STAT 103
Statistical Thinking

Exam 1 Review

Professor Mengshi Zhou

Chapter 1 Individuals and Variables

Goal: Identify the individuals and the variables in a study(dataset)

Goal: Distinguish categorical variable and numerical variables

Individuals are objects described by the (people, animals, or things)

Variables any characteristics about individuals

Categorical variables place an individual in to one of serval groups (sex, major).

Numerical variables take numerical values for which arithmetic operations make sense (weight, height).

Chapter 1 Observational studies and Experiments

Goal: Know the definition and goal of observational studies and Experiments, and the difference between an observational study and an experiment

- **Observational studies** try to gather information without disturbing the scene they are observing.

- **Experiments** assignment treatment to individuals in order to see how they respond. The goal of an **experiment** is usually to learn whether some **treatment** actually causes a certain response.

Chapter 1 Population and Sample

Goal: Identify the population and sample in a study

Population: The entire set of possible individuals about which we want information

Sample: A subset of the population from which data are collected

Chapter 1 Sample Survey and Census

Goal: Know the definition of sample survey and census, determine if an observational study is a sample survey

- A **census** is an attempt to collect data from every member of the population.
- A **sample survey** is a collection of data from a subset of the population chosen by the researcher.

Chapter 2 Biased Samples

Goal: Be able to recognize bad sampling methods and know why they cause bias

- A **convenience sample** select of whichever individuals are easiest to reach
- A **convenience sample** is biased because it favors those who are easy to access y researchers

Chapter 2 Biased Samples

Goal: Be able to recognize bad sampling methods and know why they cause bias

- In a **voluntary response sample** individuals chooses themselves by responding to a general appeal (write-in or call-in opinion polls).
- A voluntary response sample is biased because people who feel strongly about a topic are more likely to respond to voice their feelings.

Chapter 2 Simple Random Sample

Goal: Know the definition of a simple random sample and SRS can avoid bias

- The deliberate use of chance in producing data is one of the big ideas of statistics. **Random samples** use chance to choose a sample, thus avoiding bias due to personal choice.
- The basic type of random sample is the **simple random sample**, which gives all samples of the same size the same chance to be the sample we actually choose.

Chapter 2 Simple Random Sample

Goal: Know the steps to generate a simple random sample

To select a simple random sample:

1. Label each population element with as few digits as possible, making sure each label is the same length.

2. Use the table or software to select random numbers.

Chapter 3 Proportion

Goal: Calculate sample proportion

Proportion

$$Proportion = \frac{Number\ in\ the\ category}{Total\ number}$$

The symbol for a sample proportion is $\hat{p}$ and is read as p-hat. The symbol for a population proportion is p.

Chapter 3 Parameter and Statistic

Goal: Identify of parameter and statistic in a study

Parameter: A number that describes a population. It is a fixed number, but in practice we don't know the actual value of this number.

Statistic: A number that describes a sample. This is a known value when we have taken a sample, but it can change from sample to sample. It is often used to estimate an unknown parameter.

Chapter 3 Bias and Variability

Goal: Understand Bias and Variability and know how to reduce Bias and Variation

Bias: When the design of a statistical study systematically favors certain outcomes. Can be reduced using simple random sample

Variability: Describes how spread out the values of the sample statistic are when we take many samples. Can be reduced using larger sample size

Chapter 3 Margin of Error

Goal: Know what "margin of error" means

margin of error: margin of error says how close the sample statistic lies to the population parameter **(Only measure variability).**

Chapter 3 Proportion and Quick Method

Goal: Know the quick method for finding the margin of error AND when the method is appropriate

We can roughly approximate the margin of error of population proportion for 95% confidence based on a **simple random sample** of size *n* by the formula $1/\sqrt{n}$.

95% confidence interval = $\hat{p} \pm \frac{1}{\sqrt{n}}$

Chapter 3 Confidence statement

Goal: Know the 2 parts of a confidence statement and how to interpret a confidence statement

confidence statement: We are 95% confidence that the **true parameter** is between **statistics** plus or minus **margin error**

Most often, this margin of error is **for 95% confidence level**. **That is, if we chose many samples, the truth about the population would be within the margin of error 95% of the time.**

Chapter 4 Sampling Errors

- **Goal: Know the types of sampling errors**

- **Sampling errors** come from the act of choosing a sample. **Random sampling error, bad sampling methods** and **undercoverage** are common types of sampling error.
- **Undercoverage** occurs when some members of the population are left out of the **sampling frame**, the list from which the sample is actually chosen.

Chapter 4 Non-Sampling Errors

Goal: Know the types of nonsampling errors

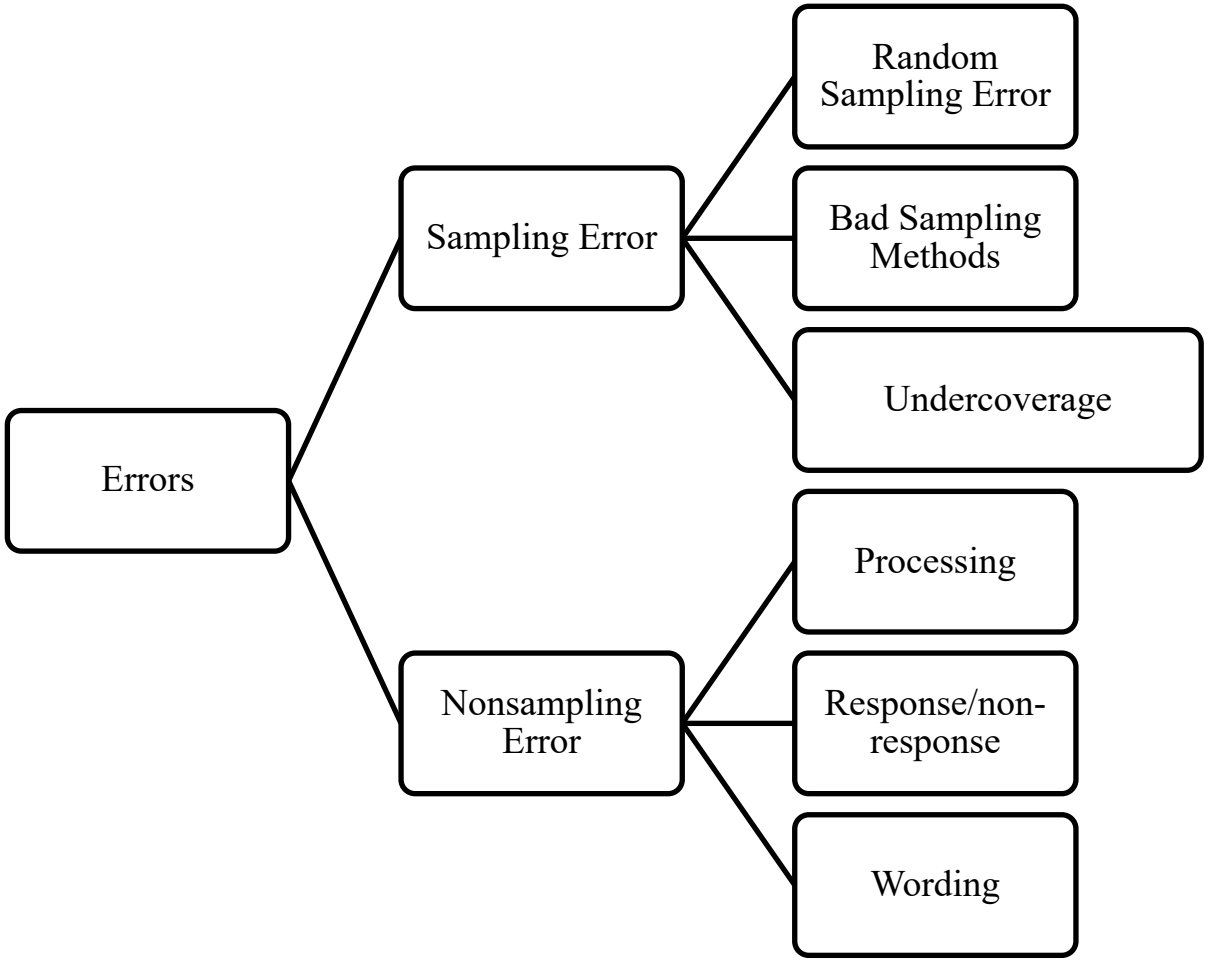
The most serious errors in most careful surveys, however, are **nonsampling errors**. These have nothing to do with choosing a sample—they are present even in a census.

The single biggest problem is **nonresponse**: subjects can't be contacted or refuse to answer.

Mistakes in handling the data (**processing errors**) and incorrect answers by respondents (**response errors**) are other examples of nonsampling errors.

Finally, the exact **wording of questions** has a big influence on the answers.

Chapter 4 Distinguish two types of errors



Margin of Error only covers random sampling error!

Chapter 4 Stratified Random Sample

Goal: Know what a stratified sample is and how to identify strata

Stratified random sample: A sample in which the sampling frame is first divided into various **strata** (groups). A simple random sample is then taken in each of these **strata**, with those selected combined to form the complete sample:

Chapter 5 response variables, explanatory variables, treatments

Goal: Identify the response variables, explanatory variables, and treatments in an experiment.

explanatory variable A variable that we think explains or causes changes in the response variable.

response variable A variable that measures an outcome or result of a study.

Treatments: Any specific experimental condition that is applied to the subjects. If an experiment has several explanatory variables, this is a combination of specific values of these variables.

Chapter 5 lurking variables and confounding

Goal: Be able to identify a lurking variable that is confounding a study

Lurking variable = has important effect on response variable but is NOT an explanatory variable (sometimes called a confounding variable or a third variable)

Two variables are **CONFOUNDED** when their effects on a response variable cannot be distinguished from each other.

Chapter 5 Randomized Comparative Experiment

Goal: Know the definition of a Randomized Comparative Experiment

Randomization--randomly assigning cases to different levels of the explanatory variable (e.g., different treatment groups).

An experiment that involves randomization may be referred to as a **randomized comparative experiment.**

Chapter 5 Logics for good experiments

Control Group--A group receive no treatment or a placebo

Placebo Group--A group that receives placebo (e.g., a sugar pill in a medication study)

If we want to compare multiple treatments, we do not need placebo group

Chapter 5 Logics for good experiments

Goal: Fully understand the logic of experimental design

Identify control groups, randomization, placebos, blinding, and use enough subjects in experiments and explain why each is used.

Chapter 5 Logics for good experiments

Control the effects of lurking variables on the response, most simply by comparing two or more treatments.

Randomize – use impersonal chance to assign subjects to treatments.

Double-Blind –Research study in which neither the participants nor the researchers interacting with them know which cases have been assigned to which treatment groups

Use enough subjects in each group to reduce chance variation in the results.

Make sure experimental units **represent** the others not in the experiments