

AS 210: RESEARCH METHODOLOGY – 2 CREDITS

Prof. Chenyambuga

Course content

- The research process.
- Preparation of a research proposal.
- Literature review and search.
- Principles in conducting animal experiments.
- Experimental designs, experimental layout, data collection, analyses and interpretation of experimental results for different livestock species and experimental data.
- Methods for participatory livestock research.
- Presentation of research findings.
- Scientific paper writing.

Learning outcomes

- At the end of the course the students shall be able to:
 1. Design and conduct experiments which are related to animal science.
 2. Summarize, analyze and interpret data.
 3. Prepare scientific presentations and publications.

Introduction to the research process

Introduction

- Research can be defined as a scientific and systematic search for pertinent information on a specific topic.
- Some define research as a careful investigation or inquiry for searching new facts in any branch of knowledge.
- Research is the systematic process of collecting and analysing information (data) in order to increase our understanding of the phenomenon with which we are concerned or interested.

- A research comprises defining a problem(s), formulating hypothesis or suggesting solutions, collecting data, organizing and evaluating data, testing the findings to determine whether they fit the formulated hypothesis, making deductions and reaching conclusions.

- A research originates with a question or problem.
- Requires a clear articulation of a goal.
- Follows a specific plan of procedure.
- Usually divides the principal problems into more manageable sub-problems (hypotheses), which guide the research.
- Accepts certain critical assumptions.
- Requires collection and interpretation of data to answer original research question

- The purpose of research is to provide answers to questions or to find out the truth which has not been discovered through scientific procedures.
- Generally the purposes of research fall into the following categories:
 - To gain familiarity with a phenomenon or to achieve new insights into it.
 - To portray accurately the characteristics of a particular population or situation.
 - To determine the frequency with which something occurs.
 - To test a hypothesis of a causal relationship between variables

Motivation to undertake research

- Desire to get a degree along with its consequential benefits
- Concern over practical problems
- Curiosity about new things
- Desire to get intellectual joy of doing some creative work;
- Desire to get respectability

Types of research

- **Descriptive research:**

- The purpose of descriptive research is description of the state of affairs as it exists at the moment.
- The main characteristic of this type of research is that the researcher has no control over the variables; he/she can only report what has happened or what is happening.
- The methods used in descriptive research are survey methods of all kind.
- Although the data description is factual, accurate and systematic, the research cannot describe what caused a situation. Thus, descriptive research cannot be used to create a causal relationship, where one variable affects another.

- **Experimental research:**
 - This involves deliberate manipulation of certain factors under highly controlled conditions.
 - Experimental research adhere to scientific method i.e. replication, randomization and local control.
 - The purpose is to identify causal connections through keeping the levels of some variables constant and manipulating others.
 - This type of research involves introduction of a treatment, use of a control group for comparing subjects who don't receive treatment with those who do and judging differences between units treated differently in the light of observed variation in units treated alike.

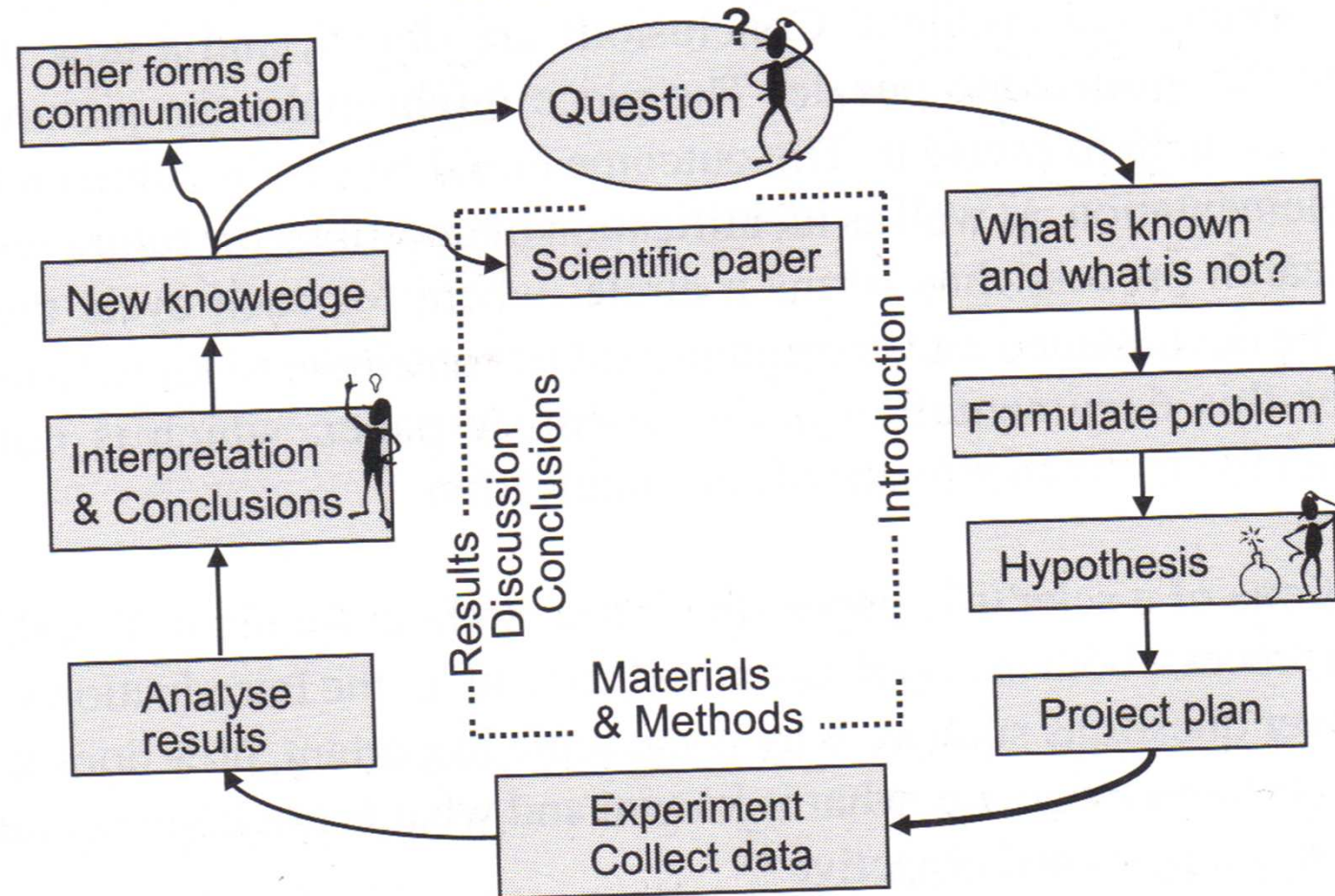
- Experimental research can further be divided into two groups: Applied research and basic research.
 - Applied research aims at finding a solution for an immediate problem facing a society or an industrial/business organization.
 - Basic research is mainly concerned with generalizations and formulation of a theory.
 - The aim of applied research is to discover solutions for some pressing practical problems whereas basic research aims at finding information that has a broad base of application and thus adds to the already existing body of scientific knowledge.

RESEARCH PROCESS

- Research process consists of a series of actions or steps necessary to effectively carry out the research and the desired sequencing of these steps.
- The research process consists of the following main stages. Problem identification, Review of the literature, Hypothesis formulation, Research Design preparation, execution of the study and Data Collection, Data analysis, Generalisation and interpretation and presentation of the results.

- The research process is cyclic in nature. It begins with a problem and ends with generalisation. The generalisation ending one cycle is the beginning of the next cycle (research process). This cyclic process continues indefinitely reflecting the progress of scientific discipline.

The research process



- Awareness of the real/physical/social world in which we exist gives rise to questions as the basis for research studies or investigations
- The question must be answered using available and established scientific research techniques and procedures.
- The guiding principle is that the research should be conducted in accordance with the scientific method.
- Data should be derived from direct, systematic and careful observation and experimentation (as opposed to speculation, intuition, opinions, feelings etc.)

- Experiments are basically about measuring phenomena and collecting accurate and reliable data which are used for analysis and evaluation.
- Accuracy – Correctness of the Measurement
- Reliability – Consistency of the Measurement

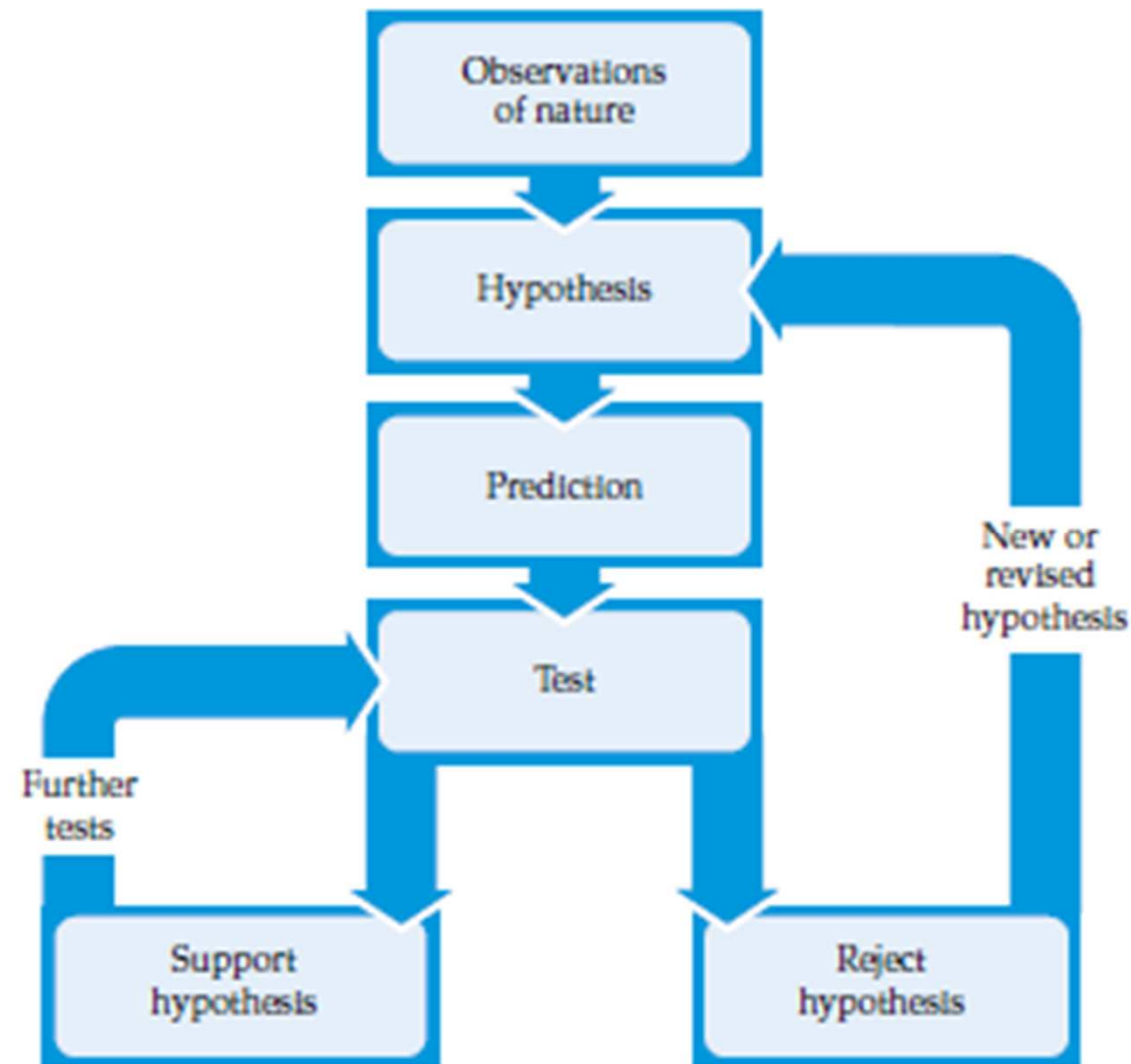
- Statistical analysis is typically used to quantitatively analyze data acquired in research studies
- Conclusions should be based on the results of the analysis conducted, and used to support or refute a hypothesis.
- Conclusions from a research should only be based on the available data and not broadened to include statements which are not supported by the data

Scientific method for doing research

- Science is a method of inquiry, a rational way of discovering truths about the natural world.
- The natural world is extremely complex. To deal with this complexity, scientists attempt to explain the natural world by developing a simplified concept or model of how some part of nature works.
- To construct such a simplified model of nature, scientists typically begin by observing, describing, and asking questions about the natural world.

- The next step in the scientific process is to come up with a possible explanation—known as a hypothesis of the phenomenon being studied. The hypothesis must have logical consequences that can be proved true or false.
- If the results of the test match the prediction, the hypothesis is supported.
- If the test does not uphold the prediction, the hypothesis must be rejected or changed.
- In this way progress is made in science: Hypotheses are constantly being tested; the good ones are kept, the bad ones rejected.

The Scientific method



Preparation of Research proposal

- The research proposal is a document that provides a logical presentation of a research idea, reveals the significance of the idea and illustrates the activities of the proposed research work.
- The research proposal is a way of showing that you have an idea that is of value and can contribute important knowledge to the specific field of your choice.
- Generally, a research proposal should contain all the key elements involved in the research process and include sufficient information for the readers to understand the proposed study.

- Regardless of the research area and the methodology to be used, all research proposals must address the following questions: What you plan to accomplish, why you want to do it and how you are going to do it.
- The proposal should have sufficient information to convince your readers that you have an important research idea, that you have a good grasp of the relevant literature and the major issues, and that your methodology is sound.

Essentials of a research proposal

- Title
- Summary
- Background information (Introduction)
- Justification (problem statement)
- Hypotheses and Objectives
- Literature review
- Research plan (Materials and methods)
- Time schedule
- Expected outputs
- Budget
- Literature references

SUA FORMAT FOR MSc/PhD RESEARCH PROJECT PROPOSALS

1. Title of thesis/dissertation – front page
2. Name of student and Department – front page
3. Name(s) of supervisor(s) – front page
4. Introduction, justification and objectives
5. Literature review
6. Materials and methods
7. Location and duration
8. Source(s) of funds and budget

9. Schedule of activities

10. References

11. Signature of student and date

12. Signature(s) of supervisor(s) and date

Maximum number of pages should not exceed
10.

The proposal should be written using Times New Roman, the font size fixed at 12 points and double spaced.

Title

- The title should be informative, specific and concise, emphasizing the main point(s) of your planned research.
- The title must be short, simple and easy to be understood by a person who is not an expert in the field.
- Avoid waste words in the title e.g. a study of, studies on, investigation on, and assessment of Such words can be removed and the remaining words can still reflect the contents of the research.

Summary

- Summarise the key information of your proposal, i.e. the problem you want to address, the objectives, the significance and potential contribution of your proposed research, the design of the research study, brief description of the methods you will use and the resources needed for the project.
- This is the last section to be written. It is written after finishing all sections of the proposal.

Background information

- Give the background to the proposed research and give an up to date summary of scholarly knowledge in the field of the proposal.
- The purpose of this section is to establish the significance of your research ideas.
- This should summarize the background information of the problem to be researched and the context within which it will be studied.

- The background information is a concise review of the main research works and current issues in the specific subject area.
- The background information should flow into the statement of problem so that the two sections together make a clear and unambiguous statement about the importance of the research.

Justification/problem statement

- A problem statement is the one which is supposed to increase the understanding of the significance of the proposed research and to show the best solution for a given problem. It should present the facts and evidence that support the need for the research project.
- State the problem your proposal seeks to address, taking into consideration some practical concern or some scientific or intellectual interest.

- Present a brief literature review to show what has been done already and identify gaps in knowledge.
- Explain why your planned research is important and how it will contribute to solving the problem or to advancement of knowledge or bring about socio-economic development.
- State the usefulness of the research results by making clear the rationale for doing it and how the results will help to improve the previous situation or contribute to formulation of interventions to address the problem more effectively.

- The problem selected for research should be feasible so that the required research materials and resources are within one's reach.
- The research must fall within the budget the researcher can afford.
- The researcher should examine all available literature to get himself/herself acquainted with the selected problem. He/she may also discuss the problem with colleagues or those having expertise in the subject matter.

Hypotheses

- A research proposal should contain a specific statement of prediction (Hypothesis) that describe in concrete term what a researcher expects will happen in his/her study.
- State the hypotheses to be tested i.e. your assumptions that you intend to test in the research.
- Research hypotheses are the specific testable predictions made about the outcome of a research or an experiment.

- Usually the literature review has given background material that justifies the particular hypotheses that are to be tested. Hypotheses are couched in terms of the particular independent and dependent variables that are going to be used in the study.
- Hypotheses must be stated in two ways, which are **null** (describe possible outcomes) and **alternative** (support research prediction).
-

- For the null hypothesis a researcher makes prediction that in the general population, no relationship or no significant difference exists between groups on a variable.
- The wording is “ there is no significant difference (or relationship) between the groups.
- In the alternative hypothesis the researcher makes a prediction about the outcome, basing this prediction on prior literature and studies on the topic that suggest a potential outcome. For example, the values will be higher for group A than for group B or group A will change more than group B.

- Hypotheses should be very specific and limited to the proposed research because they have to be tested.
- A hypothesis reflects the type of data required and the methods of data analysis to be used.

Objectives

- Objectives are statements of the results that should be measurable and observable at the end of a research process.
- Indicate the objectives of the research, which might be divided into overall (general) and specific objectives.
- The overall objective states the strategic goal(s) to which the solution of the stated research problem(s) shall contribute.
- The Specific objectives tell exactly what the researcher expects to do.

- Specific objective must be SMART - Specific, Measurable, Achievable, Realizable and Time bound in order to facilitate monitoring and evaluation of the project.
- The specific objectives form a guide to the research methodology, data analysis and presentation of results.
- Concrete verbs and not abstract ones should be used. Verbs like determine, compare, list, find, identify are relevant in stating objectives. Verbs like understand, appreciate, study, know, think, and see are irrelevant in stating the objectives since they are abstract.

LITERATURE REVIEW

- The purpose of literature review is to show that the researcher is aware of the traditions and 'state of the art' research in the respective field and demonstrate that his/her study makes a significant contribution to the field.
- It deepens the understanding of the research problem, allows a researcher to learn from and build on what others have already done as well as identify gaps to be filled.
- In a proposal this section should give a brief description of what is already known in terms of studies done in the area, policy statements and their implications.

- It should indicate how existing scientific knowledge contribute to the proposed research and visa versa, how the proposed research moves beyond existing information.
- The review of the most recent literature on the topic is important because good reviews contribute to scientific knowledge by bringing data from different sources together so that new or more definite conclusions can be drawn.

- The researcher should ensure the following:
 - reviewing literature supporting his/her hypothesized solutions;
 - reviewing critical literature opposing his/her hypothesized solutions and indicating his/her position on the contradictory literature;
 - reviewing findings of previous similar research areas; and
 - reviewing theories explaining the hypothesized relationships of variables.

Research plan/Materials and methods

- Relate the experiment or survey to the objectives.
- Describe adequately and justify the choice of the location and population where the study is going to be conducted.
- Provide sufficiently detailed descriptions of your methodologies to convince the reader that you have the necessary expertise to do the research. But avoid too much details.
- Also describe the facilities, equipments and other resources to be used.

- Describe the sampling frame, determine the sample size, describe how the sample size was determined, describe the sampling procedures.
- Describe and justify adequately study methods and laboratory analysis procedures (if applicable give references).
- In case of using questionnaires, give key information that will be collected.
- Provide detailed description on how the collected data shall be managed (coding) and analyzed.
- The statistical package to be used in data analysis should be mentioned.
- The experimental design, statistical models and the type of statistical tests to be performed need to be mentioned.

Time schedule

- Present a time schedule for the activities to be performed and milestones to be achieved i.e. a time-delivery flow chart of achievements.
- Describe the entire period of the study including preparation of the proposal, submission and approval, training (where necessary), pre-testing (of the study tools), data collection, data analysis, report preparation, and dissemination of the findings. A gaunt chart may also be drawn.

Expected outputs

- Specify the expected outputs and possible application of your research results.
- The likely impact of the results should be discussed with regard to a specific sector, to society or to the development of scientific knowledge in the research area.

Budget

- List the items you request funding for in the budget and their cost.
- The budgetary items should include such items as equipments (type and costs), expandable supplies (chemicals, drugs, glassware, stationery) literature (books, journals, internet service), travel (fares and places to be visited), and personnel cost (researchers, technicians, assistants, casual labourer).
- Describe their function and defend their use in the research project.

References

- Reference citations in the text should be typed by writing only the author's surname and the year of publication as follows:
- Leguminous tree leaves and their pods can be used as alternative protein sources to replace fishmeal and soybean in fish diets (El-Sayed, 1999; El-Saidy and Gaber, 2003). OR
- According to El-Sayed (1999) and El-Saidy and Gaber (2003) leguminous tree leaves and their pods can be used as alternative protein sources to replace fishmeal and soybean in fish diets .

- For three or more authors, include only the name of the first author plus *et al.*
- e.g. Leguminous tree leaves and their pods can be used as alternative protein sources to replace fishmeal and soybean in fish diets (Richter *et al.*, 2003; Kaushik *et al.*, 2004). OR
- According to Richter *et al.* (2003) and Kaushik *et al.* (2004) leguminous tree leaves and their pods can be used as alternative protein sources to replace fishmeal and soybean in fish diets .

- Provide a list of all references cited in the text of the proposal.
- Make sure that all references cited are listed in the references and that all references listed are cited in the text of the proposal.
- The references should be arranged alphabetically.

Example of writing reference list

1. Citation of papers published in Journals

- El-Saidy D.M.S.D. and Gaber M.M.A. (2003). Replacement of fish meal with a mixture of different plant protein sources in juvenile Nile tilapia (*Oreochromis niloticus*, L.) diets. *Aquaculture Research* 34 (13):1119-1127.
- El-Sayed A.F.M. (1999). Alternative dietary protein sources for farmed tilapia, *Oreochromis spp.* *Aquaculture* 179 (1-4):149-168.

- Kaushik S.J., Covès D., Dutto G. and Blanc D. (2004). Almost total replacement of fish meal by plant protein sources in the diet of a marine teleost, the European seabass, *Dicentrarchus labrax*. *Aquaculture* 230: 391– 404.
- Richter N., Siddhuraju P. and Becker K. (2003). Evaluation of nutritional quality of moringa (*Moringa oleifera* Lam.) leaves as an alternative protein source for Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* 217 (1-4): 599 – 611.

2. Citation of Books

- Payne, W.J.A. (1990). An introduction to animal husbandry in the tropics. Fourth edition. Longman Scientific & technical, John Wiley & Sons, Inc., New York, US. pp 881.

3. Citation of individual chapters in multi-authored books

- Green, B.N.W., K.L. Veverica and M.S. Fitzpatrick. (1997). Fry and fingerlings production. In: Dynamics of Pond Aquaculture (editors Egna, H. and C. Boyd), pp: 215 – 243. CRC Press, Boca Raton, Florida, USA.

4. Citation papers in Proceedings of conferences

- Mboera, L.E.G. and Kitaly, J.I. (1994). Diseases of small ruminants in central Tanzania. In: Proceedings of the second biennial conference of the African Small Ruminants Research Network, 7-11 December 1992, AICC Arusha, Tanzania, pp 117-120.

5.Citation of thesis

- Ngoly, T.M. (2009). Assessment of heavy metals and agrochemical residues in ruminant feeds and milk. MSc. Dissertation, Sokoine University of Agriculture, Morogoro, Tanzania. pp 74.

6. Citation of papers from electronic sources

- Lin, C.K. 1996. Clarias and Tilapia interaction in polyculture.
http://pdacrsp.oregonstate.edu/pubs/nops/full_rprts/96-94.pdf. Accessed on 26 January 2011.
- Chenyambuga, S. W., Waiswa ,C., Saimo, M., Ngumi, P. and Gwakisa, P. S. (2010). Knowledge and perceptions of traditional livestock keepers on tick-borne diseases and sero-prevalence of Theileria parva around Lake Victoria Basin. Livestock Research for Rural Development. Volume 22, Article # 137 .
<http://www.lrrd.org/lrrd22/7/chen22137.htm>
Retrieved on 16 August 2010.

SURVEY METHODS IN ANIMAL PRODUCTION RESEARCH

An Overview of Research Methods

Descriptive (Qualitative)

- Household Survey
- Group interviews
- Focus Groups
- Observation

Experimental (Quantitative)

- True Experiment

Experimental Research: True Experiment

- Introduction of a treatment
- Use of a control group for comparing subjects who don't receive treatment with those who do-
- Judge differences between units treated differently in the light of observed variation in units treated alike
- Adherence to scientific method - replication, randomization and local control

Survey methods

- The survey is a non-experimental, descriptive research method.
- Surveys can be useful when a researcher wants to collect data on phenomena that cannot be directly observed (such as opinions/attitudes on importance of particular livestock species).
- Surveys are used extensively in animal science research to assess characteristics of livestock farmers (age, gender, education, income etc), and their opinion or preference for technologies

Basic types of surveys: cross-sectional surveys and longitudinal surveys

- Cross-sectional surveys are used to gather information on a population at a single point in time. An example of a cross sectional survey would be a questionnaire that collects data on how farmers feel about the adaptability of a breed.
- Longitudinal surveys gather data over a period of time. The researcher may then analyze changes in the population and attempt to describe and/or explain them.

Advantage of survey research

- + An efficient means of gathering large amounts of data
- + Can be anonymous and inexpensive
- + Can collect a wide range of information e.g. attitudes, values, beliefs, feelings
- + Relative easy to administer

Disadvantages

- Validity of results depend on motivation and memory of participants
- Participants may not answer honestly
- Feedback often incomplete
- Wording of instrument can bias feedback
- Details often left out

Questions to ask before doing survey research

- The advantages and disadvantages of different survey instruments
- Planning your survey
- Administering your survey
- Analyzing your data

Planning your survey

- Do you have a clear objective and hypothesis?
- Do your questions focus on the objectives/hypotheses?
- Which area and whom to involve?
- Will participants answers provide accurate answers to your questions?
- Do you have adequate resources? (human, transport and finance)
- To whom will your results apply?

What type of population are you sampling?

- Very important consideration before one begins to write questions
- Consider number of qualities respondents possess:
 - Level of Education (specifically reading level)
 - Limits of Attention (avoid fatiguing respondents)
 - Motivation (why is respondent going to participate)

Different Survey Instruments

- Written Instruments
- Interviews

Written Instruments

1. Self-administered questionnaires - subject responds to printed questions.

Advantages

- Cheap, easy to distribute to large sample
- Anonymous may promote honest responses

Disadvantages

- Non-response is big problem
- Careful questionnaire design is required
- Good reading and writing skills by respondents are needed
- Very little quality control over administration

2. Investigator-administered questionnaires -
Subject responds to questions directly posed by
researcher (e.g., interview, phone survey)

Advantages

- Higher response rates
- Opportunity to answer respondent
questions and ensure quality of data (e.g.,
probe adequate answers, answer all
questions)
- Rapport and confidence building possible

- Disadvantages
 - Cost and time requirements
 - Adequate training of staff
 - Accessibility of sample
 - May hurt sense of anonymity and thus decrease honesty of responses

What type of response format is appropriate (for each question)?

- A. Open-ended Questions - permits subject freedom to answer question in own words (without pre-specified alternatives)

e.g. Why do you keep chicken?

Advantages

- Obtain unanticipated answers
- May better reflect respondents' thoughts\beliefs
- Appropriate when list of possible answers is excessive

Disadvantages

- Flexibility in responses difficult to code and analyze
- Provide incomplete or unintelligible answers

B. Close-ended Questions - subject selects from list of pre-determined, acceptable responses

- Types of closed-ended questions

1. Checklists - respondent selects certain number of pre-specified categories.

Purpose of keeping chicken:

_____ Meat for home consumption

_____ Eggs for home consumption

_____ Cash from sales of live birds and/or their products

_____ Manure

_____ Investment

2. Two-way (Forced Choice) -
respondent must select between two
alternatives.

Do you give supplementary
feeds to your local chicken?

_____ Yes _____ No

3. Ranked - respondent must place items in order of importance or value.

Rank in order of importance the traits for your local chicken:

- _____ Size
- _____ Shape
- _____ Meat quality
- _____ Egg numbers
- _____ Disease tolerance

4. Multiple-Choice (Likert) – respondent selects between range of alternatives along pre-specified continuum.

e.g. As a strategy to reduce pollution emanating from poultry farms in urban areas, all poultry keepers should build modern houses and at most keep 300 birds.

Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
1	2	3	4	5

Advantages of Closed-ended Questions

- Obtain more reliable answers
- Meaning of responses more meaningful to researcher
- Straight-forward analysis

Disadvantages of Closed-ended Questions

- Answers relative to response scale provided
- Respondent's choice not among listed alternatives
- Choices listed communicate kind of response wanted

Important considerations in designing questionnaire

- Before one sets to construct a questionnaire he/she should make sure that :
 - the objectives of the study are very clear and the information required from the field is clearly known.
 - the subheadings of the questionnaire are established, each for collection of particular kind of information, answering a certain objective.
- Draft the questions, check on their relevance into answering the study objectives, correctness of the style and clearness in being understood.

- Go through the questionnaire to make sure that no information has been left out nor any repetition has been included.
- However for some kind of respondents, it might be necessary to include crosschecking questions to confirm some answers. For example a question asked somewhere in the questionnaire as “What is your monthly income from sales of milk?” may be crosschecked by two questions
 - i) How much milk do you sell per day on average?
 - ii) What is the price of milk in your area?

The researcher should make sure that the crosschecking questions are not asked close to the key ones. The respondent should have somehow forgotten about the key question when the crosschecking one is asked.

- While it is important to ensure that all information has been collected by the questionnaire, questionnaires ought not to be too long.
- Too long questionnaires make both the enumerator and respondents bored and tired during the interviews.
- The danger of that is that the respondent may respond to some question with less seriousness, hence full of imprecision.
- Handling of too large questionnaires may also be tiresome.

Pre-testing survey

- Questionnaires should be given a preliminary trial. This may reveal unexpected faults and difficulties.
- Pre-testing will serve to determine whether the questionnaire is in the form most suitable for eliciting the required information from the respondents.
- Pre-testing will provide information on whether the questions and associated instructions are understood by, and within the capability of, the investigator.

Household surveys

- A household survey involves collecting data from a random sample of households chosen from among all households meeting a specified criteria.
- The larger the sample as a fraction of the whole, the more accurate the survey will be as an estimator of the target group.
- Information is obtained via interviews held face to face with household members.
- The interviews are based on a questionnaire, which is structured depending on the objectives and circumstances of the particular survey.

Advantages of household surveys

- Opportunities to link the sampling frame directly to objectives of the survey.
- The data collected are easier to quantify, standardize, pool and analyze.
- If the sampling frame is designed properly, household surveys allow good control over bias
- Opportunities to determine the precision of the survey by choosing the appropriate sample size.
- Opportunities to conduct interviews with both male and female livestock keepers.
- Opportunities to probe more personal issues such as income.

Disadvantages of household surveys

- Need longer time
- Have high costs

Group interviews

- Group interviews may include community meetings, to which the whole community is invited or small focus groups of 8 to 10 people selected on the basis of some similarity in their background.
- Community meetings offer opportunities for collecting general information on the area and the community; identify social norms and accepted views, identifying special interest groups, cross-checking information from other sources, identifying collective views and feelings.

- Focus groups enable the views of particular social groups to be explored.
- Compared to community meetings, focus group may provide an easier environment for women, members of minority groups or people of low socio-economic status to make their views known.
- It is important that in all group interviews, attention be paid to social and group dynamics among the participants. It should not be assumed that everyone in the group will be equally willing to express his/her opinions in a group context.
- Attention should be paid not only to what is being said, but to who is saying it.

Pros and cons of group interviews

- Advantages
 - Opportunities to involve larger numbers of people in the survey in a relatively short period of time.
 - Opportunities to foster community ownership of the surveying process
 - Helps to identify and fill gaps in current knowledge such as perceptions, attitudes, feelings, etc.

- The weaknesses include:
 - The possibility that participants will be reluctant to discuss some topics in a group context.
 - The risk that social dynamics within a group may be overlooked.
 - Does not give statistics
 - Analysis subjective

Key informants

- Key informants are individuals who are targeted by a survey because of their particular knowledge about some aspect of the location and production system under consideration or they have broad knowledge that can be drawn upon as an alternative or complement to conducting individual interviews of livestock keepers.
- Key informants may be community leaders, religion leaders, extension officers, representatives of NGOs, researchers who have worked in the relevant production system, livestock keepers with particular knowledge or older people with experiences and information on management practices and changes that have taken place during their lifetimes.

Pros and cons of using key informants

- Advantages
 - The potential to obtain a lot of information from limited number of interviewees, hence saving on time and costs
 - The potential to obtain specific and detailed information within the key informants' areas of expertise.
- Disadvantages
 - The possibility that the key informants are insufficiently knowledgeable about the situation on the ground.
 - The risk that the knowledge and opinions of the livestock keepers, particularly marginalized groups may be overlooked.

Rapid appraisal

- Rapid appraisal is a method used to gain qualitative information about an area in a fast and succinct way. It uses existing information about the area and key local people to obtain the views of their community.
- Rapid appraisal is a method of data collection that involve interaction with livestock keepers and/or other knowledgeable stakeholders, but are not based on formal sample-based survey.
- Rapid appraisals often involve obtaining information both from individuals and from groups, particularly knowledgeable individuals referred to as key informants.

- Rapid appraisal interviews are semi-structured, that is they are sufficiently flexible to allow the interests and concerns of the interviewees to be raised and explored while being sufficiently structured to ensure that the topics of interest are covered.
- Rapid appraisals involve the use of techniques that are intended to allow local people to teach outsiders about their livelihoods, their problems and their knowledge.

The advantages of rapid appraisals

- Speed relative to household surveys
- Low cost relative to household surveys
- Opportunities for greater involvement of the local communities
- Opportunities for discovering new and surprising information

Disadvantages of rapid appraisals

- Greater difficulty in obtaining objective quantitative information than in household surveys.
- Greater difficulty in standardizing and pooling data.

Sampling design

Sample vs. Population

Types of Samples

Sample Size - Considerations

Sample vs. Population

Population = all individuals of a particular group

- citizens of the United Republic of Tanzania
- Villagers in a district
- Households in a village

A survey of an entire population is a census

Why Survey the Entire Population?

- You have a relatively small population.
- You have reason to expect a relatively low response rate.
- You want to have enough respondents and be able to do sub-group analyses.

Sampling

- Sample - a representative portion of the population.
 - The size of the sample is referred to as “n”.
 - The entire population is referred to as “N”.
- Sample design – the technique or procedure that is used to select items for the sample.
- Sampling is a procedure that allows us to estimate/infer the characteristics of a population by surveying only a group/portion of the population.

Sample vs. Census

- Sampling saves resources (labour, money)
- Less time consuming: Shorter time in the field decreases the chance that external factors will influence results.
- More accurate: Doing fewer surveys increases control over the entire process (interviewing, data entry, etc.) and therefore decreases error.
- Samples can lead to statistical inference about the entire population.

Sample Warning...

- Findings from a survey sample are only representative of the group from which it was taken.
- In other words, You can't use your results to say anything about the individuals who didn't have a chance to be included.

Steps in sampling

- Type of population – define clearly the set of objects (the population) to be studied. The population can be finite or infinite.
- Sampling unit - decide on the type of a sampling unit. A sampling unit can be a geographical area e.g. district, village, etc. or it may be a social unit e.g. family, association, etc.
- Sampling frame - A complete list of all units from which a sample is drawn. Examples: Lists of villages in a district, lists of households in a village, etc. The list should be comprehensive, appropriate, reliable and representative of the population.

- Size of sample – decide on the number of items to be selected from the population to constitute a sample. When deciding the size of sample, you must determine the desired precision as well as acceptable confidence level for the estimate.
- Parameter of interest – determine the population parameters which are of interest e.g. proportions and mean of the population.
- Sampling procedure – decide on the technique that will be used in selecting the items for the sample.

Types of sampling

- Purposive sampling
- Simple random sampling
- Systematic sampling
- Stratified sampling
- Cluster sampling

Purposive sampling

- In this type of sampling, items for the sample are selected deliberately by the researcher. He/she purposively chooses particular units of the population for constituting a sample on the basis that the sample will be representative of the population.
- In this design, there is always a danger of bias as there is a great chance of personal element entering into the selection of the units of the sample.
- Sampling error cannot be estimated as there is no assurance that every element has some specifiable chance of being included.

Simple random sampling

Definition: Every element has the same probability of selection *and* every combination of elements has the same probability of selection.

Probability of selection: n/N , where n =sample size;
 N = population size

List the elements in the sampling frame and use a random number table to select elements.

This is possible only in a finite population where lists are available and items can readily be numbered.

In this case you can select some items for the sample haphazardly without a purpose.

Systematic sampling

- **Definition:** Every element has the same probability of selection, but not every combination can be selected.
- Used when drawing simple random sample is difficult and when the list of elements is long & not computerized

Procedure:

- List the elements in the sampling frame and take every N^{th} element, based on how many needed in sample. Use random number to pick up the unit with which to start with.
- Warning: Be attentive to any inherent patterns in how sampling frame is ordered.

Stratified sampling

- Population is divided into subgroups (strata) that are individually more homogeneous than the total population.
- Draw appropriate proportion of elements from each stratum (homogeneous sub-group) to constitute a sample.
- The strata should be formed based on the basis of common characteristics of the items to be put in each stratum e.g. livestock farmer by species kept or gender.
- Selection of items for the sample - Simple random sampling with items having the same probability of selection from each stratum.
- Main use is in comparison of subgroups
- Probability of selection needs to be higher for smaller stratum to be able to compare subgroups.

Cluster sampling

- **Population divided into non-overlapping clusters .**
- **The researcher randomly selects n clusters to include in the sample.**
- **Each element of the population can be assigned to one, and only one, cluster.**
- **One-stage sampling: All of the elements within selected clusters are included in the sample.**
- **Two-stage sampling. A subset of elements within selected clusters are randomly selected for inclusion in the sample.**

Reasons for cluster sampling

- Reduction in cost - **Sometimes, the cost per sample point is less for cluster sampling than for other sampling methods.**
- No satisfactory sampling frame available - **Constructing a complete list of population elements is difficult, costly, or impossible**

Cluster Sampling: Disadvantage

- **Assuming the sample size is constant across sampling methods, cluster sampling generally provides less precision than either simple random sampling or stratified sampling -This is the main disadvantage of cluster sampling.**
- **Cluster sampling should be used only when it is economically justified - when reduced costs can be used to overcome losses in precision.**

Determining the Size of a Sample

- Sample size is NOT related to representativeness.
- Sample size, however, is related to *accuracy*. *How close the sample statistic is to the actual population parameter (e.g. sample mean vs. population mean) is a function of sample size.*
- *Sample accuracy*: refers to how close a random sample's statistic (e.g. mean, variance, proportion) is to the population's value it represents (mean, variance, proportion)

Sample Size - Considerations

- Amount of sampling error that is tolerable - how accurate do our measurements need to be?
- Resources (mostly money)
- Number of categories by which data are to be analyzed (e.g. ethnic group, age, village, sex, etc.) - the more categories, the larger the sample needed

Sample Size - Considerations

- It takes fewer people to constitute a good sample of a fairly homogeneous population than it does to get a good sample of a heterogeneous population
- Samples should not be thought of as necessary proportions of populations. A sample of 382 people will result in the same degree of accuracy for a population of 50,000 as for a population of 100,000

Calculating sample size

- The formula requires that we
 - specify the amount of confidence we wish to have,
 - estimate the variance in the population, and
 - specify the level of desired accuracy we want.
- When we specify the above, the formula tells us what sample size we need to use....n

The sample size formula for estimating a proportion (also called a percentage or share):

$$n = \frac{z^2(pq)}{e^2}$$

where

n = the sample size

z = standard error associated with the chosen level of confidence (1.96)

p = estimated percent in the population

$q = 100 - p$

e = acceptable sample error

How to estimate variability (p and q shares) in the population

- **Expect the worst case ($p=50\%$; $q=50\%$)**
- **Estimate variability: results of previous studies or conduct a pilot study**

How to determine the amount of desired sample error

- Convention is $\pm 5\%$
- The more important the decision, the less should be the acceptable level of the sample error
- $e = z\sqrt{p(100-p)}/\sqrt{n}$
- Where:
- e = sampling error (the proportion of error we are prepared to accept)
- n = the sample size
- z = the number relating to the degree of confidence you wish to have in the result
- p = an estimate of the proportion of people falling into the group in which you are interested in the population

How to decide on the level of confidence desired

- The higher the desired confidence level, the larger the sample size needed
- Convention is 95% confidence level ($z = 1.96$ which is ± 1.96 s.d.'s)
- The more important the decision, the more likely we will want more confidence. For example, a 99% confidence level has a $z = 2.58$.

The sample size formula for estimating a proportion a mean

Sample size formula for a mean
$$n = \frac{s^2 z^2}{e^2}$$

Where:

n = the sample size

z = a number relating to the degree of confidence you wish to have in the result. 95% confidence is most frequently used and accepted. The value of 'z' should be 2.58 for 99% confidence, 1.96 for 95% confidence,

s = the standard deviation

e = the error you are prepared to accept, measured as a proportion of the standard deviation (accuracy)

Estimating “s” in the Formula to Determine the Sample Size Required to Estimate a Mean

How to estimate s?

- Use standard deviation of the sample from a previous study on the target population.
- Conduct a pilot study of a few members of the target population and calculate s.

Other Methods of Sample Size Determination

- ***Arbitrary “percentage rule of thumb” sample size:***
 - **Arbitrary sample size approaches rely on erroneous rules of thumb (e.g. “n must be at least 5% of the population”).**
 - **Arbitrary sample sizes are simple and easy to apply, but they are neither efficient nor economical.**

Data management

- Data management refers to any activity concerned with
 - Looking after field data sheet
 - Entering data into computer files
 - Checking data entry errors and preparing for analysis
 - Maintaining records of the processing steps
 - Archiving the data for future use

Coding

- List all variables
- Code for the responses, e.g.

Variable code	Description	Coding instruction
Sex	Sex	1 = Male, 2 = Female
AFC	Age at first calving	Number

Data structure

- Data are measurements or observations of a variable comprising of numbers, words or images.
- A data structure is a way of storing data in a computer so that it can be used efficiently.
- The data set ready for statistical analysis should be a flat rectangular file, i.e.
 - A column for each question (variable)
 - A row for each respondent (observation)

Missing values

- Missing values may occur because of:
 - Skip procedure
 - Failure to respond
- How to handle
 - Provide codes that are out of the range of legitimate value for each variable e.g. 888, 999 etc.

Example of a data spreadsheet

ID	Name of farmer	District	Village	Age	Sex	Cattle number	Grazing system

SIMPLE LINEAR REGRESSION

SIMPLE LINEAR REGRESSION

- In regression analysis the interest is to determine how changes of some variables influence the change of values of other variables, e.g. how increasing the protein level in a feed affects daily weight gain.
- The relationship between variables can be described with a function, e.g. a function of protein level to describe daily weight gain. A function that explains such relationships is called a regression function and analysis of such problems and estimation of the regression function is called regression analysis.
- Regression analysis includes a set of procedures that is used to study statistical relationships among variables, whereby one variable is defined as dependent upon others defined as independent variables. In the example above, daily weight gain is dependent variable and the protein level is independent variable.

- By using regression the cause-consequence relationship between the independent and dependent variables can be determined.
- The dependent variable is denoted by y and the independent variable by x .
- When the change of the dependent variable is described with one independent variable and the relationship between them is linear, the procedure is called simple linear regression. Multiple regression procedure is used when the change of a dependent variable is explained by changes of two or more independent variables.

- The main applications of regression analysis are:-
 - i) to describe the linear relationship between Y and X ,
 - ii) to determine how much of the variation (uncertainty) in Y *can be explained by the linear* relationship with X and *how much of this* variation remains unexplained, and
 - iii) to predict new values of Y *from new values* of X .
- A regression that explains linear change of a dependent variable based on changes of one independent variable is called a simple linear regression.
- For example the weight of cattle can be predicted by using measurements of heart girth. Heart girth is the independent variable and weight is the dependent variable.
- To estimate the function a sample of animals are chosen and heart girth and weight of each animals are measured. In other words, pairs of measurements of the dependent variable y and independent x are needed.

- If it is assumed that the relationship between the x and y variables is linear the statistical model is:
- $Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, where
- Y_i = dependent variable
- x = independent variable
- β_0, β_1 = regression parameters
- ε_i = random error.
- The parameter β_1 is called the regression coefficient and explains the slope.
- The random error describes deviations from the model due to factors unaccounted for in the equation.

- Model assumptions

- (i) $E(\varepsilon_i) = 0$, mean of errors is equal to zero.
- (ii) $Var(\varepsilon_i) = \sigma^2$, variance is constant for every ε_i , that is variance is homogeneous.
- (iii) $Cov(\varepsilon_i, \varepsilon_{i'}) = 0, i \neq i'$, errors are independent, the covariance between them is zero.
- (iv) It is assumed that ε_i are normally distributed,
 $\varepsilon_i \sim N(0, \sigma^2)$

The estimated expectation (mean) of the dependent variable y for a given value of x , denoted by $E(y/x)$ is a straight line.

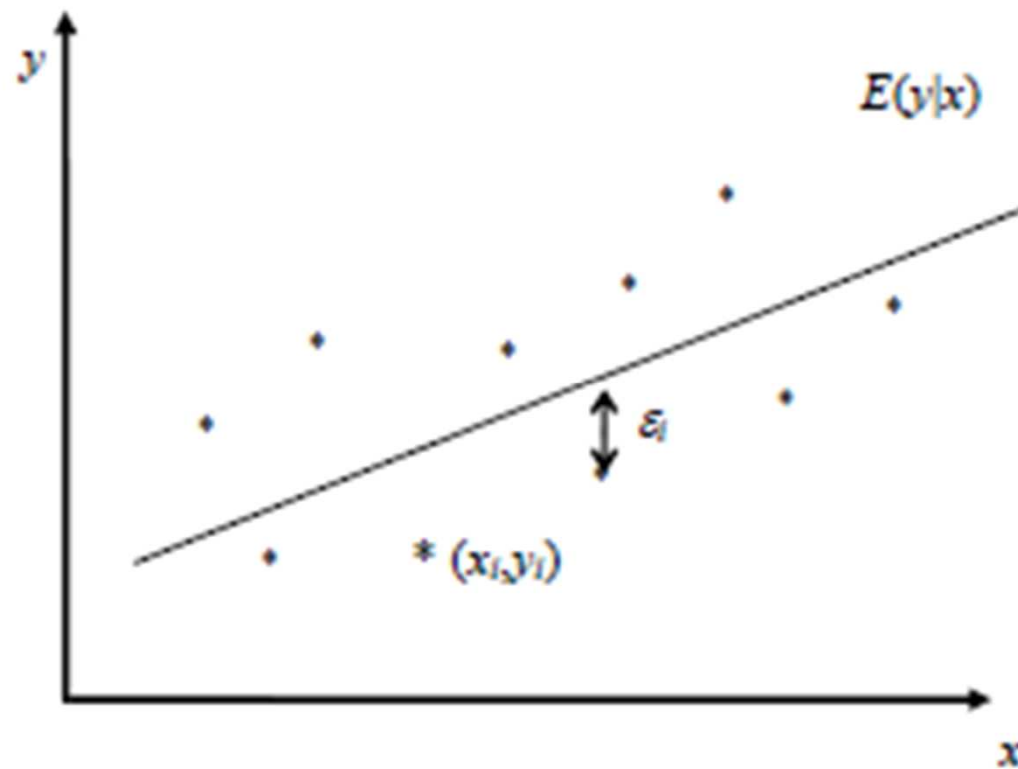


Fig. 1: Linear regression. Dots represent real observations (x_i, y_i) . The line $E(y/x)$ shows the expected value of dependent variable. The errors, ε_i , are the deviation of the observations from their expected values

- A simple linear regression can be positive or negative. A positive regression, $\beta_1 > 0$, is represented by an upward sloping line and y increases as x is increased (figure 2a). A negative regression, $\beta_1 < 0$, is represented by a downward sloping line and y decreases as x is increased (figure 2b). A regression with slope $\beta_1 = 0$ indicates no linear relationship between the variables.

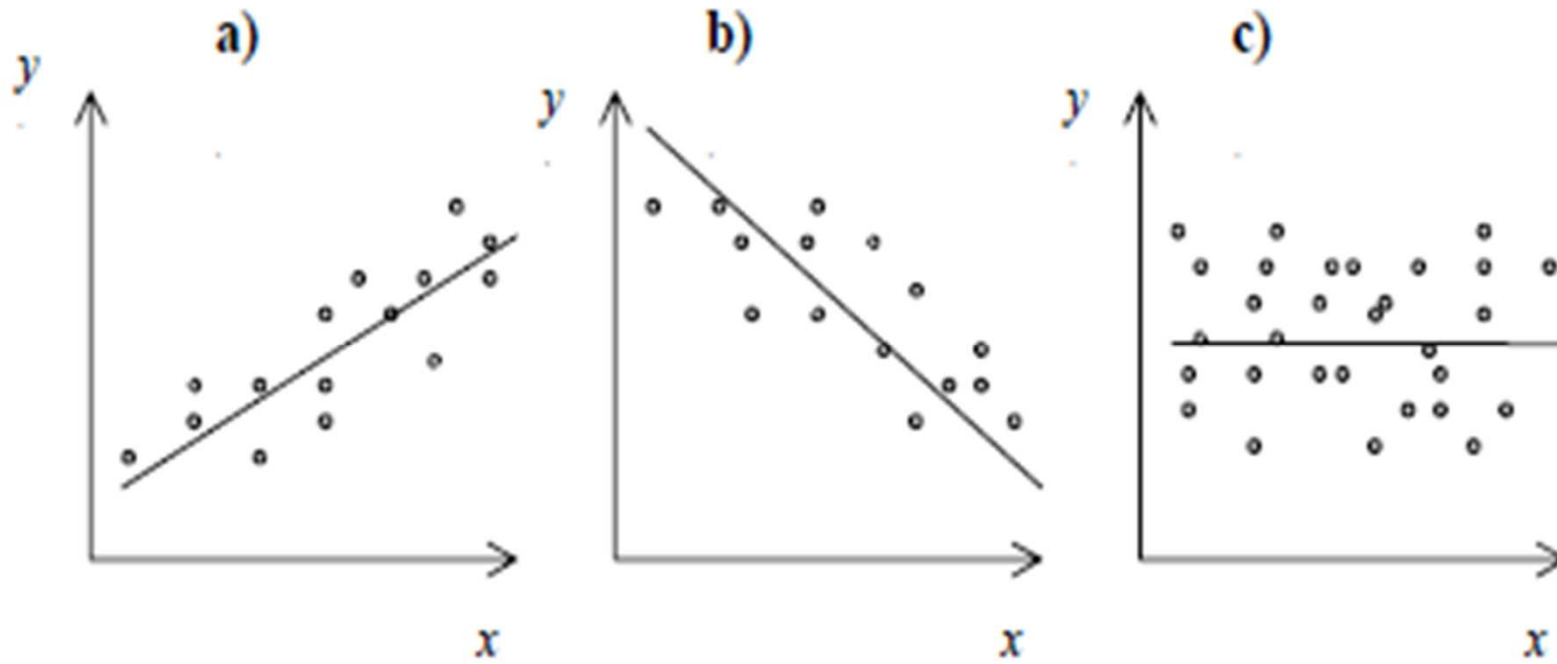


Figure 2: (a) Positive regression, (b) negative regression (c) no linear relationship

Estimation of regression parameters

- Regression parameters of a population are denoted by β_0 and β_1 and they are unknown and estimated from data collected from a sample of a population.
- The parameter estimators of the sample are denoted by b_0 and b_1 . The aim is to determine a line that will best describe the linear relationship between the dependent and independent variables given data from the sample.
- The regression line $E(y/x)$ is unknown, but can be estimated by using a sample with $\hat{Y} = b_0 + b_1 x_i$
- This line is called estimated or fitted line, or more generally the estimated regression line or estimated model.

The estimators, b_1 and b_0 are computed as follows

$$b_1 = \frac{SS_{xy}}{SS_{xx}}$$
$$b_0 = \bar{y} - b_1 \bar{x}$$

where:

$$SS_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y}) = \sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n} = \text{sum of products of } y \text{ and } x.$$

$$SS_{xx} = \sum_i (x_i - \bar{x})^2 = \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n} = \text{sum of squares of } x.$$

n = sample size

$\bar{y}$ and $\bar{x}$ = arithmetic means

$\hat{y}_i = b_0 + b_1 x_i$ describes the estimated line.

The difference between the measurement y_i and the estimated $\hat{y}_i$ is called the residual and is denoted with e_i .

$$e_i = y_i - \hat{y}_i$$

Each observation in the sample can be written as

$$Y_i = b_0 + b_1 x_i + e_i$$

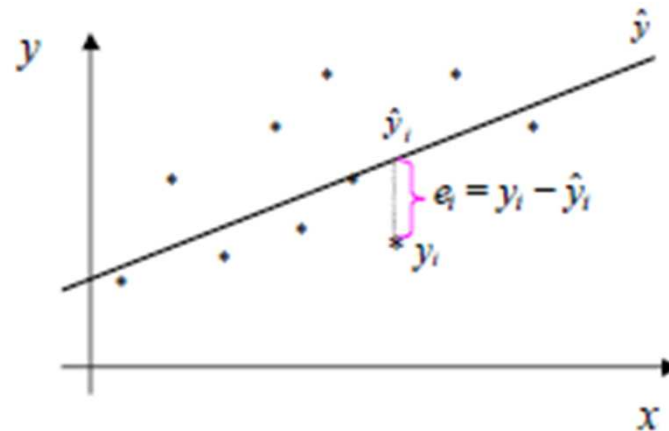


Figure 3: Estimated or fitted line of the simple linear regression

Example: Estimate the simple regression line of weight on heart girth of cows using the following sample:

Cow	1	2	3	4	5	6
Weight (y):	641	633	651	666	688	680
Heart girth (x):	214	215	216	217	219	221

	Weight (y)	Heart girth (x)	x^2	xy
	641	214	45796	137174
	633	215	46225	136095
	651	216	46656	140616
	666	217	47089	144522
	688	219	47961	150672
	680	221	48841	150280
Sums	3959	1302	282568	859359

$$n = 6$$

$$\sum_i x_i = 1302$$

$$\sum_i x_i^2 = 282568$$

$$\sum_i y_i = 3959$$

$$\sum_i x_i y_i = 859359$$

$$SS_{xy} = \sum_i x_i y_i - \frac{(\sum_i x_i)(\sum_i y_i)}{n} = 859359 - \frac{(1302)(3959)}{6} = 256$$

$$SS_{xx} = \sum_i x_i^2 - \frac{(\sum_i x_i)^2}{n} = 282568 - \frac{(1302)^2}{6} = 34$$

$$b_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{254}{34} = 7.53$$

$$b_0 = \bar{y} - b_1 \bar{x} = -974.05$$

The estimated line is $\hat{y}_i = -974.05 + 7.53x_i$

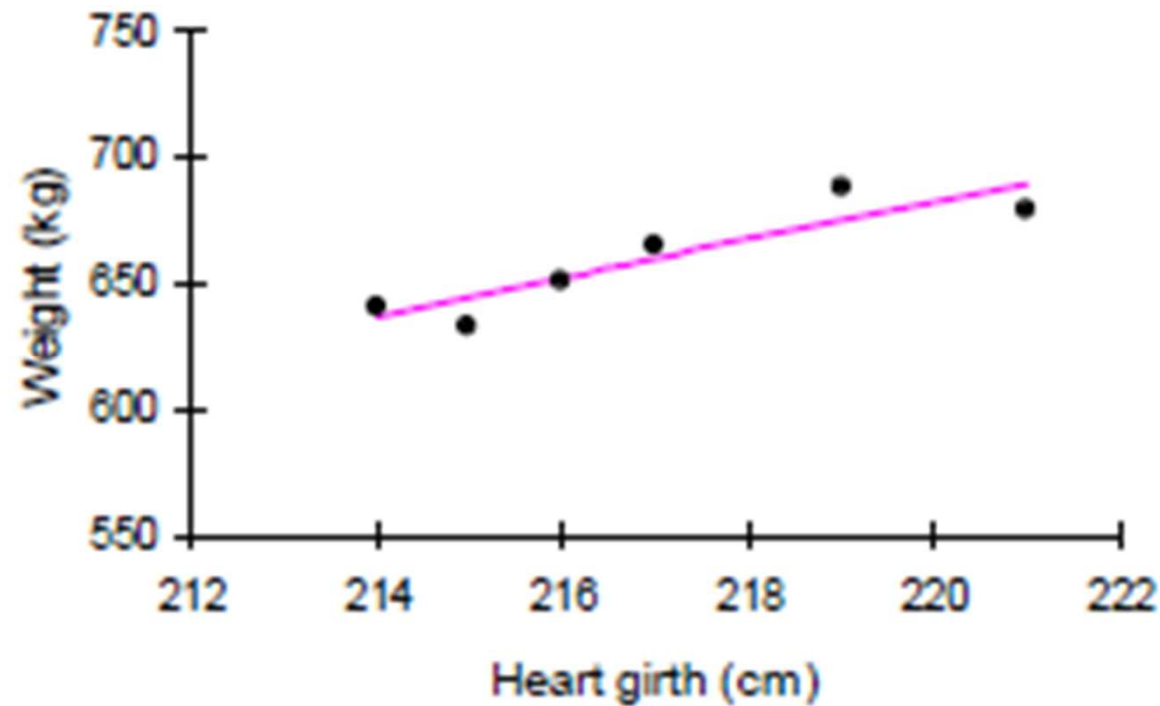


Figure 4: Regression of weight on heart girth of cows

Estimation of residual sum of squares

- The residual is defined as $e_i = y_i - \hat{y}_i$
- The residual sum of squares is
- $SS_{RES} = \sum_i (y_i - \hat{y}_i)^2$
- The residual mean square (MS_{RES}) is denoted by s^2 which equals the residual sum of squares divided by its associated degrees of freedom
- $MS_{RES} = S^2 = SS_{RES}/n - 2$, where $(n - 2)$ is the degrees of freedom. The $MS_{RES} = S^2$ is an estimator of the error variance in a population, σ^2 .
- The square root of the MS_{RES} is called a standard deviation of the regression model,
- $s = \sqrt{s^2} = \sqrt{SS_{RES}/n - 2}$.

Example of calculating the residuals,
squares of residuals and sum of squares

	y	x	$\hat{y}$	e	e^2
	641	214	637.25	3.75	14.099
	633	215	644.77	-11.77	138.639
	651	216	652.30	-1.30	1.700
	666	217	659.83	6.17	38.028
	688	219	674.89	13.11	171.816
	680	221	689.95	-9.95	99.022
Sum	3959	1302	3959.0	0.0	463.304

The residual sum of squares is:

$$SS_{RES} = \sum_i (y_i - \hat{y}_i)^2 = 463.304$$

The estimate of the variance is:

$$s^2 = MS_{RES} = \frac{SS_{RES}}{n-2} = \frac{463.304}{4} = 115.826$$

The variance of b_1 can be estimated by $s^2_{b_1} = s^2/SS_{xx}$

A standard error of b_1 is

$$s_{b_1} = \sqrt{\frac{s^2}{SS_{xx}}}$$

Testing hypotheses about the parameters

- If changing variable x effects change in variable y , then the regression line has a slope, that is the parameter β_1 is different from zero.
- To test if there is a regression in a population, the following hypotheses about the parameter β_1 are stated:
- $H_0: \beta_1 = 0$
- $H_1: \beta_1 \neq 0$
- The null hypothesis H_0 states that slope of the regression line is not different from zero and there is no linear association between the variables. The alternative hypothesis H_1 states that the regression line is not horizontal and there is linear association between the variables.

- Assuming that the dependent variable y is normally distributed, the hypotheses about the parameter β_1 can be tested using a t distribution.
- The test statistic is $t = \frac{b_1 - 0}{\sqrt{s^2/SS_{xx}}}$
- The test statistic has a t distribution with $n - 2$ degrees of freedom under H_0 .
- The null hypothesis H_0 is rejected if the computed t -value from a sample is larger than the tabular value at a given level of significance.
- Example: Test the hypotheses about the regression of the example of weight and heart girth of cows where $b_1 = 7.53$, $SS_{RES} = 463.304$, $SS_{xx} = 34$

$$s^2 = \frac{SS_{RES}}{n-2} = MS_{RES} = 115.826$$

The standard error of the estimator b_1 is:

$$s_{b_1} = \sqrt{\frac{s^2}{SS_{xx}}} = \sqrt{\frac{115.826}{34}} = 1.845$$

The calculated value of the t statistic from the sample is:

$$t = \frac{b_1 - 0}{\sqrt{s^2/SS_{xx}}} = \frac{7.53}{1.845} = 4.079$$

- The tabular value is $t_{\alpha/2, (n-2)} = t_{0.025, 4} = 2.776$
- The calculated $t = 4.079$ is larger than the tabular value (2.776), thus the estimate of the regression slope $b_1 = 7.53$ is significantly different from zero and regression in the population exists.

Partitioning total variability

- An intention of using a regression model is to explain as much variability of the dependent variable as possible.
- The variability accounted for by the model is called the explained variability. Unexplained variability is remaining variability that is not accounted for by the model.
- The total variability of the dependent variable is the variability of measurements y_i about the arithmetic mean $\bar{y}$ and is measured with the total sum of squares.
- The unexpected variability is variability of y_i about the estimated line ($\hat{y}$) and is measured with the residual sum of squares.

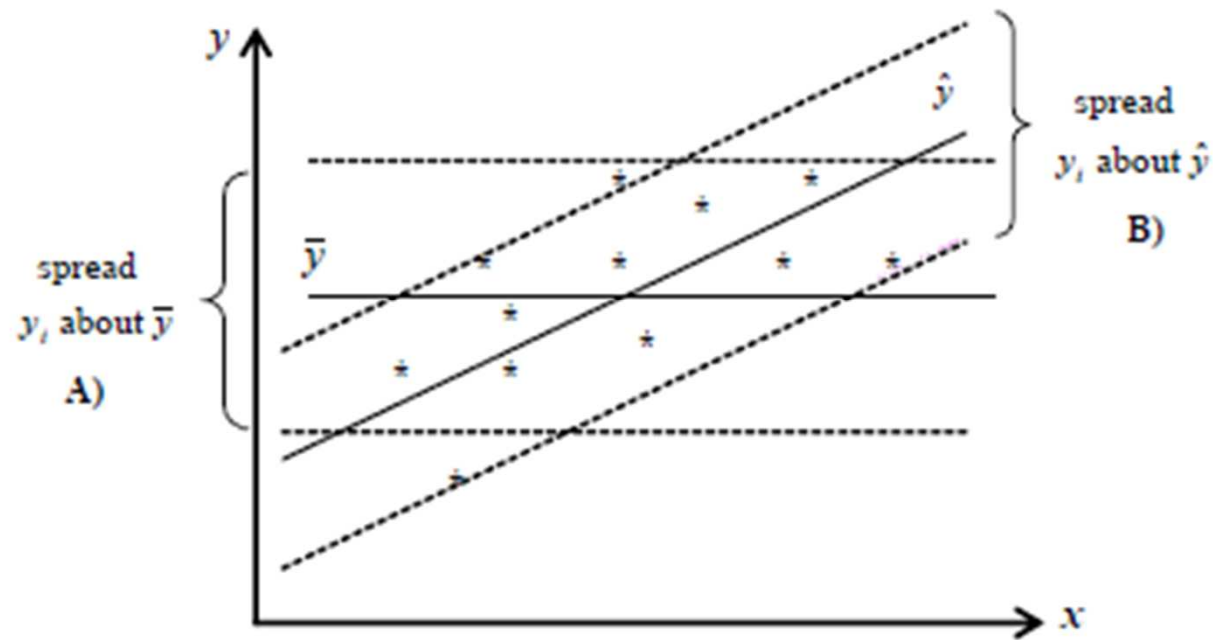


Figure 5: Distribution of variability about the arithmetic mean and estimated regression line:

A: Measured with the total sum of squares,

$$SS_{\text{TOT}} = \sum_i (y_i - \bar{y})^2$$

B: measured with the residual sum of squares,

$$SS_{\text{RES}} = \sum_i (y_i - \hat{y})^2$$

- Along with total and unexplained variability there is a variability explained by the regression model, which is measured with the regression sum of squares, $SS_{\text{REG}} = (\hat{y}_i - \bar{y})^2$.
- The total variability can be partitioned into variability explained by regression and unexplained variability,
- $SS_{\text{TOT}} = SS_{\text{REG}} + SS_{\text{RES}}$

The sums of squares can be calculated using shortcuts:

1. The total sum of squares is the sum of squares of the dependent variable:

$$SS_{TOT} = SS_{yy}$$

2. The regression sum of squares is:

$$SS_{REG} = \frac{(SS_{xy})^2}{SS_{xx}}$$

3. The residual sum of squares is the difference between the total sum of squares and the regression sum of squares:

$$SS_{RES} = SS_{yy} - \frac{(SS_{xy})^2}{SS_{xx}}$$

- Example: Compute the total, regression and residual sums of squares for the example of weight and heart girth of cows.
- The sum of product is $SS_{xy} = 256$ and the sum of squares is $SS_{xx} = 34$. The total sum of squares is the sum of squares for y .

The sums of squares are:

$$SS_{TOT} = SS_{yy} = \sum_i y_i^2 - \frac{(\sum_i y_i)^2}{n} = 2390.833$$

$$SS_{REG} = \frac{(SS_{xy})^2}{SS_{xx}} = \frac{(256)^2}{34} = 1927.529$$

$$SS_{RES} = SS_{TOT} - SS_{REG} = 2390.833 - 1927.529 = 463.304$$

Test of hypotheses with F test

- If the SS_{REG} and SS_{RES} are independent the F test can be applied for testing the hypotheses

$$H_0: \beta_1 = 0, \text{ vs } H_1 \neq 0.$$

- For the F test, mean squares are obtained by dividing the sums of squares by their corresponding degrees of freedom.
- If n is the number of pairs of observations, the degrees of freedom can be determined as follows:
 - Total df = $n - 1$
 - Residual df = $n - 2$
 - Regression df = 1

- Regression mean square, $MS_{REG} = SS_{REG}/1$
- Residual mean squares , $MS_{RES} = SS_{RES}/n-2$
- The F statistic is $F = MS_{REG}/MS_{RES}$
- The F statistic has an F distribution with 1 and $n-2$ degrees of freedom if H_0 is true.
- The sources of variability, sums of squares (SS), degrees of freedom (df) and mean squares (MS) can be written in an ANOVA table.

Anova table

Source	SS	df	$MS = SS / df$	F
Regression	SS_{REG}	1	$MS_{REG} = SS_{REG} / 1$	$F = MS_{REG} / MS_{RES}$
Residual	SS_{RES}	$n - 2$	$MS_{RES} = SS_{RES} / (n - 2)$	
Total	SS_{TOT}	$n - 1$		

Example: Test the regression hypotheses using an F test for the example of weights and heart girth of cows.

$$SS_{TOT} = SS_{yy} = \sum y_i^2 - \frac{(\sum y_i)^2}{n} = 2390.833$$

$$SS_{REG} = \frac{(SS_{xy})^2}{SS_{xx}} = \frac{(256)^2}{34} = 1927.529$$

$$SS_{RES} = SS_{TOT} - SS_{REG} = 2390.833 - 1927.529 = 463.304$$

The degrees of freedom for total, regression and residual are $(n - 1) = 5$, 1 and $(n - 2) = 4$, respectively.

The mean squares are:

$$MS_{REG} = \frac{SS_{REG}}{1} = \frac{1927.529}{1} = 1927.529$$

$$MS_{RES} = \frac{SS_{RES}}{n-2} = \frac{463.304}{4} = 115.826$$

The value of the F statistic is:

$$F = \frac{MS_{REG}}{MS_{RES}} = \frac{1927.529}{115.826} = 16.642$$

Anova table

Source	SS	df	MS	F
Regression	1927.529	1	1927.529	16.642
Residual	463.304	4	115.826	
Total	2390.833	5		

The tabular value of the F distribution for $\alpha = 0.05$ and 1 and 4 df is $F_{0.05, 1, 4} = 7.71$.

Since the calculated value $F = 16.642$ is greater than the tabular value, H_0 is rejected.

Coefficient of determination

- The coefficient of determination is used as a measure of the correctness of a model, i.e. how well a regression model will fit the data.
- The coefficient of determination represents the proportion of the total variability explained by the model.
- $R^2 = SS_{\text{REG}}/SS_{\text{TOT}}$

Example: Compute a coefficient of determination for the example of weights and heart girth of cows.

$$SS_{REG} = \frac{(SS_{xy})^2}{SS_{xx}} = \frac{(256)^2}{34} = 1927.529$$

or

$$SS_{REG} = b_1 SS_{xx} = (7.529)^2 (34) = 1927.529$$

$$SS_{TOT} = SS_{yy} = 2390.833$$

$$R^2 = \frac{SS_{REG}}{SS_{TOT}} = \frac{1927.529}{2390.833} = 0.81$$

Thus the regression model explains 81% of the total variability or variation in heart girth explains 81% of the variation in weight of these cows.

Correlation analysis

- Correlation analysis is used when there is interest in determining the degree of association among variables, but when they cannot be defined as dependent and independent. For example we may wish to determine the relationship between weight and height, but do not consider one to be dependent on the other.
- The coefficient of correlation measures the strength of the linear relationship between two variables.

- The coefficient of correlation (ρ) is defined as
- $\rho = \sigma_{xy} / \sqrt{\sigma_x^2 \sigma_y^2}$
- Where σ_y^2 = variance of y
- σ_x^2 = variance of x
- σ_{xy}^2 = covariance between x and y.
- Values of the coefficient of correlation range between -1 and 1 (i.e. $-1 \leq \rho \leq 1$).
- For $\rho > 0$, the two variables have a positive correlation. This means that as values of one variable increase, increasing values of the other variable are observed.

- For $\rho < 0$, the two variables have negative correlation, meaning that as values of one variable increase, decreasing values of the other variable are observed.
- The value $\rho = 1$ or $\rho = -1$ indicates an ideal or perfect linear relationship.
- $\rho = 0$ means that there is no linear association.

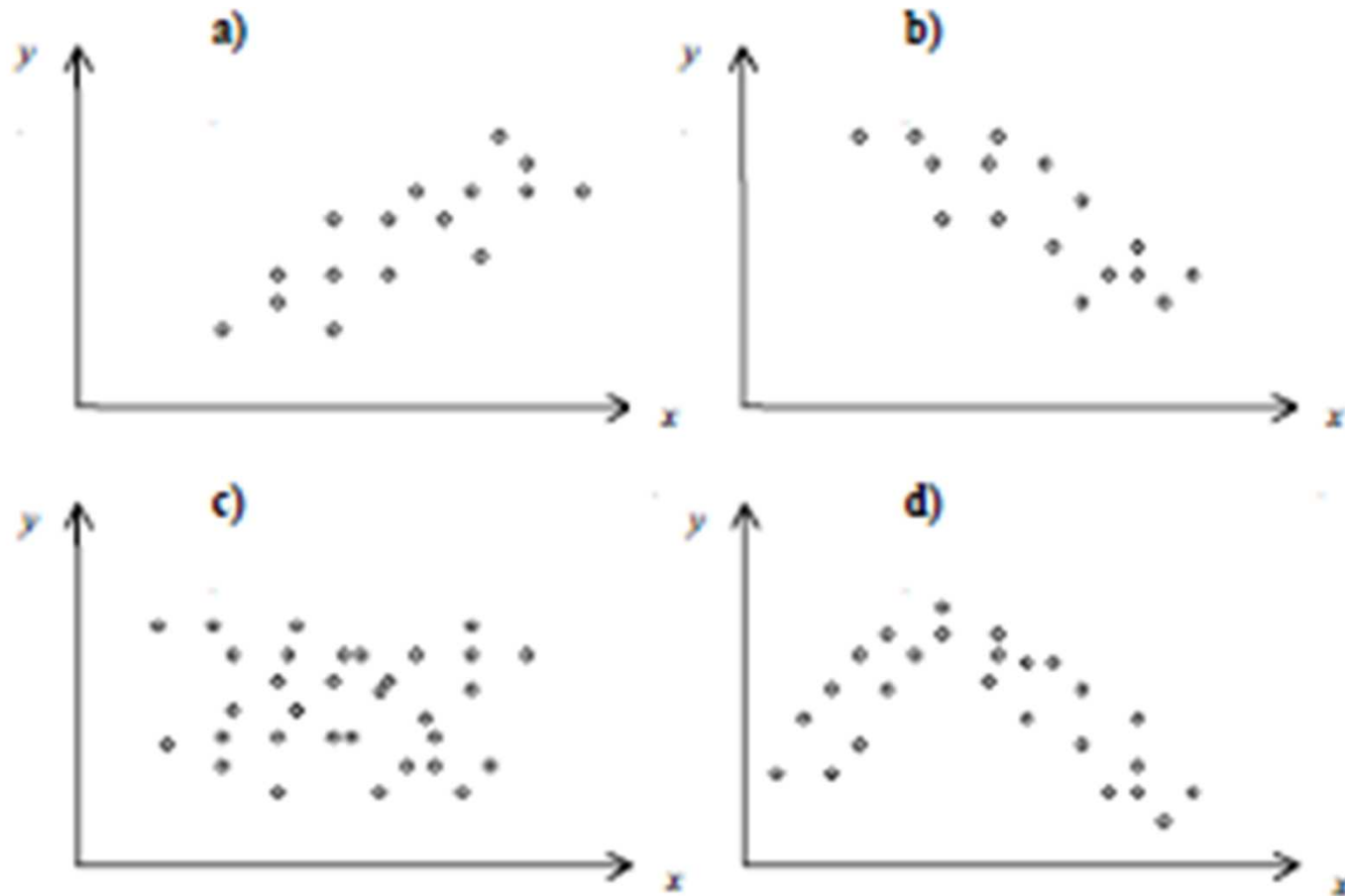


Figure 6: (a) Positive correlation, (b) negative correlation, (c) no association and (d) there is association, but not linear

- The coefficient of correlation is estimated from a random sample by a sample coefficient of correlation (r).
- $r = SS_{xy} / \sqrt{SS_{xx} SS_{yy}}$
- Where $SS_{xy} = \sum_i (x_i - \bar{x})(y_i - \bar{y})$ = sum of products of y and x
- $SS_{xx} = \sum_i (x_i - \bar{x})^2$ = sum of squares of x
- $SS_{yy} = \sum_i (y_i - \bar{y})^2$ = sum of squares of y
- To test the significance of a correlation in a sample, the null and alternative hypotheses about the parameter ρ are:
 - $H_0: \rho = 0$
 - $H_1: \rho \neq 0$

- The null hypothesis states that the coefficient of correlation in the population is not different from zero, that is, there is no linear association between variables in the population. The alternative hypothesis states that the correlation in the population differs from zero.
- In hypothesis testing a t distribution can be used and the t statistic is $t = r/s_r$.
- The statistic has a t distribution with $n-2$ degrees of freedom.
- $s_r = \sqrt{1-r^2/n-2}$, s_r is the standard error of the coefficient of correlation.
- $t = r - 0/\sqrt{1-r^2/n-2}$
- It is simplified to $t = r\sqrt{n-2}/\sqrt{1-r^2}$

Example: Weight was measured in kg and heart girth in cm on 10 cows. Is there a linear association between weight and heart girth in the herd of cows below?

Cow	1	2	3	4	5	6	7	8	9	10
Weight (y):	641	620	633	651	640	666	650	688	680	670
Heart girth (x):	205	212	213	216	216	217	218	219	221	226

- The computed sums of squares and sum of products are:-
- $SS_{xx} = 284.1$
- $SS_{xy} = 738.3$
- $SS_{yy} = 4218.9$
- $r = SS_{xy} / \sqrt{SS_{xx} SS_{yy}} = 738.3 / \sqrt{(284.1)(4218.9)} = 0.67$
- the calculated $t = r\sqrt{n-2} / \sqrt{1-r^2} = 0.67\sqrt{10-2} / \sqrt{1-0.67^2} = 2.58$
- The tabular value of t with 5% significance level and 8 df is $t_{\alpha/2,8} = t_{0.25,8} = 2.31$
- The calculated value is larger than the tabular value, hence, H_0 is rejected. There is linear association between weight and heart girth in the population.

SCIENTIFIC WRITING

Communicating research results

Introduction

- The aim of research is to contribute to knowledge.
- Every new result adds to the previous state of knowledge, forming a basis for new thinking and interpretation, new implications and identification of needs for further research.
- Research results do not contribute to knowledge and development unless they are communicated effectively.
- Communicating research results is as important to the scientific process as designing, conducting experiment and analyzing the data. 2

- Communication is needed, not only to spread research results, but also to articulate results.
- Forms of scientific communication:
 - Papers in scientific journals
 - Research reports
 - Conference papers
 - Theses and dissertations
 - Popular science and newspaper articles,
 - Computer-mediated information
 - Poster presentation
 - Oral presentation

- Communicating science usually means communicating new knowledge or summarizing the present state of knowledge.
- It is important that what is written or said is unambiguous, so that the audience understands the message.
- Science communication should be accurate, audience-adapted, brief and clear.
- When preparing to write a paper or a report or to make an oral or poster presentation, ask yourself the following questions: who are you addressing?, why is your message important?, what are your main findings or take-home messages?, how can you best deliver your message and satisfy the audience's needs?.

Major sections of a scientific paper

- Most scientific papers follow a standard format: Title, Abstract, Introduction, Materials and Methods, Results, Discussion, Conclusions and References.
- Acknowledgement and appendices are optional.
- Sometimes the two sections, Results and Discussion, are combined into one.
- For all types of papers, follow the guidelines (often called instructions for authors) set by the journal or publisher to which you will submit the paper.

Title

- The title tells what the paper is about, with the main purpose of encouraging people to read the paper.
- The title should be informative, specific and concise as it will be read more than any section of the paper.
- Journals give instructions to authors as to the length of the title. Generally limit the title to 10 – 15 words.
- Make sure that the title accurately reflects the content of the paper.
- Use technical terms only if they are familiar to most readers and do not use abbreviations in the title.

- Avoid using waste words (words that say nothing) e.g. observation of...., studies of ..., Investigations of ..., Examination of ..., A note on
- If you are writing the title for an oral or poster presentation at a scientific conference, remember that the title is often the primary basis for placing a presentation in a certain session of the scientific programme.
- Vague or uninformative titles increase the risk that the presentation might be allocated to an inappropriate session or might even be rejected.

Abstract

- The abstract is important because a large number of people will read it, and it will determine whether they read the entire paper.
- Readers should be able to understand the abstract without reading the details of the research methodologies and results (i.e. the abstract should stand alone).
- The abstract can be either indicative or informative.
- An indicative abstract indicates objectives of the research and suggests results in general terms.
- An informative abstract states the objectives and supports the conclusions with data.

- The abstract should describe the problem, state the objective and summarize the methods and techniques employed and the main results.
- The abstract should end with one or two sentences that emphasize important conclusions and implications.
- Do not cite references, tables and figures in the abstract.
- Do not use abbreviations, except those which do not require definitions (e.g. DNA).
- The abstract is written in the past tense for the objectives, materials and methods and results; and in the present tense for the justification, interpretations of the results and the conclusions.

Introduction

- The introduction persuades the reader that the topic is important and that the research is justified.
- The introduction should provide the necessary background and explain the rationale for the study by:
 - Summarizing the relevant literature.
 - Telling what has been done and what is lacking or what is the gap of knowledge in the literature.

- The introduction should specify the question to be answered or the hypothesis to be tested and state clearly the objectives.
- Use the present tense to write the motivation and justification e.g. Feed is the largest expense in dairy cattle production.
- In the review of literature use the past tense e.g. studies showed that Or in the present perfect tense, if it is common knowledge e.g. studies have shown that...
- The objective is written in the past tense e.g. the objective was....

Materials and Methods

- The Materials and Methods section gives details on how the study was done by providing a clear and complete description for all experimental procedures and statistical analyses.
- The materials and methods section should be organized into sub-section logically.
- It should include enough information so that another researcher can repeat the procedures and expect to get the same results.
- However, it is not necessary to describe in detail a procedure already published; just cite the original reference but explain the modification.

- Tell what type of study you conducted e.g. simulation study, a survey, a field study, an observational study, a designed experiment.
- The design of the experiments or the sampling protocol of surveys should be explained clearly and concisely, e.g. the randomization, replication and blocking structure should be described.
- Describe clearly and fully the experimental location and organisms and their numbers, sexes and ages.
- Describe the treatments and relevant environmental conditions and measurements taken.

- Describe the experimental units and how they were chosen and what was the rationale for doing so?
- State the assumptions you made and write measurements in standard units.
- It is not sufficient to merely mention the software used to perform statistical analysis, but specify the statistical models, methods and tests used to analyze and summarize the data.
- Use the past tense to write the materials and methods e.g. the treatments were..... and the experimental period was...

Results

- The results section describes your findings, usually with the support of tables and figures.
- The results should contain only results of your research and not comparison with the literature.
- Before writing the section for results prepare tables and figures.
- From among the tables and the figures select those are the most important or the most representative.

Table 1: An example of a table layout

Breed	Treatment	Variable 1 (units)	Variable 2 (units)	Variable 3 (units)
A	1			
	2			
	3			
B	1			
	2			
	3			

Table 1: Reproductive performance of local goats in Iramba and Kongwa districts

Parameter	Districts	
	Iramba	Kongwa
Buck life time (years)	7.7 ± 0.28^a	5.8 ± 0.23^b
Doe life time (years)	7.4 ± 0.25^a	7.0 ± 0.21^a
Number of kids per doe life time	8.8 ± 0.33^a	10.2 ± 0.29^b
Average age at first kidding (months)	16.8 ± 0.55^a	14.2 ± 0.48^b
Kidding interval (months)	8.2 ± 0.23^a	7.6 ± 0.21^a
Weaning age (months)	4.7 ± 0.18^a	5.0 ± 0.16^a
Litter size	1.2 ± 0.04^a	1.2 ± 0.03^a

^{a,b} Means with different superscript within a row differ significantly ($P < 0.01$)

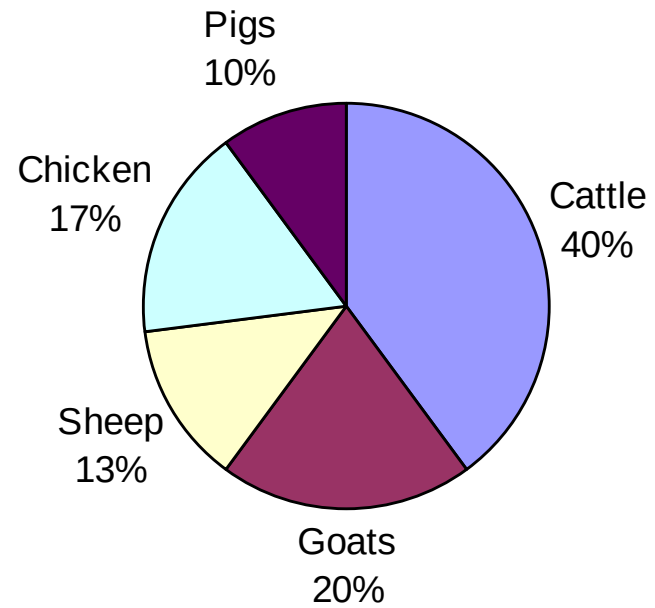


Figure 1: An example of a pie chart

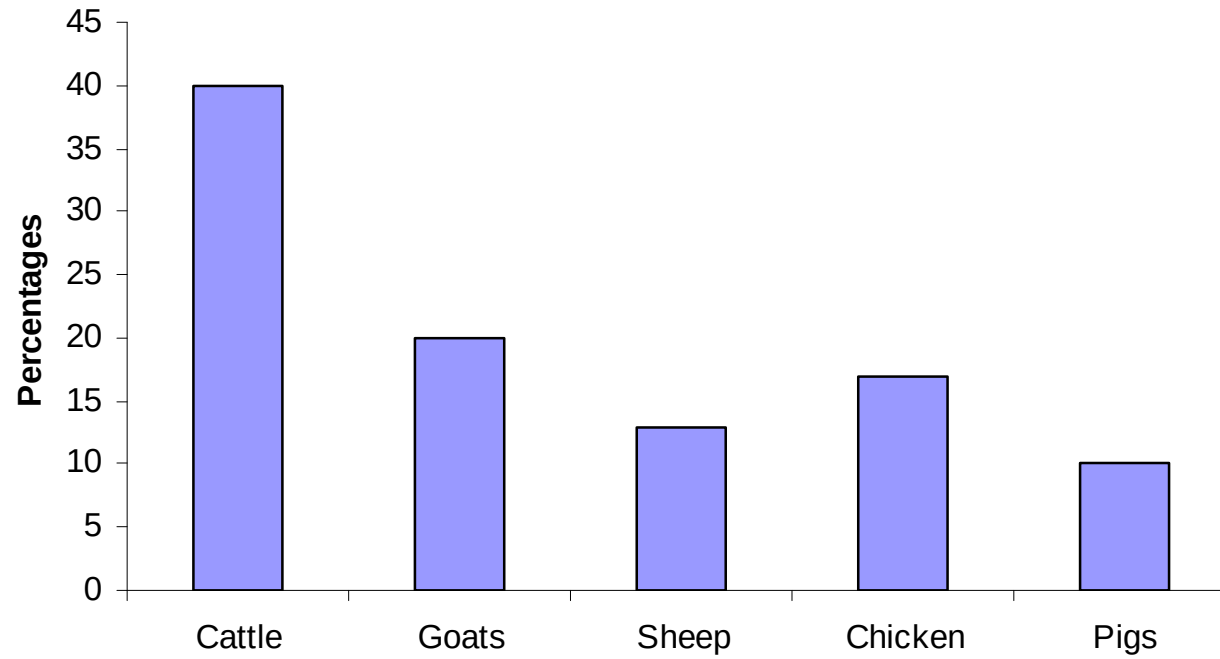


Figure 2: An examples of a bar chart

- Use tables to summarize numerical values.
- The numbers can be interpreted logically in the text of the paper.
- A table should be self-contained, so as to stand alone.
- Every table must have a title that explains what the table shows. The title should be descriptive enough.
- In the text refer to the table using the same words as in the title of the table.
- Use figures to focus on relationships among numbers.
- Do not repeat numerical values that already appear in a table.
- Figures should be understood independently (stand alone) without reference to the text or to tables.

- The results section should explain or elaborate on your findings by logically summarizing and illustrating relevant data.
- Sufficient data should be presented to allow the reader to interpret the results.
- In the text describe only the most important results that appear in a table or figure.
- Do not merely repeat numbers from the table in the text e.g. the mean of individuals on treatment A was 12 kg, whereas mean of those on treatment B was 15 kg.
- Instead describe and support your findings with numbers from the table or figure e.g. individuals on treatment A (12 kg) averaged 3 kg less than those on treatment B (15 kg).
- Use the past tense to write your results.

Discussion

- The discussion section should interpret the results clearly, concisely and logically.
- Results or references to tables or figures already described in the results section should not be repeated.
- Start by reminding the reader the objectives.
- Describe how your results meet the objectives.
- Identify the significant results and recognize the importance of negative or non-significant results.

- Cite evidence from the literature that supports or contradicts your results; explain the contradictions.
- Describe the limitations of your research e.g. limitations with regard to the design of the experiment, the number of observations or the statistical analysis.
- Use the past tense to refer to your results, but use the present tense to interpret your results and to refer to results of others.

Conclusions

- Conclusions summarize the main results of the research and describe what they mean in general.
- Do not mention new results for the first time or refer to the literature.
- Although some speculation is allowed, avoid over-extrapolating the results.
- Provide an interpretation of the impact of the research results.
- Suggest future research to follow up where your research ended.
- Use the present tense to write the conclusions. ²⁴

Acknowledgement

- Acknowledgement are of two types: general and specific.
- General acknowledgement include those for an institution, a laboratory and source of funds.
- Specific acknowledgement include those for colleagues and technicians.

References

- Citations of published literature in the text as well as in the reference list should follow the instructions set by the journal or institution where it is to be published.
- Be certain that all references listed are cited in the text of the paper and that all references cited in the text are listed in the references.
- The purpose of the reference list is for others to be able to find the papers or books you are referring to.

- For citing references within the text, the authors' last names and year of publication are given in the text, e.g. experimental evidence suggests that the sky is blue (Smith, 1999) or Smith (1999) observed that the sky is blue.
- If a reference has two authors write the last names of both authors e.g. (Smith and Brown, 1999) or Smith and Brown (1999).
- When a reference has more than two authors, write the last name of the first authors plus *et al.* e.g. (Smith *et al.*, 1999) or Smith *et al.* (1999)

- When an author or a group has more than one publication in one year, the publications are distinguished by adding a letter to the year e.g. Smith,1999a; Smith 1999b. When several references are cited together in the text, they should be listed in chronological order e.g. (Jones, 1998; Black, 2000; Green, 2002).
- All references cited in the text should be in the reference list at the end of the paper.
- The references are listed in alphabetical order of first author's last name.
- The format of reference list is different for every journal or institution. For this reason you should read the instructions for authors and follow them.

Examples of writing references in the reference list

1. Papers in journals

Chenyambuga, S.W., Hanotte, O., Hirbo, J.,
Watts, P.C., Kemp, S.J., Kifaro, G.C., Gwakisa,
P.S., Petersen, P.H., and Rege, J.E.O.
(2004). Genetic characterization of indigenous
goats of sub-Saharan Africa using microsatellite
DNA markers. *Asian-Australasian Journal of
Animal Sciences* 17(4) : 445-452.

2. Papers in proceedings

Mucuthi M. M. and Munei K. (1994). Some constraints to small ruminant production among small-scale farmers in Laikipia West. In: *Proceedings of the third biennial conference of the African Small Ruminant Research Network*, UICC, Kampala, Uganda, 5 – 9 December, 1994. pp 83 – 86.

3. Papers published online

Chenyambuga S. W., Waiswa C., Saimo M., Ngumi P. and Gwakisa P. S. (2010).

Knowledge and perceptions of traditional livestock keepers on tick-borne diseases and sero-prevalence of *Theileria parva* around Lake Victoria Basin. *Livestock Research for Rural Development*. Volume 22, Article no. 137 .

<http://www.lrrd.org/lrrd22/7/chen22137.htm>

Retrieved August 16, 2010.

4. Books

Devendra C. and Burns M. (1983). Goat production in the tropics. Commonwealth Agricultural Bureau. Farnham house, Farnham royal, Slough SL2 3BN UK. pp183.

5. Theses

Madubi M. A. (1996). Physical characterization of three strains of Small East African goats in Tanzania. *MSc. Dissertation*, Sokoine University of Agriculture, Morogoro, Tanzania.

Sequence of chapters for dissertations/theses at SUA

1. Title and name of the student
2. Abstract
3. Declaration
4. Copyright
5. Acknowledgement
6. Dedication
7. Table of contents
8. List of tables
9. List of figures
10. List of appendices

11. List of abbreviations and symbols
12. Introduction
13. Literature review
14. Materials and methods
15. Results
16. Discussion (Results and discussion may be combined).
17. Conclusions and Recommendations
18. References
19. Appendices

Oral presentation

Introduction

- New research results are communicated orally at scientific meetings such as national and international conferences.
- Oral presentations are also given at workshops and University seminars.
- It is worthwhile making an effort to perform a good oral presentation, not only because audiences remember both the good and bad presentations, but also because the message is conveyed most efficiently in the good ones.
- A presentation should arouse interest in the topic, stimulate discussion and should make it easy for the audience to understand the main points quickly.

- Think ahead about your audience: their background, interests and capability. The listeners have to understand what you say immediately, they cannot go back as they can in written text.
- Oral presentations must be attractive, interesting, easy to understand and focused on the main message.
- When preparing an oral presentation, think of how best to raise and maintain the audience's interest, and how to structure the topic so that the audience will understand and remember your main messages.

- Audience attention varies during a presentation, it is highest at the beginning of your presentation, because the audience is curious about what will come, is lowest in the middle, and is high again at the end, in anticipation of your conclusions.
- Plan your talk so that you use the times when audience attention is at the highest level to get across your main take-home messages.
- Write down a number of questions that you expect your audience might have with regard to your topic, and incorporate your answers to these questions into your presentation.

Preparation of the oral presentation

- Make a presentation outline
- Before developing an outline, you must know how much time has been allotted for your talk. Most presentations are restricted to 10 – 15 minutes. Plan to present one slide per minute.
- Consider the language for your presentation, and prepare your presentation directly in that language.
- Plan your presentation to be simple and clear.
- A general structure for a presentation outline could be: title, your name and organization, introduction, objectives, materials and methods, results and discussion, conclusions and recommendations.

- For a proposal the structure should be: title, your name and organization, Introduction, problem statement/justification, objectives, materials and methods, work plan (schedule of activities), expected output and budget.
- Support your talk with visuals to effectively communicate the message and stimulate interest.
- People absorb more information, and retain that information longer, when it is communicated to them in both words and pictures, compared with words or pictures alone.
- It is essential that the visuals used are seen clearly and are meaningful to the audience.

- The common visuals used in oral presentations at scientific meetings are electronic presentation (powerPoint and an LCD projector, overhead transparencies (acetate sheets), slides and a slide projector.
- Before preparing your presentation, make sure that the room has the appropriate equipments.

Creating visual displays

- All types of visuals mentioned above can be created on a computer in various ways.
- You can use autolayouts and design templates that help you build displays in a standard format or you can start from a blank page and give your visuals a personal design.
- Keep the display simple as there is a great risk that the audience will not grasp the main messages.
- Use a large font size that can be read from a distance by everyone, including those sitting at the back of the room. Usually the minimum font size is 28 points for the text and 40 points for the headings.

- Don't put too much information or long sentences in a slide.
- Tables of research results for oral presentation should be converted to figures, which make it easier for the audience to grasp the message quickly.
- If the precise data are important, use a table. The tables should not be complex i.e. three to five rows and/or columns are enough. The font size should be as large as for the text and rounding of numbers adds to clarity.
- Tables and figures must be simple, clear and not overloaded.
- Choose colours and background that make a good contrast between text and background e.g. dark text on a light background.

Performing the oral presentation

- An oral presentation can be performed in different ways. You can stand by the lectern, you can stand away from it to use more of your body language or you can even move around a little.
- Choose an approach that fits your personality and that makes you feel comfortable. However, at all times, remember the people in the audience.
- Keep eye contact with your audience and make sure that you don't turn back on the audience.
- Don't stare at the computer screen. Just have a quick glance at the screen when needed, and immediately reestablish eye contact with your audience.

- Make sure your body doesn't block the audience's view of your visuals.
- Speak to be heard and understood. It is essential that everyone in the audience can hear you.
- Dress appropriately.
- It is useful to have some kind of manuscript with keywords, but you should not read your talk from a fully written manuscript. Reading makes it hard to maintain eye contact with the audience.
- Rehearse on your presentation and ask a colleagues to listen to you. You can also use video or a tape recorder.
- Make sure that you will finish the presentation within the time allotted for your presentation.

- Prepare a presentation that will last about 80% of the allotted time, because it takes longer to give the actual presentation and you don't want to feel rushed.
- Feeling nervous just at the start of a presentation is natural, don't worry about it. Remember you are the expert on your topic, nobody in the room knows more about what you will say than you.
- In most scientific meetings, there will be opportunity for the audience to ask questions or to raise issues for discussion. Write down a few keywords for each question.
- Your answers to the questions should be short and avoid dialogue with the questioner, direct your answers to the entire audience.
- If you don't know the answer, say so, and offer to find out.