

Inferential Statistics

Lecture 4: Samples and sample functions

Jérémy Hervelin

Pforzheim University
j.hervelin@lb.hs-pforzheim.de

12 October 2026 15:30–17:00
Winter semester 2026/27

Adapted from the teaching materials of Prof. Dr. Thilo Klein.

Our 90-minute route

- First 15 minutes: retrieve prerequisites and introduce the question.
- Next 35 minutes: worked examples, with short checks after each concept.
- Next 25 minutes: paired activity and interpretation.
- Final 15 minutes: synthesis, questions and an applied discussion.

Learning objectives

- Describe the population and sampling frame before analysing observations.
- Compare simple random, stratified and cluster sampling.
- Distinguish a statistic from its realised value and its sampling distribution.
- Reproduce the original four-person example.

The target comes before the sample

Suppose we want the employment rate of all young job seekers entering a programme this year.

- Target population: everyone meeting that definition.
- Sampling frame: the list from which we can select people.
- Sample: those selected; respondent sample: those observed.

A precise estimate from an incomplete frame can still be biased.

Parameters, estimates and tests

A population mean μ or proportion p is a parameter. Use θ when speaking generally.

- Point estimation: one numerical estimate.
- Interval estimation: uncertainty about a parameter.
- Hypothesis testing: evidence against a specified claim.

Each requires a justified connection between the sample and the population.

Classroom check – Q1

Concept check

1–2 minutes

In a survey, the population employment rate is a . . .

- A. Statistic;
- B. Parameter;
- C. Realisation of the sample mean

Think individually, then compare your reasoning with a neighbour.

Lecture 4, classroom quiz Q1. Same question number as the printed quiz.

The original four-person population

The heights are 168, 172, 176, 180 cm, each with probability $1/4$.

$$\mu = 174, \quad \sigma^2 = \frac{36 + 4 + 4 + 36}{4} = 20.$$

This is the complete finite population, so the denominator is four.

With a sample of two, the sample mean will vary even though the population stays unchanged.

Classroom check – Q2

Exam-style

1–2 minutes

The population is 168,172,176,180. Find its mean.

Think individually, then compare your reasoning with a neighbour.

Lecture 4, classroom quiz Q2. Same question number as the printed quiz.

Sampling with replacement

Two independent uniform draws give $4^2 = 16$ equally likely ordered samples.

First/second	168	172	176	180
168	168	170	172	174
172	170	172	174	176
176	172	174	176	178
180	174	176	178	180

Each cell contains the sample mean. Repeated selection of the same person is allowed in this mathematical experiment.

Sampling without replacement

Selecting two distinct people gives $4 \times 3 = 12$ equally likely ordered samples, or $\binom{4}{2} = 6$ unordered pairs.

The second draw depends on the first: the selected person is no longer available.

Equal marginal selection probabilities do not imply independent draws. A realised random sample need not mirror every population characteristic.

Coverage and response can create bias

In the four-person example, drawing only from 168, 172 gives expected sample mean 170, below the target 174.

In an employment survey:

- an online-only frame can exclude people without access;
- employment status may affect whether people respond;
- volunteers may differ from the rest of the population.

A larger sample from the same selected group does not automatically remove bias.

Classroom check – Q3

Concept check

1–2 minutes

A very large voluntary survey necessarily removes selection bias.

- A. True;
- B. False

Think individually, then compare your reasoning with a neighbour.

Lecture 4, classroom quiz Q3. Same question number as the printed quiz.

Probability sampling and weights

Simple random sampling is one design, not the only valid design.

Unequal but known inclusion probabilities can be used with appropriate weights.

Frame coverage, nonresponse and measurement still matter.

Before applying an i.i.d. formula, ask whether the sampling design supports independence and identical distributions.

Stratified sampling

Divide the population into non-overlapping strata and sample within each.

Original example: strata $\{168, 172\}$ and $\{176, 180\}$; draw one person from each.

$$\bar{x} \in \{172, 174, 174, 176\}.$$

Applied example: sample vocational and general-education graduates separately.

When stratum population shares differ, use those shares to weight the stratum means.

Classroom check – Q4

Concept check

1–2 minutes

Sample separately from education groups and combine using population shares.
This is. . .

- A. Cluster sampling;
- B. Stratified sampling;
- C. A census

Think individually, then compare your reasoning with a neighbour.

Lecture 4, classroom quiz Q4. Same question number as the printed quiz.

Cluster sampling

Select groups, then observe all or some people within each selected group.

Original example: choose one of $\{168, 172\}$ and $\{176, 180\}$ with equal probability. The sample mean is 170 or 178.

An employment survey might sample training centres. People within a centre may have similar outcomes, reducing the information relative to independent observations.

Strata and clusters are different

Stratification	Cluster sampling
Sample within every stratum	Select a subset of clusters
Can improve precision when strata differ	Can reduce collection costs
Account for population weights	Account for within-cluster dependence
Neither label alone determines precision; the design and within-group similarity matter.	

A statistic is a function of the sample

Before sampling, $X_1, \dots, X_n$ are random. After sampling, we observe $x_1, \dots, x_n$.

$$T = g(X_1, \dots, X_n), \quad t = g(x_1, \dots, x_n).$$

Example:

$$\bar{X} = \frac{1}{n} \sum_i X_i, \quad \bar{x} = \frac{1}{n} \sum_i x_i.$$

The statistic $\bar{X}$ has a sampling distribution. The observed $\bar{x}$ is one number.

The sampling distribution with replacement

$\bar{x}$	168	170	172	174	176	178	180
$16P(\bar{X} = \bar{x})$	1	2	3	4	3	2	1

$$E(\bar{X}) = 174, \quad \text{Var}(\bar{X}) = 10.$$

The values in the second row are counts among 16 equally likely ordered samples; divide by 16 to obtain probabilities.

The sampling distribution without replacement

$\bar{x}$	170	172	174	176	178
$6P(\bar{X} = \bar{x})$	1	1	2	1	1

$$E(\bar{X}) = 174, \quad \text{Var}(\bar{X}) = 20/3 \approx 6.6667.$$

The extreme sample means 168 and 180 are impossible because the same person cannot be selected twice.

Finite population correction

For a simple random sample without replacement from $N > 1$ units:

$$\text{Var}(\bar{X}) = \frac{\sigma^2}{n} \frac{N - n}{N - 1}.$$

Here $\sigma^2 = N^{-1} \sum_{i=1}^N (x_i - \mu)^2$ is the finite population variance.

For $N = 4, n = 2$, this is $20/2 \times 2/3 = 20/3$.

When n/N is small, the correction is close to one. Unknown N is not a reason to physically sample people with replacement.

Comparing the four designs

Design	$E(\bar{X})$	$\text{Var}(\bar{X})$
With replacement	174	10
Without replacement	174	6.6667
One per stratum	174	2
One whole cluster	174	16

All four are unbiased in this example, but their precision differs.

Activity: calculate the last two variances from the possible means. Explain the role of similarity within each group.

A population distribution is not a sampling distribution

The population distribution describes individual heights. The sampling distribution describes the means of repeated samples.

In our example:

$$\text{SD}(X) = \sqrt{20}, \quad \text{SD}(\bar{X}) = \sqrt{10} \quad (\text{with replacement, } n = 2).$$

The second quantity is the standard error of the mean. Next lecture: how it changes with sample size.

Classroom check – Q5

Exam-style

1–2 minutes

Standard error describes. . .

- A. Dispersion of individuals;
- B. Sampling dispersion of an estimator;
- C. A data entry mistake

Think individually, then compare your reasoning with a neighbour.

Lecture 4, classroom quiz Q5. Same question number as the printed quiz.

Design exercise: a youth-employment survey

- Define the population and frame for a six-month follow-up survey.
- Explain how nonresponse might depend on employment.
- Compare sampling individuals with sampling whole training centres.
- Explain why random assignment to a programme and random sampling from a population answer different design questions.

Independent practice

- Practise the five online questions for this lecture; 80% provides access to the next practice module.
- Course materials remain accessible regardless of your practice progress.
- Explain your reasoning before comparing answers.
- Review the method, assumptions and interpretation as well as the calculation.

Sources and teaching notes

Adapted from Prof. Dr. Thilo Klein's supplied Inferential Statistics materials.

Revised teaching edition: Jérémy Hervein, Pforzheim University, winter semester 2026/27.

New youth-employment, programme and interview-score examples are synthetic illustrations.