Part 1
What is statistics?

Natalia Levshina © 2015

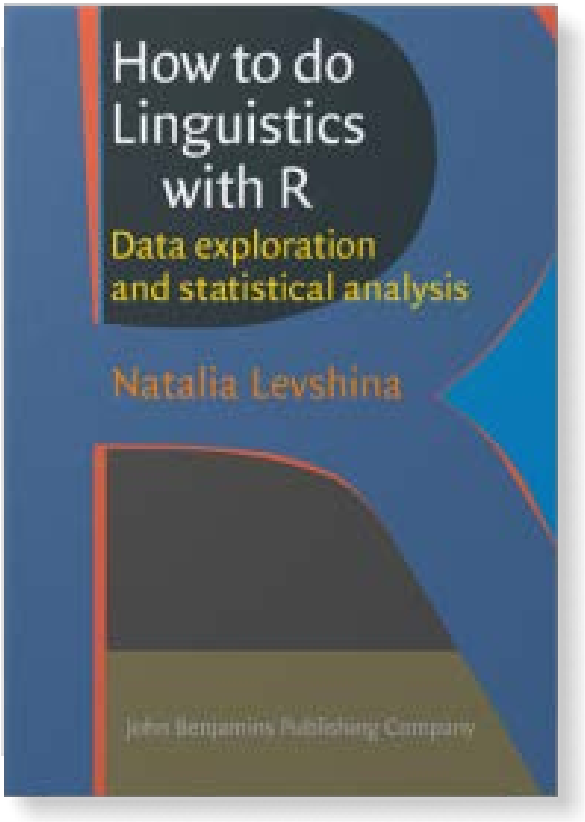
Part of the course taught at the University of Mainz, Germany
26-28 May 2015

Practicalities

- The slides (pdf) are downloadable from http://natalialevshina.com/statistics.html

- We will use R, free statistical software. R code can be copied from the slides and pasted into R

More information here:



Scheduled for August 2015

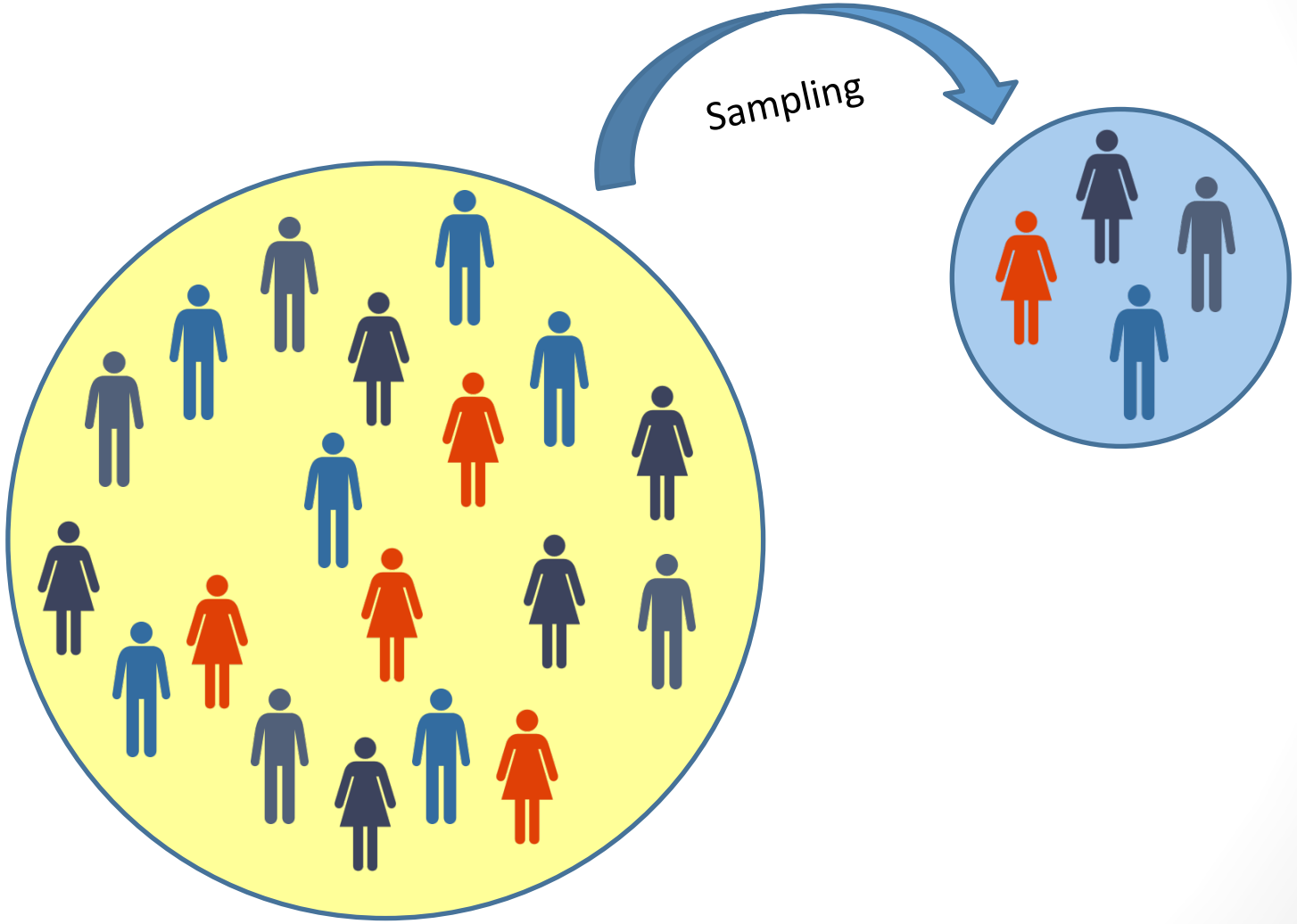
Goals of the course

- To learn what statistics is about
- To learn how to perform basic operations with data in R
- To learn two popular methods in variational and sociolinguistics: Multidimensional Scaling and logistic regression analysis

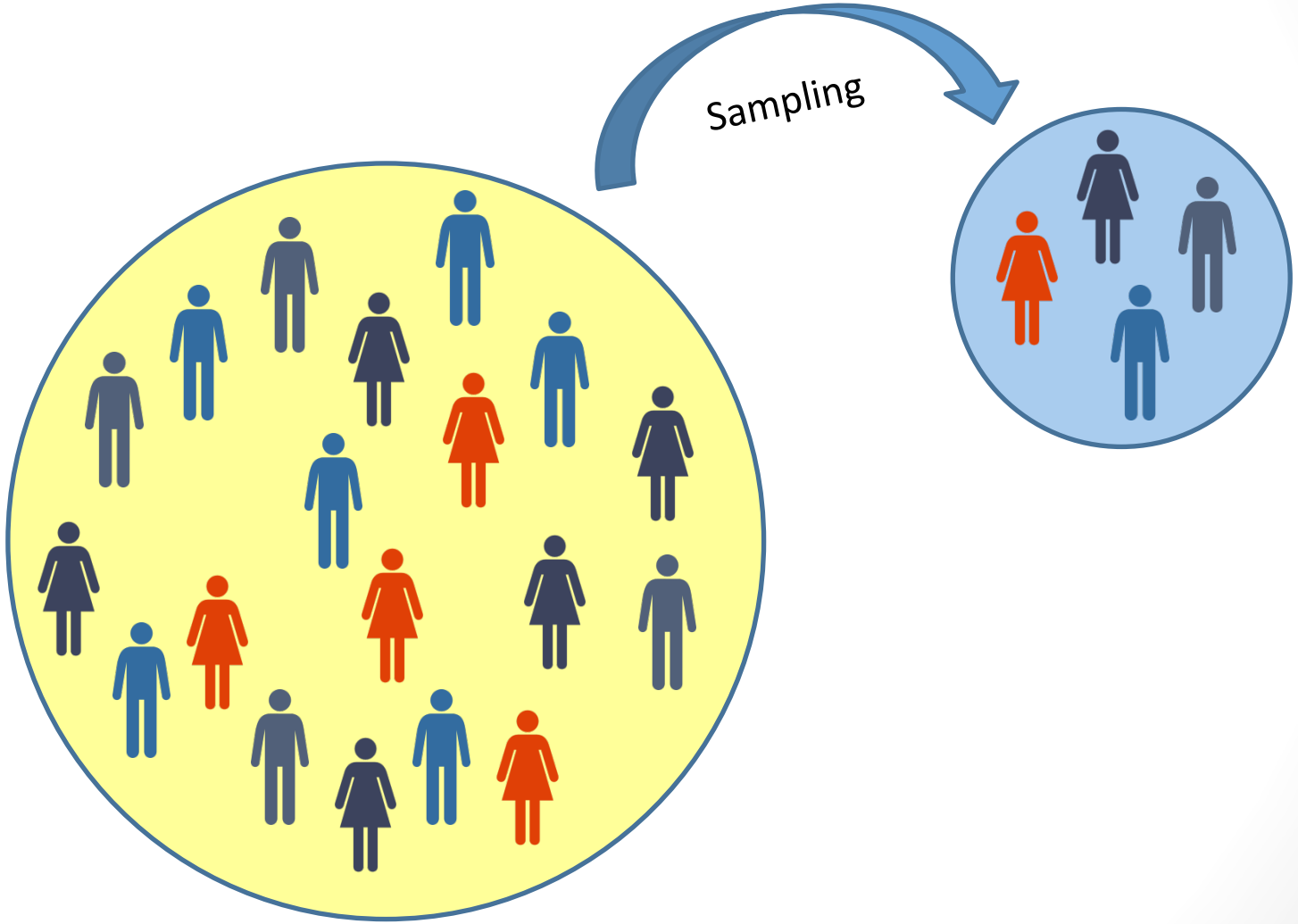
Outline of the course

1. What is statistics?
2. Introduction to R
3. Basic descriptive statistics and graphs
4. T-test
5. Logistic regression
6. Multidimensional Scaling

Population and Sample



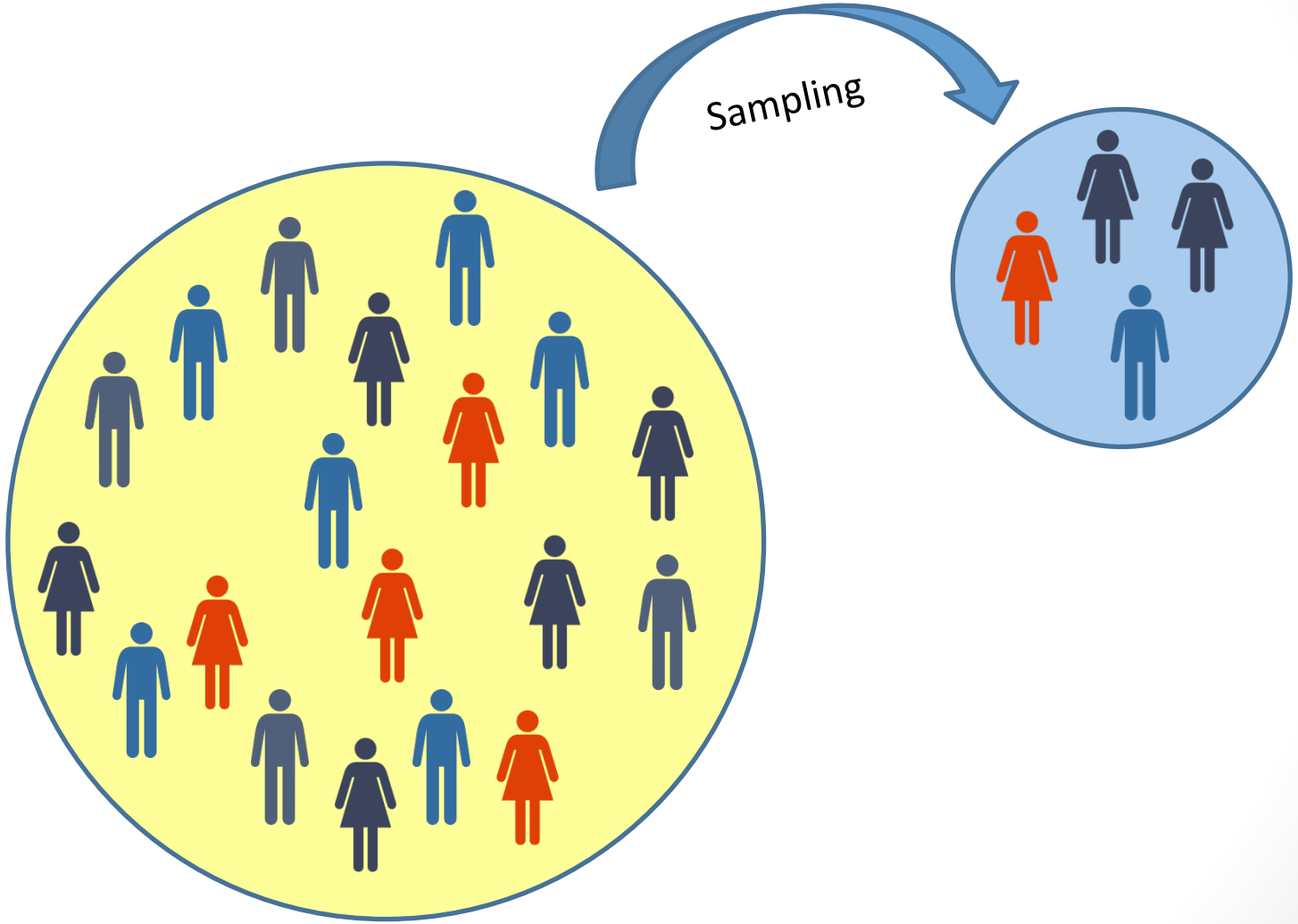
Population and Sample



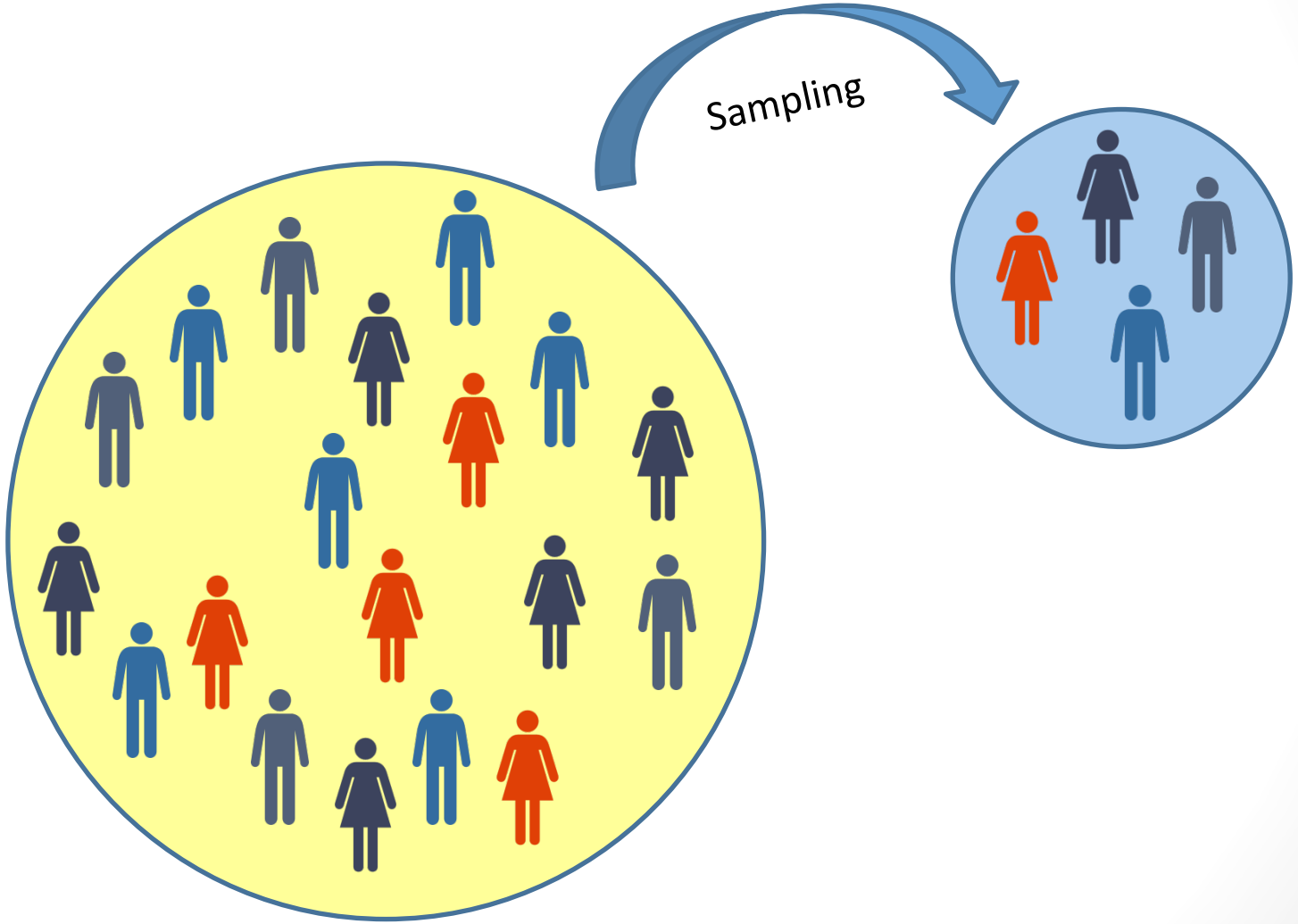
Population parameters (mean, variance, etc.)

Sample statistics (mean, variance, etc.)

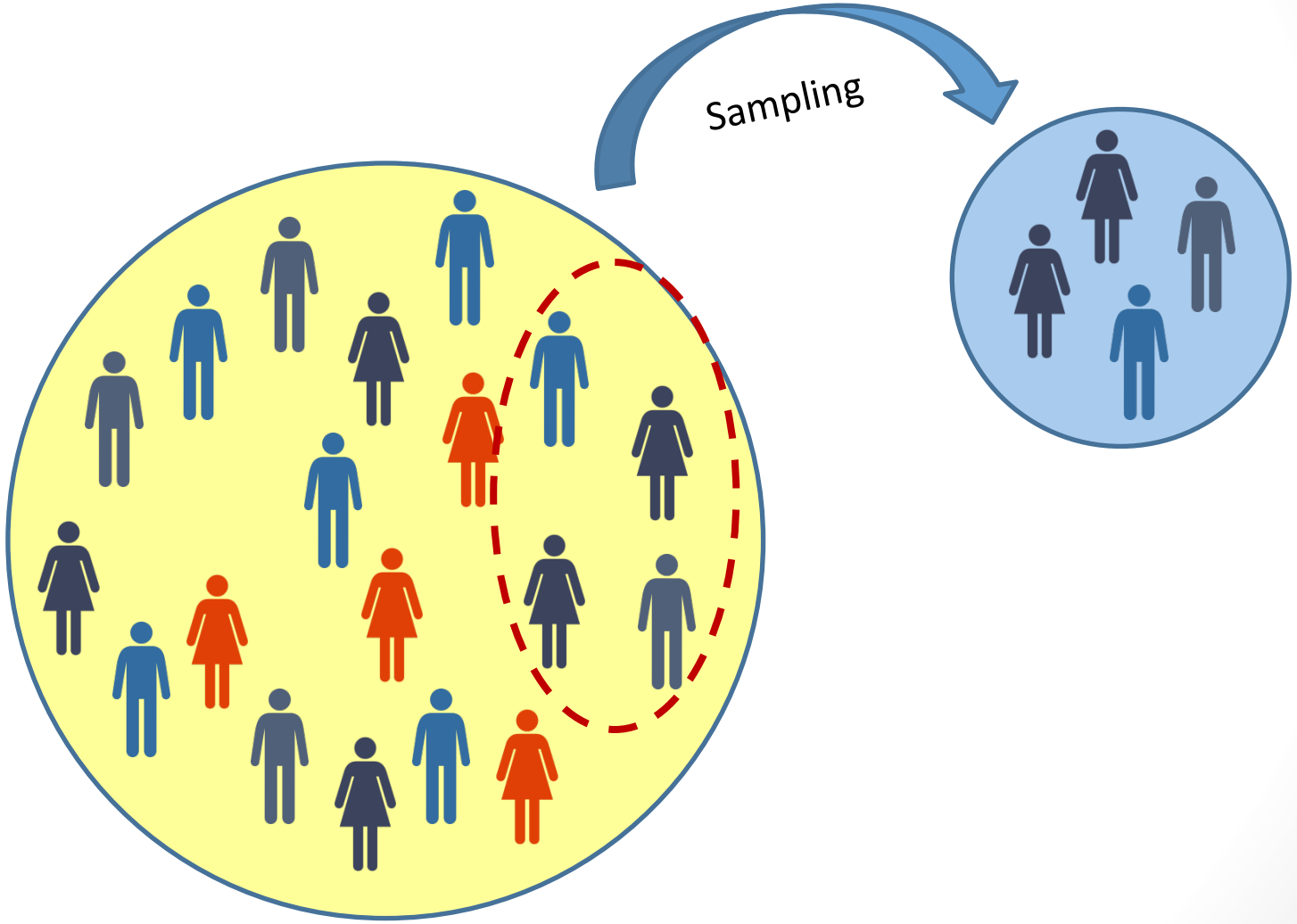
Random sampling



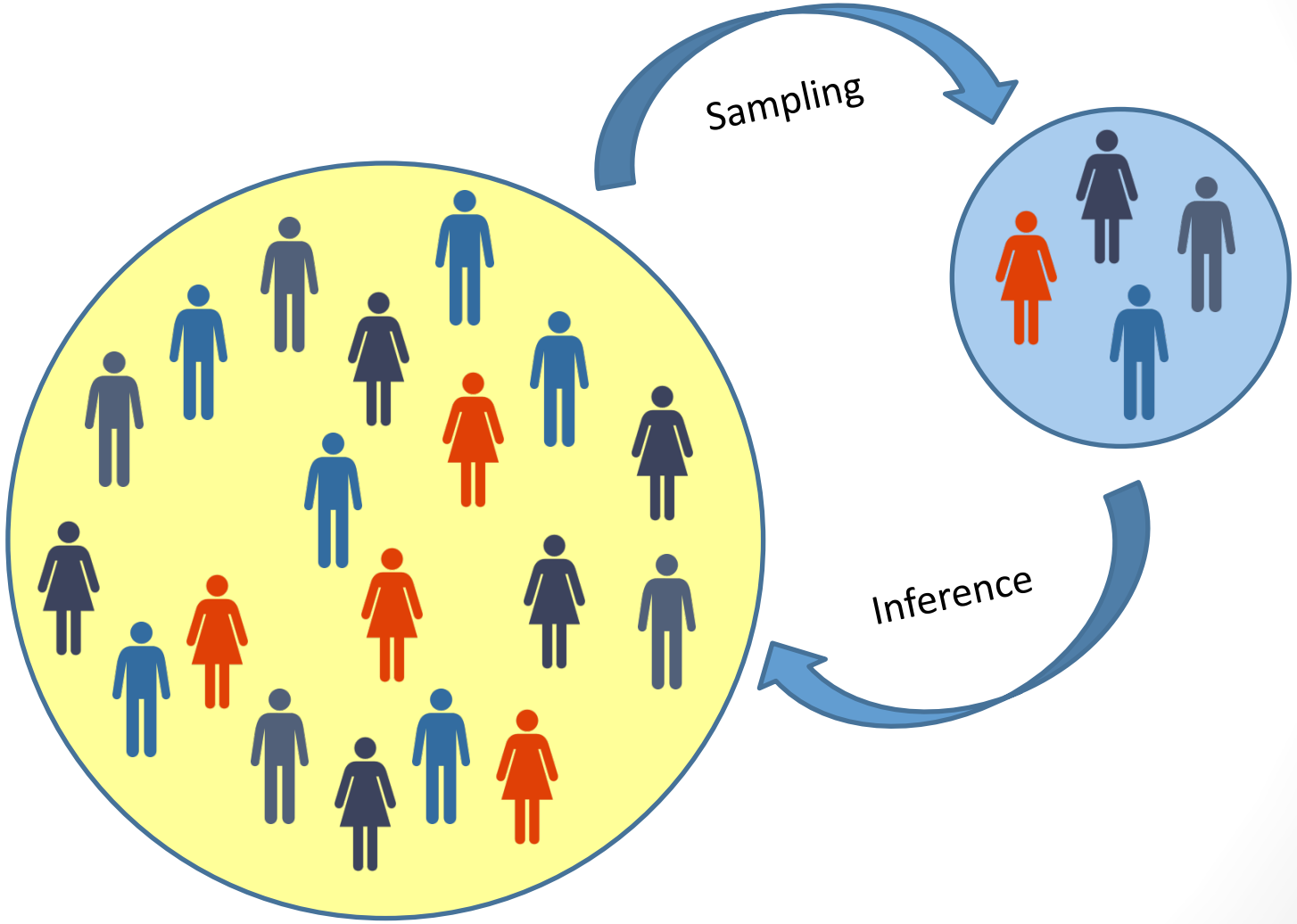
Representative sampling



Convenience sampling



Inferential statistics



Logic of hypothesis testing

- Alternative hypothesis (your research idea: difference between groups, association between variables)

  - directional (e.g. group 1 is GREATER/LESS than group 2; there is a POSITIVE/NEGATIVE correlation between variables A and B)

  - non-directional (some difference, some association)

- Null hypothesis (no difference between groups, no association between variables, etc.)

Example 1

H_0 (the null hypothesis): There is no difference in the number of lexemes that denote snow in Eskimo and Yucatec Maya.

H_1 (the alternative hypothesis): There are more lexemes that denote snow in Eskimo than in Yucatec Maya.

Is H_1 directional or non-directional?

Example 2

H_0 (the null hypothesis): there is no relationship between the frequency of a word and how fast it is recognized in a lexical decision task.

H_1 (the alternative hypothesis): the more frequent a word, the faster it is recognized in a lexical decision task.

Is H_1 directional or non-directional?

Example 3

H_0 (the null hypothesis): there is no difference in the relative frequencies of metaphoric expressions used by men and women when they speak about sex.

H_1 (the alternative hypothesis): there is a difference in the relative frequencies of metaphoric expressions used by men and women when they speak about sex.

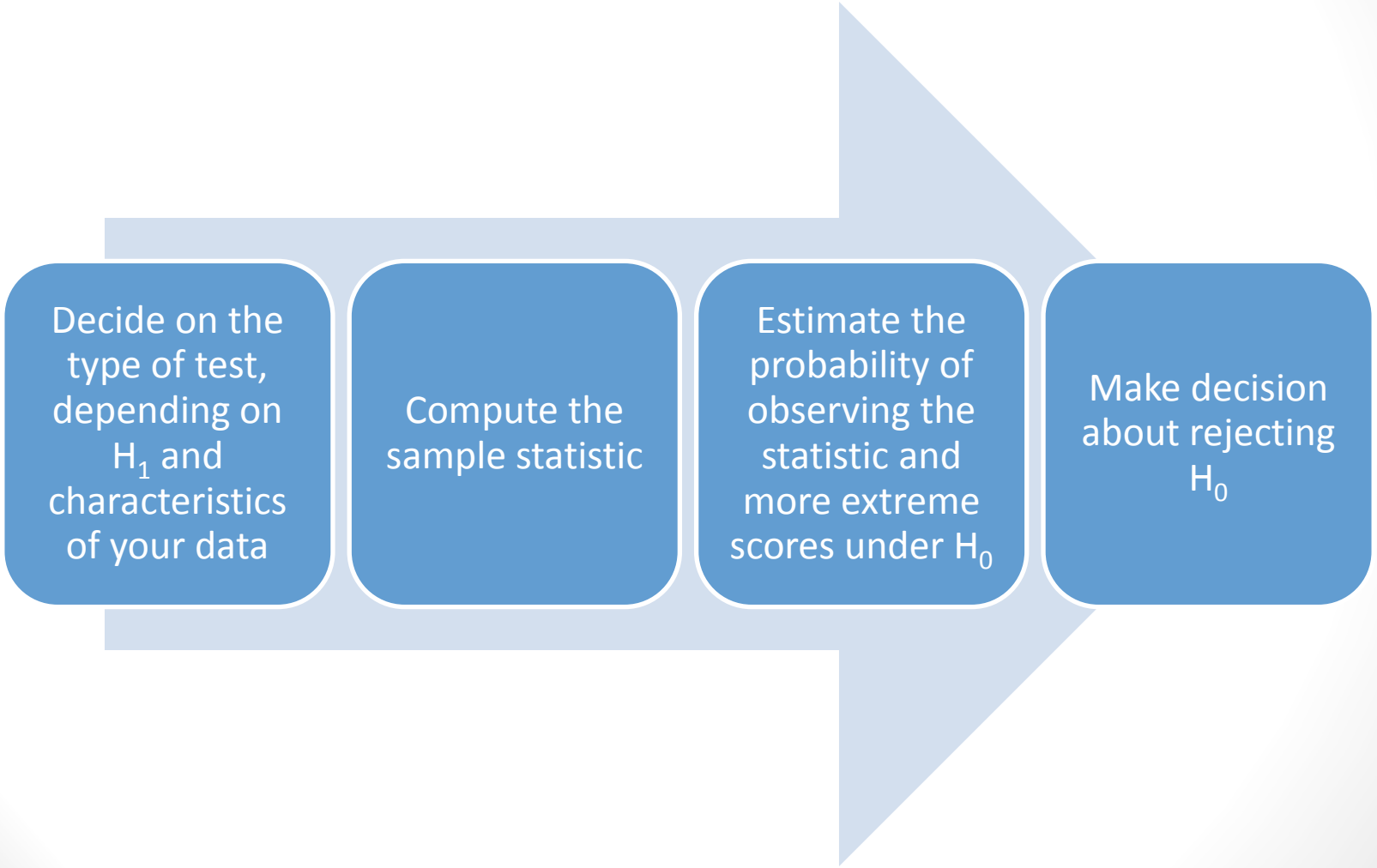
Is H_1 directional or non-directional?

Exercise

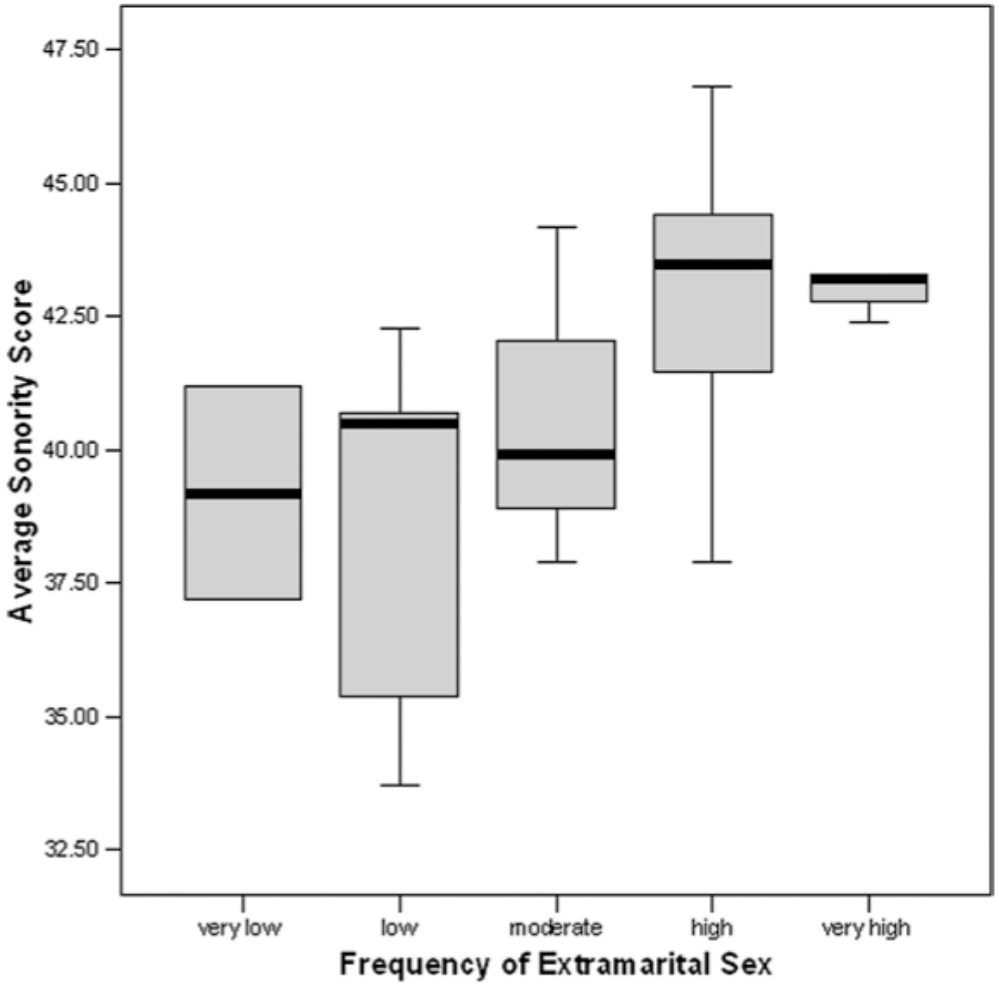
Think about two research questions and try to formulate

a) a null hypothesis and a non-directional alternative hypothesis;
b) a null hypothesis and a directional alternative hypothesis.

Hypothesis testing algorithm

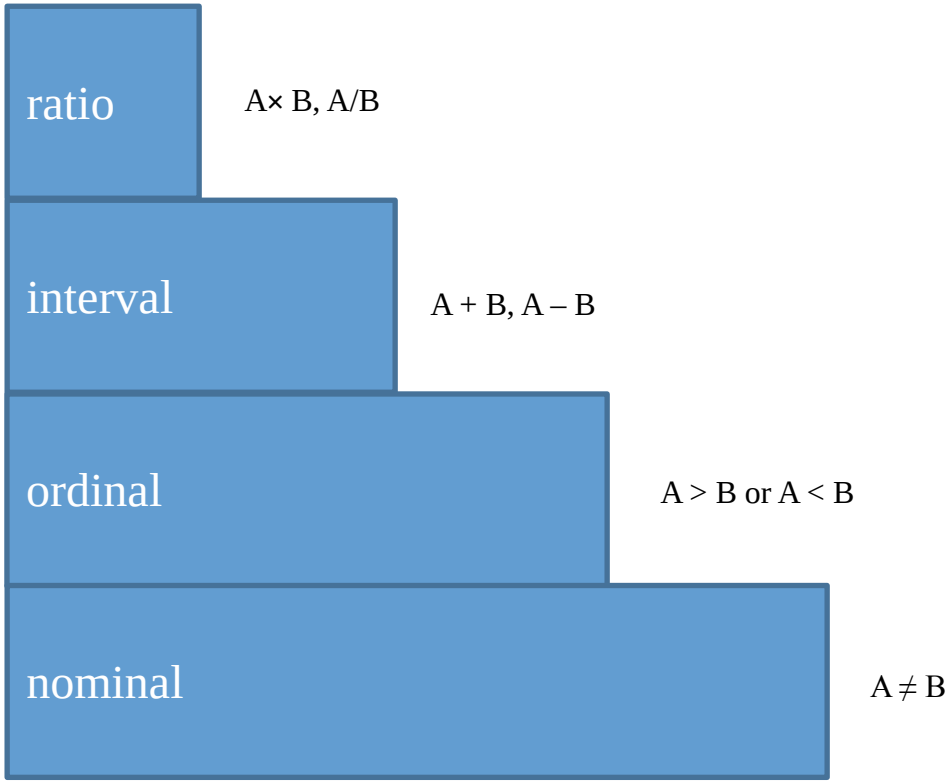


Correlation is NOT causation



From Ember & Ember 2007

Scales of measurement



Exercise

Give examples of variables on the nominal, ordinal, interval and ratio scale of measurement.