1375)-(75)^2] = [965-1102.5] / [6875-5625] = -137.5/1250 = -0.11. Interpretation: each additional weed plant per m2 reduces crop yield by 0.11 t/ha.
17. a = y-bar - b(x-bar) = 2.94 - (-0.11)(15) = 2.94 + 1.65 = 4.59. Regression equation: y = 4.59 - 0.11x.
18. r = 0.987: very strong positive linear relationship between fertiliser rate and maize yield. r2 = 0.974: 97.4 percent of the variation in maize yield is explained by fertiliser application rate.
19. No. Correlation measures the strength of a linear association, not causation. Both the number of tractors and crop yields in a village may both be driven by a third factor such as overall farm investment or farm size; more tractors do not necessarily cause higher yields directly.
20. Two limitations: (1) predictions are only reliable within the observed range of x; extrapolation is unreliable if the relationship changes beyond that range. (2) linear

regression assumes a straight-line relationship; if the true relationship is curved (e.g. diminishing yield returns at high fertiliser rates), a linear model is a poor fit.

Part 5.3

16. Six elements: (1) descriptive title above the table; (2) column headings with units in parentheses; (3) row labels in the first column; (4) data values aligned by decimal point; (5) SE and LSD rows or significance letters; (6) footnote explaining abbreviations and significance notation.
17. Results section: describes what was found (data, statistical test outcomes, significance levels); no interpretation of why. Discussion section: interprets the findings, compares with previous work, proposes explanations for patterns, acknowledges limitations, and draws practical conclusions.
18. Grain yield differed significantly among the three cowpea varieties ($F=8.2$; $p<0.01$; Table 1). IT97K produced the highest mean yield (2.1 t/ha), which was significantly greater than the local variety (1.5 t/ha; $p<0.05$). IT90K (1.8 t/ha) did not differ significantly from either IT97K or the local variety ($p>0.05$).
19. Error bars show the variability of the data (standard error of the mean or standard deviation); they are important because they allow the reader to assess whether the differences between bars are likely to be statistically significant and whether the data are precise or highly variable.
20. Three common errors: (1) stating opinions instead of data ('results were good'); (2) repeating all table values in the text instead of summarising key findings; (3) omitting units or significance levels from reported values.

Part 5.4

16. (1) Abstract: brief summary of objectives, methods, results, and conclusions. (2) Introduction: background, review of relevant literature, and statement of study objectives. (3) Materials and Methods: experimental design, treatments, location, measurements, and statistical analysis used. (4) Results: findings presented as tables and figures with significance information. (5) Discussion: interpretation of results, comparison with published data, implications, and limitations.

17. Questions to ask: (1) What was the baseline yield of the local variety? (a 300% increase from 0.5 t/ha gives only 2.0 t/ha, which may still be low). (2) Is the claim based on a replicated, statistically significant trial or a single observation? (3) How many seasons and locations was the trial conducted in? (4) Is the improvement economically meaningful after accounting for seed cost and input changes?
18. Statistical significance: the probability that the observed difference could occur by chance is below 5%; the t-test or F-test gives $p < 0.05$. Practical significance: the actual difference in yield or profit is large enough to justify adopting the new practice after accounting for its additional cost; a statistically significant difference of 50 kg/ha may not cover the extra seed cost of an improved variety.
19. Three ways: (1) field days, where farmers observe trial results in the field and discuss findings with researchers in their local language; (2) radio extension programmes in Hausa, Yoruba, or Igbo that present yields in familiar units (bags per acre) and state the naira profit advantage; (3) extension leaflets with pictograms comparing yields of improved and local varieties, requiring minimal reading ability.
20. No. $r = 0.82$ indicates a strong positive association between farm size and income, but does not prove causation. Larger farms may also have more capital, better access to credit, and more family labour, all of which increase income independently of farm size. Correlation cannot establish which variable causes the other, or whether both are driven by a third factor.

References and Attributions [Series 5]

Textual References (Books and External Sources)

- Gomez, K. A., & Gomez, A. A. (1984). Statistical procedures for agricultural research (2nd ed.). Wiley-Interscience.
- Snedecor, G. W., & Cochran, W. G. (1989). Statistical methods (8th ed.). Iowa State University Press.
- Steel, R. G. D., & Torrie, J. H. (1980). Principles and procedures of statistics: A biometrical approach (2nd ed.). McGraw-Hill.
- Day, R. A., & Gastel, B. (2012). How to write and publish a scientific paper (7th ed.). Cambridge University Press.

Figures, Plates, Tables, and Boxes

- Figure 5.1: Model results table for a maize variety trial showing LSD letter groupings. Attribution: AI generated. Suggested OER search string: “agricultural research results table LSD letter grouping variety trial maize format OER”.
- Box 5.1: Checklist for evaluating statistical claims in Nigerian agricultural reports. Attribution: AI generated. Suggested OER search string: “evaluating statistical claims agricultural research report checklist critical reading OER”.