


浙江大学爱丁堡大学联合学院

ZJU-UoE Institute

Sampling

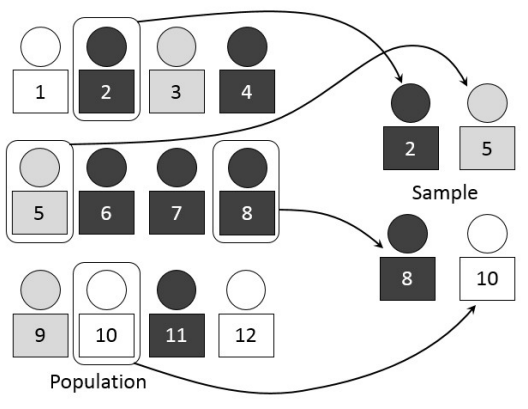
Nicola Romanò - nicola.romano@ed.ac.uk

Based on slides by Duncan McGregor

Raise your hand if you love statistics!

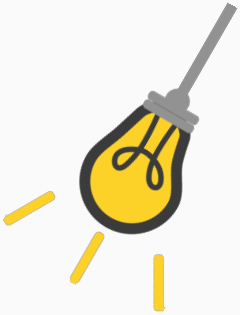
This lecture is about...

How to **sample** from a **population** and avoid common types of **sampling bias**.



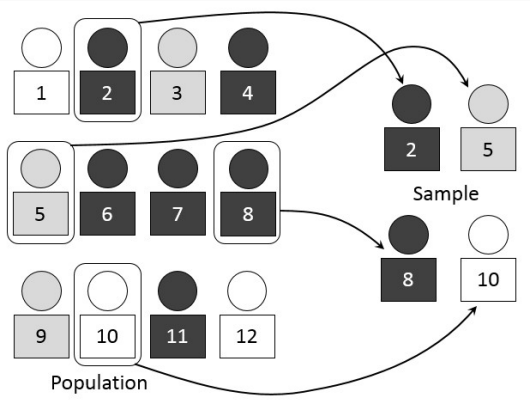
Learning objectives

- Explain the relationship between a population and a sample
- Explain the concept of sampling bias
- Give examples of sampling biases that can occur
- Design data collection procedures that avoid sampling bias

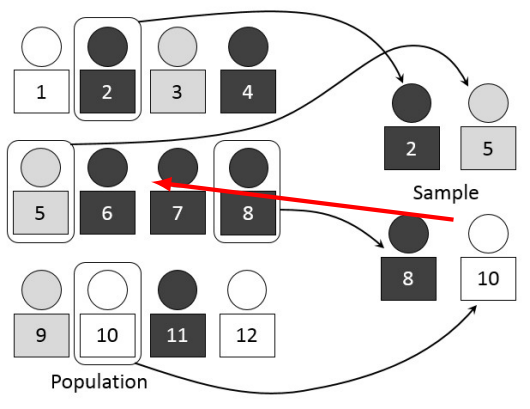


Samples and populations

We can only access the population by taking samples…



. . . but we want to draw conclusions about the population



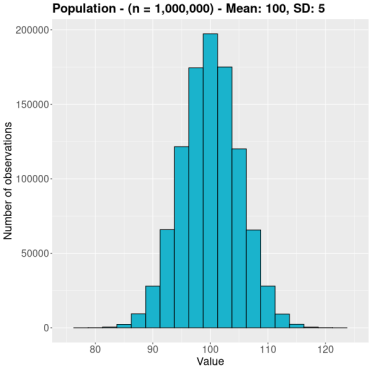
Inferential statistics

We need to draw conclusions about the population from knowledge of a sample. We do this using **inferential statistics**.

EXAMPLES	
Sample estimate	**Population parameter**
BMI of student volunteers for a study	BMI of all undergraduate students
Neuronal activity of 10 rats after a learning task	
	Response of stroke patients to medication XYZ

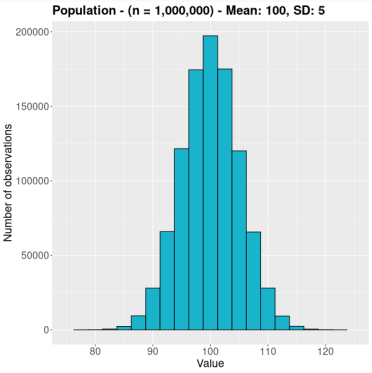
Sampling example

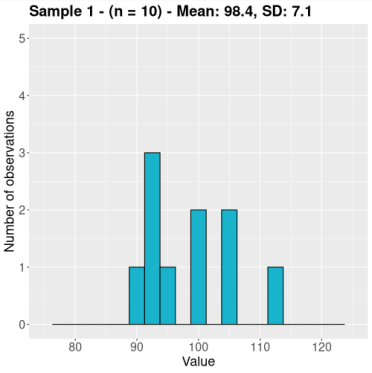
Let’s sample from a population. What features of the sample will be similar/different from the population?



Sampling example

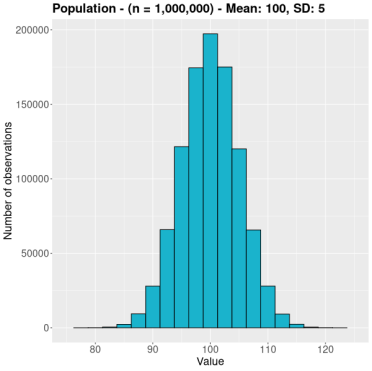
Let's sample from a population. What features of the sample will be similar/different from the population?

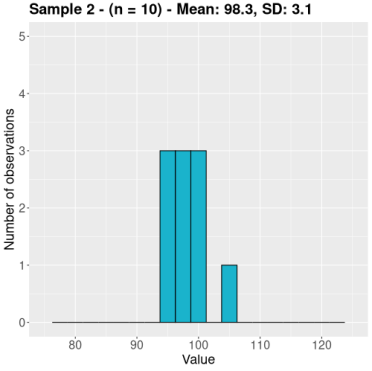




Sampling example

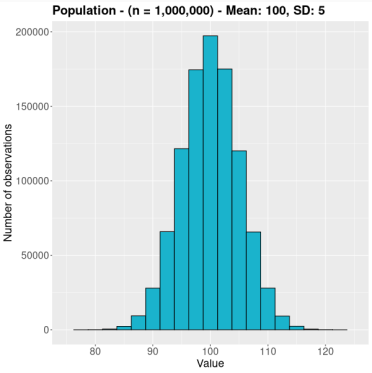
Let's sample from a population. What features of the sample will be similar/different from the population?

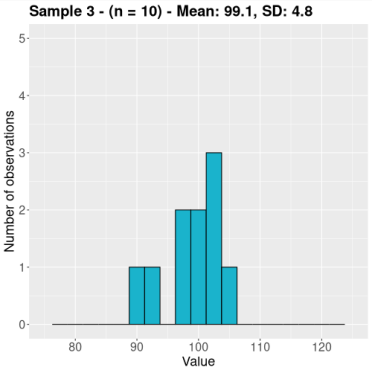




Sampling example

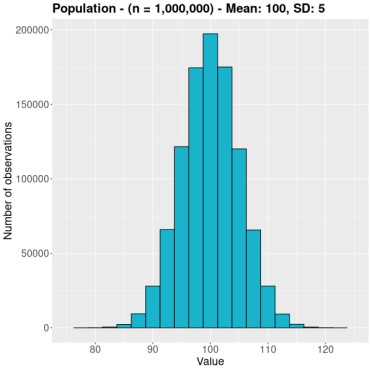
Let's sample from a population. What features of the sample will be similar/different from the population?

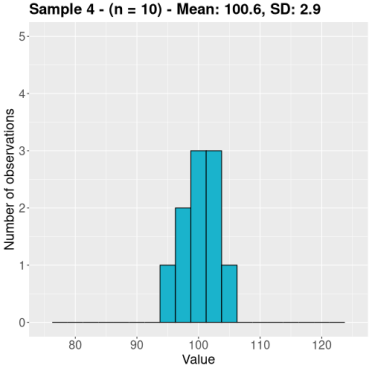




Sampling example

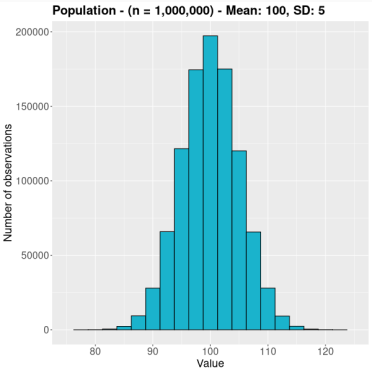
Let's sample from a population. What features of the sample will be similar/different from the population?

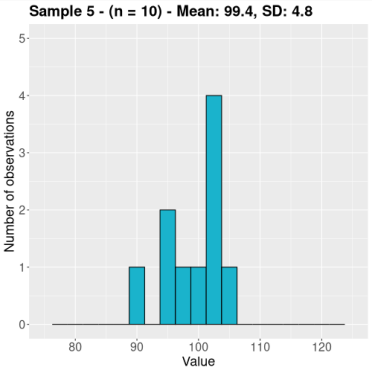




Sampling example

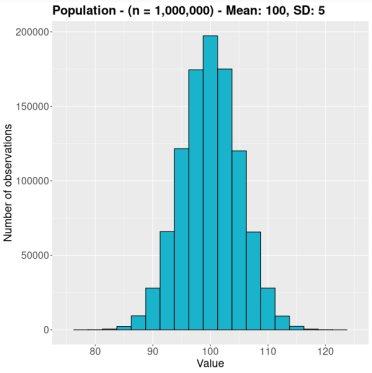
Let's sample from a population. What features of the sample will be similar/different from the population?

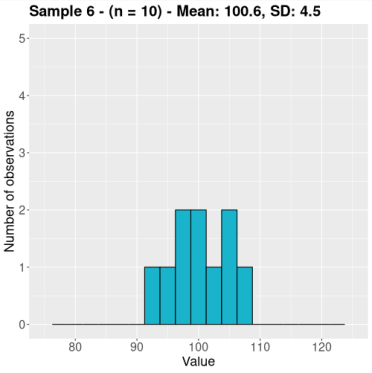




Sampling example

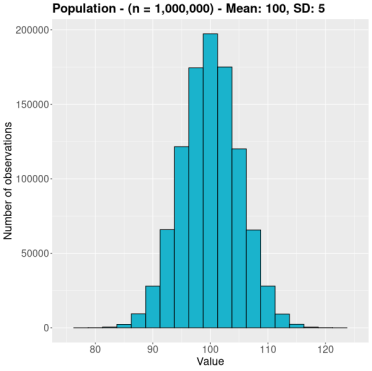
Let's sample from a population. What features of the sample will be similar/different from the population?

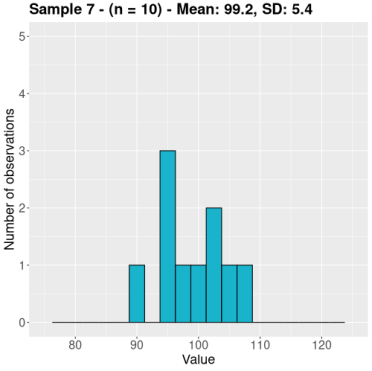




Sampling example

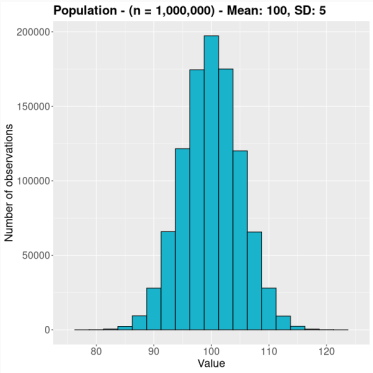
Let's sample from a population. What features of the sample will be similar/different from the population?

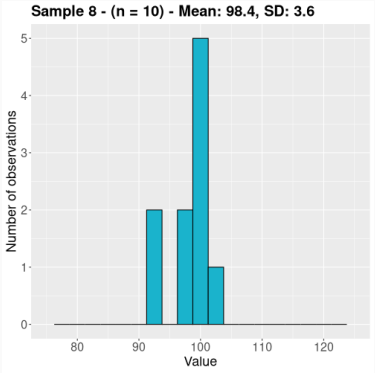




Sampling example

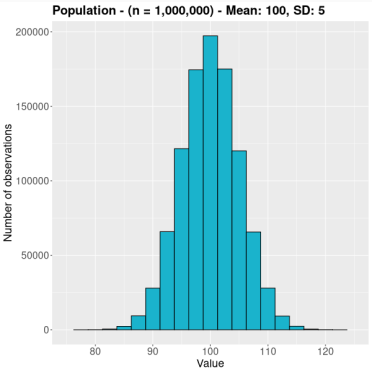
Let's sample from a population. What features of the sample will be similar/different from the population?

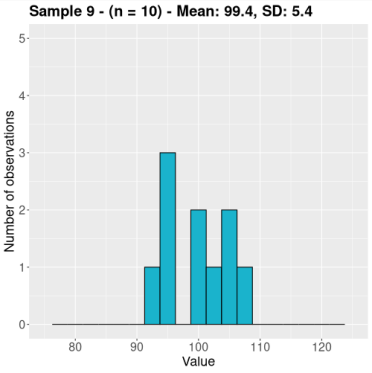




Sampling example

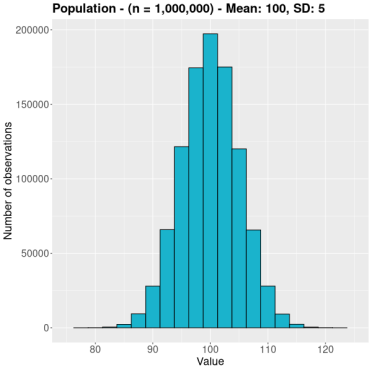
Let's sample from a population. What features of the sample will be similar/different from the population?

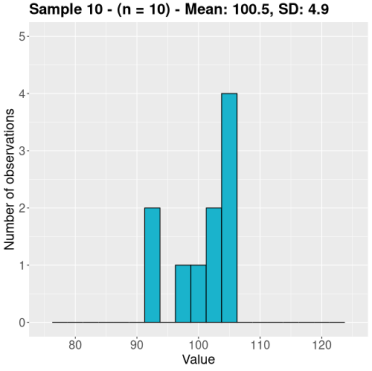




Sampling example

Let's sample from a population. What features of the sample will be similar/different from the population?





Sampling error

Sampling error

1. Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
2. The difference between sample and population is called *sampling error* (“error” here means “difference from the true population”, not “mistake”)
3. How does the *sampling error* depend on *sample size*?

Sampling error

1. Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
2. The difference between sample and population is called *sampling error* (“error” here means “difference from the true population”, not “mistake”)
3. How does the *sampling error* depend on *sample size*? **More about this next week!**

Sampling bias

Sampling bias

Ideally, a sample is random and each individual in the population has an equal probability of being sampled. If this is not the case (i.e. if some parts of the population have a higher chance of being sampled than others), there is a **sampling bias**.

Sampling bias - example

1948 presidential election

Sampling bias - example

1948 presidential election

Thomas E. Dewey against Harry S. Truman

Sampling bias - example

1948 presidential election

Thomas E. Dewey against Harry S. Truman

Chicago Daily Tribune wanted to predict outcome of the vote.

Sampling bias - example

1948 presidential election

Thomas E. Dewey against Harry S. Truman

Chicago Daily Tribune wanted to predict outcome of the vote.

Telephone poll to find out about voting intentions

Check your understanding!

What is the difference between sampling error and sampling bias?

Recognising and avoiding sampling bias

Avoiding sampling bias

Questions to ask

- Who is invited to sign up for the study?
- Where are we looking for our sample?
- How is the sample collected?
- Who/what is included, who/what is excluded?
- How are treatment and control group determined?
- Who might drop out, and why?

Recognising sampling bias

Prof. Getafix wants to study the prevalence and frequency of drug use among young people in Scotland. Here are possible ways of collecting a sample:

1. Calling a local hospital and asking them for records about young people who have been admitted because of a drug-related health problem.

Recognising sampling bias

Prof. Getafix wants to study the prevalence and frequency of drug use among young people in Scotland. Here are possible ways of collecting a sample:

1. Calling a local hospital and asking them for records about young people who have been admitted because of a drug-related health problem.
2. Standing on a busy square in the city centre, stopping every young person who comes by and saying: “Excuse me, I am studying drug use. I need volunteers for my study. Do you take drugs?”

Recognising sampling bias

Prof. Getafix wants to study the prevalence and frequency of drug use among young people in Scotland. Here are possible ways of collecting a sample:

1. Calling a local hospital and asking them for records about young people who have been admitted because of a drug-related health problem.
2. Standing on a busy square in the city centre, stopping every young person who comes by and saying: “Excuse me, I am studying drug use. I need volunteers for my study. Do you take drugs?”
3. Putting up a notice on the University Library noticeboard asking for volunteers to fill out an anonymous online survey about drug use and containing a QR code that people can use to complete the survey.

Recognising sampling bias

Prof. Getafix wants to study the prevalence and frequency of drug use among young people in Scotland. Here are possible ways of collecting a sample:

1. Calling a local hospital and asking them for records about young people who have been admitted because of a drug-related health problem.
2. Standing on a busy square in the city centre, stopping every young person who comes by and saying: “Excuse me, I am studying drug use. I need volunteers for my study. Do you take drugs?”
3. Putting up a notice on the University Library noticeboard asking for volunteers to fill out an anonymous online survey about drug use and containing a QR code that people can use to complete the survey.

Recognising sampling bias

Prof. Getafix wants to study the prevalence and frequency of drug use among young people in Scotland. Here are possible ways of collecting a sample:

1. Calling a local hospital and asking them for records about young people who have been admitted because of a drug-related health problem.
2. Standing on a busy square in the city centre, stopping every young person who comes by and saying: “Excuse me, I am studying drug use. I need volunteers for my study. Do you take drugs?”
3. Putting up a notice on the University Library noticeboard asking for volunteers to fill out an anonymous online survey about drug use and containing a QR code that people can use to complete the survey.

Which of these methods may result in sampling bias, and why? Which method would you choose?

Any questions?

You now should be able to:

- Explain the relationship between a population and a sample
- Explain the concept of sampling bias
- Give examples of sampling biases that can occur
- Design data collection procedures that avoid sampling bias

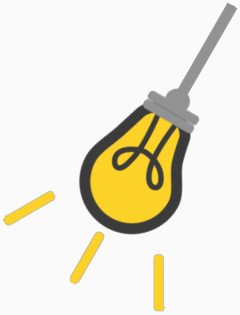


Image credits

- Student raising her hand. By Thomas Taylor Hammond (1920-1993) CC BY-SA 4.0, via Wikimedia Commons
- Random sample. By Dan Kernler CC BY-SA 4.0, from Wikimedia Commons
- Histograms of a population and randomly drawn samples from it. My own work using RStudio, 2023. CC BY-SA 4.0 (code on Learn!).
- Dice by OFour via Flickr - CC BY-NC 2.0.
- President Truman holding a copy of the Chicago Daily Tribune. By Associated Press photo by Byron Rollins, Fair use, via Wikimedia Commons.
- Two students discussing. By Yuuki Guzman (Okinawa Institute of Science and Technology), 2016.