



浙江大学爱丁堡大学联合学院
ZJU-UoE Institute

Sampling

ADS 2, Lecture 2

Melanie Stefan - melanie.stefan@ed.ac.uk

Semester 1, 2019/20

Before we start . . .

Before we start ...

We need class reps!



Raise your hand if you love statistics



How is this “study” biased?



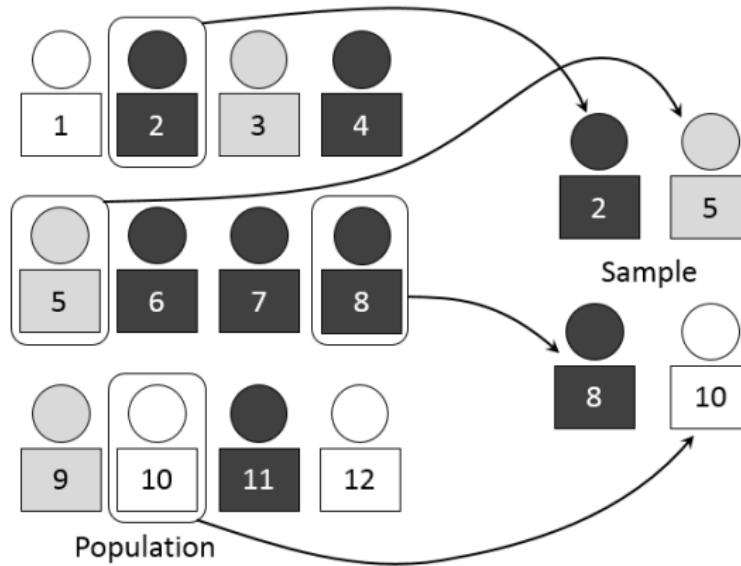
How is this “study” biased?



- Leading question, image, and body language
- Raising hand is more difficult than not raising hand
- Survey is non-anonymous
- Power imbalance/conflict of interest
- Is this a representative sample?

This lecture is about . . .

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



Learning Objectives

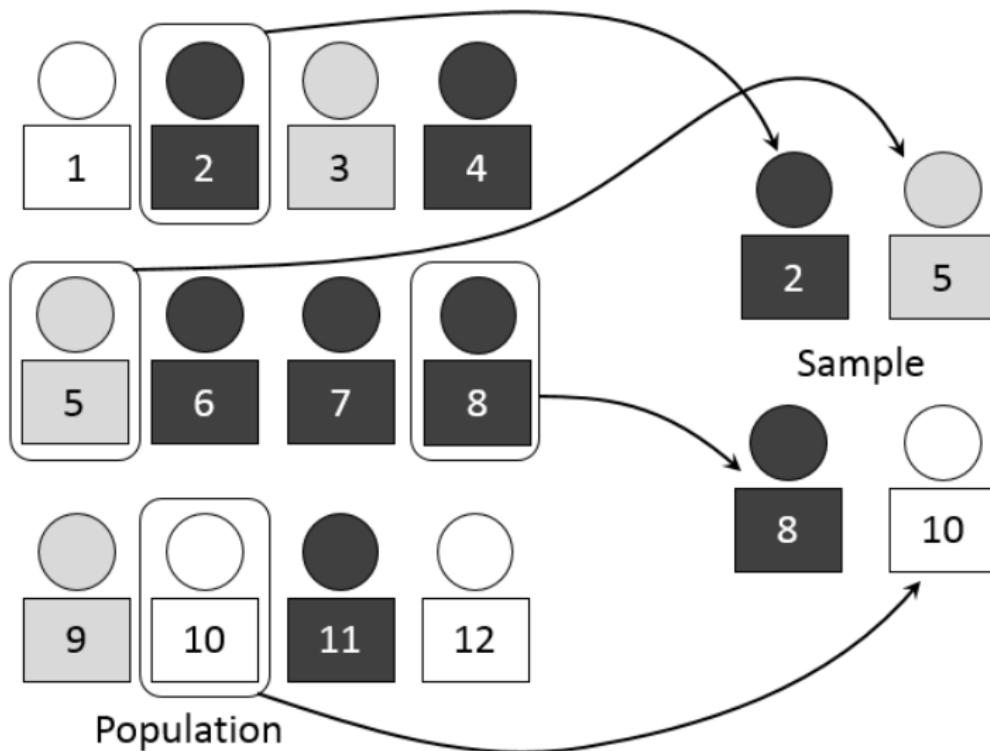
After this lecture, you should be able to ...

- 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

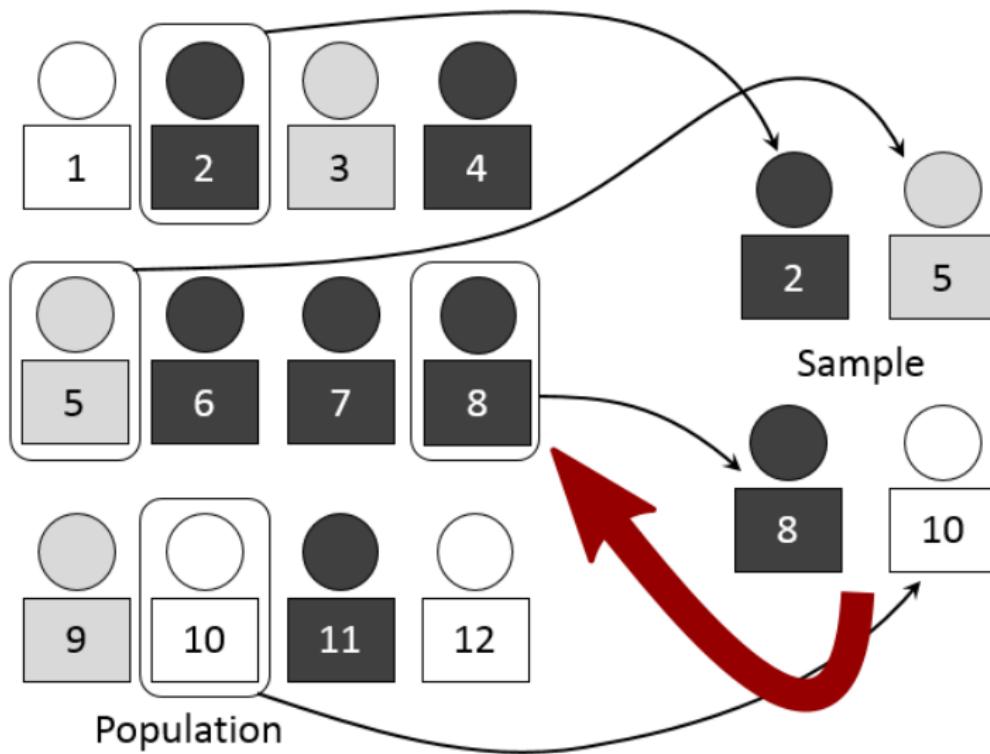
Outline

- 1 Samples and populations
- 2 Sampling error
- 3 Sampling bias
- 4 Recognising and avoiding sampling bias

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

Inferential statistics

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

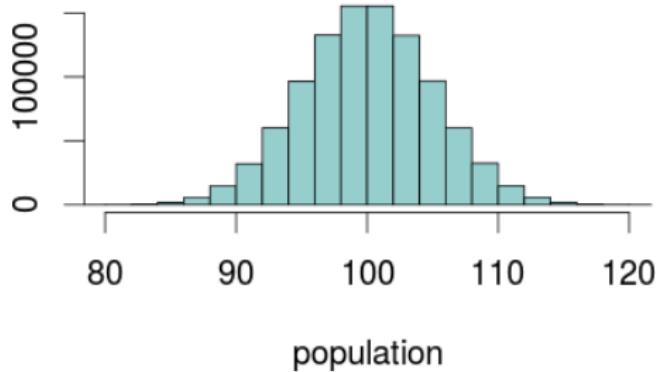
Examples

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	Neuronal activity of rats after a learning task
Response to stroke patients in one clinical trial to medication XYZ	Response of stroke patients to medication XYZ

Sampling example

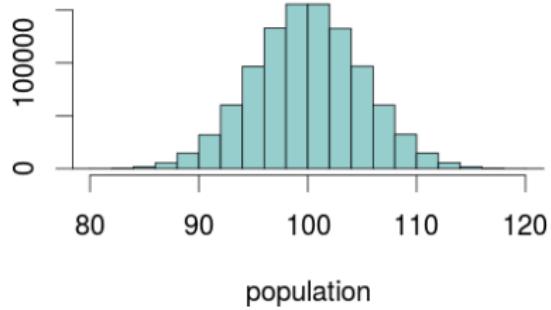
Let's sample from a population. What features of the sample will be similar/different from the population?

mean = 100, sd = 5

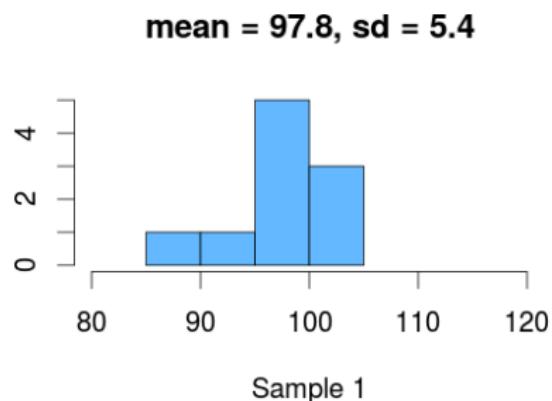
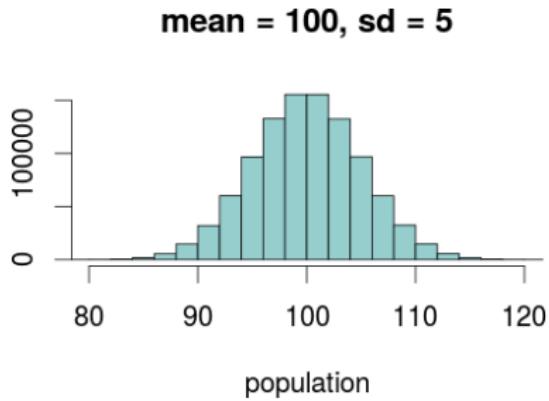


Sampling example

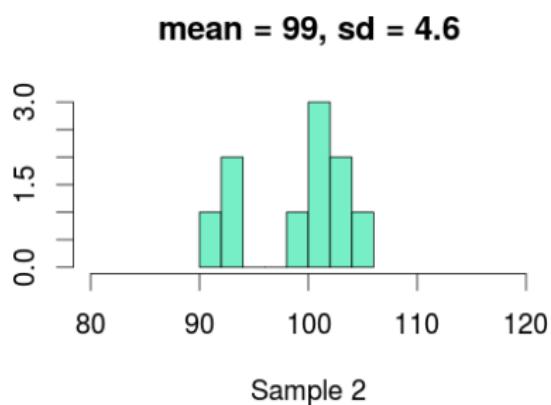
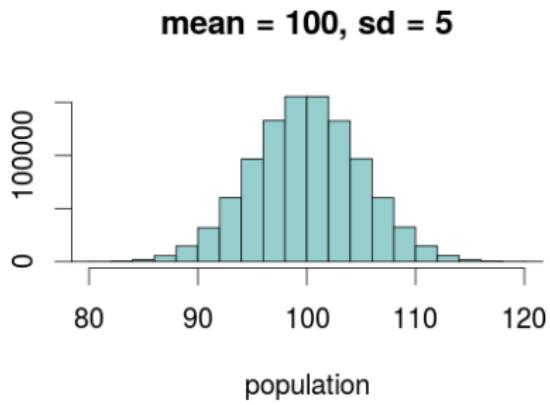
mean = 100, sd = 5



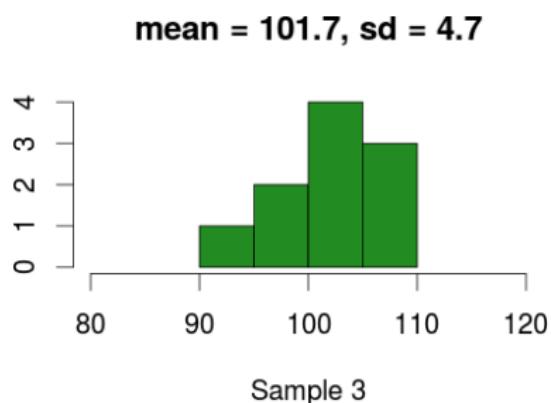
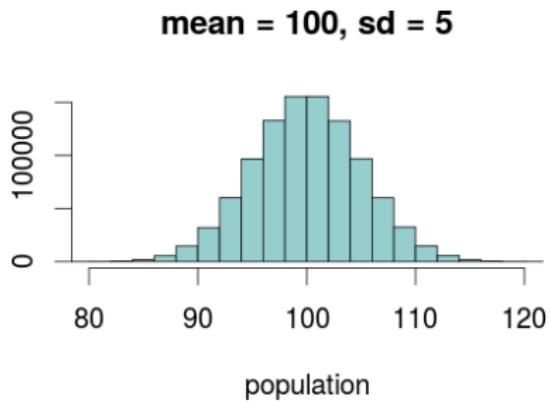
Sampling example



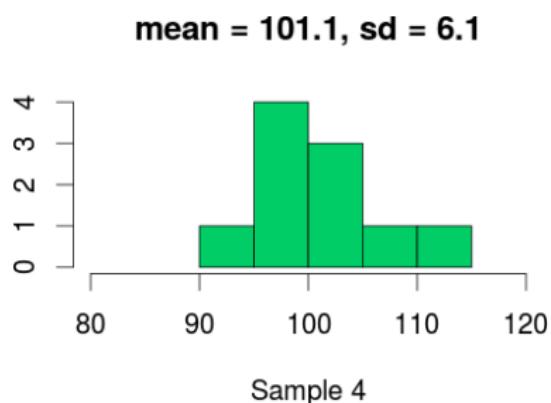
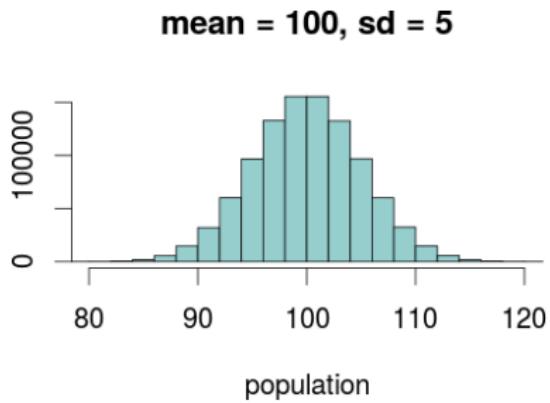
Sampling example



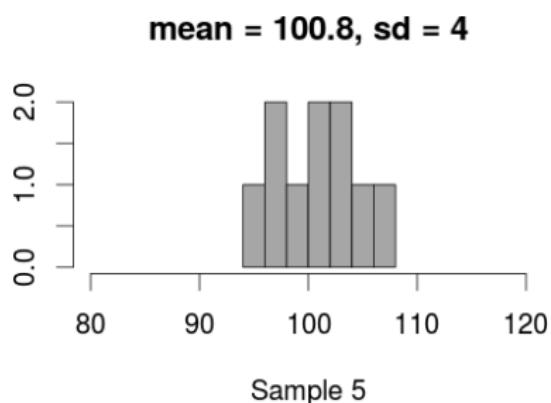
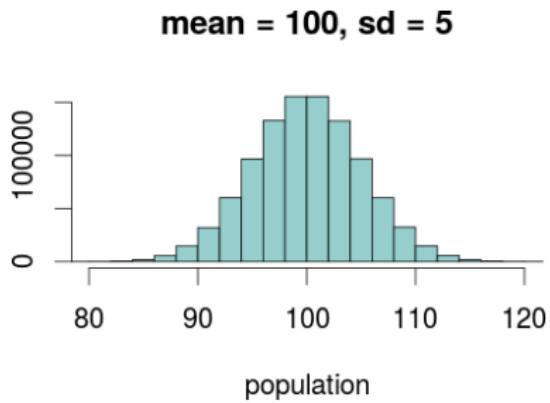
Sampling example



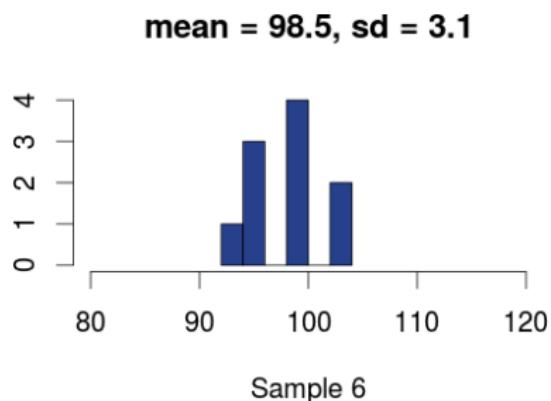
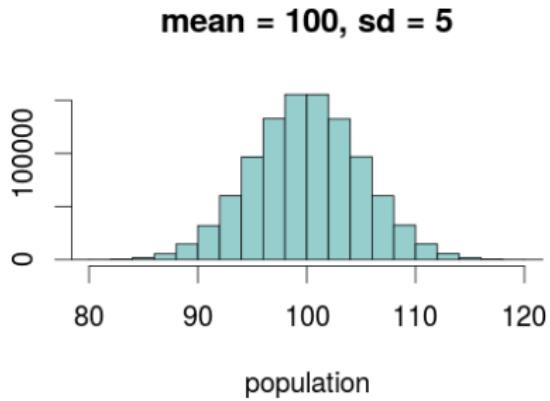
Sampling example



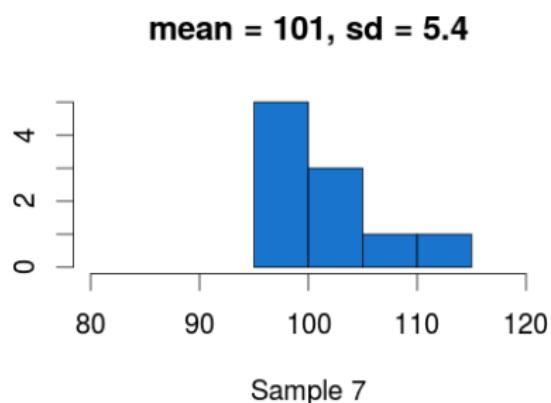
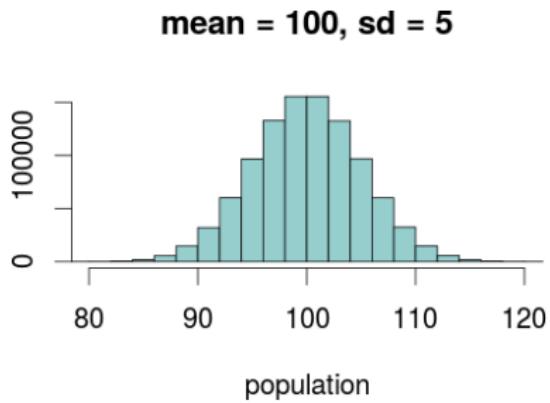
Sampling example



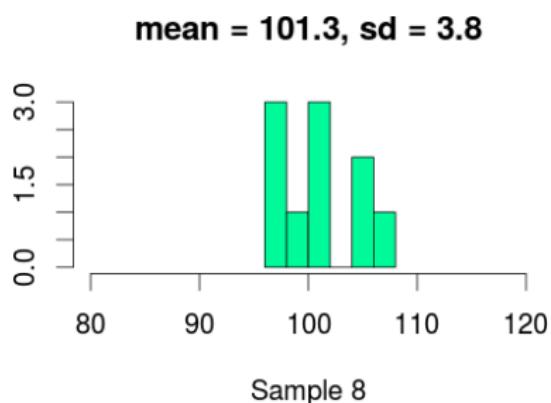
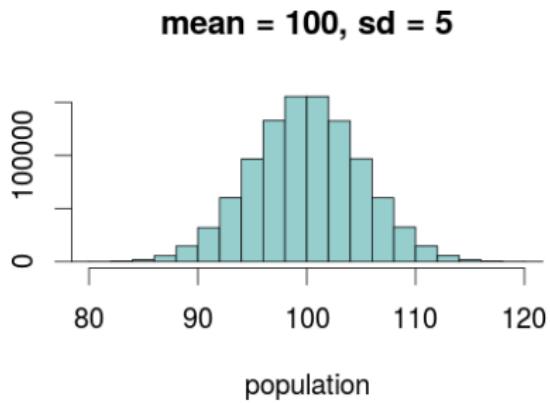
Sampling example



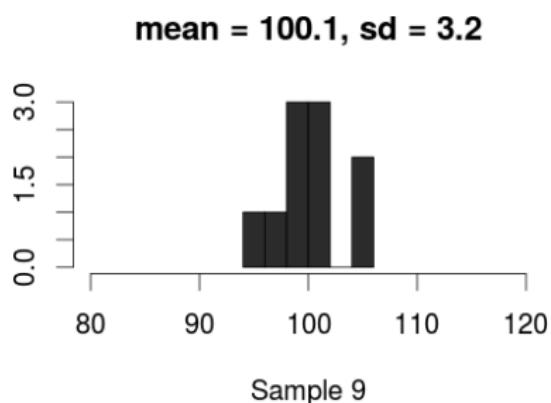
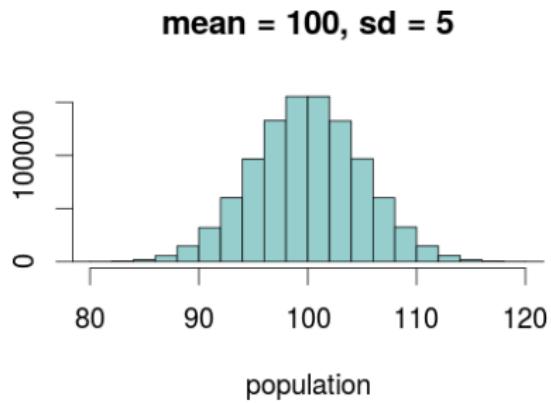
Sampling example



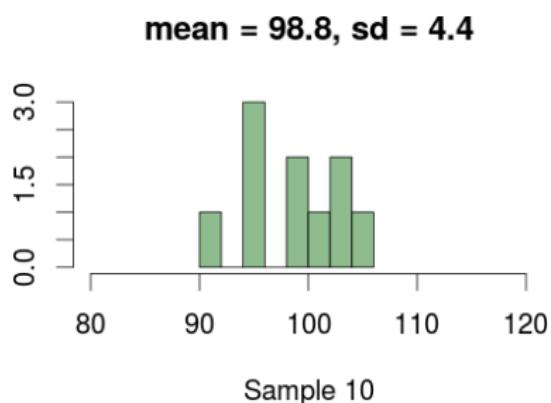
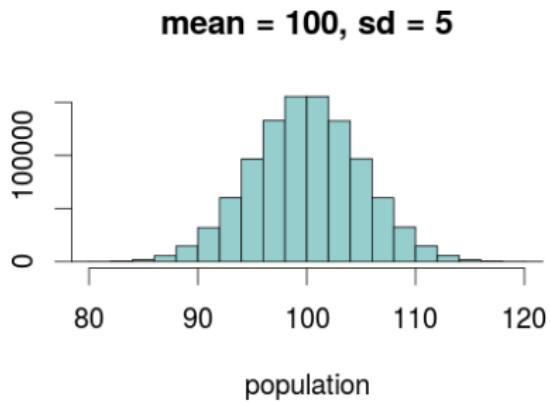
Sampling example



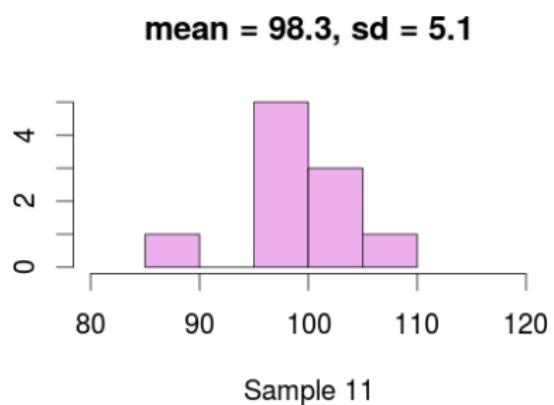
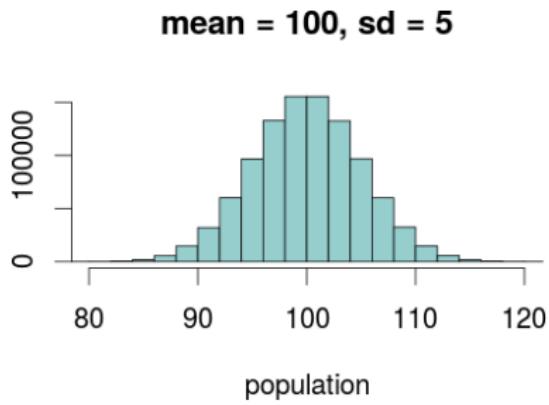
Sampling example



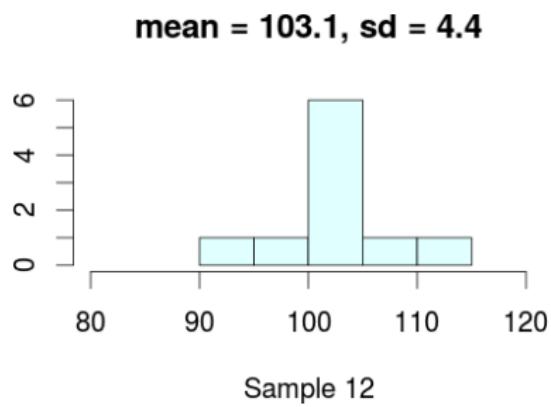
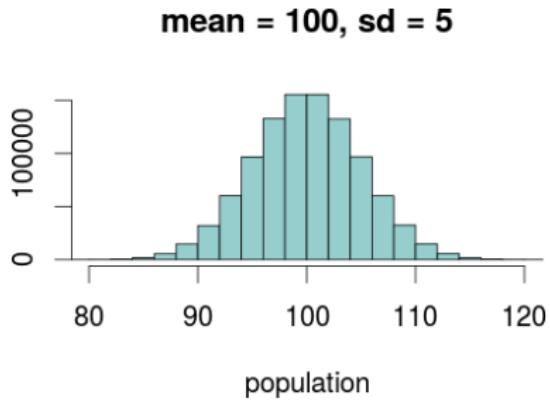
Sampling example



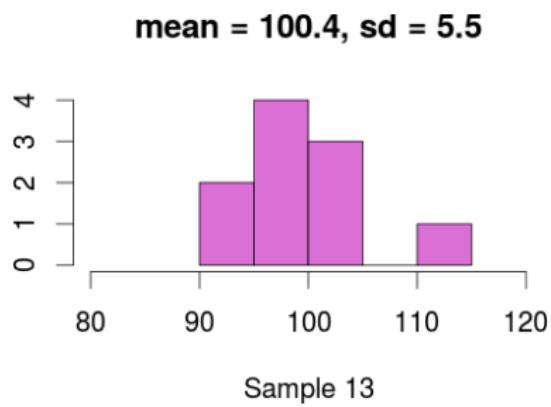
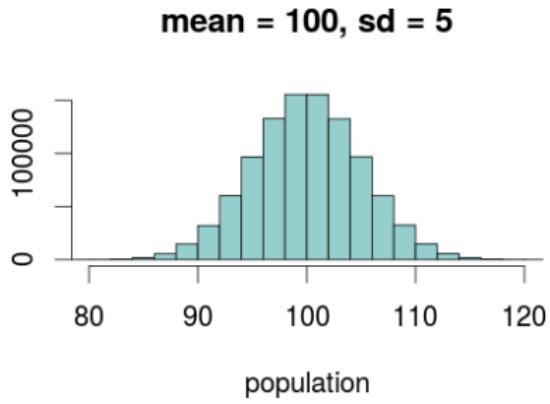
Sampling example



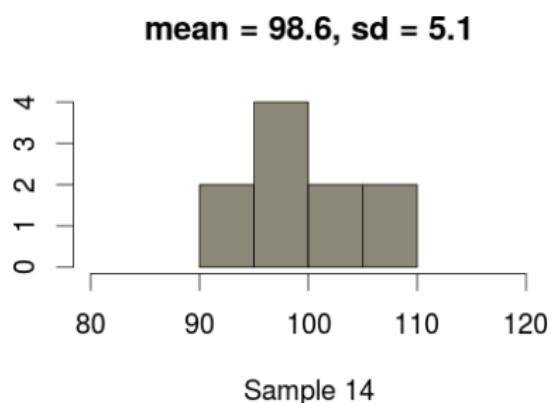
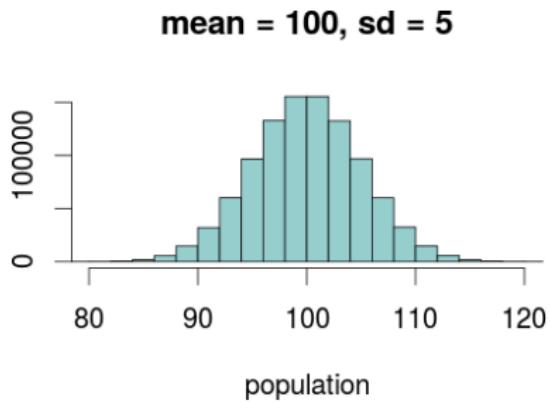
Sampling example



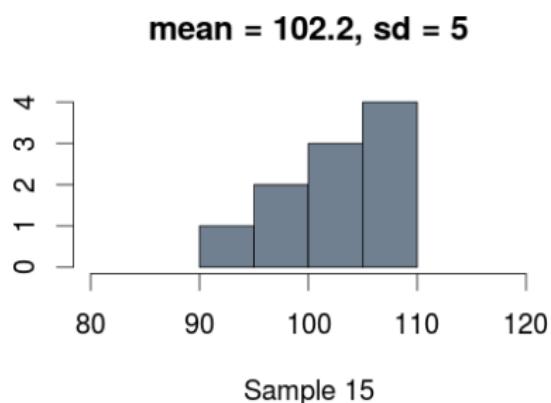
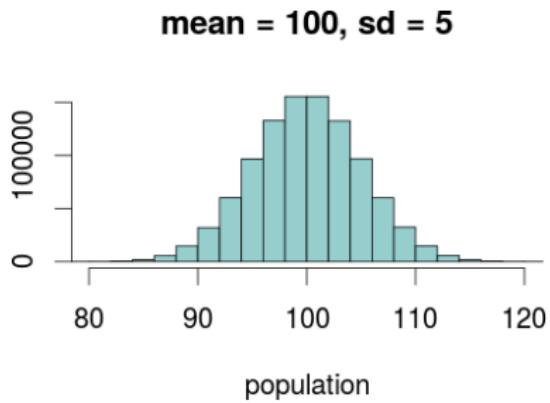
Sampling example



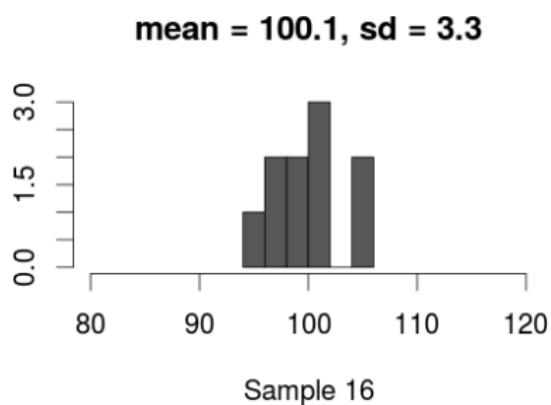
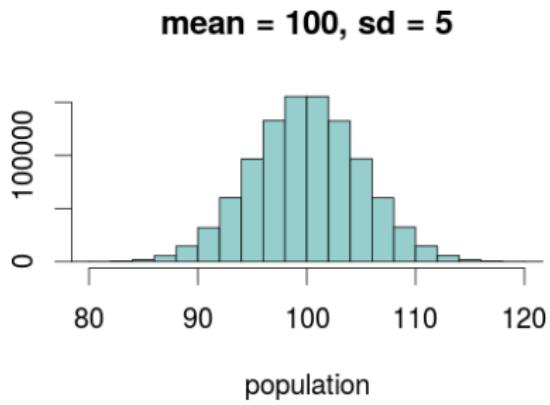
Sampling example



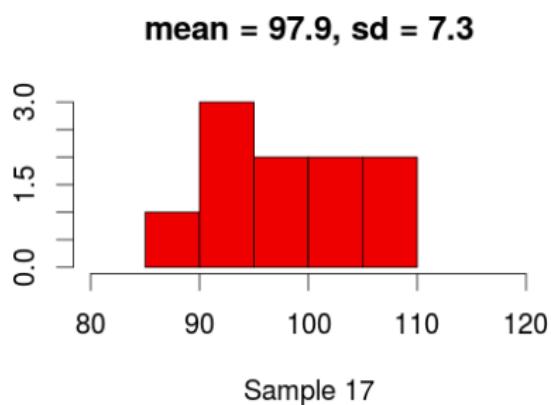
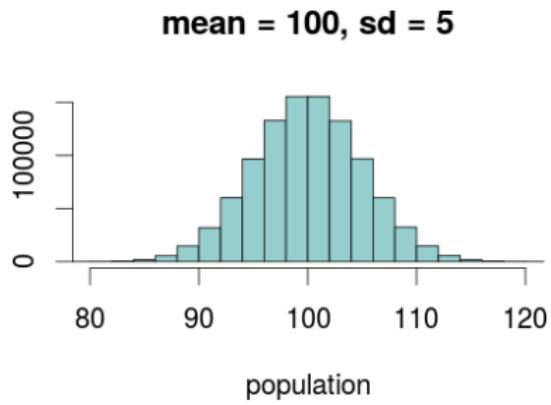
Sampling example



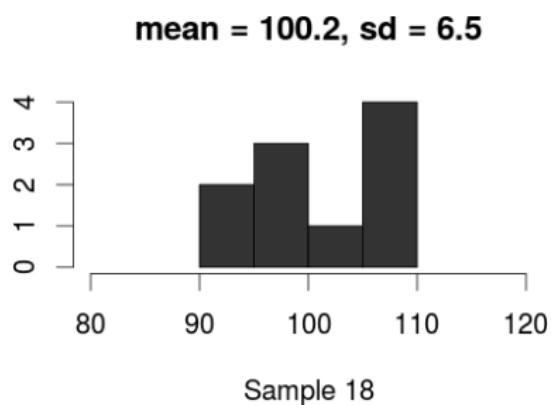
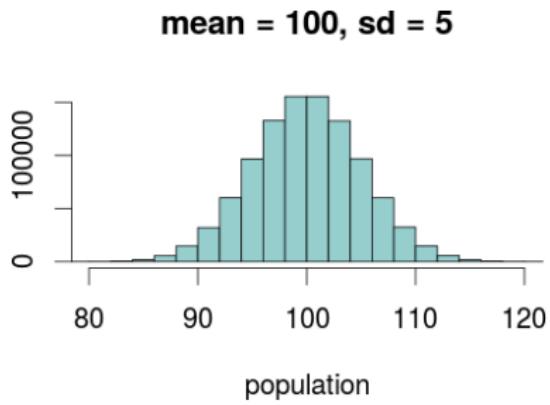
Sampling example



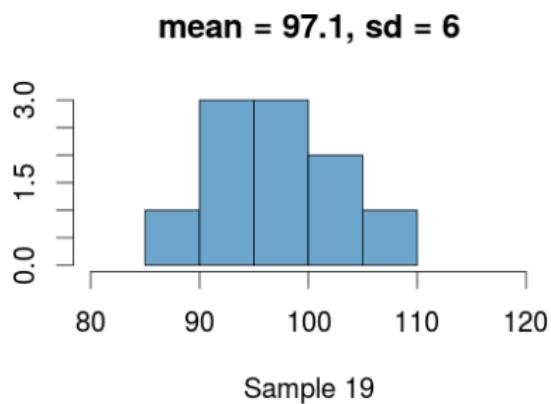
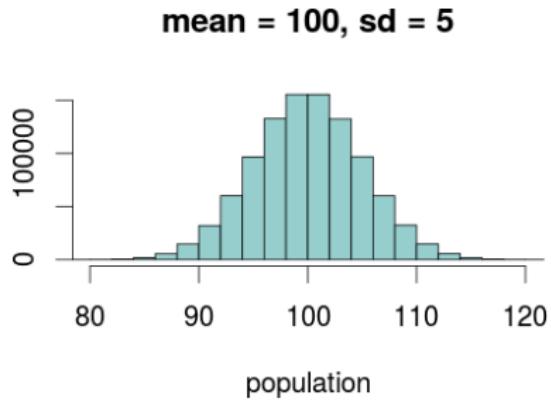
Sampling example



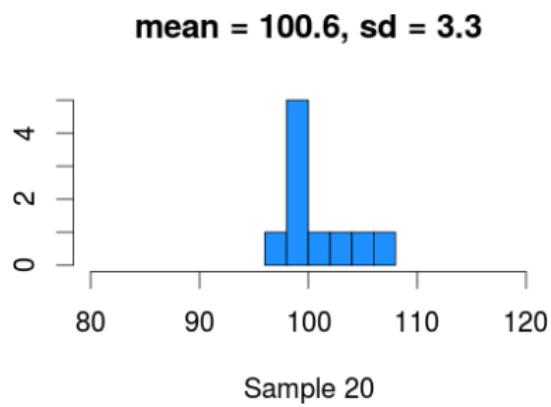
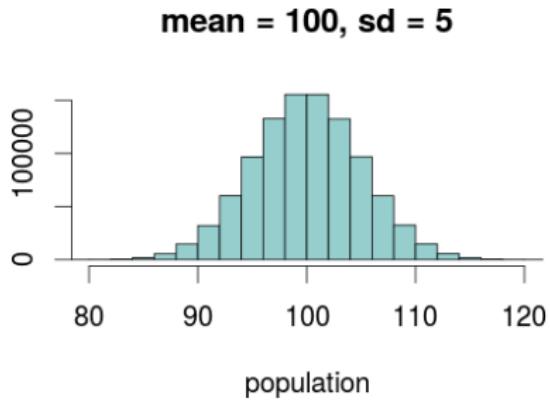
Sampling example



Sampling example



Sampling example



Outline

- 1 Samples and populations
- 2 Sampling error
- 3 Sampling bias
- 4 Recognising and avoiding sampling bias

Sampling error

- Even if sampling is perfectly done, the sample will never *exactly* reflect the population.

Sampling error

- Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
- The difference between sample and population is called **sampling error**

Sampling error

- Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
- The difference between sample and population is called **sampling error**
- (“Error” here means “difference from the true population”, not “mistake”)

Sampling error

- Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
- The difference between sample and population is called **sampling error**
- (“Error” here means “difference from the true population”, not “mistake”)

How does the sampling error depend on sample size?

Sampling error

- Even if sampling is perfectly done, the sample will never *exactly* reflect the population.
- The difference between sample and population is called **sampling error**
- (“Error” here means “difference from the true population”, not “mistake”)

How does the sampling error depend on sample size?

We will talk more about this next week!

Outline

- 1 Samples and populations
- 2 Sampling error
- 3 Sampling bias
- 4 Recognising and avoiding 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 US presidential elections:

- 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

Sampling bias - Example

1948 US presidential elections:

- 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!

Check your understanding!

What is the difference between sampling error and sampling bias?



Outline

- 1 Samples and populations
- 2 Sampling error
- 3 Sampling bias
- 4 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.
- 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?

What questions do you have?

You should now 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

- Four students sitting around a table. By Yuuki Guzman and Agoston Tyll (Okinawa Institute of Science and Technology), 2015.
- President Truman holding a copy of the Chicago Daily Tribune. By Associated Press photo by Byron Rollins, Fair use, <https://en.wikipedia.org/w/index.php?curid=701015> via Wikimedia Commons.
- Random sample. By Dan Kernler [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)], from Wikimedia Commons
- Student raising her hand. By Thomas Taylor Hammond (1920-1993) [CC BY-SA 4.0 (<https://creativecommons.org/licenses/by-sa/4.0/>)], via Wikimedia Commons
- Two students discussing. By Yuuki Guzman (Okinawa Institute of Science and Technology), 2016.