EGERTON UNIVERSITY

Department of Crops, Horticulture Soils (CHS)

AICM 703: RESEARCH METHODS

Published by RUFORUM
Plot 151 Garden Hill, Makerere University
P. O. Box 7062 Kampala, Uganda
Tel. : 256 414 535939
Fax: 256 414 543153
Email: secretariat@ruforum.org
Website: www.ruforum.org

AICM 703: **RESEARCH METHODS**

Course Instructor: Prof. Bockline Omedo Bebe (Male)

Department: Department of Animal Sciences

Email: bbebe@egerton.ac.ke, obebeb@yahoo.com, omedobebe@gmail.com

Office: Faculty of Agriculture Building

Credit Factor: 3.0

Lecture Hours: 30

Practical Hours: 15

Tutorials: 15

Table of content

Course Description

Role and characteristics of research in the development of scientific knowledge; AICM research approaches, Research process; Principles of scientific writing; Quantitative and qualitative research; Development of research proposals and thesis reports; Major areas of research in AICM: Agricultural information user studies, Access to agricultural data, information, and knowledge, Agricultural information retrieval research; Agricultural information systems research, Agricultural communications research; Principles and procedures of Sampling; AICM research data collection; Communicating AICM/agricultural research; Collaborative research project management.

Prerequisite: AICM 702: *Statistical Methods for AICM*

Course Aims

Research Methods for AICM will provide to learner the insight into systematic approach to implementing research process. The course will develop the analytical thinking in learners through comparisons and contrasts in conceptualising, designing, planning and implementing research projects in Agricultural Information and Communication Management.

Learning Outcomes

By the completion of this course, the learner will be able to:

a. Discuss the importance of research in the development of scientific knowledge
b. Describe the different research approaches
c. Identify researchable issues in AICM
d. Formulate a research problem and the study objectives, research questions and hypothesis to guide the study.
e. Design conceptual model of the research questions/hypothesis.
f. Design qualitative and quantitative research studies and data collection process
g. Design data collection and database management systems
h. Perform statistical modelling
i. Communicate scientific results appropriately

Instructional Methodology

1. Lectures and tutorials
2. Review of case studies, research projects, theses and Journal articles
3. Group discussions
4. Writing a research project under supervision
5. Class presentations
6. Practical use of computer software for data management and statistical analysis

Course Outline

1. Research in AICM
1.1. Researchable issues in AICM

1.2. Structure of research proposal and thesis

2. Research Process
2.1. Components of research process
2.2. Research implementation strategy
3. Conceptualisation of Research Problem
3.1. Problem tree analysis
3.2. Log frame development

4. Formulating Research Objectives, Hypothesis and Research Questions
4.1. Research objectives
4.2. Research hypothesis and research questions

5. Theoretical Framework and Conceptual Model Development
5.1. Why present theoretical framework and conceptual model in research project
5.2. Defining theoretical and conceptual model for research problem

6. Study Designs
6.1. Study designs
6.2. Sampling techniques
6.3. Randomization process
6.4. Blocking and replication concepts
6.5. Sample size determination

7. Data Collection Methods
7.1. Observations for data collection
7.2. Tests for data collection
7.3. Questionnaires for data collection
7.4. Interviews for data collection
7.5. Focus Group Discussion for data collection
7.6. Collecting secondary data

8. Data Measurements
8.1. Guide to variables to measure in research process
8.2. How often to measure the variables (Frequency of measuring variables)
8.3. Data recording

9. Database Management
9.1. Principles of database design
9.2. Data processing procedure

10. Data Exploration and Analysis
10.1. Data exploration goals
10.2. Checking for normality in data distribution
10.3. Descriptive statistics
10.4. Decisions to make before data analysis
10.5. Data analysis

11. Reporting Results

11.1. Scientific report
11.2. Ethics in scientific reporting

Assessment

- Class assignments 10%
- Research project proposal development 15%
- CATS 15%
- Final written exam 60%

Course Evaluation

- Process monitoring for every lesson
- Monkey survey at end of the course

Topic 1: Research in AICM

Introduction

The World Bank recognizes that agricultural development depends to a great extent on how successfully knowledge is generated, shared and applied. Despite much investment in agricultural systems, the desired impacts are yet to be realized in enhanced productivity and value. This is because knowledge generated remains unknown and ineffectively transmitted to the target primary beneficiaries along research impact pathways. In Eastern and Central African region (ECA), adoption of modern information communication technologies (ICTs) lags behind that of the rest of the world. The gap observed between knowledge and action in agriculture remain huge because knowledge management approaches limit the extent to which primary target beneficiaries along research impact pathways are reached by the research knowledge.

Learning Objectives

By the completion of this topic, the learner should be able to:

- State the importance of research in AICM
- Identify researchable issues in AICM
- Outline possible research topics in AICM
- Prioritise research project topics for a thesis

Key Concepts

- Agricultural knowledge products
- Knowledge management
- Exploratory research
- Descriptive research
- Explanatory research
- Predictive research
- Demonstration research

1.1. Researchable Issues in AICM

The role of research in AICM is to generate knowledge products and innovations for application in enhancing sustainable use of resources in agricultural systems, advancing agricultural development and promoting economic growth, fighting poverty and eradicating hunger. Research can bridge the gap between generation and utilization of research information, design appropriate communication channels for disseminating targeted agricultural knowledge and improving knowledge management including use of information and communication technology (ICT). The AICM research projects could be on issues of availability, accessibility and applicability of agricultural knowledge.

Types of research in AICM can be classified into exploratory research, descriptive research, explanatory research, predictive research, and demonstration research. Exploratory research can generate knowledge products or innovations. The descriptive research usually is about characterisation of some aspects of AICM. Explanatory research

tries to explain observed causality to show how and why a phenomenon operates as it does. The predictive research involves empirical approaches that quantify the change in response due to certain factors or intervention. The demonstration research is designed to influence the application of research results and to attain impact. Products of AICM research include scientific knowledge product or an innovation.

For instance; the knowledge management and upscaling programme of ASARECA (define the acronym) has identified what hampers effective knowledge management in Eastern and Central Africa to include:

a. Inadequate analysis of agricultural sector communication, stakeholders, their knowledge needs, attitudes and practices to knowledge management
b. Poor identification of needs and utility of knowledge products and services in the agricultural sector
c. Inadequate mechanisms for capturing, systematizing and sharing available knowledge;
d. Use of ineffective media and channels for communicating with different stakeholders;
e. Weak monitoring and evaluation of knowledge management systems

1.2 Structure of Research Proposal and Thesis

AICM postgraduate students undertake research as a partial fulfilment of the requirements for the award of a degree. AICM research strives to enhance effective knowledge management. The constraints to this goal have been identified by ASARECA (section 1.1), which should be priority research areas in AICM.

The generic and detailed format of a research proposal and thesis can be found in the library or graduate school. The structure reflects the systematic steps involved in conducting research. The first chapter is introduction which has the background information, statement of the problem, the objectives, the research questions or hypotheses corresponding to the specific objectives, the expected outputs. The second chapter is traditionally literature review for the researcher to gain deeper understanding of what is known and is not known including approaches used in studying a related hypothesis or research questions. Literature review is guided by the objectives, the hypothesis or research questions of the study. The third chapter details the methodological approaches and procedures to data collection and data analysis. The next chapters are about results and discussion if in a thesis, otherwise it is about the details or work plan and budget for implementing the research if it is a project proposal. The details of what is involved in preparing these sections are covered in AICM 797 Seminars.

Learning Activities

1. In groups, sample at least three postgraduate theses to familiarise with the structure and formats. Write down your findings are present to the entire class.
2. In groups, sample at least three scientific papers to familiarise with the structure and formats. Write down your findings are present to the entire class.

3. Individually use library and web search to identify journals of Agricultural Information and Communication and submit a report to through email to facilitator. These findings will be consolidated and shared with the entire class.
4. Individually visit the online discussion forum that has been set up for you to discuss the importance of agricultural research

Assignments

1. From lessons learnt in AICM 701: Fundamental of AICM, identify researchable issues in AICM
2. Summarise AICM research objectives and topics from interviews with postgraduate students, lecturers, Library staff, Research and Extension staff, and from the websites of KARI, ILRI, ICRAF, ASARECA, RUFORUM and FAO.
3. Compare the structure and formats of postgraduate proposals, theses and scientific papers
4. With use of key words, do an electronic search of journals that publish agricultural information and communication research and for each journal identified summarise research topics and publication policy statements.
5. Synthesise research topics in MSc AICM theses and indicate where they fall within the types of research in 1.1 above

References and Additional Reading Materials

1. AICM 701: Fundamentals of AICM course
2. AICM 753: Farming systems and rural livelihoods
3. Graduate school guidelines for research proposals and thesis presentations MSc and PhD research proposals and theses at the Graduate School.

Useful links

http://www.ruforum.org for summary and abstracts of postgraduate research projects

Topic 2: Research Process

Introduction

Research is a process of investigation for the unknown; searching for an answer to a question of concern about something. An investigation is conducted in order to know why things happen, why there are differences between interventions we make, how to explain the observed responses in a system. Research is scientific if investigation is done in a systematic process that runs through circle with several interdependent steps. Each step has a specific role in contributing to body of existing knowledge.

Learning Objectives

By the completion of this topic, the learner should be able to:

a. Describe the systematic implementation of research process
b. Explain the importance of components of a research process
c. Develop a research implementation strategy
d. Critique research implementation strategies in research reports.

Key Concepts

Research Issue: From personal experience, literature, clients demand, directed by lecturers, informed by collogues

Problem: Study objectives, hypothesis, research questions, theoretical framework, and conceptual framework.

Study designs: Data collection through experiments, observations, surveys.

Data management: Variables coding, entry and cleaning

Data analysis; Descriptive, analytical model, hypothesis testing, interpretation of results, presentation of results

Reporting: Thesis, journal articles, conference paper, policy brief

2.1. Components of Research Process

Research is a process of scientific investigation implemented in systematic approach. There are interdependent components in both organization and content. Figure 1 illustrates these components and their interdependency.

The first step in research process is a question about the unknown for which an investigator attempts to understand the present state of knowledge on the issues of concern. The present state of knowledge about the issues can be found from literature about what is already known and documented. The literature review forms the basis for finding out what is known and the gaps in knowledge about the issue of concern.

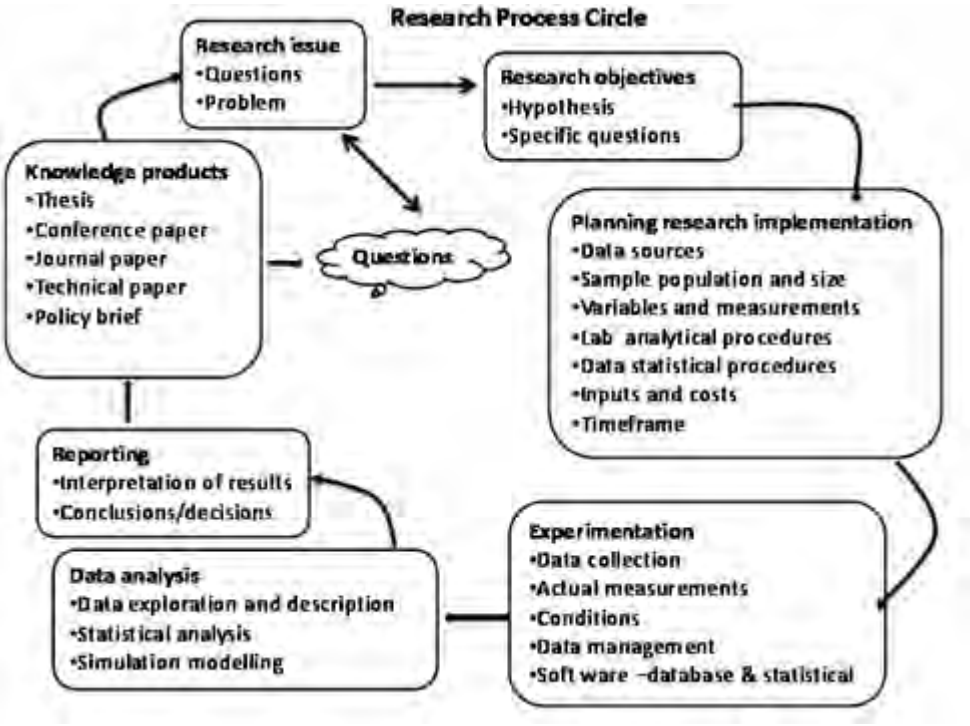


Figure: Research process circle (Source: Bebe B.O. Lecture notes)

- The gap in knowledge may be sometimes documented in the literature, but often is a result of information synthesis by the investigator. From synthesis of existing information about knowledge of what is known, the investigator identifies the knowledge gap. The knowledge gap forms the basis on which to formulate specific problem to study, therefore requires paying attention to details as reemphasized in the shaded text . Analyze literature relevant to deeper understanding of the problem
- Use the literature to show how the research will contribute to greater understanding of the subject
- Synthesise and summarise the literature in a way that links together and highlights and resolves conflicts between different approaches
- Conduct a critical review to show the reader its relevance to the study.

Specific problem definition informs formulation of the research questions and/or hypotheses that will guide the subsequent process. Research questions and/or hypotheses are the basis for data collection hence the design of research, and data collection tools and the materials needed to obtain the relevant data. Data collection for hypothesis testing is a systematic process on its own. It may be done by making observations, carrying out an experiment, using field data or interviewing people.

The data obtained through data collection process is subjected to appropriate statistical modeling procedures to obtain information in form of results. The obtained results are evidence on which the investigator interprets and builds a discussion to know why

things happen as observed, why there are differences between interventions made and how to explain the observed responses in a system. In other words, results provide the answers to the questions the investigator was concerned about before implementing the research.

The interpretation and discussion of the results is in view of existing body of knowledge. The new knowledge contributed is the value of the research undertaken, therefore must be reported in a way that a wider audience will understand and use.

2.2. Research Implementation Strategy

A research project must have a goal that direct mode of action with clarity about what is to be done by the researcher and partners involved in the implementation of the research. Coherency in activities from the beginning provides researcher with insight into the various steps involved in preparing and implementing the research project. A well thought-out and planned approach is of essence. So what should be prioritized in planning research project?

Research planning has conceptual and technical design issues to deal with. In conceptual design, the researcher determines *what, why and how much* will be studied. These are issues of concepts and have four elements: objective, framework, research issues and definition of concepts. The technical design issues in research are decisions to make about *where and when* to conduct the research. There are three elements to prioritize: research materials, strategy and planning.

Research Objectives are formulated in the context of the problem. Problem definition is informative when specific in order to inform objectives that will be feasible to research on towards finding solutions for or solving the problem. Objectives are therefore a result of delineation of what is feasible within the acceptable timeframe and logistics.

Research Framework is a schematic visual representation of the steps to be taken to realize the research objectives. This is useful tool to focus the research to its goals.

Research Issues are the set of research questions and/or hypotheses to provide useful information for realizing the research objectives. The answers to the research questions and hypotheses are the information to obtain from conducting research. Formulation of research questions and hypotheses inform data collection and study design necessary for obtaining reliable information to understand the underlying cause of the problem of concern.

Definition of Concepts is a delineation process to improve clarity and inform empirical approach to undertake.

Research Materials are the subjects needed to generate valuable and reliable information on the problem of concern. Choice of materials has implications on quality of data generated and the cost involved in generating that data. In turn, materials used in research have implications on the costs. Therefore choice of research materials presents logistical issues to the researcher.

Research Strategy is decision about breadth and depth necessary in achieving the research objectives and the appropriate study design to adopt.

Research Planning is timeframe for performing specific activities to achieve the objectives of the research.

Summary of Topic

Put simply, research process is the systematic steps of *what* (research objectives and research issues), *how* (research materials and research strategy) and *when* (sequences and periods in which research activities are to be performed). The steps in implementation of research project are:

i. Exploring the context of research and isolating feasible research objectives
ii. Constructing research framework that indicates the individual stages in systematic manner for completion of the research objectives.
iii. Examining what information will be useful or necessary in form of research questions and hypothesis for achieving the research objectives.
iv. Determining key concepts and operationalise definitions in context of the research problem.
v. Examining type of materials needed for each research question and/or hypothesis that is needed to provide reliable data and answers.
vi. Determine research strategy that will guide the implementation process.
vii. Drawing work plan of activities with time frame for implementation.

Learning Activities

1. Individually review the concepts involved in the research process
2. Identify the individual steps in research process from research proposals and theses provided and from journal papers of your choice. Make a table to show the differences between these scientific reports in terms of what step is included and excluded. Explain the observed differences. The journal papers accessed must be identified fully by declaring the authors, year of publication, the journal, and the pages of the journal where the paper is.
3. Describe the specific roles of the individual steps of research process you have identified
4. With the use of a flow diagram illustrate the interdependency of these steps in a research process

References and Additional Reading Materials

Visit Egerton university library, theses and dissertations for additional reading materials
Useful links

http://thesisnotes.com/category/research-topic/ Accessed on: 24th June, 2011

http://thesisnotes.com/conceptual-framework/sample-conceptual-framework/ Accessed on: 24th June, 2011

Topic 3: Conceptualisation of Research Problem

Introduction

Research problem indicate specifically the researchable issues of concern for the study. It outlines why addressing that particular issue is important to academic enquiry to some primary targeted clients or to society in general. It informs the research questions for which answers are needed or the hypotheses to test for which decision will be made. Critical analysis of a problem is essential to get to cause-effect relationships that the research will target to find answers, solve the problem or improve better understanding about the observed phenomenon.

Learning Objectives

By completing this topic, the learner should be able to:

- Analyse researchable issues in AICM
- Formulate statement of the problem for researchable issues
- Develop a logframe from problem analysis process

Key Concepts

- Problem tree
- Objective tree
- Logframe

3.1. Problem Tree Analysis

Problem tree analysis is an analysis of a researchable issue to gain insights about its possible cause-effect relationships. This analysis helps in identifying the critical areas where an intervention would provide a solution to the problem of concern. The analysis can be completed by the researcher alone applying knowledge from literature review. For development projects, problem analysis is more effective if the primary beneficiaries are involved in a participatory process with people knowledgeable about the issue of concern.

The process of problem analysis is focused on identifying the existing situation that the research will be designed to address by reversing or changing the primary causes of the core problem. Note that absence of a solution is not a problem, but rather an existing negative situation. For instance, stating that there is *no pesticide available is not a problem*, but stating that *harvest is pest infested* is a practical existing problem demanding for a solution. Through the process, researcher identifies the effects of the core problem impacting on people, environment or development. The process of problem tree analysis produce linkages of issues with the middle being the situation and issues above it present the effects while issues below present the cause-effect. The causes and effects are expressed in negative conditions and are presented singly in order to give attention to specific issue so identified.

The problem tree can be translated into objective tree and eventually into logical framework. This is a systematic analysis of the situation, trends and its causes and consequences. The objective tree is expressed in positive condition and then translated into logframe. A case study is used to illustrate formulation of research problem.

Case Study 1: Formulating a research problem

In Molo peri-urban area, there is stiff competition for milk between the formal and informal market outlets, yet many smallholder milk producers sell milk at low prices, not competitive enough to ensure positive returns to external inputs they use to support intensive milk production. It is suggested that market information flow between milk producers and market actors impact on the price that farmers receive for their milk, and access and use of market information could explain why some are able to sell competitively while others are not. Farmers may not be seeking for market information among the market outlets, if they do, there could be information barriers in access, timeliness and reliability. Knowledge of the flow and use of market information on milk prices would be useful in designing effective price information dissemination strategies to help farmers sell their milk at profitable prices and realize positive returns to their investment for improved income and food security from dairy production.

This statement of the problem has four important informative characteristics that define research problem or the researchable issue.

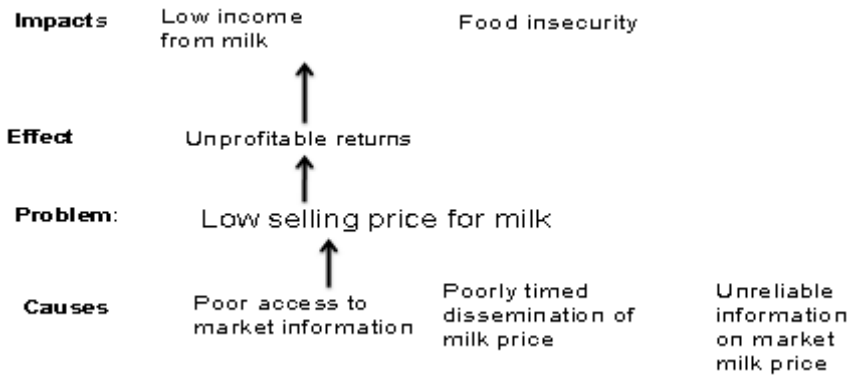
a. The problem/issue of concern is specific: *many producers sell milk at low prices*
b. The effects/impacts directly resulting from the problem are clear: *Milk production not competitive enough to ensure positive returns to external inputs used to support intensive milk production.*
c. The hypotheses/research questions to address the problem can be linked directly with the problem statement:
    - Hypothesis*: That milk producers that source market information are able to sell their milk at competitive prices than those not sourcing for market information*
    - Research questions: *Do milk producers sourcing market information sell their milk at more competitive price than the producers who do not source market information?*
d. The gap in knowledge, possible solution and the expected results from implementing the suggested research are well articulated:
    i. Barriers in information flow and access, timeliness and reliability between milk producers and market actors
    ii. The influence of knowledge of the flow and use of market information on milk prices that farmers obtain in the market
    iii. Closing the knowledge will inform the designing of effective price information dissemination strategies to help farmers sell their milk at profitable prices and enable them realize positive returns to their investment for improved income and food security from dairy production

From the statement of the problem, the problem tree can be presented and translated into objective tree then turned into logframe as now illustrated subsequently. First is

problem tree then objective tree which is then turned into log frame. To translate problem tree into objective tree, do

- Transform all negative conditions in problem tree into positive conditions that are desirable, realistically achievable
- The cause-effect relationship turns into a means-ends relationship,
- Check validity and completeness of the cause-effect relationship, and the hypothesis
- Objectives seemingly not expedient or necessary may be deleted
- May rephrase statements to attain clarity

Problem tree



Objective tree

3.2. Logical framework development

The objective tree transformed into logframe will indicate the goal, the purpose, the outputs, the activities and the inputs in rows with columns indicating the verifiable indicators, sources of verification and assumptions.

Table -: Model logical framework

Objectives	Objectively verifiable indicators	Means of verifications	Assumptions and risks
Results expected	Measure of whether results have been achieved	Where to source information or evidence that achievement has been made	Impediments expected not to stand on the way the way, If true, then specific objectives will be achieved

The objective tree turned into logframe for the above case studys presented next. Note that the indicators for the outputs/objectives are **S**pecific, **M**easurable, **A**chievable, **R**ealistic and **T**ime-bound, usually termed **SMART**.

Logical framework

In this logframe, complete the log frame by stating the missing verifiable indicators for the output two using the approach used for the output one.

Narratives	Verifiable indicators	Sources of verification	Assumptions
Goal: Milk production profitable			
Purpose: Milk sold at competitive price			
Outputs: **1.1 Reliable market milk price better accessed**			
Activities **1.1 Disseminate market milk price timely**			

Summary of Topic

Characteristics of well formulated research problem are:

- Specific problem/issue of the study can be identified by a reader
- Impacts/effects resulting from presence of the problem identified
- The hypotheses/research questions that guide the research study are articulated
- Hypotheses and research questions are stated in a way that clearly distinguishes the dependent and independent variables
- The gap in knowledge, solution needed and the expected results from implementing the research are articulated well for a reader.

Learning Activities

One

In groups read the following two research problem statements and answer the following questions for each and presentation:

a. Point out the characteristics of well formulated research problem that the statement has captured adequately.
b. Identify the specific researchable issues of concern for the study
c. Translate the researchable issues into problem tree and objective tree
d. From your problem and objective trees, transform the researchable issues into a logframe.

Exercise 1:

Pesticides and chemical fertilizers are expensive and frequently unobtainable. Some smallholders in Western Kenya have been successful in increasing incomes and maintaining soil fertility by using an integrated pest management (IPM) scheme and crop rotation. However other farmers have been less successful. It has been suggested that soil type has a large impact on the viability of these management techniques, and that might explain varying adoption. If we knew which practices were better adapted for which soils, dissemination could be targeted and further research may develop techniques for problem soils.

Two

From lessons learnt in AICM 701 Fundamental of AICM, identify a problem in agricultural information and communication management experienced in African agriculture, then:

- Conduct problem analysis to produce a problem tree
- Translate the problem tree into an objective tree
- Translate the objective tree into a logframe

Three

In groups, critically review at least two MSc and PhD proposals and theses to identify how well a problem statement articulates the good characteristics of a research problem. Discuss areas of general weakness in the formulation of research problem you find the MSc and PhD proposals and theses.

Critique problem statements in any two selected MSc / PhD proposals and theses based on the how well the problem statements articulates the researchable issues for the study.

References and Additional Reading Materials

MSc and PhD research proposals and theses

GEAR CD supplied by RUFORUM

Useful link

URL: http://www.ruforum.org for the GEAR, summary of postgraduate theses

URL: http://www.wur.nl for the Researchable issues in AICM and postgraduate theses

http://www.mande.co.uk/docs/DFID1997CARELogFrameGuide.doc *for logframe*

http://www.slideshare.net/rexcris/beginners-guide-to-logical-framework-approach-bond

Topic 4: Formulating Study Objectives, Research Questions, Hypothesis

Introduction

Study objectives are formulated to direct implementation of research study. Objectives directly emanates from the problem statement of the identified researchable issues. The objectives reflect the cause-effect identified in the problem tree and therefore inform the formulation of hypothesis and research questions for the study. Clarity in objectives enhances clarity of hypothesis and research questions, subsequently the conceptual model and data collection needed to address the research issue of concern. The objectives define the limits of the research and describe the expected outputs when the objectives are achieved. The objectives define the needed data and form the links the problem statement with data collection procedure in the research design.

Learning Objectives

Upon completion of this topic, the learner should be able to:

a. Distinguish null hypothesis from alternative hypothesis
b. Distinguish research question from research hypothesis
c. Write well understood research (SMART) objectives
d. Write well understood research questions and hypothesis
e. Critique objectives, research questions and hypotheses in research reports.

Key Concepts

- Research objective
- Research hypothesis
- Research questions
- Developmental hypothesis

4.1. Research Objectives

Research objectives are the achievements a researcher can point out to show success made in implementing the research project. Objectives describe the endpoints that a researcher will be accountable for.

The characteristics of well stated statements of objectives are:

- Logical consequence of the background and problem statement
- Are achievable with data to be collected from surveys, observations and experiments
- Have active verbs such as:
    - Determining...
    - Measuring how much...
    - Identifying...
    - Establishing...
    - Evaluating...
    - Assessing...

- Are not statement of the methods:
  To carry out a survey...
  To compare treatments...
  because the methods are developed to meet the objectives, not the other way around.
- Declare the relationships to be investigated, identifying independent and dependent variables
- Make sense to an informed reader without additional information

4.2. Research Hypothesis and Research Questions

Hypothesis is a statement of the expectation that a researcher states about the population characteristics for making statistical decisions on the basis of sample data. Statistical decisions to make is rejecting or accepting the hypothesis within a specified level of certainty. Hypothesis formulation is done at the stage of developing the proposal to guide collection of appropriate data.

There are two approaches of formulating hypothesis:

- The Statistical approach
- The developmental approach

Statistical hypothesis and population parameters

Statistical hypothesis approach is inferential; based on estimation of population parameters from a random sample to describe population characteristics. It is stated in mathematical/statistical terms that make it possible to calculate the probability of possible samples assuming the hypothesis is correct. It is comparative in nature for factor effects of interest. The hypothesis testing may be for one or more of the population parameter:

- Mean μ
- Median **M**
- Standard deviation σ
- Proportion π

Statistical hypothesis can be stated in the Null or Alternative forms, and Non directional or directional form.

The Null hypothesis expects equality: **Ho: $\mu_1 - \mu_2 = 0$ or Ho: $\mu_1 = \mu_2$**

The Alternative expects non equality: **Ha: $\mu_1 - \mu_2 \neq 0$ or Ho: $\mu_1 \neq \mu_2$**

Directional and non-directional form of hypothesis is about the area of rejection of the hypothesis in the distribution function. Directional hypothesis has rejection area to one tail of the distribution:

Ho: $\mu_1 = 5$ a specified value: **Ha: $\mu_1 > 5$ or Ha: $\mu_1 < 5$**

While Non directional hypothesis has rejection area to either of the tail of the distribution

Ho: $\mu_1 - \mu_2 = 0$; **Ho:** $\mu_1 = \mu_2$
Ha: $\mu_1 - \mu_2 \neq 0$; **Ho:** $\mu_1 \neq \mu_2$

An example of hypothesis for testing single mean in non directional form is stated as:

Null **Ho**: $\mu_1 - \mu_0 = 0$ Alternative **Ha**: $\mu_1 - \mu_0 \neq 0$

In the distribution function, rejection area is to either of the tail, hence the term two sided or two tailed test. In directional form, the rejection is within one specified tail area, hence the term one sided or one tailed test.

Ho: $\mu_1 > \mu_0$ Ha: $\mu_1 < \mu_0$

Stating hypothesis for testing two means in non directional form:

Ho: $\mu_1 = \mu_2$ Ha: $\mu_1 \neq \mu_2$

Stating hypothesis for testing two means in the directional form:

Ho: $\mu_1 \leq \mu_2$ **Ha:** $\mu_1 > \mu_2$

The statistical approaches used to test hypothesis are addressed in AICM 702: Statistical methods and includes:

1. Confidence Interval (CI) which define the range of values within which the true population mean (ì, ð) lies with a certain probability (99%, 95%, 90%), and may be estimated for large or small sample size with the formula:

Large population size: $CI = \bar{y} \pm Z\alpha \times SE$

Small sample size:) $$CI = \bar{y} \pm t_{\alpha}^{n-1} \times \frac{S}{\sqrt{n}}$$ $CI = \bar{y} \pm t_{\alpha}^{n-1} \times \frac{S}{\sqrt{n}}$

The decision to reject or not to reject the null hypothesis is based area where the CI estimate falls, which is illustrated in here.

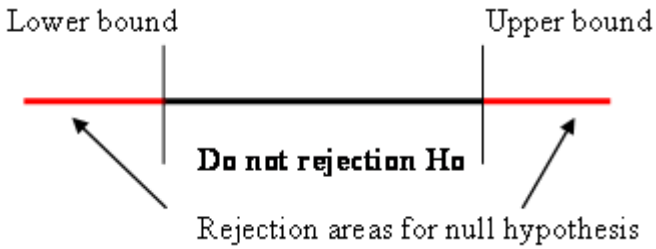


Fig- Rejection areas for null hypothesis

The illustration shows the area where the estimated CI is rejected when falling outside or is not rejected when falling within the area of expectation/assumption if the null hypothesis were true.

2. Test statistics is based on statistical procedures appropriate for the sample data distribution function to test the stated hypothesis. A general formula for test statistics is:

Test = Estimate – Hypothesis / SE

Commonly used test statistics for testing hypothesis include:

a. z - test
b. t – test, the student t test
c. F – test
d. χ^2 - chi square test
e. Sign rank test
f. Other specialized tests statistics

Developmental hypothesis

Developmental hypothesis may not be statistically tested. The hypotheses are on objectives relating to macro development goals such as the Millennium Development Goals or national development goals. These are the higher level goals in the logical framework. The definition can be developed through participatory problem analysis with the primary beneficiaries in identifying entry points for development intervention, analysis of the objectives and activities and analysis of important assumptions that is likely to be barriers to the attainment of the stated objectives. Therefore developmental hypothesis is more relevant to development projects rather than academic projects such as thesis.

Summary of Topic

Hypothesis is formulated in comparative statements that:

- Compare the value of parameter estimates
- Compare the effects of factors/ treatments
- Compare the association between factors

Hypothesis is stated at the time of developing the research concept in order to:

- Aid design of questionnaire
- Guide planning of data collection process
- Aid choice of appropriate analytical procedures

Learning Activities

Case study for the problem of low milk sale price earlier introduced in the previous chapter to illustrate stating of hypothesis:

Case study I: Formulating statement of the problem

In Molo peri-urban area, many smallholders sell milk at low prices despite stiff competition for milk between the formal and informal market outlets. Consequently, farmers have low returns to external inputs, which they use to support intensive milk production. The access, timeliness, reliability and use of market information between producers and traders could explain why some producers are sell milk competitively while others are not. Knowledge of the flow and use of market information on milk prices would be useful in designing effective price information dissemination strategies to help farmers sell their milk at profitable prices and realize positive returns to their investment for improved income and food security from dairy production.

Overall objective:	To enable milk producers access competitive market price in order to produce milk profitably and increase income and food security for the households
Specific objectives:	To determine the influence of market information sourcing on the milk sale price
Research question:	Do producers sourcing market information sell their milk at more competitive prices than those who do not?
Hypothesis Ha:	Milk producers that source market information about milk price sell milk at more competitive prices than those not sourcing for market information
Hypothesis Ho:	Market milk sale price is not different between producers sourcing market information and those not sourcing

Exercise

Applying the knowledge and skills gained on the above case, present the overall study objective, the specific objectives and their corresponding research questions and hypotheses stated in both Null and Alternative form for the assignment 1 in the last chapter. The first case was about unsuccessful application of an integrated pest management (IPM) scheme and crop rotation by some farmers in western Kenya to increase incomes and maintain soil fertility. The second case was about failure by many smallholder milk producers in Molo peri-urban area to sell milk at competitive price despite the competition for their milk

Assignments

1. An ICT Company has been experiencing revenue decline for the last two years. The company has identified that the problem is emanating from the declining number of people using their ICT products. Their analysis points to negative perception of the company by the public. Based on this brief, in groups, define the research objective and research question for this case.

2. In groups, review at least five MSc and PhD proposals and theses to evaluate the stated objectives, hypothesis and research questions if these are clear on what is being tested and measurements. Identify the relationships hypothesised between dependent and independent variables
3. Individually conduct a web search and identify three journals that publish ICM research. Scan to identify 3 titles published within the last five years that are closest to a topic of research you are interested in for your MSc thesis. From among the three paper titles you selected, choose one and from its contents, extract the excerpts on:

a) The researchable issue addressed
b) The objectives of the study
c) The research questions answered
d) The hypotheses tested
e) The conceptual framework
f) The data collection approaches used
g) The statistical methods used

References and Additional Reading Materials

MSc and PhD research proposals and theses.

GEAR CD supplied by RUFORUM

Journals of communication and information management

Trochim, W. (1989). Concept mapping: Soft science or hard art? In W. Trochim (Ed.) A Special Issue of Evaluation and Program Planning, 12, 87-110.

Trochim, W. Reliability of Concept Mapping. Paper presented at the Annual Conference of the American Evaluation Association, Dallas, Texas, November, 1993

Useful links

www.ilri.org/rmg/index.php?option=com_content *:* Customised courses for specific groups; On-line *research methods* resource materials; Biometrics and *Research Methods* Teaching Resource"

books.google.co.ke/books?isbn=0415223121: Catherine Hakim - 2000 - Social Science - 256 Covering both theoretical and policy research; this is a practical overview of the central issues involved in the design of social and economic research.

http://thesisnotes.com/conceptual-framework/sample-conceptual-framework/

Topic 5: Theoretical Framework and Conceptual Model Development

Introduction

A theoretical framework specifies which key variables influence a phenomenon of interest and therefore what variables to measure and the rationale for relationships between the variables. This will inform what statistical relationships on which to focus the tests. Theory guides every aspect of research, from formulation of the research question through operationalization and discussion. Research has two realms: the theory and observation. Theory is the scientific basis to explain an observed phenomenon which the researcher applies to gain better understanding of the phenomenon. On the other hand, observation is what occurs in the real world or measures. Researchers conduct research on the basis of theory about what is observed.

A conceptual framework is an illustrated representation of an idea or body of knowledge based on individual understanding of the relationships between the variables. The illustration shows the variables and the hypothesized relationship between them.

Learning Objectives

Upon completion of this topic, the learner should be able to:

- Explain the relationship between Theoretical framework and Conceptual model
- Develop a well understood conceptual model relating to research questions and hypotheses of the study
- Critique theoretical frameworks and conceptual models in research reports.

Key Concepts

- Theoretical framework
- Conceptual model

5.1. Why Present Theoretical Framework and Conceptual Model in Research Projects

Presentation of the theoretical framework and conceptual model is to aid in explaining the underlying assumptions for relationships between the variables, choice of model and to adequately specify the model parameters. They build upon existing knowledge and relevant theory; forms basis for the hypotheses and choice of research methods. Presenting these allows for articulation of why and how questions beyond simply describing a phenomenon observed to generalizing about various aspects of that phenomenon.

Researchers present these frameworks to identify the limits to the generalizations made. Their presentation facilitates identification of variables likely to have most influence on the phenomenon and how those variables might differ between the different populations.

5.2. Describing Theoretical Framework and Conceptual Model

Theoretical framework and Conceptual model are related and complements conceptualisation of the research regarding the objectives, hypotheses and research questions, and the independent and dependent variables. The theoretical basis is founded on known principles and the variables from literature are used for developing Theoretical framework and Conceptual model . Conceptual model is a mental illustration of how the variables are likely to relate to each other.

From literature and possibly consultation with experts, identify the theory that best explains the relationships between the variable regarding the observed phenomenon and the relevant assumptions. Theory is selected on the basis of how best it can explain the relationships among the variables and should reflect a connection between the theoretical framework, conceptual framework, operationalization and instrument.

The conceptual model of a research is a schematic illustration of how the relationships between the variables are visualized. We use two case studies to illustrate this.

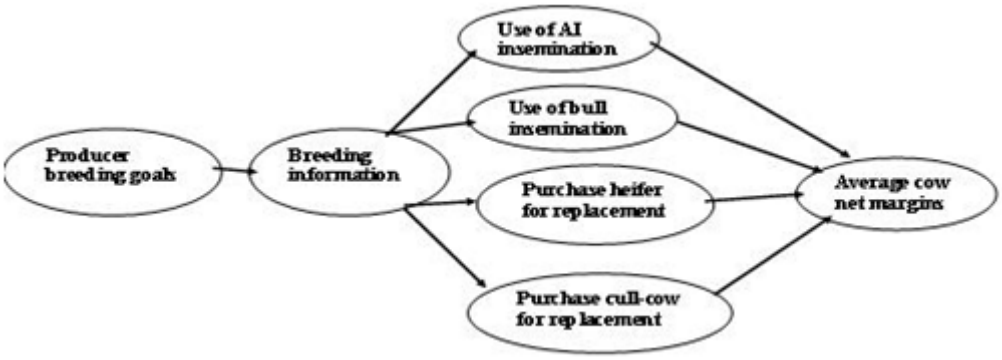
Case study 2: Research objectives

The study was to test the hypothesis that farmers' breeding goals have substantial influence on whether they seek for breeding information to inform their insemination decision (bull or Artificial Insemination) and replacement (purchase of a heifer or a cull cow for breeding). The breeding decisions made will subsequently impact on profitability of the dairy herd.

The specific objectives of the study were stated as:

i. To compare breeding information that smallholders seek for in choosing and purchasing a bull or AI when inseminating own cows.
ii. To compare breeding information that smallholders seek for when purchasing heifers or cull cows for herd replacements.
iii. To compare the information barriers in the flow of breeding information about insemination and replacement between farmers of Rift valley and their trade partners in Nyanza and Western provinces.
iv. To compare estimated net margins associated with insemination using a bull or Artificial Insemination and with purchase of a heifer or a cull-cow.
v. To compare estimated net margins associated with inseminations using a bull or Artificial Insemination and with purchase of a heifer or a cull-cow in dairy herds of Rift Valley supplying breeding stock to smallholders of Western and Nyanza provinces.

For implementing this research, an investigator could conceptualise the associations of breeding information and decisions with the net margins, the proxy for profit, in a dairy herd with the illustration presented. Breeding goal of the producers is hypothesized to influence seeking of breeding information for decisions on insemination (bull or Artificial Insemination) and replacement (purchase of a heifer or a cull cow). The breeding decisions subsequently impact on the realized profitability in the dairy enterprise.



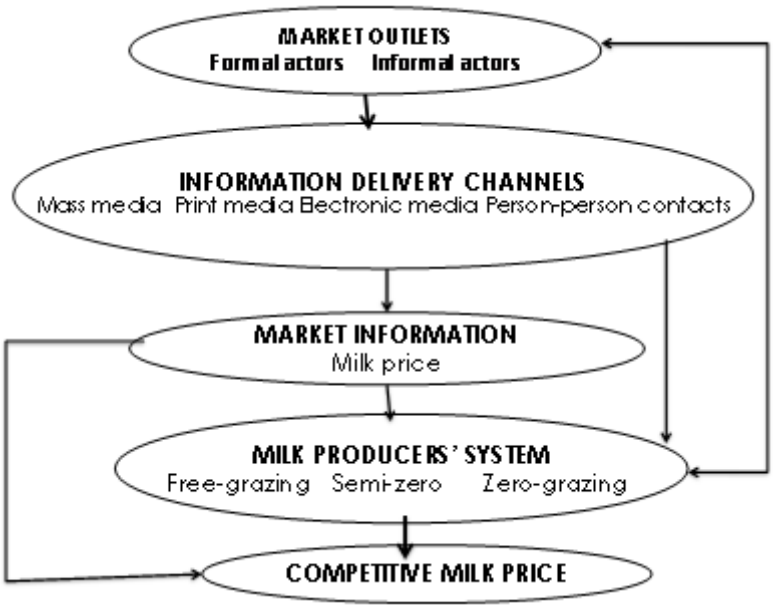
Case study 3: Problem statement

In Molo peri-urban area, smallholders sell milk at prices not competitive enough to ensure positive returns to external inputs they use to support intensive milk production. Market information flow between milk producers and market actors could be impacting on the price that farmers receive for their milk, therefore access, timeliness and reliability and use of market information could explain why some are able to sell competitively while others are not. Knowledge of the flow and use of market information on milk prices would be useful in designing effective price information dissemination strategies to help farmers sell their milk at profitable prices and realize positive returns to their investment for improved income and food security from dairy production.

For this research issue the research questions that could interest a researcher in finding answers are possibly:

i. Is sourcing of market milk price associated with sale of milk at competitive price?
ii. Are milk producers and traders effectively communicating on milk price information?
iii. To what extent is frequency of using a channel of information delivery correlated with reliability, accuracy and timeliness of that channel in the communication of milk price information?
iv. What are the information needs on market milk price communication by producers and market actors?

The problem described and the research questions raised could be conceptualized in the model illustration presented. Study the conceptual model and number appropriately the lines of connections corresponding to the each specific research question.



Summary of Topic

Conceptual model is a mental illustration of how the variables are likely to relate to each other. Theoretical framework and Conceptual model are related and complements conceptualisation of the research regarding the objectives, hypotheses and research questions, and the independent and dependent variables. The theoretical basis is founded on known principles and the variables to select for developing Theoretical framework and Conceptual model are identified from literature.

Learning Activities

One

For the two conceptual models illustrated, state the corresponding:

- Overall study objective
- Specific objectives of the study corresponding to the research questions of the study
- In the conceptual model, indicate by number (RQ1 to RQ5) the arrow line depicting the question to be answered.

Two

For selected researchable issues in AICM towards the student thesis project, present the following sections:

a) Title of the research
b) Statement of the problem
c) overall and specific objectives of the study

d) The research questions corresponding to the specific objectives
e) The hypotheses corresponding to the specific objectives
f) The conceptual framework clearly indicating the hypotheses /research questions
g) Theoretical framework for the conceptual model

Useful link

http://thesisnotes.com/category/theoretical-framework/ Accessed 24th June 2011.

http://www.aginternetwork.net/whalecomonlinelibrary.wiley.com/whalecom0/doi/10.1111/j.1574-0862.2010.00462.x/abstract

http://: www.elsevier.com/locate/im < Information & Management>

Topic 6: Study Designs

Introduction

A study design is a plan of systematic data collection process to meet the objectives for concluding whether a null hypothesis can be rejected or not, or providing valid answers to the research questions. The process in design includes sampling techniques, the measurements and application of randomization, replication and blocking concepts. Data is collected on factors, treatments and response variables identified in the hypothesis and research questions to enable statistical testing. Planning and implementation of research is based solving the problem defined, finding valid answers to research questions and making decision to reject or not to reject the hypothesis. The planning builds on the conceptual mode and theoretical framework explaining the associations between the variables.

Learning Objectives

By completion of this topic, the learner should be able to:

a. Describe different study designs
b. Identify the replicates and blocks in a research design
c. Apply randomization process in research implementation
d. Calculate appropriate sample size for a given research design.

Key Concepts

- Population
- Sample
- Sampling unit, Experimental unit, Observational unit
- Survey designs, Experimental designs, observational design
- Participatory Rapid Appraisal
- Sample size determination
- Randomization process
- Replication and blocking concepts

6.1. Study Designs

There are several approaches to design a research study applicable to AICM research projects to collect the needed data from the study subjects. The type of study designs reflects how the response variable is obtained from the subjects. This may be either through planned intervention to induce the response or non intervention on the subject. Another aspect the design reflects is the number of subjects used in the study to obtain measurements of the same variable at one time of measurement and the number of times the same measurement is taken from the subject for the period of data collection process. The study subjects on which measurements are made or observed are referred to as experimental units, observational units or sampling units, depending on the type of study design used. The general scheme defining types of research designs is illustrated below.

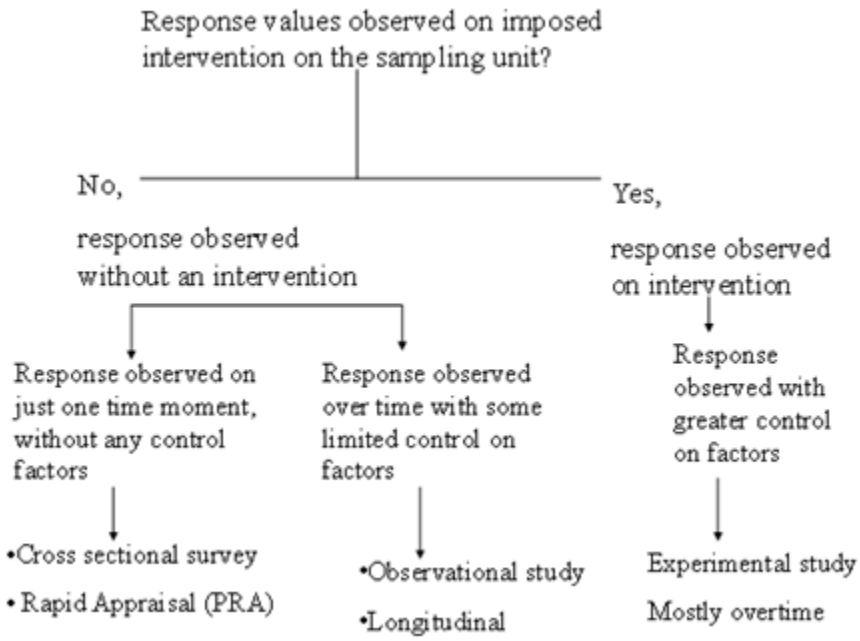


Figure 1: Classification of study designs (Source: Bebe B O. Lecture Notes

There are statistical terminologies used to define the considerations in a research study, which are introduced first to give clarity in the explanations that will follow.

Population: A set of subjects or individuals having certain characteristics which a researcher wishes to make a statement about by means of empirical investigation of a sample

Sample: The subset of the population actually selected for taking or observing measurements

Experimental unit: Individual or subject which receives the intervention (factor, treatment) that the researcher choices to give, or assign them to, to take measurement (response) *Sampling unit* in a survey, *observational unit* in observational study: are equivalent to experimental units in experimental study designs.

Location of the study: This is where you obtain the appropriate study population with set of values of the characteristics which the research objective is focusing on

The types of study designs are now described.

a) Experimental study designs

An experimental design is where a researcher imposes an intervention or treatment on the experimental units then measures the response. The Researcher measures the effect of intervention which forms factors/treatments in the hypothesis/research questions, usually in multiples or replicates. Researcher compares response to the

factors /treatments corresponding to the hypothesis. The example of this in AICM research may be:

Case study 4:

A researcher teaching farmers how to use several ICT facilities to improve decision making in input use in farming then after some period of time determine the association between the frequency of use and the farm productivity, or comparing farm productivity between those trained and untrained to use ICT facilities.

The conditions that define an experimental design are:

- Researcher can control some conditions to reduce experimental error
- Assessment can only be made on the factors imposed on the subjects or the experimental units. Effect of the treatment/intervention is observed to detect whether the subjects express change in response to answer the question of whether the subjects respond differently to the same treatment.
- A control treatment is included and this is the baseline to which change in response of other treatments will be compared against. The researcher chooses the factors / treatments relevant to the stated hypothesis/ research questions
- The subjects are randomly and independently selected from the target population then are randomly allocated to treatments/ interventions
- Measurements are undertaken under controlled conditions. The response is random variable while Factors / treatments are fixed variables. Variables are fixed when researcher selects them prior to start of the data collection and remains in the same state throughout the period of data collection.

b) Survey study designs

Survey is the process of using a questionnaire or interview protocol to collect data from the study subjects (respondents). Surveys may be cross sectional if measurement is taken only at one time on the respondent. Survey is longitudinal if measurements on the respondent repeated over time.

Cross sectional surveys are cheap, quick to implement, no risk on sampling units and allows for evaluation of multiple factors. However, there are a lot of noise factors not under the control of the researcher as is the case in experimental designs.

The characteristics describing survey designs are:

- Researcher impose no intervention on the sampling units (subjects)
- Researcher obtains information from sampling units on the existing situation now, in the past or in the future, for instance adoption, attitudes, behaviour, preferences, practices
- Can capture response on basis of recall, perception or observation on as is basis
- Sampling units are randomly selected from the target population

c). Participatory Rapid Appraisal

Participatory Rapid Appraisal (PRA) methodologies are techniques for rapid analysis of the situation on the ground through interviews with key informants within the community. The informants are those knowledgeable of the situation being investigated within the community. PRA are popular for collecting supportive secondary data used to strengthen research process, verification of the data collected and for direct engagement of the stakeholders in research process. The many uses of PRA include community action plans, spatial mapping and institutional analysis and mapping including SWOT analysis. Some of the PRA research approaches are illustrated now.

i) Stakeholder Analysis

This is analysis of roles of stakeholders and their contribution to development of the community and the linkages currently existing in serving the community.

Step 1: identify the stakeholders, their roles and associated challenges and strengths in the roles they play. An example is illustrated for FAO and ASARECA as stakeholders in agricultural information management.

Stakeholder	**What they do**	**Challenges faced**	**Strengths recognized**
FAO	Generation and of agricultural knowledge	Bridging gap in knowledge generation and application by farmers	Application of ICT in knowledge management
ASARECA			

Step 2: Pair wise ranking of strengths towards contribution made to the community development

STAKEHOLDERS							SCORE	RANK
	A	**B**	**C**	**D**	**E**	**F**		
A		B	C	A	A	A	3	1
B			B	B	E	F	2	2
C				D	C	F	1	3
D					E	F	0	4
E						E	1	3
F								

This analysis identifies stakeholder A to play greater developmental roles to the community

Step 3: Mapping linkages between the institutions

This involves representing an institution with an oval circle and those with a linkage are crossed over, indicating a shared role.

ii) SWOT analysis

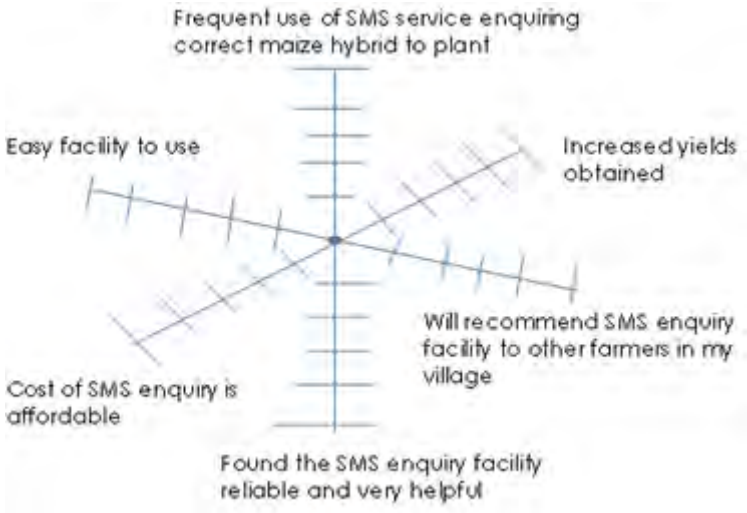
This analysis can be conducted to assess the capacity of stakeholders and identify areas that need strengthening and those to maintain, the threats and how to overcome them.

Internal in origin	Strengths	Weaknesses
External in origin	Opportunities	Threats

iii) Monitoring and Evaluation wheel

This M&E wheel is constructed to scale in spider web shape to provide a quick visual index helpful in assessing issues being monitored. The scale measurement indicates strengths and weakness for a given issue that stakeholders respond to. On the scale, stakeholders rate their agreed impact, which when joined form the web. The protruding spikes indicate the areas of strengths and dented spikes indicate weak areas based on community perspectives. An illustration for community evaluation of adoption of SMS facility for enquiring about the correct hybrid maize seed for a given with details on agronomic practices is presented.

M&E wheel for evaluating impact of using SMS facility for hybrid maize planting



d). Observational study designs

Observational study designs are where the researcher:

- observes sampling units for factors/treatment of interest
- has limited control on what is done on the observation or response and the factors / treatment to include in the study

- Observation may be made on one time or overtime

a) Case control study
- Sampling is intentionally targeted at the factor of interest
- Sampling units exposed to disturbed and undisturbed conditions,
- response then compared between disturbed and undisturbed conditions

b) Cohort study- groups
- sample units grouped then one group exposed and the other not exposed to factor/ treatment of interest
- Selected sampling units from each group observed overtime for response to the factor/ treatment
- Time of observation for response may be in prospective or retrospective
- Sampling units in prospective observation selected on basis of current exposure to the factors
- Sampling units in retrospective observation are selected on basis of past exposure to the factors

6.2. Sampling Techniques

Sampling is the process of selecting a sample from the population on which to do measurements or allocate treatments (the subjects being sampling, observational or experimental) in the research. A representative is obtained through random selection to yield valid results for making inferences about the population characteristics. In this way sampling helps to reduce cost in terms of human time and labour inputs and financial inputs, facilities requirements like lab, meeting halls. Sampling aims at obtaining a representative of the target population.

The sample selected contains only part of the information about the population, therefore true values of population parameters is not obtained but is estimated. The estimate is informative about the population when sample is selected through probability process, which yields unbiased, accurate and precise estimate.

In some instances, sample may be obtained through non probability process in purposive and convenient sampling. This sampling approach does not allow subjects an equal chance for inclusion in the study. This leads to biased, inaccurate, imprecise estimates but may be used when the target population is unique or when there is need for a detailed case study. It is possible to obtain sample purposively or conveniently if there are access barriers to obtaining sampling units in a random manner or resources are limited and have to be used efficiently.

The approaches to enable a researcher obtain an informative, unbiased, accurate and precise estimates are application of the concepts of randomization, replication and blocking or stratification. These concepts are discussed in next sections within this chapter.

The various approaches to obtain a sample through probability process are:

i. **Simple random sampling** in which all sampling units have equal chance of being selected for the study

ii. **Systematic random sampling** in which each sampling unit is randomly selected at regular intervals to evenly spread their distribution within the sampling frame

iii. Multistage sampling

- Divide population into distinct groups, based on priori information
- Then sampling process in e"2 stages
- For instance; Zone, farming systems, households, individual members of the household

In this case the at first stage select k groups randomly from n groups and at the next stage you sampling units within each group randomly

iv. Stratified sampling

- Sampling frame divided into groups: strata
- Next apply systematic or simple random sampling to obtain sample units from each strata
- Useful in heterogeneous population to group homogenous individuals together for a specified criteria
- Allows for unequal proportion of sampling units within each strata

v. Cluster sampling

- Population divided into cluster units, then random sample of clusters selected among the clusters
- All sampling units within selected clusters are then examined
- Clusters are equivalent to strata
- In cluster sampling, clusters are randomly selected, but not necessarily random in stratified sampling
- Sampling units within strata are all sampled in cluster, but randomly selected within each strata in stratified sampling
- Multistage sampling is an extension of Cluster sampling, with e" 2 stages of sampling

6.3 Randomization Process

Randomisation is a process of randomly selecting sampling units from the population and randomly allocating sampling units to treatments/factors. Each sampling unit has an equal chance of being selected or being allocated to any treatment. This process allows for making statements about the target population characteristics. The process does remove bias in the allocation of sampling units to treatments/factors and enhance valid statistical testing / answer to research question based on normal population distribution.

The principles of randomization require that:

a. All sampling units have equal chance of being allocated to any of the treatments / factors

b. The allocation of one sampling unit to a particular treatment has no influence on the allocation of the next sampling unit to any of the treatments/factors
c. It is not known in advance the treatment that a particular sampling unit will be allocated to.

Researcher may conduct randomization process with the use of

- Statistical random number tables
- Scientific calculator
- Statistical software

Randomization procedure using random number tables

i. List sampling frame, numbered from 1 to n
ii. Generate random numbers from the tables
iii. Rank the generated random numbers from lowest to highest
iv. Match the samples with their ranked random numbers
v. The random number that the sample unit is matched to becomes the allocated treatment /factor, giving the selected units.

Randomization procedure using scientific calculator

i. Number the sampling units in the sample frame from 1 to n
ii. Generating random numbers with the calculator
iii. Rank the generated random numbers from lowest to highest, ignore the first zero
iv. Allocate sampling units to treatments/factors by matching sample unit number to ranked random number

Randomization procedure of using Statistical software
SPSS statistical software

Open data base and enter variable names for sampling frame
Number the subjects in sampling frame from 1 to n
Select from the menu bar:

Data>select cases> random sample cases> sample size: Approx %; Exactly n cases> continue: delete, Filtered>Ok

Selected cases made new variable, columns with filtered $ and are labelled 1 while unselected are labelled cancelled /

6.4. Blocking and Replication Concepts

Blocking and stratification are essentially same concept. Blocking or stratification is a procedure of grouping study subjects into more homogenous groups where the study subjects exhibit large heterogeneity in characteristics. Blocking is important in experiments in which experimental units are heterogeneous while stratification is important in surveys where sampling units are heterogeneous. In practice blocking and stratification are used interchangeably in research designs.

Blocking or stratification in research serves the purpose of:

- Grouping the sampling units together in more homogenous groups of units before assigning them to treatments eliminates sources of variability / noise factors, thereby reducing the size of the experimental error.
- Partitioning sources of variability due to blocks or strata and due to treatments to improve precision of the estimate of the treatment effect compared to when blocks or strata are ignored
- administrative convenience in conducting a survey if estimates for each stratum are required
- reduce the total number of units that need to be sampled because the choice of blocks or strata leads to smaller variation among sampling units for within block / strata than among block/strata,

The decision to block or stratify is based on sampling units being heterogeneous, the response variable being subject to noise factors and researcher having prior knowledge of the effect of noise factors and characteristics of the target study population.

An example of blocking a population of beef farmers who use SMS marketing information systems facility for identifying market outlet offering completive price

Block / strata		**Factor levels**
Farming system	Market access level	
Rain-fed	Low Medium High	3
Irrigated	Low Medium High	3

The table illustrates:

- The blocking or stratification of the sampling units is by farming systems and by market access
- sampling units within a block or strata are of less random variability than the general population
- Random variability is smaller within a block or strata than between the blocks
- Treatments randomly allocated within each of the blocks or strata gives stratified randomization

The sampling units in a survey can be selected within strata. When the sampling units are randomly selected, the focus of the analysis is on the variations within and between the blocks/ strata. If blocking or stratification is successful in producing more homogeneous groups, then the residual variance will be less than if the study had been designed as a completely randomised design. For instance, having *b* blocks/strata with

equal assignment to *t* treatments from within blocks/strata, the analysis of variance structure for this study designs becomes:

Source of variation	Degree of freedom
Block /strata (e.g *Market access in low, medium and high*)	b - 1
Treatment (e.g *use of mobile phones in marketing)*	t - 1
Residual	(b - 1) (t - 1)
Total *(e.g number of farmers samples)*	n -1

Replication concept

Replication means repeating measurements at the level of sampling units at the same time. Replication may be in duplicates, triplicates or quadruplicates, for instance. In experiments, replicates would be three animals receiving the same treatment at the same time within the same block. Replicates are rarely used in a survey study designs, but would be measurements of same variable say feeding regime on more than one member of the household taken at same time and obtaining a mean of the measurements to represent the value estimate for them. Another example of replicate in survey is when taking say three measurements on a farmers' field at the same time and obtaining a mean of the three measurements for the response measured. Otherwise, cross sectional surveys have one replicate because measurement is done once on the head of the household. Here, a replicate does not represent sampling unit!!

Why replication?

- Replication in an experiment is used to yield balanced treatment designs by having the sampling units subjected to equal number of measurements, such that the measurement are equal at any time of observation. This leads to balanced design, called orthogonal which makes interpretation of statistical results easy.
- Increase variance and precision of the estimate
- Increase detection of effects of the treatment/factor through segregation of within units variability. The mean of replicates for each sampling unit gives the within variability.

6.5 Sample Size Determination

Sample size is the number of sampling units to obtain for the measurements in a study. This means determining how large of a sample should be taken to make an accurate estimation. The question of how large a sample size is appropriate for a study is often a challenge in research. To deal with this challenging question in research, the importance of sample is explained first.

Importance of sample size

1. Enhances probability of detection of differences between factors / treatments investigated in the study

2. Ensures that results are obtained within specified parameter:
   (a) Acceptable level of type 1 error. This is the probability of inferring a difference between factors/ treatments when actually there is none.
   (b) Acceptable test power. This is based on: **$1 - \beta$,** the probability of correctly inferring a difference. Conventionally, **$\beta = 4 \times \beta$** , that is : **$4 \times 0.05 = 0.2$;** therefore **$1 - \beta = 1 - 0.2 = 0.8$**, which represent the minimum test power that is acceptable.
   (c) The level of precision of the estimate on mean, median or proportion of the population. Precision is defined by Standard Error of the Mean: **$SE = SD / \sqrt{n}$.** Precision is improved by increasing n or reducing standard deviation
   (d) Choice alternative hypothesis, which can be specified as either one tailed or two tailed test.

Information required determining sample size

Calculation of the needed sample size for a given study is based some prior knowledge about the study population. To determine the sample size required you need knowledge of:

a. Number of samples required: one sample, two samples, three or more sample population?
b. Previous estimates of population parameters obtained from the target population, such as mean, standard deviation, coefficient of variation
c. Setting the precision levels for the estimate at:
i. α =0.05 (95% CI), 0.001(99% CI), 0.1 (90% CI)
ii. test power $1 - \beta$: 0.8

Calculating sample size

There are several approaches to calculating sample dependent on the information available about the population. One general formula for calculating sample size is:

$$n = 2s^2 t2 / d2 \qquad \textbf{(i)}$$

Where

- n is the required sampling units for each level of the factor to be compared,
- s^2 is an estimate of the residual variance,
- t is the value of the Student t-test at the required level of significance: P=0.05 =1.960, approx 2: P=0.01=2.576 approx 2.6
- d is the difference between the two mean responses that needs to be shown to be statistically significant

Example:

The sample size is being determined for dairy smallholder households in the Kenya central highlands with three distinct land use systems. The sample size was obtained from estimating the number of observations potentially needed to distinguish between three land use systems a difference of 20% in some of the important household variables.

For a desired confidence interval of 95%, and using population parameters of: a coefficient of variation of 68%, previously observed in the target population, the minimum sample size in each land-use zone was calculated from:

$$n = 2\left[\frac{ZC}{d}\right]^2$$

Where

z = 1.96 for 95% confidence interval, c is coefficient of variation = 68% and d is level of difference set at 20%. The formula yields a sample size of (88.82) 89 in each land-use zone. In principle, the sample size eventually used reflects the concept of optimum sample size, which is:

- Large enough to have a good chance of detecting differences between the factors of interest,
- Yet not so large to be wasteful in terms labour, finance, facilities and time required to obtain access all the sampling units
- Not so small to yield unrepresentative sample, invalid information and estimates,
- A compromise in statistics and reality. Sometimes the determination of an optimal sample size may justify the need to rethink the original objectives and corresponding hypotheses. A decision needs to be made between what is desirable in terms of numbers of units from a statistical point of view, and the number of units that can be managed logistically.

In general, the sample size determination formulas are based on formula for the maximum error of the estimates then solving for n (http://people.richland.edu/james/lecture/m170/ **)**

Summary of Topic

In summary, each of these different methods of sampling, namely random, representative, convenience and purposive, may feature within study design. As far as possible samples should be drawn at random to ensure that sample estimates can be extrapolated to the population that they represent. Sampling procedure has implications on the results and on the inferences that can be made from the data.

Learning Activities

One

Individually using the formula given, calculate sample size needed to distinguish between three milk marketing systems a difference of 20% in some of the important milk marketing variables of the household within 95% confidence interval for a population that past studies estimated a coefficient of variation of 68% in the amount of milk sold daily, applying the formula:

$$n = 2\left[\frac{ZC}{d}\right]^2$$

Two

For at least four research reports, compare the sample size estimate with the actual sample size obtained, and make a general statement.

Reference and further reading materials

GEAR Research methods resources in http://www.**ruforum**.org/

http://people.richland.edu/james/lecture/m170/ for sample size determination

Topic 7: Data Collection Methods

Introduction

Data collection is a systematic process of collecting information about objects of study (people, objects, phenomena for instance) in order to test the hypothesis or answer the research questions of the study. Being systematic in data collection is a prerequisite to testing the hypothesis or answering the research questions in a conclusive way. Data collected haphazardly is difficult to analyse. The data collected is analysed and interpreted to turn it into information and results or findings. An empirical research can apply one or more methods to collect the data desired for the study. The approaches available for data collection in AICM research includes:

- Testing respondents about their level of knowledge, skills or information on reliability and validity
- Questionnaires administration directly to respondents
- Interviews with the respondents
- Focus group discussion
- Observation of the actual present situation and what people actually do
- Extraction of existing data, usually termed secondary data

Learning Objectives

By completion of this topic the learner should be able to:

a. Describe various data collection techniques and state their uses
b. Explain activities in the data collection process in research
c. Explain the strengths and weaknesses of data collection methods
d. Identify data collection approaches in published papers and theses
e. Identify various sources of bias in data collection and ways of preventing bias.
f. Plan data collection for a research
g. Design a questionnaire for data collection

Key Concepts

- Test
- Questionnaire
- Focus Group Discussion
- Observation
- Secondary data

7.1. Observations for Data Collection

Observation is a method of data collection where the researcher observes participants in natural and/or structured environments. Observation can be carried out in a laboratory environment or just natural environment in real-world settings. Data to collect may be qualitative or quantitative. Observation may involve standardization procedures to generate quantitative data. The standardization can be in respect of who and what is

observed, when and where the observation takes place, and the instruments used, the time interval of observation and procedure of sampling.

Strengths of observational data

- Allows one to directly see what people do without having to rely on what they say they do.
- Provides firsthand experience, especially if the observer participates in activities.
- Can provide relatively objective measurement of behaviour when using standardized observations.
- Observer can determine what does not occur.
- Observer may see things that escape the awareness of people in the setting.
- Excellent way to discover what is occurring in a setting.
- Helps in understanding importance of contextual factors.
- Can be used with participants with weak verbal skills.
- May provide information on things people would otherwise be unwilling to talk about.
- Observer may move beyond selective perceptions of people in the setting.
- Good for description.
- Provides moderate degree of realism when done outside of the laboratory.

Weaknesses of observational data

- Reasons for observed behaviour may be unclear.
- Reactive effects may occur when respondents know they are being observed
- Investigator effects may occur from personal biases and selective perception of observer
- Observer may “go native” if turn over-identifying with the group being studied.
- Cannot observe large or dispersed populations.
- Some settings and content of interest cannot be observed.
- Collection of unimportant material may be moderately high.
- More expensive to conduct than questionnaires and tests.
- Data analysis can be time consuming.

7.2. Tests for Data Collection

Tests may be used in research that wants to measure personality, aptitude, achievement, and performance on specific knowledge, skills, behaviour, or cognitive activity and also can be used to complement other measures where necessary. Developed standard tests can be found in the internet, which researcher may consider using if relevant to the hypothesis or research questions of the study. There are strengths and weaknesses in using test methods in data collection now outlined below.

Strengths of using standardized tests approaches in data collection

- Can provide measures of many characteristics of people.
- Standardized measure in subjected all respondents to same treatment.
- Allows comparability of common measures across research populations.
- Strong psychometric properties yielding high measurement validity.
- Availability of reference group data.

- Many tests can be administered to groups which saves time.
- Can provide “hard,” quantitative data.
- Tests are usually already developed.
- A wide range of tests can be used to tap more content.
- Response rate is high for group administered tests.
- Ease of data analysis because of quantitative nature of data.

Weaknesses of using tests approaches in data collection

- Can be expensive if test must be purchased for each research participant.
- Reactive effects such as social desirability can occur.
- Test may not be appropriate for a local or unique population.
- Open-ended questions and probing options are not available.
- Tests are sometimes biased against certain groups of people.
- Non response from the respondents to selected items on the test.

7.3. Questionnaires for Data Collection

A questionnaire is an instrument for data collection in survey studies which may involve person to person interview, telephone interview or web based for participants to fill out. A questionnaire is composed of questions and/or statements. Here is an example of a questionnaire designed to collect data on the adoption of AICM tools in marketing

Questionnaire on Adoption of ICM tools in marketing dairy products

Objectives : Collecting information on the adoption of ICM tools in marketing of dairy products by smallholder dairy farmers.

1.1. RESPONDENT'S NAME_____**1.2**. Date interviewed [___ ___ ___ ___]

1.3. LOCATION [___ ___ ___ ___ ___ _____ ___ ____]

A/1. Characteristics of the enterprise owner

Gender 1 = Male 2 = Female	Age (years)	Primary occupation	Years of experience in dairyprocessing	Education level
[___]	[__ ___]	[___]	[__ ___]	[___]

PRIMARY OCCUPATION
1 = Farmer; 2 = Employed; 3 = Businessman; 4 = Retiree

EDUCATION LEVEL

0 = No formal education	4 = Technical college (Agric., Teacher's etc)
1 = Primary school	5 = Adult literacy education
2 = Secondary school ('O' level)	6 = University_____
3 = Post secondary school ('A' level)	7 = Other (specify)

Page 1

B/1. Which dairy product do you process? [_____]

DAIRY PRODUCT

1 = Sour milk; 2 = Butter; 3 = Cheese; 4 = Yoghurt; 5 = Ghee; 6 =other (specify) ____

B/2. What processing technology do you use for the dairy products in B1 above? [____]

DAIRY PROCESSING TECHNOLOGY USED

1 = Traditional; 2 = Improved traditional; 3 = Modern technologies

B/3. Do you have difficulties selling your dairy products? [__] 1 = Yes, 2 = No

B/4. Rank three main reasons for searching for ICM tools for marketing your products?

[First _____] [Second _____] [Third _____]

MAIN REASON REASONS FOR SEARCHING FOR ICM TOOLS FOR MARKETING YOUR PRODUCTS?

1.= Not applicable; 2 = Find better price; 3 = Find a single buyer; 4 = Want more buyers
5 = Find a "regular" buyer; 6 = Others (specify) _____

Page 2

dairy products. This is closed questionnaire in which possible options of responses are pre determined, which enables coding of the data to facilitate data entry into an electronic database. Open ended questions may be preferred for qualitative data where opinion of the respondent can be captured. Open-ended questions are common in exploratory research and closed-ended questions are common in confirmatory research.

The guidelines for developing questionnaires

Divide the questionnaire into sections corresponding to the objectives of the study to ensure complete capture of all the data needed and to ease verification of completeness. The language use must be simple in away clearly understood by the respondents otherwise use a translator familiar with the language of the respondents to be interviewed. It is therefore necessary to consider the demographic and cultural characteristics of the potential study population.

Write the questions in clear, precise and short sentences in neutral wording. The responses will be misleading if the respondents cannot understand the questions well, which make the data collected invalid for the research questions or hypothesis being tested. Questions more easily understood are less stressful to and minimise emotional reaction or responses from the respondents. Avoid use of leading or loaded questions in order to get true position from the respondent. Equally avoid double-barrelled questions and double negative wordings. Such questions combine two or more issues in a single question. A question is likely to be double-barrelled if has the word **"and"** within the sentence. The question will be ambiguous because two or more ideas are confounded and the interviewer will not be clear of what the respondents are specifically answering, one or the other item or both items in the question.

The questionnaire can be designed to capture both objective and subjective measurement. The objective measurements can involve direct measurements while subjective measurements are commonly through rating and ranking in likert numerical scale. It is better to pilot the questionnaire before administering to respondents for data collection.

The strengths of questionnaires

- Good for measuring attitudes and eliciting other content from research participants.
- Can provide information about participants' internal meanings and ways of thinking.
- Can administer to probability samples.
- Quick turnaround.
- Can be administered to groups.
- Perceived anonymity by respondent may be high.
- Well constructed and validated questionnaires yields moderately high measurement validity and reliability
- Closed-ended items can provide exact information needed by researcher.
- Open-ended items can provide detailed information in respondents' own words.
- Ease of data analysis for closed-ended items.
- Useful for exploration as well as confirmation.

The weaknesses of questionnaires

- Reactive effects may occur when interviewee try to show only what is socially desirable
- Non response to selective items is common
- People filling out questionnaires may not recall important information and may lack self-awareness
- Response rate may be low for mail and email questionnaires
- Open-ended items may reflect differences in verbal ability, obscuring the issues of interest.
- Data analysis can be time consuming for open-ended items.
- Measures need validation, accuracy level is low.

7.4. Interviews for Data Collection

Interviews involve an interviewer asking the respondent questions either in-person or over the telephone. Interviewer uses structured or open ended questionnaire. Accuracy and validity of the data depends on the trust and rapport with the respondent. An interviewer sometimes needs to probe the respondent to reach clarity or gain additional information.

Interviews may be quantitative or qualitative. In quantitative interviews, the same information is provided to everyone, so making quantitative standardized. Qualitative interviews use closed-ended questions. There are three types of qualitative interviews

- Informal Conversational Interview, which is spontaneous, loosely structured because the no interview protocol is used.
- Interview Guide Approach which is more structured and has a protocol unlike the informal conversational interview. The interviewer can use a list of open-ended

questions about the items to capture. The questions can be asked in any order; wording can be changed when deemed appropriate.
- Standardized Open-Ended Interview in which open-ended questions are written on an interview protocol, and are asked in the exact order given on the protocol. In this case the wording of the questions cannot be changed.

The strengths of using interviews in collection of research data

- Good for measuring attitudes and most other content of interest.
- Allows probing and posing of follow-up questions by the interviewer.
- Can provide in-depth information.
- Can provide information about participants' internal meanings and ways of thinking.
- Closed-ended interviews provide exact information needed by researcher.
- Telephone and e-mail interviews provide very quick turnaround.
- Well constructed and tested interview protocols yields moderately high measurement validity and reliability
- Often relatively high response rates are attainable.
- Useful for exploration as well as confirmation.

The weaknesses of using interviews in collection of research data

- Can be expensive and time consuming if in-person interviews are used.
- There is possibility of reactive effects if interviewees may try to show only what is socially desirable.
- Investigator effects may occur if conducted by untrained interviewers who distort data because of personal biases and poor interviewing skills.
- Interviewees may not recall important information and may lack self-awareness.
- Perceived anonymity by respondents may be low.
- Data analysis can be time consuming for open-ended items.
- Measures need validation.

7.5. Focus Groups Discussions

A focus group is a situation where a focus group moderator keeps a small and homogeneous group, usually less than 15 people, focused on the discussion of a research topic or issue. Focus group sessions may be recorded using audio and/or videotapes. The method is useful for exploring ideas and obtaining in-depth information about how people think about an issue.

The strengths of focus groups

- Useful for exploring ideas and concepts.
- Provides window into participants' internal thinking.
- Can obtain in-depth information.
- Can examine how participants react to each other.
- Allows probing.
- Most content can be tapped.
- Allows quick turnaround.

The weaknesses of focus groups

- Sometimes expensive.
- May be difficult to find a focus group moderator with good facilitative and rapport building skills.
- Reactive and investigator effects may occur if participants feel they are being watched or studied.
- May be dominated by one or two participants.
- Difficult to generalize results if small, unrepresentative samples of participants are used.
- May include large amount of extra or unnecessary information.
- Measurement validity may be low.
- Usually should not be the only data collection methods used in a study.
- Data analysis can be time consuming because of the open-ended nature of the data.

7.6. Collecting Secondary Data

Secondary data is data that was originally used for a different purpose while primary data is original data collected specifically for the research study. The most commonly used secondary data are documents, physical data, and archived research data.

The documents may be personal if data was written or recorded for private purposes such as letters, diaries or family pictures. Alternative documents are official documents which were written or recorded for public or private organizations such as thesis, newspapers, annual reports, yearbooks or minutes. Physical data are any material thing created or left by humans that might provide information about a phenomenon of interest to a researcher. Archived research data are the research data collected by other researchers for other purposes and can be saved in tapes, CD, databases, etc that are accessible to others.

The strengths of documents and physical data:

- Can provide insight into what people think and what they do.
- Unobtrusive, making reactive and investigator effects very unlikely.
- Can be collected for time periods occurring in the past (e.g., historical data).
- Provides useful background and historical data on people, groups, and organizations.
- Useful for corroboration.
- Grounded in local setting.
- Useful for exploration.

The strengths of archived research data:

- Archived research data are available on a wide variety of topics.
- Inexpensive.
- Often are reliable and valid (high measurement validity).
- Can study trends.
- Ease of data analysis.
- Often based on high quality or large probability samples.

Weaknesses of archived research data:

- May not be available for the population of interest to you.
- May not be available for the research questions of interest to you.
- Data may be dated.
- Open-ended or qualitative data usually not available.
- Many of the most important findings have already been mined from the data

Weaknesses of documents and physical data:

- May be incomplete.
- May be representative only of one perspective.
- Access to some types of content is limited.
- May not provide insight into participants' personal thinking for physical data.
- May not apply to general populations.

Summary of Topic

There are several techniques for collecting data and corresponding tools for data collection that a researcher may use to obtain the needed information from the study subjects. Techniques for data collection include observations, interviews, administering questionnaire and mining available information. The tools for data collection include checklist, datasheet or forms, questionnaire, tape recorder and specialised equipments for measurement taking. These techniques and tools can complement each other. It is therefore advisable for researchers to explore using a combination of different techniques that can reduce the chance of bias in order to give a more comprehensive understanding of the issue of the study.

Learning Activities

1. Select at five research reports and do the following:
a. Identify the study designs used
b. Identify the data collection techniques used
2. As individuals, state the study designs and data collection techniques to your research project
3. Design a questionnaire for collecting data you would need to address the research problem described in case study 1
4. Design a questionnaire for collecting the data you need for your proposed research project

References and further reading materials

http://www.idrc.ca/cp/ev-56606-201-1-DO_TOPIC.html for techniques and tools of data collection explanation and exercises

Topic 8: Data Measurements

Introduction

Data measurements are values associated with the variables to be analysed for purposes of testing the hypothesis or answering the research questions. Variable are measurements that take values which vary from one individual to another or from one sampling unit to another. The important considerations about data measurements are

- What variables to measure / observe?
- How to measure the variables?
- How often to measure the variables?

Learning Objectives

By completion of this topic the learner should be able to:

a. Identify variables in a conceptual model, hypothesis
b. Identify data measurements scales
c. Identity t data measurements used in research reports
d. Apply the data measurement techniques in data collection for research process

Key Concepts

a. Variables
b. Scales of measurements
c. Nominal and ordinal scale
d. Discrete, discontinuous and continuous scale

8.1. Identifying Variables to Measure in Research Process

In a research what variables to measure will be guided by

- Hypothesis and Research questions
- Theoretical and Conceptual frameworks
- Literature review
- Analysis of the problem to identify the causes and effects relationships

The question of how to measure the variables relate to the scales of measurements to use. The scales are:

- Qualitative or categorical scales
- Quantitative or continuous scales

The qualitative or categorical scales are measurements taking distinct categories, binary or dichotomous form such as male or female and yes or no responses. The measurements are in multi categories for instance education levels defined as: primary, secondary, diploma, graduate.

The scales of measurements can be nominal or ordinal.

- i) Nominal scale: distinct categories which define the variable are unordered. Each category can be assigned a name: Colour; white, black, red, etc.
- ii) Ordinal scale: the categories which constitute the variable have some intrinsic order, but have no consistent and defined intervals between the various categories: poor, average, good, excellent. The scales can be assigned values reflecting increasing magnitude: 1=poor, 2=average, 3=good, 4=excellent using Likert scale rating or relative ranking measurement. The Likert scales can be treated as quantitative variables in data analysis when the assumptions applicable are declared.

The quantitative or continuous scales consist of numerical values on well defined scale which may be discrete / discontinuous scale or continuous scale.

- Discrete scale take only particular integer values, typically counts: parity
- Continuous scale can take any value within lower and upper bounds: weight
- Interval between two adjacent points on the scale are of same magnitude:
- Magnitude between: 1 to 2 kg interval same to 3 to 4 kg interval
- Measurement process determines the accuracy: weighing scales- digital vs manual

Example1 8.1: In a survey of farm households to quantify uptake of ICT in herd recording

Measurement /response is: uptake of ICT herd recording

Experimental /sampling/observational unit is: Individual households, the level where uptake of ICT is measured

8.2. How Often to Measure the Variables

This is guided by

- The study design
- The objectives
- The resources available
- The desired level of accuracy
- The nature of the variable, for instance variables that values change within a short time like biological measures may require repeated measurements while variables that do not change within a short time like soil properties may be taken once.

Measurements are taken only once in cross sectional and PRA but in replicates in experimental and in observational study designs. When measurements can be taken in replicates, the sample size used is usually small unlike in the cross sectional surveys in which there are no replicate measurements. In this, a large sample size is preferred for better estimates.

8.3. Data Recording

Data collected must be recorded accurately and permanently for future use. Data may be recorded in datasheets, questionnaires or with use of electronic recording equipment. The guidelines for achieving this goal in research in general terms, depending on whether using datasheet, questionnaire or an electronic recorder, are:

i. Identify observational units
ii. Partitioned appropriately into sections corresponding to objectives/hypothesis to be tested
iii. Leave adequate spacing for clear recording of data/information in the paper
iv. Design coding for the variables
v. Record date of data being collected
vi. Uniquely identify the individual observational units
vii. Write in permanent ink to avoid easy rub off
viii. Maintain consistency in the number of decimal places used; realistic to accuracy of measurement used
ix. Enter data directly from sheet/questionnaire to electronic database
x. Avoid copying data from sheet of paper to another to minimise errors
xi. Variables that can be calculated /derived from raw data be computed within the database facilities to minimise errors.

Summary of Topic

- Concentrate on collecting data relevant to the objectives and hypotheses of the study.
- Record date
- Uniquely Identify the individual observational units
- Write in permanent ink to avoid easy rub off
- Maintain consistency in the number of decimal places used, realistic to accuracy of measurement used
- Avoid copying data from sheet of paper to another to minimise errors
- Variables that can be calculated /derived from raw data be computed within the database facilities.

Learning Activities

1. Select at least two research reports then do the following:
   Identify the variables in the reports
   Identify scales of measurement used
   Identify data measurement techniques or tools
2. Explain how you will measure variables in your research project
3. State the data measurement techniques relevant to your proposed research project
4. Pilot data collection for your proposed research project with the use of data tool you have designed.

Useful links

http://www.tardis.ed.ac.uk/~kate/qmcweb/m1.htm

Topic 9: Database Management

Introduction

Database is the data processed into computer readable form for computation with aid of computer software for analysis. Management of the database is an important aspect of research process. Collecting data is expensive and time consuming, but receive less attention. Yet, this is a process from start to the end of research process. Poorly managed research data will:

- Compromise scientific analytical quality of the data and information generated
- Lead to wrong conclusions,
- Waste resources: time, money, labour, facilities

This chapter presents the basic principles in database management to achieve quality and integrity for valid research results.

Learning Objectives

By completion of this topic, the learner should be able to:

a. Explain principles of database design
b. Design a database
c. Describe database management processes

Key Concepts

a. Database
b. Database software

9.1. Principles of Database Design

The design of database and questionnaire are much alike in principle

- Identify observational units
- Design coding for the variables
- Matrix format design
- Variables entered in columns for each observational unit entered in the rows
- Identity of observational units entered in the first columns followed with the measurements
- Data entry for missing values be consistent with the software design; May be: left blank, dot, BUT NOT zero value

Design may be simple to complex depending on the data structure

- Hierarchical structure design
- Multi level structure design
- Relational structure design

The database design software could be statistical packages, spreadsheet packages or special database management software. The advantages and weaknesses of these are outlined below.

Statistical packages: SPSS, Genstat, SAS, STATA

Advantages	Disadvantages
Data management and analysis possible	Usually unsuitable for multi-level data structures
Do most that a spreadsheet package does	Lacks security of a relational database management system.
Usually have programming capabilities	Graphical facilities not that quality

Spreadsheet: Excel

Advantages	Disadvantages
Good statistical exploratory facilities	Lacks security of a relational database management system
	Limited Statistical analysis capabilities; simple descriptive

Relational database management systems: Access, dbase

Advantage	Disadvantage
Secure	Needs expertise in database development
Handle complex multi-level data structures	Graphical facilities not good
Allows screen design for data input	Analysis limited to descriptive statistics
Have standard facilities for reporting	

9.2. Data Processing Procedure

The procedure for creation of database is essentially:

- Examine raw data collected for accuracy
- Uniquely identify the individual observational units, all other data in the dataset are linked to it
- Enter data directly from sheet/questionnaire to electronic database to reduce possible errors when transferring from one sheet to another

- Edit database and calculate new variables so desired when all row data has been entered
- Manipulate the data when necessary for deriving new variables
- Run descriptive statistics to recheck correctness of data entry and derived variables.

The minimum and maximum values and frequency are useful first checks in detecting anomalies in the dataset.

- Validate and clean the data for performing statistical analysis
- Create data files for each research objective or question for ease of data handling. These small datasets are created parallel to the main dataset.
- Run descriptive summaries to check for anomalies
- Re edit and re validate the data before eventually subjecting the cleaned data to statistical analysis. This process ensures validity of the research results.

Summary of Topic

The activities in database management can be summarised as a process from the start to the end of research, because it entails:

i. Identify the observational units
ii. Designing data collection and recording tools
iii. Budgeting and planning for data management
iv. Choosing the software for creating database
v. Planning for electronic data entry and storage
vi. Data cleaning, and validation
vii. Organizing data for analysis – data files by each objective, hypothesis
viii. Calculating variables derived from input variables
ix. Archiving data once analyzed and reported

Learning Activities

1. Select a research report then design a database for its variables using Access, Excel, SPSS or Genstat

2. Design a database for the information you captured in piloting your tool for your proposed research project using Access, Excel, SPSS or Genstat.

Reference and additional reading materials

GEAR CD supplied by RUFORUM

Manuals: Excel and Access

Topic 10: Data Exploration and Analysis

Introduction

Once database is ready, it is necessary to ensure accuracy in the dataset before formal statistical analysis in order to obtain valid results. This can be achieved by:

a. Detecting any abnormal and missing values that have to be corrected
b. Identifying definite patterns in the data that may have implications on the formal statistical analysis
c. Understanding the extent of variability in the data that may have implications on the formal analysis

Learning Objectives

Upon completion of this topic the learner should be able to:

a. Apply several techniques to detect any abnormal and missing values in a dataset
b. Identify definite patterns in the data that may have implications on the formal statistical analysis
c. Understand the extent of variability in the data that may have implications on the formal analysis
d. Interpret computer outputs of data exploration procedures

Key Concepts

- Data normality
- Histogram
- Box plot
- Normality plot
- Scatter plot distribution
- Residual plot

10.1 Data Exploration Goals

There are two steps involved in data modelling

i. Data exploration and description to understand the data better
ii. Formal statistical analysis to meet the research objectives

Data exploration involves checking of data accuracy before performing formal statistical analysis. This enables the researcher to:

a. Detect any abnormal and missing values that have to be corrected
b. Identify definite patterns in the data that may have implications on the formal statistical analysis
c. Understand the extent of variability in the data that may have implications on the formal analysis

Two methods to explore data before formal analysis are to check for normality in the data and perform descriptive statistics.

Check for normality in the dataset: Inform decision on whether

a) to perform parametric or non parametric statistical procedures
b) to analyse the data as is or to transform

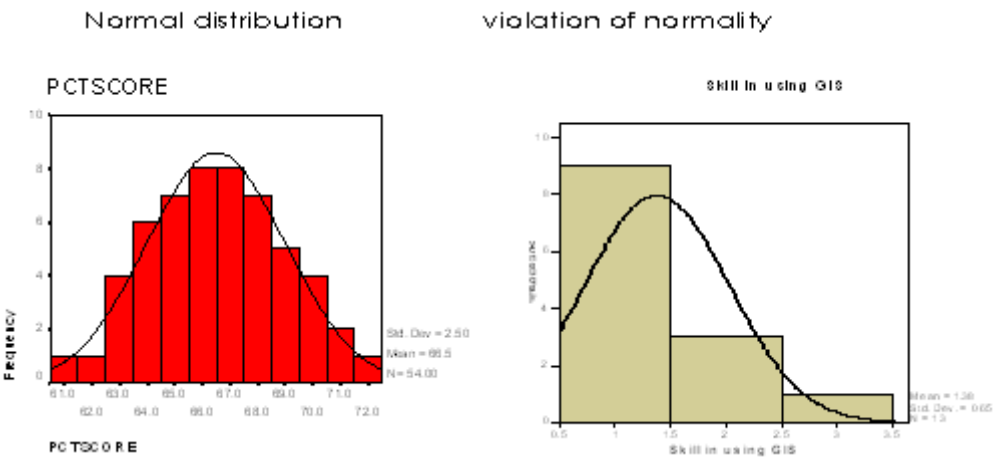
The descriptive statistics performed for data check helps the researcher:

a) to gain insight into the magnitude of the difference between samples and factors
b) to identify pattern of distribution, outliers, valid and invalid observations.

10.2. Checking for Normality in Dataset Distribution

a) Histogram approach

One way to check for normality in the dataset is use of histogram option available in statistical software. The Unimodal, Bell and Symmetrical shapes are good pictorial illustration of normality in the dataset. A histogram deviating from these shapes indicates non normality in the dataset. The direct evidence of that deviation is no coinciding mean, median and mode.

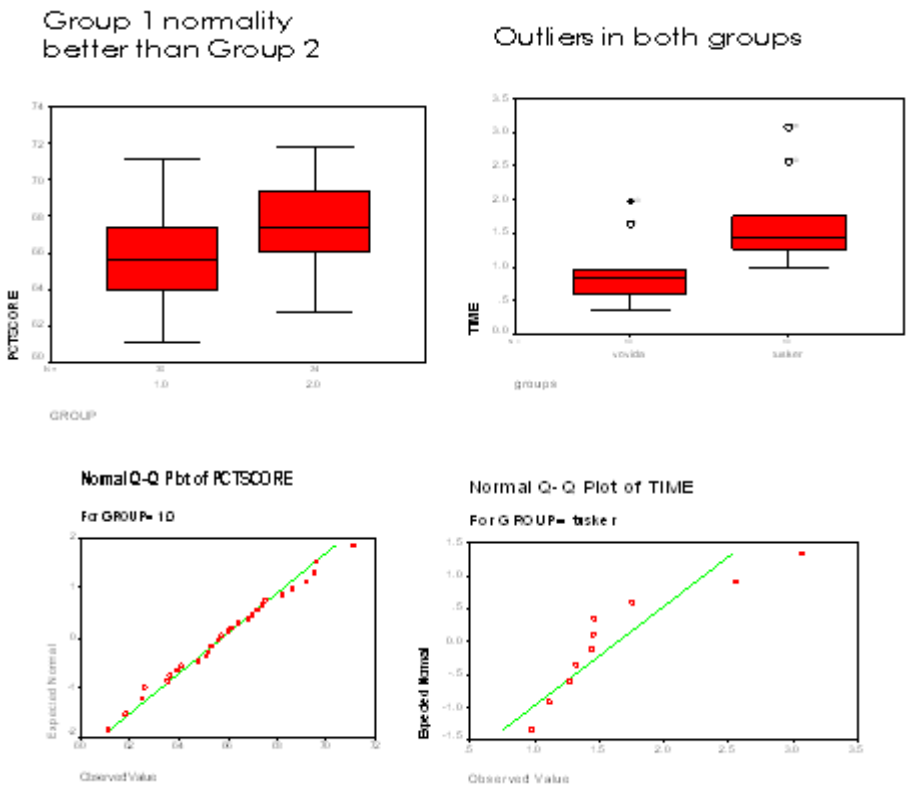


b) Box plot to check for normality

Box plot can be used to evaluate dataset for normality based on the dispersion parameters: minimum and maximum values, mean, median values, upper and lower quartiles and outliers

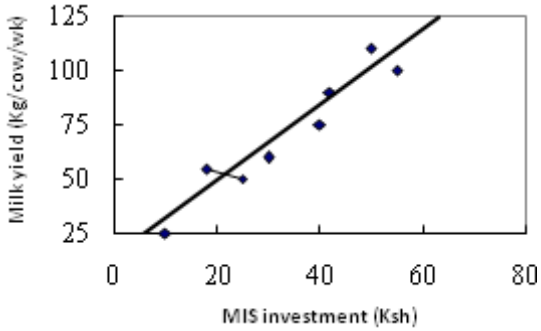
c) Normality plot to check normal distribution in data set

The normality plot in statistical software will yield values that fall along the diagonal without substantial or systematic departures if the dataset conforms to normal distribution.



d) Scatter plot distribution

Scatter plot is a display of distribution of individual values illustrating the association or correlation between two independent variables. The plot can used to present regression line. The figure here illustrates a positive association between milk yield per cow per week with investment made on Management Information Systems for recording feeding and individual animal performance



e) Using residuals to check for normality in the dataset

Residuals are the difference between the actual dependent variable value and its predicted value. Residuals are compared using studentized residuals versus the predicted value. Assumptions that residual plot checks are:

1. **Linearity pattern** of linear regression when fitting general linear models. If many predictors are fitted, the regression model can be asked to reduce standardized residual plots and there is option to produce all partial regression plots for examination

2. **Constancy** of residuals, no clear pattern of decreasing or increasing

3. **Independence and randomness** of residuals no carry over from one observation to another (dependency)

The graphical analysis of the residuals and their interpretations are summarised here to guide decision making about the normality in a dataset.

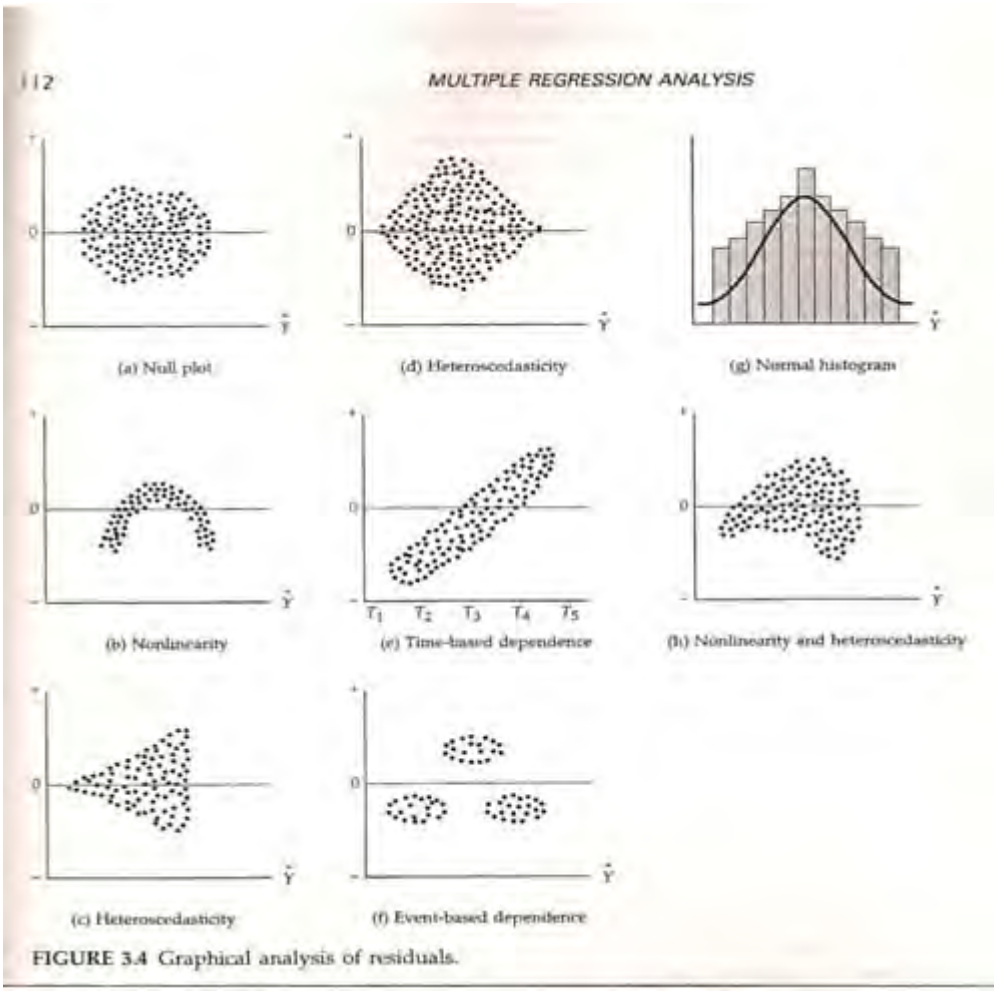


FIGURE 3.4 Graphical analysis of residuals.

10.3. Descriptive statistics

Descriptive statistics performed before formal statistics are useful for:

- gaining insight into the magnitude of the difference between samples, factors
- detecting pattern of distribution, outliers, valid and invalid observations

The descriptive statistics to guide data check against the expected values in the dataset are:

- Frequencies
- Means, Standard deviations, Standard errors
- Median, minimum and maximum observed values
- Number of observations valid
- Coefficient of variations

10.4. Decisions to Make Before Data Analysis

1. Data transformation if data violates normality
- May transform raw data using log, square root, cube root, sin functions
- The results not easy to interpret, so transform back to original scale when reporting

2. Handling of missing values
- They may be mistakes in recording, blanks during data collection, real values
- Do not confuse missing from zero values
- Each statistical software treats missing values uniquely the student must familiarise with
  - SPPS and excel: . |Genstat: * | SAS: -9999.
- May replace them with median, mode or mean, whichever best suits the data pattern
- Alternatively omit them altogether

3. Handling outlier values
- Values or observations not conforming to the general pattern in the data
- To exclude outlier data or not?
- Depend on the reason for their occurrence, exclude if mistake
- Must formulate objective criteria for exclusion and declare the criteria in the methodology

10.5. Data Analysis

The decision on how to proceed with the data analysis is guided by the hypothesis, research questions, how the variable was measured and the objective of the analysis. The options available for analysis research data includes:

a. Dependence relationships Analysis
b. Interdependent relationship Analysis
c. Confirmatory Analysis
d. Exploratory Analysis

Dependence relationship analysis can be through exploration and confirmation and testing. This kind of analysis uses information in the independent variables to explain variation in dependent variables. The analysis fits the model in form of:

Model: Data = Pattern + Residuals

Where Data is response dependent variable, Pattern is a set of independent variables to explain the response, while Residuals is error term not captured in the pattern fitted.

The interdependence relationship analysis is an exploration, confirmation and testing but for characterising patterns of associations between the independents variables to yield new set of independent variables. The model fitted is in the form of:

$$Z_j = \sum_{j=1}^{j} \beta_j X_j + \varepsilon_j$$

Where Z = standardized variable, â loading parameter for the independent variable factors.

For performing dependence relationship analysis, the statistical techniques available for exploration are Canonical correlation and Discriminant analysis. On the other hand, statistical techniques available for Confirmation and Testing when performing dependence relationship analysis include:

- t- tests
- Nonpametric tests
- Chi square test
- Analysis of variance
- Regression
- Discriminant analysis

In the interdependence relationship analysis, statistical techniques for exploration analysis include:

- Principal Component Analysis,
- Cluster analysis
- Factor analysis

Factors analysis can also be used for Confirmation and Testing statistics.

Statistical Procedures for Dependence Relationship analysis

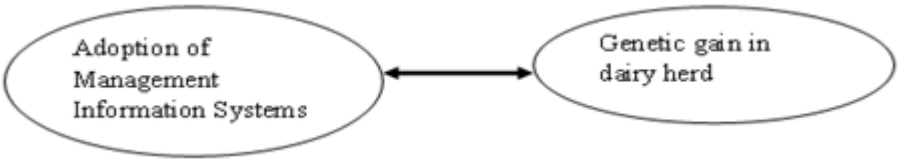
Statistical techniques for dependence relationship analysis apply some underlying assumptions:

- Normal distribution in the dataset
- Sample size is large (>40 sampling units)
- Selection of sampling units is randomly and independently
- Variables are in continuous scale measurements
- Parameter estimates from the sample have the distribution mean (μ) and variance (σ^2) that follow normal distribution. This is conventionally written as N (μ; σ^2) in statistical books.

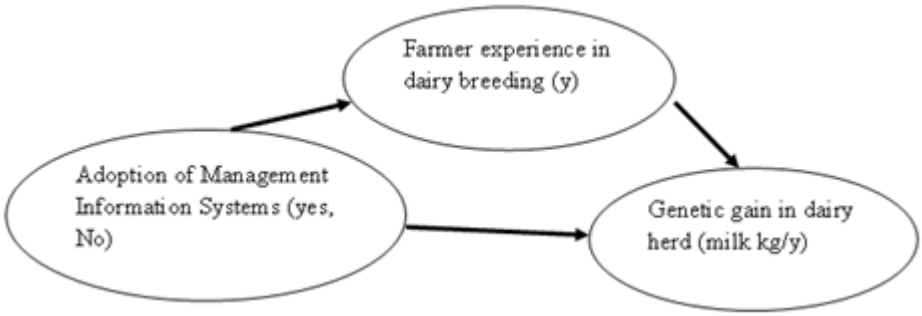
Before performing analysis of the dependence, conceptualise the relationships being investigated. The type of relationships commonly investigated in research includes:

1. Direct influence of adoption of management information systems on the genetic gain the dairy herd.

2. Bivariate relationship between adoptions of management systems with genetic gain in the dairy herd



3. Indirect influence of management systems with genetic gain in the dairy herd through experience of the farmer



t- test statistics procedures

There are several statistics to test hypothesis, varying with how samples are treated. For instance:

i. One sample mean test
ii. Paired dependent samples
iii. Independent samples with equal variance
iv. Independent samples with unequal variance

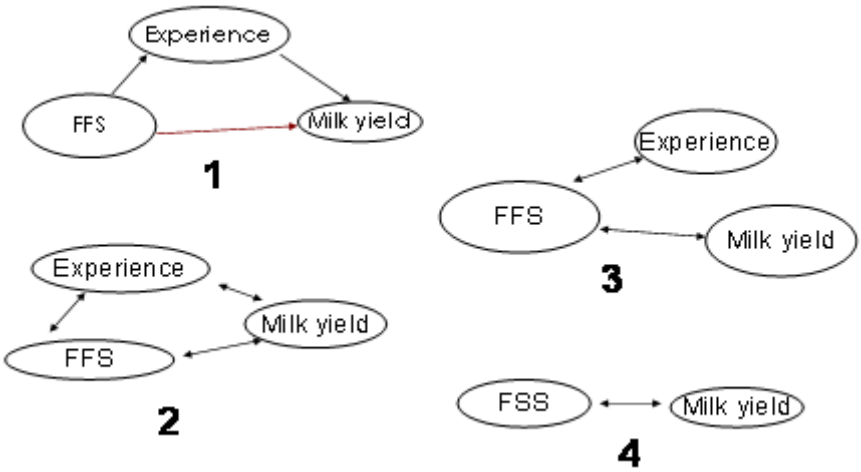
All these test statistics are based on normal distribution and apply a common formula:

$$\mathbf{t-test=Estimate-\frac{Hypothesis}{\frac{SD}{\sqrt{n}}}}\quad \mathbf{t-test=Estimate-\frac{Hypothesis}{\frac{SD}{\sqrt{n}}}}$$

The application of these test statistics is illustrated for a random sample of dairy farmers participating in Farmer Field School (FFS) and in conventional extension (NFFS) approaches compared for their:

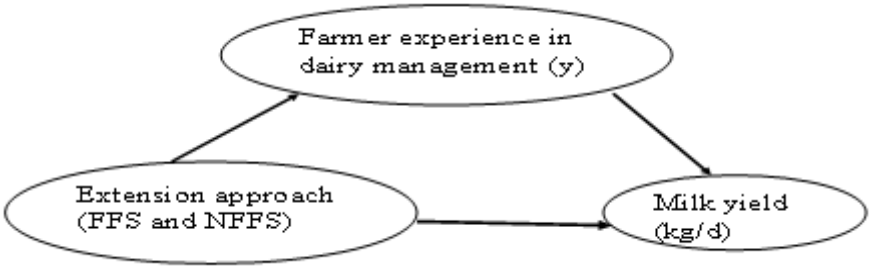
Conceptualising Relationships

Is there any difference in relationship depicted by figure 1, 2. 3 and 4?



- Knowledge and skills in application of information management technologies
- Years of experience in dairy management
- Dairy productivity levels attained in the herd

This problem can be conceptualised as illustrated below for the subsequent analysis performed to different research questions.



The analysis uses SPSS software for the hypothesis testing, assuming one sample, paired dependent samples and two independent samples with equal and unequal variance to illustrate the different statistical techniques corresponding with the assumption about the sample.

a) Tests for One sample mean

The hypothesis being stated is whether milk yield in these dairy herds averages 10 kg per cow per day: **Ho: $\mu_1 = 10.0$** **Ha: $\mu_1 \neq 10.0$**

Subjecting the data to SPSS procedure for one sample mean t test specified in the shaded text will give the outputs reproduced:

SPSS procedure:

Analyse>compare means>One –sample T test, test value (Ho=10 litres/day)

The results in the first table gives the observed mean milk yield averages 7.75 kg, which is lower than (P=0.000) the hypothesised 10 kg per cow per day. This was one sample, so the observed estimate is compared against the stated 10 kg litres of milk.

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Average daily milk production	166	7.7449	4.57571	.35514

One-Sample Test

	Test Value = 10					
	t	df	Sig. (2-tailed)	Mean Difference	95% Confidence Interval of the Difference	
					Lower	Upper
Average daily milk production	-6.350	165	.000	-2.25512	-2.9563	-1.5539

b) Tests for paired- dependent sample mean

Paired sample are dependent because a pair is treated as one sample. The hypothesis is whether milk yield was different between FFS and NFFS farmers compared:

Ho: $\mu_{11} - \mu_{12} = 0$; Ha: $\mu_{11} - \mu_{12} \neq 0$;

Subjecting the data to SPSS procedure for one sample mean t test specified in the shaded text will give the outputs reproduced:

SPSS procedure

Analyse>compare means>paired –sample T test, paired variables (milk yield FFS and milk yield NFFS)

The results in the two tables are for a pair of FFS and NFFS farmers with the FFS obtaining a mean milk yield of 9.00kg compared to 6.5 kg for NFFS. Compared, the FFS obtained higher (P=0.02) observed milk yield than NFFS.

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	milk yield on FFS farms	9.0000	73	5.24603	.61400
	milk yield on NFFS farms	6.5397	73	3.79363	.44401

Paired Samples Test

	Paired Differences					t	df	Sig. (2-tailed)
	Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1 FFS farms NFFS f	2.46027	6.64987	.77831	.90874	4.01180	3.161	72	.002

c). Tests for independent samples means

Independent samples are samples in which inclusion of one sample population is independent of the other sample population. In this case, the hypothesis about the two independent samples milk yield of FFS and NFFS farmers

Ho: $\mu_1 = \mu_2$; Ha: $\mu_1 \neq \mu_2$;

SPSS procedure

Analyse>compare means>independent samples T test (FFS: 1=FFS, 2=NFFS, milk yield)

Outputs

Group Statistics

	FFS groups	N	Mean	Std. Deviation	Std. Error Mean
Average daily milk	Yes	94	8.6622	4.90374	.50578
	No	72	6.5472	3.81970	.45016
Years of dairy experience	Yes	94	11.98	10.232	1.055
	No	71	12.31	8.776	1.042

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	95% Confidence Interval of the Difference Lower	Upper
Average daily milk	Equal variances assumed	.593	.443	3.023	164	.003	2.11601	.69956	.73371	3.49832
	Equal variances not assumed			3.124	163.945	.002	2.11601	.67709	.77805	3.45196
Years of dairy experien	Equal variances assumed	1.529	.218	-.219	163	.827	-.331	1.516	-3.322	2.660
	Equal variances not assumed			-.223	160.323	.824	-.331	1.483	-3.259	2.597

10.1 Exercise

Compare the hypothesis tested and results of the tests in paired and independent samples (sample size, mean, test statistics). How does milk yield compares for FFS and NFFS when variance assumed equal and unequal.

Statistical procedures for dependence relationship analysis when normality assumptions are violated

The normality assumptions can be violated when:

- Normal distribution is violated
- Sample is small (<20 sampling units)
- Selection of the sample is NOT randomly and independently taken
- The variables measured are NOT continuous
- Parameter estimates from the sample do not have the distribution mean (μ) and variance (σ^2) Not following normal distribution

In such cases, the non parametric statistics are applicable.

The hypothesis to test is that FFS and NFFS farmers are equally skilled in heat detection and use of mastitis testing equipment. Skills are self rated by respondent on a scale of 1 to 5, reflecting increasing level of skills.

SPSS procedure

Analyse>Nonparametric tests>2 independent sample test> tests variables (skills), grouping variable (FFS 1=yes, 2=no)>test type>Mann-Whitnney U, Kolomogorov-Smimorov Z

Outputs

NPar Tests

Descriptive Statistics

	N	Mean	Std. Deviation	Minimum	Maximum
skills in detecting heat	166	2.78	.802	1	4
skills in using a mastitis detection equipment	166	2.09	.984	1	4
FFS groups	166	1.43	.497	1	2

Mann-Whitnney U test

Ranks

	FFS groups	N	Mean Rank	Sum of Ranks
skills in detecting heat	Yes	94	94.72	8903.50
	No	72	68.85	4957.50
	Total	166		
skills in using a mastitis detection equipment	Yes	94	81.57	7668.00
	No	72	86.01	6193.00
	Total	166		

Test Statistics[a]

	skills in detecting heat	skills in using a mastitis detection equipment
Mann-Whitney U	2329.500	3203.000
Wilcoxon W	4957.500	7668.000
Z	-3.718	-.618
Asymp. Sig. (2-tailed)	.000	.537

a. Grouping Variable: FFS groups

The Non parametric test statistics, using Mann-Whitney U test for pair wise comparisons, the skills in detecting heat differ (P=0.000) between FFS and NFFS farmers. Skills in using a mastitis detection equipment was not different (P=0.537) between the two types of farmers classified by extension service.

The next table of results has Kolomogorov-Smimorov Z indicating a violation of normality (P=0.013) in skills for heat detection but not (P=1.00) skills for mastitis detection.

Test Statistics[a]

		skills in detecting heat	skills in using a mastitis detection equipment
Most Extreme Differences	Absolute	.249	.048
	Positive	.000	.048
	Negative	-.249	.000
Kolmogorov-Smirnov Z		1.589	.309
Asymp. Sig. (2-tailed)		.013	1.000

a. Grouping Variable: FFS groups

Multivariate analysis

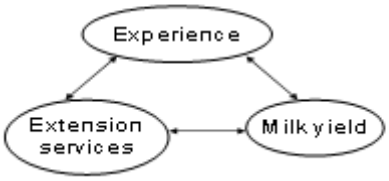
Analysing single predictor variable applies is a univariate analysis. When relating only two variables you are performing a bivariate analysis like performing correlation. The bivariate and univariate extended to more than two variables is multivariate analysis. Multivariate analysis refers to all statistical methods that simultaneously analyse multiple measurements on each individual or object that is being investigated.

Research in AICM is likely to require greater application of multivariate analysis. There are several statistical techniques to achieve this, subject to some assumptions being satisfied.

Bivariate and partial correlation for extension services, years of dairy experience, milk yield

Conceptualising relationships....

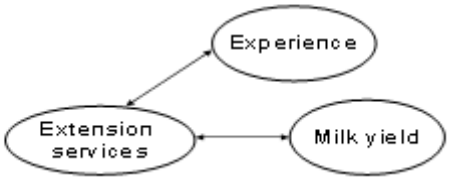
4. Partial correlation



Correlation between extension services and milk yield correcting for experience

SPSS: Analyze> Correlate>Partial

5. Semi-partial correlation



Part correlation is correlation between two variables when other variables are ignored

SPSS: Regression > Linear>part & partial correlation

Correlations

		Frequency of receiving any extension service	Years of dairy experience	Average daily milk
Frequency of receiving any extension service	Pearson Correlation	1	-.160*	-.066
	Sig. (2-tailed)	.	.040	.400
	Sum of Squares and Cross-products	56.512	-147.909	-29.046
	Covariance	.342	-.902	-.176
	N	166	165	166
Years of dairy experience	Pearson Correlation	-.160*	1	.011
	Sig. (2-tailed)	.040	.	.889
	Sum of Squares and Cross-products	-147.909	15131.576	79.106
	Covariance	-.902	92.266	.482
	N	165	165	165
Average daily milk	Pearson Correlation	-.066	.011	1
	Sig. (2-tailed)	.400	.889	.
	Sum of Squares and Cross-products	-29.046	79.106	3454.618
	Covariance	-.176	.482	20.937
	N	166	165	166

*. Correlation is significant at the 0.05 level (2-tailed).

Descriptive Statistics

	Mean	Std. Deviation	N
Frequency of receiving any extension service	1.9455	.58700	165
Average daily milk	7.7797	4.56752	165
Years of dairy experience	12.1212	9.60550	165

Correlations

Control Variables			Frequency of receiving any extension service	Average daily milk	Years of dairy experience
-none-[a]	Frequency of receiving any extension service	Correlation	1.000	-.065	-.160
		Significance (2-tailed)	.	.404	.040
		df	0	163	163
	Average daily milk	Correlation	-.065	1.000	.011
		Significance (2-tailed)	.404	.	.889
		df	163	0	163
	Years of dairy experienc	Correlation	-.160	.011	1.000
		Significance (2-tailed)	.040	.889	.
		df	163	163	0
Years of dairy experienc	Frequency of receiving any extension service	Correlation	1.000	-.064	
		Significance (2-tailed)	.	.412	
		df	0	162	
	Average daily milk	Correlation	-.064	1.000	
		Significance (2-tailed)	.412	.	
		df	162	0	

a. Cells contain zero-order (Pearson) correlations.

Correlations

			Frequency of receiving any extension service	Years of dairy experience	Average daily milk
Spearman's rho	Frequency of receiving any extension service	Correlation Coefficient	1.000	-.215**	.035
		Sig. (2-tailed)	.	.005	.653
		N	166	165	166
	Years of dairy experience	Correlation Coefficient	-.215**	1.000	-.063
		Sig. (2-tailed)	.005	.	.423
		N	165	165	165
	Average daily milk	Correlation Coefficient	.035	-.063	1.000
		Sig. (2-tailed)	.653	.423	.
		N	166	165	166

**. Correlation is significant at the 0.01 level (2-tailed).

Principal component Analysis

SPSS
Analyse>Data reduction> Factor>description (mean, SD), Extraction (PCA)> Rotation (Verimax), Score (loading)

The outputs

Total Variance Explained

Component	Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %
1	1.367	34.177	34.177
2	1.050	26.243	60.421

Extraction Method: Principal Component Analysis.

The results of the principal component analysis, two principal components are extracted, which explain 60.4% of the variance in the variables. The PC1

Rotated Component Matrix [a]

	Component	
	1	2
skills in feeding a lactating cow	.725	
Years of dairy experience	-.721	
Frequency of receiving any extension service	.564	-.404
Average daily milk		.910

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

The PC1 indicates that average daily milk yield was positively associated with increasing skills in feeding a lactating cow and increasing frequency of receiving any extension service but with negatively associated with years of dairy experience. PC2 depicts higher milk yield being negatively associated with frequency of extension service received.

Summary of Topic

Data exploration involves checking for normality in the dataset, which informs the need for transformation before formal statistical analysis can be performed. Statistical analyses apply some assumptions that require checking for when fitting the regression and multivariate models to validate the results.

Learning Activities

One

Select any two research reports and do the following:

a. Identify the types of multivariate methods used to achieve the objectives.
b. Identify the variables, their measurement scales and the model fitted
c. State the assumptions made in the analysis, the reasons for those assumption and whether data was subjected to formal check for normality before fitting the statistical model.

Two

With reasons, identify whether the statistical analysis used was:

a. Dependence relationships Analysis?
b. Interdependent relationship Analysis?
c. Confirmatory Analysis?
d. Exploratory Analysis?

Three

Explain whether the model fitted was evaluated for overall fit and validated and how that was achieved

Four

For the research project you have developed objectives, identify the statistical techniques you will use to test the hypotheses or research questions. Discuss the methodology with your supervisors for guidance.

References and additional reading materials

Joseph F. Hair, Jr; Rolph E Anderson; Ronald l Tatham; William C black. Multivariate Analysis with Readings, 4th Edition.

Chapter 1: Basic concepts in multivariate analysis
Chapter 2: Examining your data
Chapter 3: Multiple regression analysis
Chapter 7: Factor analysis

Useful links:

http://cast.massey.ac.nz/african for the CAST (Computer-Assisted Statistics Textbooks) which consists of a collection of electronic textbooks (e-books).

http://www.statsoftinc.com/textbook/stcanan.html

Topic 11: Reporting Results

Introduction

Scientist in report their findings in conferences, both orally and in written form and in peer reviewed journals. In conferences, scientist sometimes report findings in posters. Scientific reporting follows a format termed IMRAD, which is: Introduction, Materials and Methods, Results and Discussion. Reported scientific information is by practice subjected to peer review to ensure that the results and interpretations published are of good quality and valid.

Effective scientific writing is guided by the objective being clear to the effect that the readers gets the message correctly; being complete in a way that leaves no doubt in readers mind; being correct to the extent that the message is accurate; being efficient to save the reader time.

Learning Objectives

Upon completion of this topic the learner should be able to:

i. Describe content for all the sections of a research report
ii. Report scientific findings in a formal format of a research report
iii. Prepare a research report for a journal following a specified format
iv. Critique a research report

Key Concepts

- Abstract
- Title
- Introduction
- Materials and Methods
- Results
- Discussion
- Acknowledgment
- References

11.1. Sections of Research Report

A scientific research report has a structure reflecting the process of implementation strategy of research. The following sections are presented, each serving a unique communication role as explained in the text

Title is written to tell the reader what the paper is about, what is being communicated. It must be informative about the research, not so much about the results and has to attract people to the paper. Put the most important words first in the title, all the words better be less than 15 to be clear.

Abstract describes the problem, summarises the objectives, the methodology, the key findings and end with conclusions made. Being specific and concise and having the

abstract stand alone is important. Abstract is normally restricted to a page for thesis and to less than 250 words for scientific papers. To be able to capture all important areas of the paper, write the abstract last.

Introduction is the section that the author uses to persuade readers about the importance of the paper. It motivates and justifies the research. It has the problem, hypothesis and research questions, and the rationale, based on literature review.

In theses and not in peer reviewed papers, literature review is presented as standalone section. This section examines the body of knowledge existing on the subject to identify state of knowledge, the knowledge gaps and make hypothesis or research questions to follow in the research plan. This sub section of the literature review is guided by the objectives of the study.

Materials and methods provide a clear and complete description for all data collection methods and the procedure of analysis. Specify the statistical models used and the assumptions made in the analysis.

Results and discussion sections may be separated or combined. If separated, results describes what was found using tables and figures then summarise what the tables and figures presented mean. Each table and figure presented must be referred to in the text in the sequence in which they appear. This section is very important because the data is now synthesized or transformed into information. Figures illustrate trends and for few data. Tables are more detailed on quantitative values, level of significance. The text must be concise, clear and unambiguous, and allow clearer understanding of the key points. For clarity, present results in sub headings corresponding to the specific objectives. Language use is a comparative for the hypothesis tested. Tables and figures need not repeat the values. So decide what is presented in tables and what to present in figures. Tables and figures must be stand alone; use legends where necessary for making distinctions. Labelling must be informative. In the discussion, you must acknowledge authors whose work has been referred to. Read other scientific papers and familiarise yourself with the different approaches to presenting results using tables and figures.

Guides to presenting tables:

- Orientation that is easy to read; showing comparisons within columns rather than within rows
- The groups compared are presented within rows and the measurements forming basis for comparison within columns
- Reference or control group presented first in the row
- Tables and figures must stand alone, self explanatory and numbered and referred to in text in the order they appear
- Have clear titles, legends, units of measurements and clear labels
- Means are reported with SD if descriptive and with standard error (SE) if ANOVA or regression results
- Consistency in the number of decimal places be observed

The table of results here presented has inconsistency in decimal use (in color). Identify some good aspects and undesirable aspects of table presentation in the table.

Presenting results neatly

Season	e-market information user	n	Years						
			2002	2003	2004	2005	2006	2007	2008
			----------Sheep prices ----------						
Normal	users	62	2567	2750	2815	2946	3255	3550.5	4525
	Non users	78	2658	2800	2825	2835	2905	2950.5	3275
Drought	Non users	78	1957	2045.0	2250	2500	2450	2250	2195
	users	62	1970	1985.0	2295	2775	2885	2935	2905
			----------Goat prices----------						
Drought	Non users	58	1757	2045	2150	2600	2410	2050.4	2245
	users	74	1870	1985	2195	2875	2845	2735.5	2955
Normal	users	74	2567	2750	2815	2946	3255	3550.3	4525
	Non users	58	2658	2800	2825	2835	2905	2950.9	3275

Conclusions or implications describe what the results mean for the general field, interpretation of the impact of the findings. This section may suggest further research based on the gaps in the results.

Reference is a list of all cited authors in the text. This reflects what relevant literature to your work is.

Acknowledgment may be general or specific. The generals are about the institutions while the specific are about individuals that contributed to the successful implementation of the work.

Report writing is an iterative process, demanding repeated perfecting of the sentence and structure. For a thesis, first produce tables and figures of the results for each specific objective to guide your decision about what need necessarily be reported. With the tables and figures ready, the author can then add the text for the results and discussion. In practice, write the abstract after all sections are ready.

11.2. Ethics in Research Reporting

Ethics is important in scientific reporting to uphold professionalism. The general principles to check ethics when writing scientific report are outlined here:

- It is advisable being objective with a focus on the data clarity in presentation.
- Report both practical and theoretical implication of the results, whether statistically significance or not.
- Report both positive and negative results, which are interesting for the study objectives.

- Declare all data from the total sample study without selective reporting, contextualize the whole study.
- Describe accurately the methods you used to select sampling units and how randomization process was achieved.
- Report the procedures with the criteria you used in data cleaning and screening
- Report the procedure of data analysis with any assumptions applied and the bearing it has on reliability of statistical inference.
- Respect and acknowledge contributions of others you have cited.
- Decide in advance co-authors of a publication and their order. Generally the first author will be the one who owns the work, originator of the idea. Consultation is necessary between the co authors on the content.
- Authors of a scientific report have the responsibility of acknowledging all who have contributed to the work, and also the source of any material used.

Summary of Topic

The sections of a scientific paper reflect the research process. Title tells what the paper is reporting. Abstract summarises the research process implemented including key finding and conclusion. Introduction states the problem, what is known and what is unknown about the problem. Materials and methods section describes what was done to implement the research. Results are what you found and discussion is the interpretation of the findings. Conclusions are possible implications of the findings. Acknowledgement is recognition of those who contributed to the work and their contribution. References presents details of how to find the papers referred to and cited in the text. Appendix is added when there is supplementary material to show additional information and strengthened clarity and evidence of some undertaking necessary for the research results presented.

Learning Activities

One

Compare the use of *et al* within the text and in the reference section of a research report. How different is it when all the author names are given within the text?

Two

What is the meaning of $P<0.05$ and $P>0.05$ in a research report?

Three

Select at least three research reports and then compare results presented in the tables and figures. Find out if they are overlapping, repeated or otherwise and make your conclusions on how scientists use tables and figures to report results.

References and additional reading materials

Brightta Malmfors, Phil Garnsworthy and Michael Grossman, 2000. Writing and Presenting Scientific papers. Nottingham university Press, UK.

RUFORUM Secretariat, Plot 151 Garden Hill, Makerere University
P. O. Box 7062 Kampala Uganda

Tel. : +256 414 535939; Fax: +256 414 543153
Email: secretariat@ruforum.org Website: www.ruforum.org