

# Quantitative Methods

## Sequence, 18 sessions

This curriculum separates qualitative from quantitative methods but only for the purposes of planning and organising. In practice, CARTA strongly recommends that you combine these areas through an integrated and multidisciplinary approach. This sequence is best taught in tandem with qualitative approaches to research and mixed methods.

Students gain and practise skills and understanding of quantitative methods through integrated activities, in particular:

- [Posters: Applying Research Methods.](#)
- [Journal Clubs.](#)

Qualitative methodology is introduced in an earlier sequence in this curriculum:

- [Research Question and Methodology.](#)

A later sequence – once researchers have collected data – revisits the analysis of quantitative data:

- [Qualitative Data Analysis.](#)

Download [the curriculum](#) for this sequence.

## Outcomes

By the end of this sequence of sessions, together with the sessions on qualitative and mixed methods, students can:

- Select the appropriate research study design for their chosen study.
- State the limitation(s) of various research methodologies.
- Understand how to generate, manage, and analyse research data.

## Schedule

Your institution and resources will determine how you schedule this training. You might:

- Run the sequence a week or a few days at a time, to pace the input to your cohort of students and bring different disciplines together.

In whatever way you schedule and integrate this training, students need to meet certain milestones in order to move ahead on their PhD journey. Most important is to bring students together to:

- Support and motivate each other. Students who are skilled in one area assist those whose strengths lie elsewhere.
- Reinforce the value of multiple views on an issue.
- Teach certain aspects that individual supervisors would otherwise have to cover.

This sequence of sessions, together with sessions on qualitative and mixed methods, supports students to:

- Develop or strengthen their PhD research protocol.
- Understand the methods that are used in qualitative research
- Gain or strengthen core skills in data management and analysis.
- Evaluate different research methods and select the most appropriate design for their research.

## Preparation

A number of these sessions involve work in small groups. Identify appropriate skilled qualitative researchers to act as resource people in order to:

- Participate as small-group facilitators.
- Answer students' questions in open discussion.
- Provide input and guidance during group exercises.

Ensure that resource people are familiar with the participatory [approach](#), have read the relevant sessions that make up this sequence, and engage informally as well as giving lectures or instructions.

You can run this sequence of sessions as face-to-face teaching, on-line or a blend of the two. For online elements, organise an online platform where students upload and comment on exercises. Ensure that you have tech support on hand.

## Sessions

### Session 1. Introduction to the Quantitative Sessions | 2 hours

This session prepares students for the rest of this sequence on quantitative research. Here, you:

- Encourage students to think about how to link their objectives with their analysis.
- Examine the different data types.
- Consider outcome variables, exposure variables, and potential confounders.
- Highlight the importance of collecting correct, high-quality data
- Emphasise the need for a data management plan
- Preview expectations of subsequent sessions.

#### Outcomes

By the end of the session, students can:

- Identify and use different data types appropriately (quantitative/ discrete or continuous data).
- Identify their primary and secondary outcome variables and the potential confounders.
- Explain the importance of a well-structured data management plan to ensure that they have quality data, includes all required variables, for analysis.

#### Preparation

*As facilitator*

Develop or source a data set and activity.

Prepare or source an introductory presentation.

#### References

John, C. (2009). *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*, 3rd edition. University of Nebraska, Lincoln: SAGE Publications, ISBN: 1412965578

Kathryn, P. (2007). *Mixed Method Designs: A Review of Strategies for Blending Quantitative and Qualitative Methodologies*. *Mid-Western Educational Researcher*, 20(4), 35-38

Bryman, A. (2014). *Social Science Research Methods*, 5th Edition. Oxford, UK: Oxford University Press.

Gary, K., Robert, O. K., and Sidney, V. (1992). *Designing Social Inquiry*. Princeton, NJ: Princeton University Press.

#### Students

Must have their research topic and objectives.

#### Assessment

Assess and give feedback on group exercise

#### Steps

Introduce or re-visit these topics:

- Types of data.
- Outcome variables, exposure variables and potential confounders /effect modifiers.
- Data management and data entry requirements.
- The need for statistical models. Provide a data set and the related exercise. Students to work on in small groups to identify types of data, outcome and exposure variables, and data management.

### Session 2. Measures of Central Tendencies | 2 hours

This session gives students in all fields of public health a grounding in the basic concepts of epidemiology – the study of

the distribution and determinants of health and disease in different human populations and the application of methods to improve disease outcomes.

Use the reproductive-debut example to guide students to:

- Understand, use, and value simple statistical measures: mean, median, range, and variance.
- Discuss the interpretation and generalisability of data, and concepts of bias.
- Begin to develop the skills to read, interpret, and evaluate health information from published epidemiologic studies.
- Appreciate the notion of confidentiality and sensitivity in sexual and reproductive health data
- Learn ways to protect the privacy of research subjects.

### Outcomes

By the end of this session, students can:

- Analyse data and calculate simple statistical measures (range, mean, median, mode, and standard deviation).
- Discuss generalisability, and how this particular sample may influence it.
- Discuss bias (e.g., self-exclusion by not filling it in) and how this may play out in a population-based survey.
- Describe confidentiality and how/if it was adhered to in the example 'study' and how people felt about filling in this form.

### Preparation

*As facilitator*

Develop a form on reproductive information through an online tool (Google sheet for example), something like this:

[table "116" not found /]

Note that you require prior data collection. Send the form to the students the day before the session with this request:

Please fill in and return this form. For the variables marked \*, please fill in information related to the first time you were a co-parent in a pregnancy irrespective of whether you are male or female and irrespective of whether the pregnancy resulted in a live birth or not and the first time you were a co-parent for a child which was born. If you have never had sex, or never been a co-parent of a pregnancy or birth or have never been married, please leave blank.

*Students*

Fill in the form well before the session.

*References*

- John M Last (2001). [A Dictionary of Epidemiology](#). 4th edition. Oxford University Press.
- Miquel Porta (2008). [A Dictionary of Epidemiology](#). 5th edition. Oxford University Press.
- Kabacoff, R. (2015). *In Action: Data Analysis and Graphics with R*. 2nd Edition. Shelter Island, NY: Manning Publications Co.
- Longest, K.C. (2014). *Using Stata for Quantitative Analysis*. Thousand Oaks, CA: Sage.

### Assessment

#### Steps

In groups, students analyse the data from the forms.

In plenary, present findings and explain each was calculated and what they mean – focus on issues of central tendency.

Present findings in a bar graph with a distribution drawn over it – focus on measures of spread.

Interpret the data.

Discuss generalisability, and how this particular sample may influence it.

Discuss bias (e.g., self-exclusion by not filling it in) and how this may play out in a population-based survey.

Introduce anonymity and privacy in relation to this 'study'.

Discuss confidentiality and how/if it was adhered to in this 'study': how did participants feel about filling in this form?

## Session 3. What is My Exposure, What is My Outcome? | 2 hours

This session supports students in framing their research question in such a way that they can clearly identify the exposure

and outcomes. It enables students to understand a range of concepts concerning exposure and outcome variables, to refine their research questions, and to apply this knowledge in their own research process.

### Outcomes

By the end of the session, students can:

- Describe the exposure(s) that relate to their research question.
- Describe the outcome(s) that relate to their research.
- Appraise and critique their own and other students' research questions.

### Preparation

*As facilitator*

Create or source a presentation to introduce the concepts.

*Students*

Must have their research questions and objectives.

### Steps

Introduce the relevant concepts and the exercise for work in groups. Individually and in groups of four, students discuss the exercise: to develop and refine their research question. Ensure that you and co-facilitators are available to provide support as peers review each other's questions, ask questions of clarification, and provide constructive criticism.

## Session 4. Statistical Bias | 90 minutes

Introduce and discuss the concept of statistical bias: a feature of a statistical technique or of its results in which the expected value of the results differs from the true quantitative parameter being estimated. Cover:

- The definition of statistical bias.
- Types of statistical bias.
- The effect of statistical bias on the research results.
- Strategies to control these types of biases.

### Outcomes

By the end of these steps, students can:

- Discuss the concept of statistical bias.
- Identify types of statistical bias.
- Explain the sources of statistical bias.
- Describe how to avoid statistical bias.

### Preparation

*As facilitator*

Create or source a presentation to introduce the concepts.

Create or source scenarios for students to work on in groups.

*References*

Grimes, D. A., & Schulz, K. F. (2002). [Bias and causal associations in observational research](#). *The Lancet*, 359(9302), 248–252. (Access through your institution.)

Krause, M. S., & Howard, K. I. (2003). [What random assignment does and does not do](#). *Journal of Clinical Psychology*, 59(7), 751–766. (Not open access.)

### Assessment

Assess individual participation and group assignments.

### Steps

Explain the concepts and invite questions and discussion. Students work on scenarios in small groups. Individuals apply the concept to their proposed research studies.

In plenary, groups present their work for discussion.

## Session 5. Confounding and Effect Modification | 90 minutes

Study results can be considerably distorted by the presence of an extraneous factor (a confounder) and effect modification by a third factor (interaction). In this session:

- Introduce the concepts of confounding and effect modification in epidemiology.
- Describe methods of controlling them in both study design and data analysis.

Effect modification and confounding are difficult concepts to understand and distinguish from each other.

Confounding is defined as a distortion in an association that is seen when the exposal factor of interest is muddled with other factors that related to the outcome. The word 'confounding' is derived from Latin "confundere" meaning to mix or muddle.

Effect modification is seen when various effects are brought about among different subgroups by an exposure and this can be handled by doing stratification. Effect modification is associated only with the outcome of the study, but not the exposure. In this session the theory as well as the practical side of these issues will be discussed. It will cover the definition of both topics, approaches to control for confounding and effect modification.

### Preparation

*As facilitator*

Create or source a presentation to introduce concepts.

Identify research examples where effect modification and confounding are present.

### References

Shapiro, S. (2008). [Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy. Part 1. Principles of causality in epidemiological research: time order, specification of the study base and specificity.](#) J Fam Plann Reprod Health Care. 34: 83-7.

Shapiro, S. (2008). [Causation, bias and confounding: a hitchhiker's guide to the epidemiological galaxy Part 2. Principles of causality in epidemiological research: confounding, effect modification and strength of association.](#) J Fam Plann Reprod Health Care. 34:185-90.

Pearce, N. and Greenland, S. (2014). Confounding and Interaction. In: Handbook of Epidemiology. Ahrens and Pigeot I, eds. New York: Springer, pp 659-684.

Grimes and Schulz (2002). [Bias and causal associations in observational research.](#) Lancet 2002; 359:248-52.

Greenland, S. and Morgenstern, H. (2001). [Confounding in health research.](#) Annu Rev Public Health. 22:189-212. (Request pdf.)

John, M. L. (2014). A Dictionary of Epidemiology. Oxford University Press:4th Edition: P. No. 14, 37, 57

Kahlert, J., Gribsholt, S.B., Gammelager, H., Dekkers, O.M., Luta, G. (2017). [Control of confounding in the analysis phase – an overview for clinicians.](#) Clin Epidemiol. 2017 Mar 31;9:195-204. doi: 10.2147/CLEP.S129886. PMID: 28408854; PMCID: PMC5384727.

### Assessment

#### Steps

Introduce the concepts of random and systematic error, and then cover the concept of confounding in detail, using examples (such as smoking and lung cancer; maternal age and Down Syndrome; the relationship between obesity and cardiovascular disease, confounded by age).

Discuss methods of controlling for confounding during the design and analysis stages of research. Invite students' participation and input.

Work through examples for testing for confounding are worked through. Introduce the concept of effect modification.

Discuss methods of controlling for effect modification during the analysis stage of research. Introduce and discuss ways to distinguish between confounding and effect modification, using an example.

To conclude, summarise the differences between confounding and effect modification.

## Session 6. Validity and Reliability | 90 minutes

“Any research can be affected by different kinds of factors which, while extraneous to the concerns of the research, can invalidate the findings.”- Seliger and Shohamy, 1995.

Every researcher wants to be certain that their research findings are precise, valid, and reliable. But there are many threats to validity and reliability. This session covers

- The meaning of validity and reliability and the differences between them.
- Threats to validity and reliability.
- Measurement of validity and reliability.
- Measures to ensure high validity and reliability.

### Outcomes

By the end of the session, students can:

- Describe types of validity and reliability and their importance.
- Differentiate validity from reliability.
- Describe measures of validity and reliability.
- Describe threats to validity and reliability.
- Estimate reliability measures using Stata.

### Preparation

#### References

Braimoh, B., Danuta, K., Dick, H., Kerry, W. (2010). [Time-to-pregnancy and pregnancy outcomes in a South African population](#). BMC Public Health. 10:565.

Pay attention to the following sections:

- Data collection – paragraph 3
- Statistical analysis
- Results: Questionnaire reliability
- Discussion on reliability: Paragraphs 5 to 7.

Antoinette, F. D., Martin, J.R., Luke, M., Inocencio, M., and John L. [Test-retest stability of patient experience items derived from the national GP patient survey](#). Campbell

Pay attention to the following sections:

- Measure of reliability for categorical variables
- Measure of reliability for numerical variables

### Steps

Cover these aspects in plenary and group exercises:

- Understanding reliability vs validity.
- Rationale and purpose of validity and reliability.
- Types of reliability (Test-retest, interrater, internal consistency).
- Types of validity (content, construct, face, criterion...).
- Deal with validity and reliability in quantitative research.
- Threats to validity and reliability.
- Estimate measures of reliability using Stata.

## Session 7. Quantitative Research Study Designs | 90 minutes

### Outcomes

By the end of the session, students can:

- Describe the different study designs.
- Explain the factors that determine which study design is appropriate for a particular research question.

### Preparation

*As facilitator*

Create or source an introductory presentation.

#### References

Creswell, J.W. (2014). *Research design: qualitative, quantitative, and mixed methods approaches*. (4th ed.). Thousand Oaks: SAGE Publications. ISBN 978-1-4522-2609-5.

Claybaugh, Z. "Research Guides: Organizing Academic Research Papers: Types of Research Designs".

Wright, S., O'Brien, B.C., Nimmon, L., Law, M., Mylopoulos, M. (2016). [Research Design Considerations](#). *Journal of Graduate Medical Education*. 8 (1): 97–98. doi:10.4300/JGME-D-15-00566.1. ISSN 1949-8349. PMC 4763399. PMID 26913111.

#### Steps

1. Introduction to research designs and methodological choices
2. Experimental designs:
  - True experimental designs
  - Pre-experimental designs
  - Quasi-experimental designs
3. Non-Experimental/observational designs:
  - Cross-Sectional design
  - Longitudinal design
  - Historical design
  - Correlational/causal design
  - Cohort study design
  - Case-control design
  - Meta-analysis design
  - Action research design

## Session 8. Quantitative Data Collection Tools | 1 hour

The quality of research outputs largely depends on the quality of the data analysed. Thus, researchers should aim to collect high-quality data for their studies. One of the key factors is the appropriateness and quality of the tools used to collect them, whether the data are primary or secondary.

This session builds and strengthens participants' capacity to select, design, and develop appropriate and robust data collection tools for their research studies. They also learn how to critique a data collection tool.

#### Outcomes

By the end of this session, students can:

- Describe appropriate forms of data collection for differing research designs.
- Explain correctly the importance of layout in data collection tool design.
- Describe how data can be coded to facilitate data management.
- Critique data collection tools.

#### Preparation

*As facilitator*

Create or source an introductory presentation and group exercises.

#### References

Boynton, P., Greenhalgh T. (2004). [Selecting, designing and developing your questionnaire](#). *BMJ*; 328: 1312-5

Boynton, P. (2004). [Administering, analysing and reporting your questionnaire](#). *BMJ*; 328:1372-5

Boyton, P., Wood G.W., Greenhalgh T. (2004). [Reaching beyond the white middle classes](#). *BMJ*;328; 1433-6

Ann, Bowling (2009). *Research Methods in Health: Investigating health and health services*. Open University Press McGraw Hill International Maidenhead, Berks, UK

## Steps

In an introductory presentation, plenary discussion, and group exercises, cover these aspects:

- Data collection tools used in quantitative research for different designs.
- Design a data collection tool.
- Coding to facilitate data management.
- Evaluation of a data collection tool.

## Session 9. Sample Size Calculations | 2 hours

### Outcomes

By the end of this session, students can:

- Explain the factors to consider when deciding on the sample size for a research project.
- Carry out sample-size calculations for a descriptive study and for an analytic study in which two groups are compared.
- Describe the implications of other considerations, such as the need to adjust for confounders, the need to rule out interaction, and the need to adjust for clustering for the overall sample size.

### Preparation

*As facilitator*

Ensure that Stata, StatCalc and/or Epiinfo software is/are installed on participants' computers before the session starts.

Identify and engage skilled co-facilitators to support groups in the exercises.

Create or source a presentation to introduce the methods and exercise.

*Reference*

Bartlett, E.J., Kotrlik, W.J., Higgins, C.C.(2001). [Organizational Research: determining appropriate sample size in survey research](#). Information Technology, Learning, and Performance Journal. 19.

### Assessment

Assess students' sample-size calculations. (Group: 80%)

Assess participation in the session. (Individuals: 20%)

## Steps

Introduce and discuss:

- The problem for sample size calculations (15 minutes).
- Sample size for descriptive studies: means and proportions and use of software (30 minutes).
- Sample size for comparing two means and use of software (15 minutes).
- Sample size for comparing two proportions and use of software (15 minutes).

In a practical exercise (45 minutes), students work in groups of four to

- Determine sample size.
- Discuss and describe the implications of the other considerations: confounders, interaction clustering for the overall sample size.

Using real-world examples from students' own research, groups do the calculations in Stata and StatCalc within EpiInfo.

## Session 10. Sampling Methods | 2 hours

Sampling is concerned with the selection of a subset of individuals from within a population to estimate characteristics of the whole population and to make inferences from them. In this session, students gain a strong understanding of different types of sampling method and their application in scientific research.



## Outcomes

By the end of this session, students can:

- Explain the different sampling methods and considerations for each method.
- Identify and apply the appropriate sampling methods to different research studies.
- Choose the appropriate sampling methods for their proposed PhD research.

## Preparation

*As facilitator*

Identify and engage a trained co-facilitator if you are working remotely.

Create or source an introductory presentation.

## References

Coyne, I. T. (1997). [Sampling in qualitative research. Purposeful and theoretical sampling: merging or clear boundaries?](#) Journal of advanced nursing, 26(3), 623-630.

Latham, B. (2007). [Sampling: What is it.](#) Quantitative Research Methods, ENGL, 5377.

Altmann, J. (1974). [Observational study of behavior: sampling methods.](#) Behaviour, 49(3), 227-266. (Not open access)

## Steps

In the introductory presentation, cover these elements:

- Reasons for sampling.
- Classification of different sampling methods.
- Requirements for probability and nonprobability sampling methods.
- Advantages & disadvantages of each sampling method.
- Implication for sample size and generalizability of results /inferences.

In groups, students work on scenarios to identify the most appropriate sampling method for their proposed research studies.

## Session 11. Introduction to Stata | 2 hours

Limited knowledge and skills in statistical data analysis among doctoral students is one of the important causes of delay in completing the doctoral studies. In particular, little or no knowledge of the use of statistical software is a key component of this impediment.

This session equips students with practical knowledge to help them analyse their research data using Stata. This introductory session covers data entry, data importing, and data manipulation using Stata.

Through hands-on training and a variety of examples, students learn Stata structure and philosophy and recognise the potential of the software for analysing their own research data. They run statistical analyses and learn to interpret the Stata results correctly.

In addition, those students who have already collected some data for their doctoral study have the opportunity learn statistical analysis using their own research data.

## Outcomes

By the end of this session, students can:

- Perform data entry, editing, and handling using sort, in/by/if, drop and keep.
- Save, exporting and importing data into Stata.
- Summarize, tabulate data using Stata.
- Use Stata graphics, box plots, histogram, bar graphs, pie charts, etc.

## Preparation

*As facilitator*

Ensure that students have learning dataset loaded and Stata software installed on their computers before the session starts. Create or source presentations and a data set and instructions for the group exercise.

#### References

Germán Rodríguez. (2023). [Stata Tutorial](#). Princeton University.  
Daniels, L., Minot, N. (2020). [An Introduction to Statistics and Data Analysis Using Stata](#). Sage Publications. 392 pages.

#### Steps

Cover these elements:

- Data entry, editing, and handling, using sort, in/by/if, drop and keep in Stata. (30 minutes)
- Save, export and import data into Stata. (15 minutes).
- Summarize, tabulate using Stata. (15 minutes)
- Using Stata graphics: box plots, histogram, bar graphs, pie charts, etc. (20 minutes)

In groups and with a dataset, students conduct a practical exercise (40 minutes), to generate appropriate summary statistics and graphics according to types of indicated variables (including continuous, nominal, and ordinal variables).

## Session 12. Quantitative Data Analysis Plan | 6 hours in 3 separate sub-sessions

Here, students come to understand what is required in the development of both a data management plan and a data analysis plan for quantitative methods. They learn to specify:

- The outcome variables and important exposure variables for their study.
- How they will collect data and entered it in a study database.
- What measures they will implement for data validation.
- What is entailed in data lock.

Students need to identify – clearly and unambiguously – the study population they will analyse and then write up a data analysis plan to reflect the study objectives.

#### Outcomes

By the end of these sessions, students can:

- Specify clearly how they will collect and store data – a data management plan – to ensure they have quality data for their project.
- Specify clearly the study population they will analyse.
- Write a data analysis plan that reflects the study objectives, whether the study is a randomised controlled trial or an observational study.

#### Preparation

*As facilitator*

Create or source a presentation to cover the elements.

#### Reference

Vandenbroucke, J. P. (2007). [Strengthening the Reporting of Observational Studies in Epidemiology \(STROBE\)](#): explanation and elaboration, PLoS Medicine 4 (10): e29

#### Steps

Between your presentation and students' work on their own plan, cover these elements:

- The need for quality data – development of a data management plan.
- Specifying the population to be analysed.
- Data analysis plans for randomised controlled trials.
- Data analysis plans for observational studies.

Divide each of the three part-sessions into:

*First hour:*

You or co-facilitator present guidance. At the end of presentation, students peer-review/ critique each other's draft data analysis plans.

*Second hour:*

Each student revises their draft data analysis plans and dummy tables based on their learning from facilitator presentations and peer-review comments.

Support students to put their research thoughts into a plan of action in such a way that they can achieve their study objectives.

## Session 13. Approaching Data Analysis | 135 minutes

The choice of appropriate statistical methods for quantitative research data analysis is mainly driven by the type of data variables, research question, and study design. This session provides an introductory overview of the main types of statistical tests and their application in quantitative research studies. Equip students to determine the correct statistical test for different types of quantitative data and research questions.

### Outcomes

By the end of this session, students can:

- Describe commonly used statistical tests.
- Identify the correct statistical test to be used for a specific research question and type of data.

### Preparation

*As facilitator*

Create or source an introductory presentation.

*Students*

In preparation for this session, each student must ensure that they are able to:

- Describe different types of quantitative research questions and study designs.
- Understand different measurement scales of data (nominal, ordinal, interval, ratio) for quantitative analysis.
- Describe the basic statistical method of hypothesis testing and interpret p-values.

### References

- Nayak, B.K. and Hazra, A. (2011). [How to choose the right statistical test?](http://www.ncbi.nlm.nih.gov/pmc/articles/PMC3116565/) Indian J Ophthalmol. 59(2): p. 85-6.
- Zinsmeister, A.R., and Connor, J.T., (2008). [Ten Common Statistical Errors and How to Avoid Them](#). Am J Gastroenterol. 103(2): p. 262-266. (Request access.)

### Additional reading/ viewing

- [What statistical analysis should I use?](#) Institute for digital research and education, UCLA.
- [Selecting statistics](#). Online statistical advisor. Web center for social research methods.
- [Introduction to Statistics: Levels of Measurement](#). Youtube video.
- [Choosing which statistical test to use](#). Youtube video.

### Steps

Combine presentation and hands-on exercises on:

The statistical method. (45 minutes)

Criteria for choosing the appropriate statistical test. (60 minutes)

Common statistical errors. (30 minutes)

## Session 14. T-Tests and Chi-squared Tests | 2 hours

Strengthen PhD students' skills to apply inferential methods of t-tests and chi-squared tests to compare continuous and categorical outcomes respectively between an exposure variable with two levels. Emphasise the link between hypothesis tests and measures of effect and the corresponding confidence intervals.

### Outcomes

By the end of this session, students can:

- Carry out t-tests for two independent samples and for paired samples in Stata.
- Explain the link between hypothesis testing and confidence intervals.
- Carry out a chi-squared test and find measures of association in Stata.
- Distinguish between confounding and effect modification in an observational study.

### Preparation

*As facilitator*

Create or source an introductory presentation.

Source a data set and write tasks for the students' exercise.

### Reference

Germán Rodríguez. (2023). [Stata Tutorial](#). Princeton University.

### Steps

Allocate one hour to introducing the session, to cover:

- The concepts of sample, sampling variability and the standard error of the mean.
- The concept of a 95% confidence interval and how this can be estimated
- The concept of a hypothesis test and how this can be carried out for a single sample and its link with confidence interval.
- Compare population means based on data from two independent samples
- Tests for association between categorical variables including 2X2 and larger contingency tables. Confounding and effect modification.

For the second hour, groups analyse the data set your provided to generate appropriate analysis for different formats and data types. Each group submits their analysis to a different group for peer assessment and feedback.

## Session 15. Linear Regression and Residual Analysis | 2 hours

From this session, students gain the knowledge and practical skills to assess the suitability of the linear regression model to the data at hand with a focus on residual analysis. They learn how to use linear regression plots to assess the model adequacy.

### Outcomes

By the end of this session, students can:

- Fit a linear regression model correctly to data.
- Apply the correct order of steps to determine whether linear regression is a suitable model for a set of data based on residual analysis.

### Preparation

*References*

Montgomery, D. C., Peck, E. A5, and Vining, G. G. (2012), Introduction to Linear Regression Analysis, 4th Edition, Wiley, New York.

Penn State. (2018). [Residuals](#). Stat462

Penn State. (2018). [Residuals vs Order Plot](#). Stat462

### Steps

In the first hour, give a presentation on linear regression and residual analysis, with a focus on its suitability assessment. In the second hour, introduce a practical session, to investigate residual plots, interaction, and factors individually associated with the outcome. Finally, fit a multiple linear regression model to data and carry out assessment of model suitability.

## Session 16. When to Use Logistical Regression Analysis? | 2 hours

Logistic regression analysis is used to examine the relationship between independent variable(s) (categorical or continuous) and a categorical dependent variable. This session equips students with knowledge and practical skills related to the use of logistic regression analysis, with a focus on the binary logistic regression model. Students learn when to use logistic regression analysis and how to use it in data analysis and interpretation as well as in assessing confounding and interaction, using Stata.

### Outcomes

By the end of this session, students can:

- Determine when to use logistic regression analysis.
- Explain the concept of logistic regression and describe its application correctly.
- Explain how to build a logistic regression model.
- Apply logistic regression to assess confounding and interaction.

### Preparation

*As facilitator*

Ensure that the learning dataset is loaded on students' computers.

### References

Michael, P. L. (2008). [Logistic Regression](#). *Circulation*, 117:2395-2399.

Stoltzfus, J.C. (2011). [Logistic Regression: A Brief Primer](#). *Academic Emergency Medicine*, 18: 1099-1104.

### Assessment

Peer review of group exercise.

### Steps

In the first hour, give a presentation on the logistic regression model and its use. In the second hour, introduce the practical session. With the dataset at hand, students work in small groups to build and fit a logistic regression model. They investigate confounding, interaction, and factors associated with the outcome.

Each group submits their work to peers for assessment and feedback. You and co-facilitator/s facilitator may give collective feedback based on selected group work.

## Session 17. Selection of Predictors in Regression Models | 1 hour

The aim of this session is to transfer and strengthen knowledge, skills, and strategies to improve regression models. These include transforming both the outcome (in linear regression models) and continuous exposures (in all models) and selecting the variables to include in the final model.

Students consider three different situations in which regression models are used:

- When the overall aim is prediction.
- When the aim is to evaluate a predictor of primary interest.
- When the aim is to identify important independent predictors of an outcome.

### Outcomes

By the end of this session, students can:

- Check the assumptions and where necessary carry out a transformation in linear regression.
- Check for linear trend effects in predictors
- Understand how fractional polynomial models can be used to improve prediction.

- Apply strategies for selecting predictors in the three different situations in which regression models are fitted.

### Preparation

*As facilitator*

Create or source a presentation to explain these topics.

Prepare the hands-on exercise for students to complete in groups.

*Students read*

- Deegan, J. (1976). [The Consequences Of Model Misspecification In Regression Analysis](#). *Multivariate Behav Res.* 11(2):237-48.
- Vatcheva, K.P., Lee, M., McCormick, J.B., Rahbar, M.H., (2016). [Multicollinearity in Regression Analyses Conducted in Epidemiologic Studies](#). *Epidemiology (Sunnyvale)*. 6(2):227.

*Additional reading*

- Chowdhury, M. and Turin, T.C. (2020). [Variable selection strategies and its importance in clinical prediction modelling](#). *Fam Med Community Health*. 8(1):e000262. doi:10.1136/fmch-2019-000262
- Morozova, O., Levina, O., Uusküla, A., Heimer, R., (2015). [Comparison of subset selection methods in linear regression in the context of health-related quality of life and substance abuse in Russia](#). *BMC Med Res Methodol*. 30;15:71.
- Heinze, G., Wallisch, C., Dunkler, D. (2018). [Variable selection – A review and recommendations for the practicing statistician](#). *Biom J*. 60(3):431-449.
- Genell, A., Nemes, S., Steineck, G., Dickman, P.W. (2010). [Model selection in medical research: a simulation study comparing Bayesian model averaging and stepwise regression](#). *BMC Med Res Methodol*. 6;10:108.
- Smith, G. (2018). [Step away from stepwise](#). *J Big Data* 5, 32.
- Ratner, B. (2010). [Variable selection methods in regression: Ignorable problem, outing notable solution](#). *J Target Meas Anal Mark* 18, 65–75.

### Assessment

Participation in session: 20% (Individual)

Group exercise: 80% (Group)

### Steps

Give a lecture on the selection of predictors in multiple linear regression analysis. (30 minutes) In the practical session, students work in groups to fit a multiple linear regression model to research data and apply variable selection strategies.

## Session 18. Spatial Analysis | 2 hours

Several health phenomena exhibit an important spatial dimension. Approaches or methods that ignore the spatial dimension are prone to skewed or inaccurate results. Fortunately, with the advent of Geographic Information Systems (GIS), geo-referenced population and health data are increasingly available, and consideration of the spatial component sheds light on most public-health issues.

In this session, introduce the PhD students to spatial analytical techniques and the importance of accounting for spatial autocorrelation when analysing spatial referenced data sets. Support students with georeferenced data and a spatial component in their research to include it during data analysis.

### Outcomes

By the end of this session, students can:

- Determine the essential features of spatially referenced data, detecting spatial clustering/ autocorrelation.
- Describe types and sources of spatial data pertaining to public health.
- Presentation of spatial data using different formats using Stata or R.
- Describe methods for analysing point referenced and areal data sets.

### Preparation

### *As facilitator*

Share relevant resources, including a dataset, with students within a reasonable time period prior to the session.

Develop or source a presentation.

Prepare the practical demonstrations and a group exercise.

### *References*

- De Smith, M. J., Goodchild, M. F., & Longley, P. (2021). *Geospatial Analysis: a Comprehensive Guide to Principles, Techniques and Software Tools*.
- Leitner, M. (ed). (2013). *Cartography and Geographic Information Science*.
- Barry, J.K. (2013). *Beyond Mapping Compilation Series*.

### **Steps**

Combine slides with Stata software-based demonstrations to introduce and explain the techniques. Then, in small groups students work with the dataset you provide to generate appropriate analyses for different formats and data types.

Each group submits their analysis to another different group for peer assessment and feedback.