Introductory Statistics 1
AP Statistics

Instructors: Bakhtiyar Daukeev and Prof. Máté Fodor

Statistics – a definition

- Statistics is the science and, arguably, also the art of learning from data.
- As a discipline it is concerned with the collection, analysis, and interpretation of data, as well as the effective communication and presentation of results relying on data.
- Statistics lies at the heart of the type of quantitative reasoning necessary for making important advances in the sciences, such as medicine and genetics, and for making important decisions in business and public policy.

Variables

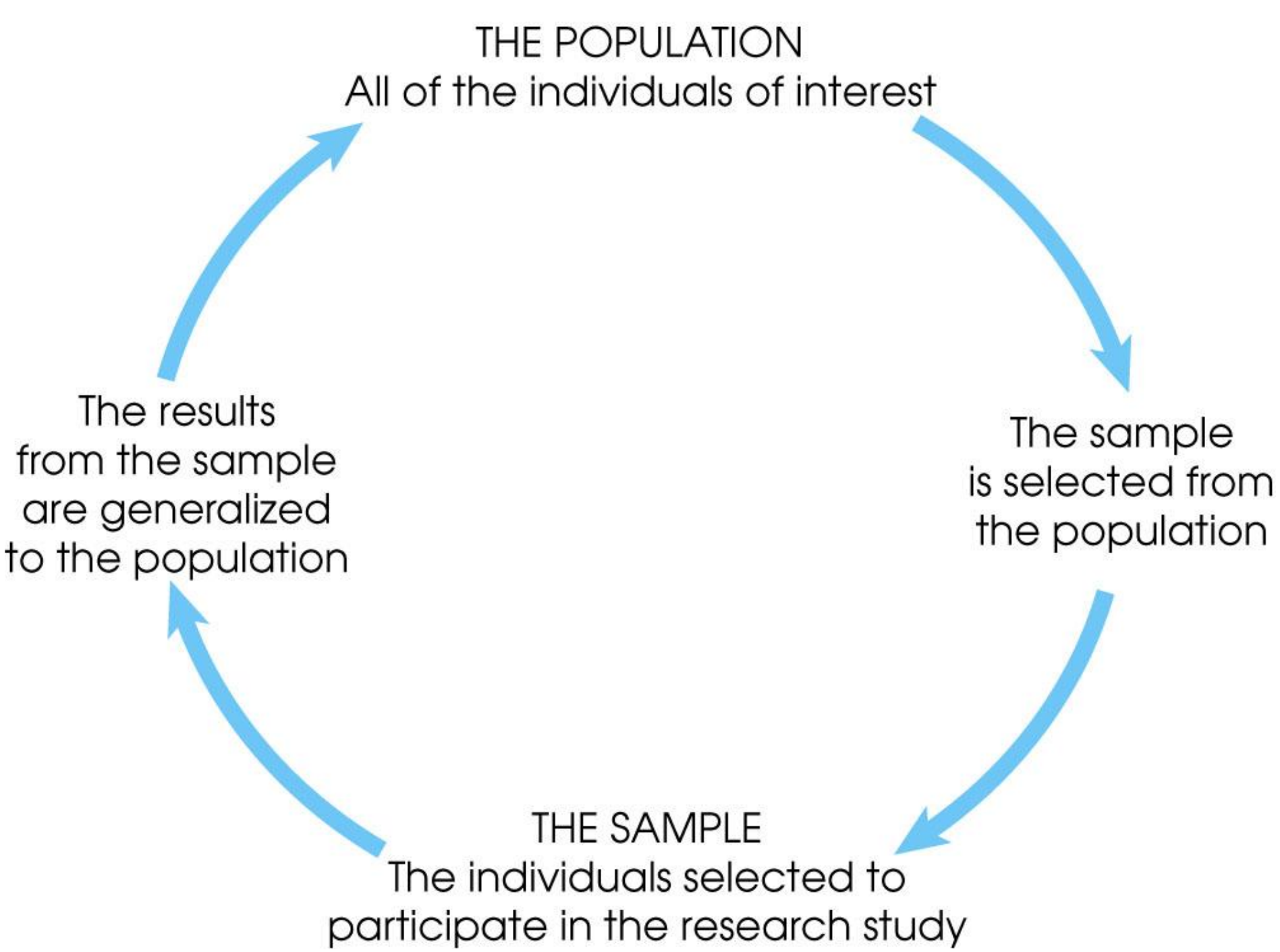
- A **variable** is a characteristic or condition that can change or take on different values.
- Most research begins with a general question about the relationship between two variables for a specific group of individuals.

Population

- The entire group of individuals is called the **population**.
- For example, a researcher may be interested in the relation between class size (variable 1) and academic performance (variable 2) for the population of third-grade children.

Sample

- Usually populations are so large that a researcher cannot examine the entire group. Therefore, a **sample** is selected to represent the population in a research study. The goal is to use the results obtained from the sample to help answer questions about the population.


THE POPULATION
All of the individuals of interest
The sample
is selected from
the population
THE SAMPLE
The individuals selected to
participate in the research study
The results
from the sample
are generalized
to the population

Types of Variables

- Variables can be classified as discrete or continuous.
- **Discrete variables** (such as class size) consist of indivisible categories, and **continuous variables** (such as time or weight) are infinitely divisible into whatever units a researcher may choose. For example, time can be measured to the nearest minute, second, half-second, etc.

Measuring Variables

- To establish relationships between variables, researchers must observe the variables and record their observations. This requires that the variables be **measured**.
- The process of measuring a variable requires a set of categories called a **scale of measurement** and a process that classifies each individual into one category.

4 Types of Measurement Scales

1. A **nominal scale** is an unordered set of categories identified only by name. Nominal measurements only permit you to determine whether two individuals are the same or different.
2. An **ordinal scale** is an ordered set of categories. Ordinal measurements tell you the direction of difference between two individuals.

4 Types of Measurement Scales

3. An **interval scale** is an ordered series of equal-sized categories. Interval measurements identify the direction and magnitude of a difference. The zero point is located arbitrarily on an interval scale.
4. A **ratio scale** is an interval scale where a value of zero indicates none of the variable. Ratio measurements identify the direction and magnitude of differences and allow ratio comparisons of measurements.

Quantitative versus qualitative variables

- Quantitative means it can be counted, like “number of people per square mile.”
- Qualitative means it is a description, like “brown dog fur.”
- A Deck of cards contains quantitative variables (the numbers on the card) and qualitative variables (Spades, Hearts, Diamonds, Clubs).

Quantitative versus qualitative variables (2)

- Simplest way to decide: can you add them? - can you rank them?
- You can rank cars by numbers sold – and number of cars sold is indeed a quantitative variable
- But you cannot rank cars by colors (even though you might have a preference of blue over red – that is just your preferences and not statistical analysis)
- The color of a car is a qualitative variable.

A little break from statistics – practical organization of course

- The course is given by two lecturers – myself Prof. Máté Fodor, and Mr. Bakhtiyar Daukeev.
- You will see me every Monday, and you will have tutorials in groups with Mr. Daukeev.
- Our teaching is harmonized, we teach the same course material.

Practical organization (2)

- Mr. Daukeev will give you homework to do
- I may also give you homework to do.
- I will test you on your homework. I will select students each class, that need to come up in front of the class – and I will ask them questions about their homework.
- To make sure you did the homework on your own.

Practical organization (3)

- I will also give surprise quizzes – be prepared all the time.
- Course material: my slides (sent to you after class via email), Mr. Daukeev's class material, your notes you take in classes, reading I give you, reading Mr. Daukeev gives you, exercises that I or Mr. Daukeev gives you and homework.
- You may be tested on any of these at any time.

Practical organization – self study

- Seek out self-study guides, and help online
- Stattrek.com – AP tutorials : extremely good help
- Wikipedia is also great for basic concepts
- Wolframalpha.com amazing for basic and more advanced calculations.

Practical organization (4)

- Your grade will depend on
  - Your presence, your participation (have nametags in front of you)
  - Your performance on quizzes
  - Your homework and your defence of homework.
- 80 percent attendance mandatory at both lectures and tutorials.
- If you miss more than that, it's an automatic F – try again next year.

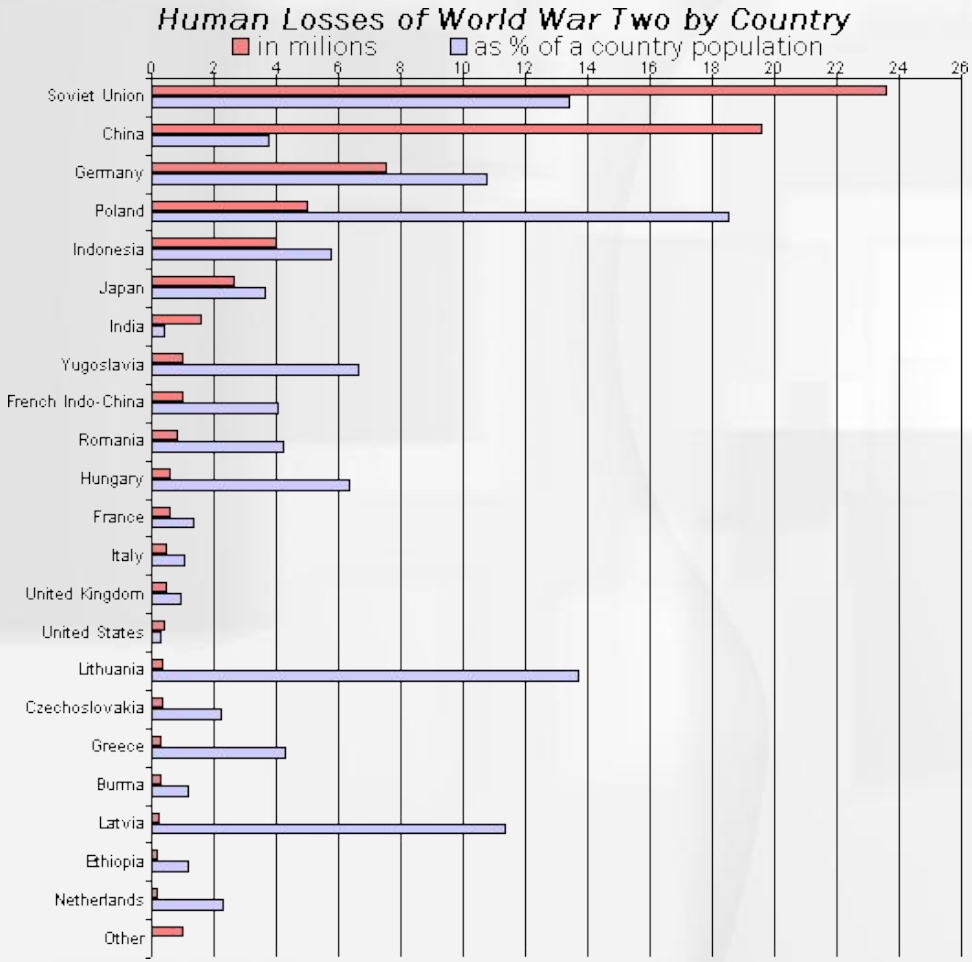
Practical organization (5)

- I do not accept doctor's notes. (I do not know if Mr. Daukeev does) If you are sick, send me an email before 9 in the morning on the day of your sickness, informing me you will be sick.
- To mate.m.fodor@gmail.com
- You may not look at your phone, wristwatch or any distracting device during class.
- Just looking at your watch – I will send you out, and you will be counted as absent (for both hours).

Back to Statistics – visual representation of data: Bar Charts

- Horizontal rectangles (bars) chart in which the length of a bar is proportional to the value (as measured along the horizontal axis) of the item (entity or quantity) it represents.
- Also called bar graph, it is used commonly to compare the values of several items in a group at a given point in time.

Bar charts (2) – an example



Further examples given on the board.

Example 1: temperature in a week

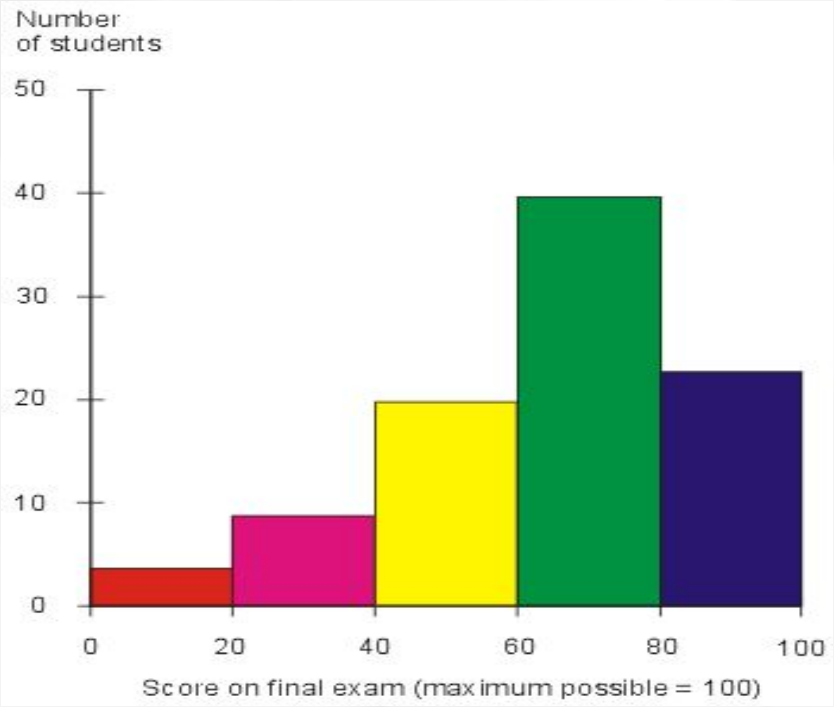
Example 2: weight of marathon runners by result

Example 3: average size of dogs by breed.

Any other examples you can think of?

Histograms

- A histogram is a display of statistical information that uses rectangles to show the frequency of data items in successive numerical intervals of equal size.



Histograms (2)

- It differs from a bar graph, in the sense that a bar graph relates two variables, but a histogram relates only one.
- To construct a histogram, the first step is to "binTo construct a histogram, the first step is to "bin" (or "bucket") the range of values—that is, divide the entire range of values into a series of intervals—and then count how many values fall into each interval.
- The bins are usually specified as consecutive, non-overlapping intervals of a variable. The bins (intervals) must be adjacent, and are often of equal size.

Histograms (3)

- Other examples of histograms are
  - The level of education of employees within a firm.
  - Value of transactions an individual makes in a week.
  - Number of drinks consumed by guests in a bar on a Friday night.

Frequency

- As you can see, histograms are a good representation of frequency.
- Definition: frequency is the times an event happens within a study.
- Say you observe a residential complex and see how people get to work.
- Some people cycle to work, some drive, some take public transport, some walk.
- If you observe 5 people walking, then the frequency of walking is simply 5.
- This is known as “absolute frequency”. Of course all alone, this does not make much sense.

Relative frequency

- Definition: how often an event happens divided by the sum of all possibilities.
  Example:

92 people were asked how they got to work:

- 35 used a car
- 42 took public transport
- 8 rode a bicycle
- 7 walked

The Relative Frequencies (to 2 decimal places) are:

- Car: 35/92 = **0,38**
- Public Transport: 42/92 = **0,46**
- Bicycle: 8/92 = **0,09**
- Walking: 7/92 = **0,08**

Cumulative frequency

- You're interested in studying a population to find out a "more" or "less" question. For example, you're thinking of opening a bargain grocery store and you want to know how many people in a particular geographic area spend up to $6000 per person per year in groceries. Your table might look like this:

Cumulative frequency (2)

Type	Freq	cumulative frequency
Up to 1000	22	22
1001-2000	45	67
2001-3000	57	124
3001-4000	97	221
4001-5000	152	373
5001-6000	241	614
6001-7000	153	767

- Cumulative frequency tells how many times an event happens up to a certain point
- when data is organized in ordered categories

See you next week

- For next week when you see me – you will need to do everything Mr. Daukeev tells you
- You will need to read pages 3 to 15 from Cliff's AP Statistics textbook (will send you an electronic version)
- Everything (including the organization of the AP exam) may be on the quiz next week.