Producing Data

Madiba Hudson-Quansah

Contents

Chapter 1

Introduction

Definition 1.0.1: Sample

A subset of a larger population

In order to reliably draw conclusions about a population, we must ensure that our chosen sample is truly representative of the population.

Definition 1.0.2: Sampling Bias

A sample that is not representative of a population due to systematic under or over estimation of the values of the variable of interest.

Chapter 2

Module 6: Sampling

2.1 Sampling Methods

2.1.1 Simple Random Sampling

Definition 2.1.1: Simple Random Sampling

The simplest probability sampling method. Each member of the population has an equal chance of being selected.

2.1.2 Cluster Sampling

Definition 2.1.2: Cluster Sampling

Used when the population is naturally divided into groups / clusters. A simple random sample of clusters is selected and we use all the individuals in the selected clusters as our sample.

2.1.3 Stratified Sampling

Definition 2.1.3: Stratified Sampling

Used when the population is naturally divide into sub populations / strata, e.g. gender. A simple random sample is selected from each stratum and the results are combined to form the sample.

Chapter 3

Module 7: Designing Studies

3.1 Introduction

Definition 3.1.1: Observational Study

The variable(s) of interest are recorded as they naturally occur, with no interference from the researchers conducting the study

Definition 3.1.2: Sample Survey

A type of observational study where subjects report variables' values themselves, frequently giving their opinions.

Definition 3.1.3: Experiment

The researchers interfere with the variable(s) of interest, assigning the values of the independent variable to the subjects.

3.2 Observational Study

There are two types of observational studies:

- Prospective
- Retrospective

3.3 Causation and Observational Studies

As a result of the nature of observational studies they come with one glaring weakness: differing factors among subjects could be lurking variables, affecting the results of the study. To contend with this observational studies often have to control of the effects of lurking variables by separating each study group by similar possible lurking variables. But even with this attempt at control it is impossible to account for a near unlimited amount of potential lurking variables, so the claims of a causation based on an observational study can never be 100% accurate.

3.4 Causation and Experiments

Definition 3.4.1: Factor

A variable whose levels are controlled by the experimenter / Independent variable

Definition 3.4.2: Treatments / *ttt*

Different imposed values of the factor

Definition 3.4.3: Control Group

A group of subjects that are not given any treatment

Definition 3.4.4: Subjects

Individuals on which the experiment is performed

Definition 3.4.5: Treatment Group

A group of subjects that are given a treatment

3.4.1 Randomized Controlled Experiments

Definition 3.4.6: Randomized Controlled Experiment

An experiment where subjects are randomly assigned to treatment groups

Under a randomized controlled experiment, the only difference between the treatment and control groups should be the treatment itself. This allows us to draw conclusions about the effects of the treatment on the subjects.

3.4.2 Blind and Double-Blind Experiments

Definition 3.4.7: Blind Experiment

An experiment where the subjects do not know which treatment they are receiving

Definition 3.4.8: Double-Blind Experiment

An experiment where neither the subjects nor the researchers know which treatment the subjects are receiving

3.4.3 Modifications to Randomization

In some cases the design of an experiment may be enhanced by relaxing the requirement of total randomization and blocking the subjects first.

Definition 3.4.9: Block

A group of subjects that are similar in ways that are expected to affect the response to the treatments

Definition 3.4.10: Matched Pairs

A type of blocking where each subject receives both treatments in a random order