STATISTICS

STATISTICAL DEFINITIONS

Statistical definitions (1)

- Multitude, population, random experiment
- Statistical sample, sample realizations
- Statistical sampling

Statistical multitude, population

A set of large number of individuals but finite number of elements, under investigation.

Target:
(1) solid characterization of the set as a whole **with a few parameters**,
(2) discovering **associations among parameters** introduced for describing individuals of the populations. It is usually not possible (resources) to obtain data on each element of the population.

- Citizens of Hungary
- Students of a university faculty
- A set of cars with current traffic licence
- A particular set of customers
- A set of viewers of a TV channel

Statistical multitude, population

Observation of a random experiment

The subject of a statistical analysis can be a random experiment, which can take place in principle any time **in the same circumstances**.

- lottery
- operation of a server
- mean temperature of January in Budapest
- an orchard yields
- effect of a new drog
- efficacy of an advertising campaign
- random selection of an element of a population

Statistical sample

Observation data on **a small subset of the population**.

The sample should be representative.

- a set of Hungarian citizens in a survey
- a set of students taking part on a lecture
- a set of cars contracted with an insurer
- a set of customers surveyed on a given day
- a set of viewing TV viewers surveyed
- mean January temperature data of Budapest

- **Types of statistical population:**

- ***Stationary population*:** status, time character; its observation can always be carried out at any given time; e.g. the number of students of the Faculty of Agriculture on September 7, 2015.
- ***Moving population*:** time-varying multitude; a process is detected; it can be observed during a time interval; e.g. the number of students have taken exams in the first semester of the study year 2015/2016 at the Faculty of Agriculture.
- ***Discrete population*:** its elements can be clearly separated, e.g. students;
- ***Continuous population*:** the elements are not separated; values are usually given at intervals, e.g. the height of students;
- ***Finite multitude:*** consists of finite number of elements, e.g. the population at any given time or space;
- ***Infinite multitude:*** consists of endless many elements, or of so many elements that should be considered infinite;
- ***Homogenuous population*:** consists of the same elements, can not be broken into parts, e.g. sand of a sand mine;
- ***Advanced population*:** its elements have not only common, but also distinctive characteristics, eg. students;
- ***Real population:*** consists of actually occurring elements;
- ***Theoretical population*:** consists of all possible outcomes for the occurrences of an event;

Statistical criterion

- ***Statistical criterion or criterion variable:*** *a* feature characterizing the elements of the statistical population (e.g. body weight, body height);
- ***Criterion variant:*** posssible values of the criteria, e.g.
  - Statistical individuum (individual): student,
  - Statistical criterion (criterion variable): gender; mark; body weight, body height;
  - Statistical variant: male, female; 1,2,3,4,5; 65 kg; 172 cm;

- ***Alternative criterion:*** if the criterion has only two versions;

- ***Common criterion*****:** determines the statistical population, based on which the units of the population are the same (e.g. university course);

- ***Distinguishing criterion*****:** according to which individuals (elements) of the population differ from each other (e.g. gender, age, address, academic average);

- **Classification of criteria according to their information content:**
  - ***Territorial criteria:*** **guiding principles** for **spatial** location of units. Their variants are usually geographic units (e.g. place of birth of students);
  - ***Temporal criteria*****: guiding principles** for **temporal** location of units. Their variants are dates and periods (e.g. date of birth of students). It can express status and duration, as well.
  - **Subject criteria**:
    - ***Quantitative criteria*****:** quantifiable characteristics of individuals.
      - continuous – it can take any value within a specified interval (e.g. body weight, body height);
      - discrete – it can take only distinct values from each other (e.g. number of students in the Faculty, those who passed secondary school, secondary technocal school, etc.);
    - ***Quality criteria*****:** the units of the population can be distinguished on the basis of some characteristics (e.g. gender, profession, hair colour);

- **Statistical data I.**
  - ***Basic data*:** obtained by measuring or counting;
  - ***Derived data*:** they are results of operations;

- **Statistical data II.**

  Numerical data:
  - ***Absolute data*:** derived from data collection, obtained by calculations (sum, difference, etc.).
  - **R*elative data:*** it can be calculated as the quotient of two statistics;
  - **Other data:** text data, they should be coded, in order to use for analysis;

- ***Data collection***

  Obtaining data for analysis, can be occurred by data collection, which may be:

  - **complete:** concerns each individual of the multitude;
  - **partial:** concerns only a part of the population;

Error of the statistical data

The data collected have usually limited accuracy.

According to its formation, the error could be**:**

- Data collection error,
- Random error.

According to its nature, the error could be:

- **Absolute error (a):** the difference between the actual data (A) and the measured data (A ').

  $$a=A-A'$$

- **Relative error (α):** quotient of the absolute error and the actual data, which is used to provide in percentage

  $$\alpha = a/A$$

When giving data we can talk about **significant digit** to. This means that the order of **only those numbers are described** (the error is written only in that order), **which is still believed to be reliable**. The majority of the data are rounded numbers.

- The **estimated absolute error** can be calculated with the following formula:

$$\hat{a} \leq \frac{10^k}{2}$$

wherein $\hat{a}$: the estimated absolute error

k: power of local value of the last significant digit (i.e. magnitude of the **estimated absolute error**)

- **The estimated relative error:** $\hat{\alpha} = \frac{\hat{a}}{A'}$

- **Task:**

Population of Hungary on January 1, 2010 was 10,013 thousand people.

$$\hat{a} \leq \frac{10^3}{2} = 500 \qquad \hat{\alpha} = \frac{500}{10013000} = 0{,}0000499 = 0{,}005\%$$

Sampling procedures

Each element of the population should have the same chance to get into the sample.

The sample size should be large enough so that our conclusions are transferred to to the population, as well.

Stratified sampling: The population is divided into groups according to specific criteria, and the proportions of the groups in the sample is retained.

Random sampling: The elements getting into the sample is selected by lottery.

Census:

- ***Probability sampling:*** all individuals of the population has the opportunity (greater than zero chance) to be included in the sample. The probability sampling includes the following sampling procedures:
  - simple random sampling;
  - systematic sampling;
  - stratified sampling;
  - multi-stage cluster sampling;

  The different sampling procedures are the same in two things:
  - every element has a greater than zero probability of getting into the sample;
  - in some respects they include random selection;

- ***Non-random sampling method:*** means a method in which there is no possibility for certain elements of the population to be included in the sample, or the likelihood of selection can not be accurately determined. The non-random sampling include the following:
  - arbitrary sampling;
  - concentrated sampling;
  - quota sampling;
  - snowball method sampling;

Statistical basics (2)

Case

An element of the sample is a row of the data matrix.

The number of elements of a given sample

The number of rows of data matrix.

Data matrix

Data of n cases and *p* variables arranged in a matrix.

Variable

A measurable characteristics of the population. A column of the data matrix.

Variables, as examples

- Citizens of Hungary: payment, age, gender, party, etc.
- Students of a university faculty: average of marks of exams, codes of electronic diary for subjects, gender, subject, number of fulfilled credits, etc.
- A set of cars: acceleration, consumption, horsepower, type;...
- A set of costumers of a particular product: opinion on the price, quality, ...
- A set of viewers of a TV channel: age, gender, index of enjoyment, education, etc.

Statistics

Calculated value of a statistics is a calculated data computed by a formula on the data of the realizations of the sample.

mean, standard deviation, median, quartiles, skewness, kurtosis, modus, frequency, test statistics, etc.

Statistical basics (3)

- ***Statistical population:*** all individuals covered by the statistical observation.
- ***Units of the population:*** individuals that constitute the multitude.
- ***Criteria:*** based on them units of the multitude can be characterized.

Examples
Data sets

- Sample
  - A group of students at the Faculty of Agriculture
  - Body weight of 20 students
  - …
- Multitude (population)
  - Students at the Faculty of Agriculture
  - Body weight of the students at the Faculty of Agriculture
  - Body weight of university students, in general
  - …

Variables and data

- **Variables (criteria):**

  „value" of the individual criteria of the population before sampling or observation, referred to herein: $X, Y, X_1, X_2, \ldots$ **(e.g. throwing a six-sided cube)**

- **Data (variant):**

  „value" of the criteria in the sample after sampling or data collection, referred to herein: $x, y, x_1, x_2, \ldots,$ **(e.g. a result of a throw of a six-sided cube)**

Variables and data, example

- The amount of rainfall for tomorrow at a given location
  - today only variable: X
  - the day after tomorrow already data, e.g. x = 8 mm

Types of criteria (variables)

- *qualitative* (can still be established) criteria
  - (e.g. ”gender”, ”colour”, ”office position”);
- *quantitative* (can already be measured) criteria
  - discrete (e.g. ”the number of schools at a given settlement”);
  - continuous (e.g. ”temperature at a given settlement and time)”;

Variants of qualitative criteria:

- **Classes, categories (these are data, as well!)**

E.g. type variants

gender **(criterion)**	man, woman **(variants)**
colour **(criterion)**	white, red, etc. **(variants)**

- **Dichotom criteria: they have only two variants**
- **Trichotom criteria: they have three variants**

Discrete quantitative variable (criterion)

- **Its possible values (variants) are**
  - **finite;**

    e.g. „the number of boys in a class of 30 people”

    it may be: 0, 1, 2, ……..,30

  - **countably infinite;**

    (practically no upper limit)

Continuous quantitative variables

- Its possible values are any value of an interval

e.g: blood sugar
ph-value
age
body weight
temperature

Data-transformations, re-scaling

Frequently, instead of measured (observed, established) data it is better to work with their „transfomed" versions.

- Sometimes quantitative data are qualified (e.g. boniting); (boniting: breeding bun for ship, in which it is observed that which males are favoured to copulate with females);

- The most common transformation of quantitative data:
  - log-transformation
  - square root transformation
  - reciprocal-making

3. Populations (multitudes)

Observation unit – Reporting unit

Totality of observation units are called statistical population.

Type	Discrete	Continuous	Aggregate
Stationary	Number of students	Flour stock of a bakery	Stock pile of a grocery
Moving	The number of guests of a hairdresser	Water consumption of a plant	Production value of a furniture factory

4. Criteria

Properties of observation units are called criteria.

Variants: possible outcomes of the criteria.

Types of criteria:

a) common - distinctive

b) quantitative (e.g. body weight, body height)

c) non-quantitative

- regional (place of birth of the students)
- temporal (year of birth of the students)
- qualitative (hair colour)

Measurement scales of criteria

- The **types of data are measured** on an adequate scale based on **arrangeability** and the **distance function among them**.

a) ***Nominal scale:*** typical qualitative scale. Its values are non-sequencable, belonging of the elements to identical or different categories can only be determined (X=Y) or (XY) (e.g. passport number).

b) ***Ordinal scale:*** a qualitative scale, on which the order of the categories can also established (X<Y) (e.g. level of education).

c) ***Interval scale:*** a qualitatve scale, on which the distance of two individuals (X-Y) can be measured. The scale has no real zero point. X=0 does not mean the lack of the criteria (e.g. temperature).

d) ***Ratio scale:*** this is a quantitative scale, where zero is true. In this scale the ratio of two values (Y / X) can be interpreted (e.g. weight).

Nominal scale (1)

- **The simplest form of measurement, an informal assign of numbers to things. All observed individuals are described with data, which are not comparable with one another according to size. The variants can only be measured on the basis of their identity or their differences. Order can not be established.**
- **Symbols or numbers assigned to the objects are only used to identify the objects or their classes (they have no other meaning!)**
- **Only distinctness is demanded, so equality relation can only be interpreted.**
- **E.g. passport number, numbering flights, shirt numbers, the name of the employee, location of birth, gender, ... , etc. are even nominal, if they are encoded by numbers: e.g. identification number of the employee. Other examples are: occupation, marital status;**
- **Calculated statistical indicators: in case of identification of classes: frequency, modus.**

Ordinal scale (2)

- **Variants are arranged not only under the same or different criteria but also according to a natural order.**
- **In addition to the relation of equality, the relation for sequencing is still valid.**
- **Units measured on the ordinal scale are not equidistant from each other!**
- **Arithmetic mean and standard deviation can not be calculated!!!!! Quantiles, median and rank correlation coefficient can be calculated.**
- **Any transformation can be performed, which preserves the original order of the scale.**
- **E.g. quality classification of products, a survey with 3-, 5-, 7-stage scale, fire hazard classes, school marks and evaluation categories (excellent, good, etc.) numbers of placement in a race, etc.**

Ordinal scale (2)

- **It means measurement because the scale differences give real information on the units of the population. Any two data are comparable.**
  - E.g.: educational level of workers.
- Characteristics:
  - Distance is no interpreted among data. *(E.g. it cannot be told that how valuable is your graduation than 8 primary.)*
  - **Only operation: sorting data** – Sorting data - a **rank statistics** can be used that can only be applied **for relative ordering of the data**. (For example, average is no point, but median and mode - these will be discussed later).

Interval scale (3)

- **It has the properties of ordinal scale + difference between any two points on the scale, that is, the distance is understandable and applicable.**
- **There is no fixed zero point, zero point of the scale and unit of measure can be selected freely.**
- **It is characterized by common and constant units and the numbers are assigned to sorted things accordingly.**
- **Any kind of linear transformation of the scale is allowed.**
- **Except for the geometric mean and relative standard deviation, all statistical indicators and properties can be calculated.**
- **E.g. temperature, calendar time, altitude**

Interval scale (3)

E.g. rank ordered properties are separated by **equal intervals**. (For example, IQ tests: 100-110, 110-120: equal distances. But: we can not say that the 150-IQ individuals are by 50% smarter than the 100-IQ people.)

- It is meaningless to talk about a ratio relative to each other, e.g. 20°C is twice as hot as 10°C. (Temperature on the Kelvin scale is not of interval type!)

- Differences of the scale of values provide actual information on the individuums of a multitude (population).

  [E.g. when comparing two people it can be said that they **differ (nominal)**, one is higher than the other **(ordinal)** and we can tell how much **(interval)**].

Ratio scale (4)

- **This is the strongest and highest order of measurement form.**
- **It has the properties of the former scales, in addition additivity requirements are also fulfilled:**

  The scale has a real zero point, and the ratio of any two points is independent of the measuring unit.

- **E.g. measurement of production, sales, income, earnings, etc.**

Ratio scale (4)

- Data that can be characterized by real numbers.
- All operations can be peformed with them that can be perfomed with the real numbers.

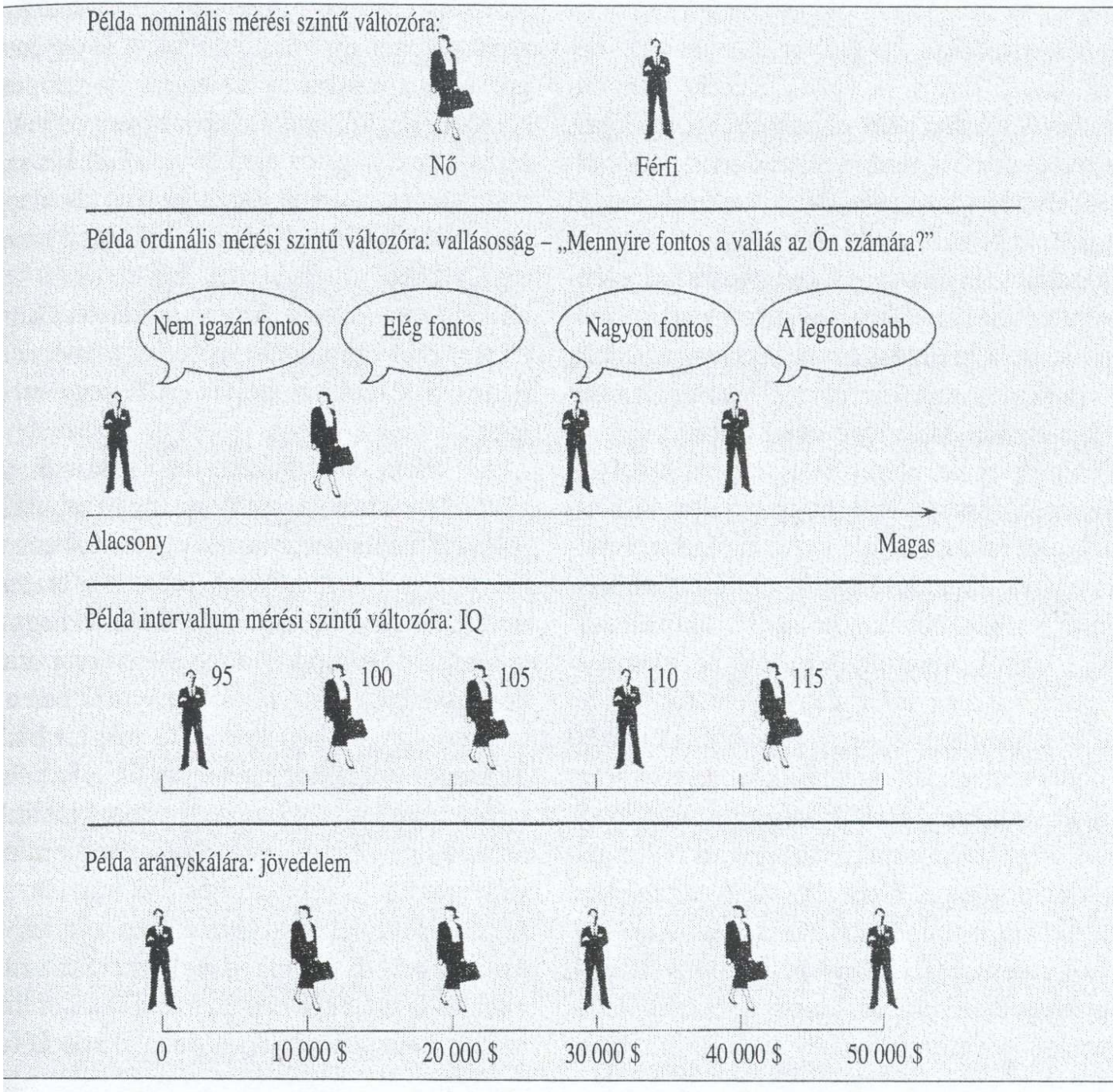
The starting point is clearly defined and fixed, so that the **proportion of the scale values** can also be determined (e.g. age, income, etc.).

Measurement scales	Property	Understandable relations	Features	Examples
Nominal	Distinction	$X_a=X_b$ or $X_a \neq X_b$ Can be calculated: Frequency, modus,	Not quantified	Name, place of birth, gender, passport number, numbering flights,
Ordinal	Distinction, order	$X_a=X_b$ or Xa$\neq X_b$ and $X_a \geq X_b$ or $X_a<X_b$ Can be calculated: Quartiles, median, rank correlations coefficient	It is hard to measure, can only be ordered	Sequences (military) rankings, education, quality classification of products, survey with 3-, 5-, 7-stage scales, fire class
Interval	Distinction, order, difference	$X_a=X_b$ or $X_a \neq X_b$ and $X_a \geq X_b$ or $X_a<X_b$ Interpretable: X_a-X_b Can be calculated All statistical characteristics, except for: geometric mean and relative standard deviation	Positive and negative values	Temperature scale (except Kelvin scale), calendar time, altitude, equidistant categories of IQ-tests
Ratio	Distinction, order, difference, ratio	$X_a=X_b$ or $X_a \neq X_b$ and $X_a \geq X_b$ or $X_a<X_b$ Interpretable: X_a-X_b and X_a/X_b Can be calculated: All operations that can be performed with real numbers.	There is theroetical minimum, identical signs	Quantities with number or intensity of production, traffic, population, age, income, earning

Properties of data of different measuerement scales

Nominal criterion (variable)

Gender



Ordinal criterion (variable)

religiosity

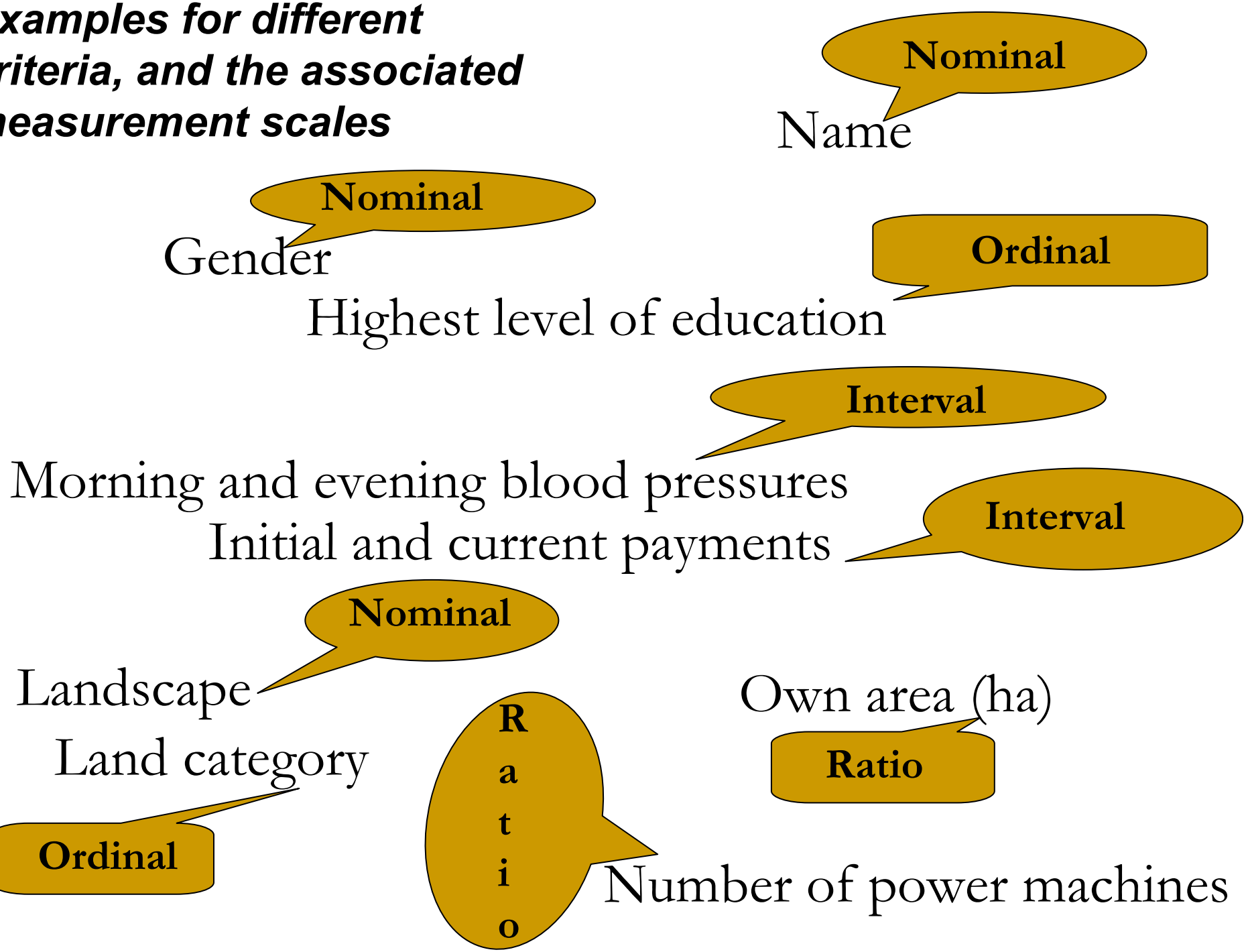
Interval criterion (variable)

IQ

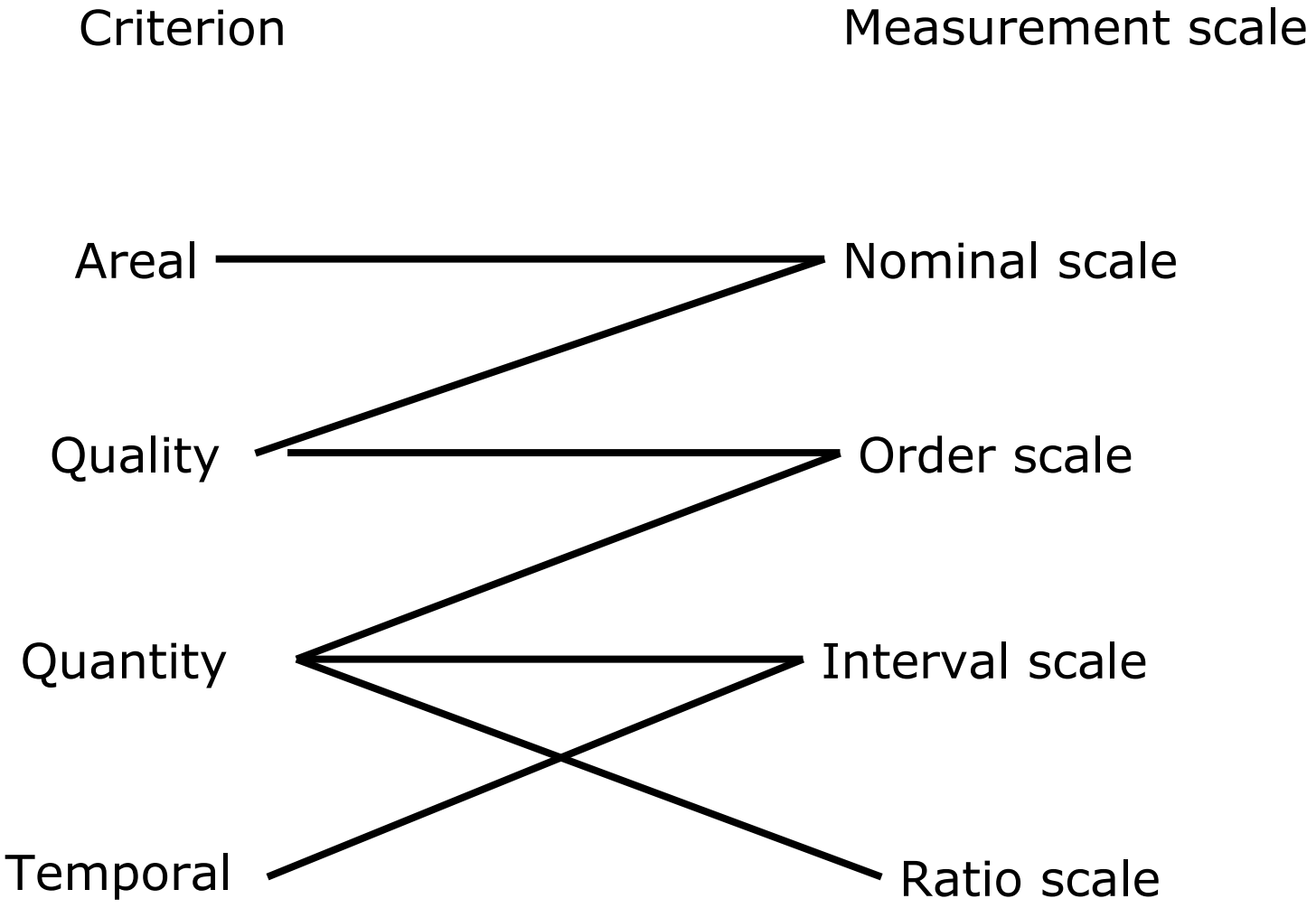
Ratio criterion (variable)

Annual income

Examples for different criteria, and the associated measurement scales



Criteria and measurement scales



Task 1.

Population	A concrete unit	Criterion	Variant	Type of criterion / measurement scale
Population of Hungary, January 1, 2007	Kiss Réka	**Date of birth**	1976	**Temporal / interval**
		Place of birth	Budapest	**Areal / nominal**
		Gender	**woman**	**Quality / nominal**
		Age	29	**Quantity / ratio**

Task 2.

Populations below are given:

- **Population of Hungary on January 1, 2006: 10 076 581 people**
- **Beer consumption of the Budapest men during the world championship, 2006**
- **Teachers of BCE on September 4, 2006**
- **Charity concerts at the Music Academy, 2006**

Task

- **Determine the types and units of the above multitudes!**

Task 3.

Define the types of the criteria below (quantitative / qualitative)!

- ❑ Gender (male, female)
- ❑ Age
- ❑ Body height
- ❑ Body weight
- ❑ Marital status
- ❑ Education
- ❑ Gross monthly salary

Statistical data and index

Information on individuals

(number or numerical characteristics)

Temporal identification

Spatial identification

Value

Unit

(measurement or calculation)

Index number

Statistical value calculated on some statistical method

For example:
(Monthly) mean salary in Hungary in 2009: 194.000 HUF/month

Statistical rows

Concise characterization of the population according to a criterion

According to the object of the row-making

grouper rows
comparative rows
descriptive rows

Real statistical rows
(From data of identical type)

Non-real statistical rows
(From different types of data)

According to variants

Temporal (duration), areal, quantity, quality + descriptive rows

Making rows: variants → values

- ***Statistical row***

  The statistical row is a list of **statistical data** in a specified context, namely it is listed as a result of **a single criterion**. The grouping provides an opportunity to study the structure of the population.

  Division of statistical rows:

- ***According to data types:***
  - ***Real rows*:** are related to a population and created from data of identical types
  - ***Non-real rows:*** give a description of any given phenomenon in many respects, and concern several populations:
    - **Descripitive row:** describes different features of the object of the observation, there is a logic relationship among data. They are sources of intensity ratios.

- ***According to their formation, real rows can be further broken down for:***

  - **Comparative rows:** include data as a list:
    - State series
    - Description row
  - **Grouping rows:** consist of sub-populations according to the variant:
    - Quality row,
    - Quantity row
    - Areal row
    - Duration time series

- ***According to the type of the criterion, real rows may include***:
  - Time series,
  - Areal row,
  - Quality row,
  - Quantity row.

- ***Time series***

  List the details of phenomena that occur during an existing period of time or a certain time.

  - **Status time series:** shows temporal change of stationary population. It is no sense of summing up data (e.g. passenger cars by year).
  - **Duration time series:** shows temporal change of a moving population during a given time interval. It reflects a dynamic situation. Summing up data has of sense. When summing up, we can form a data being characteristic for a longer time period (e.g. the number of tourists by year).

- ***Regional rows***

  - If grouping is the organizing principle, then data of a larger regional unit are classified according to smaller territorial units. It contains spatially split sub-populations of the population recorded at the same time. The order of variant is free. Summing up frequencies has sense (e.g. distribution of the population by county).
  - However, if the comparison is the guiding principle, then summing up data has no sense.

- ***Quality rows***

  It is formed through a pre-recorded grouping of quality characteristics of the population elements. It serves insight into the composition and structure of the population. The order of variants is free, summing up frequencies (number of occurrences) has of sense (e.g. distribution of the foreigners, coming to Hungary, according to the character of the travel in 2015: tourist groups, individual tourist, transit).

- ***Quantity rows***

  They are grouped according to quantitative criteria. Quantitative rows will be created. If the variants are numbers given with a concrete accuracy, then we speak about discrete quantitative row. In general, they are obtained by counting. If the variants can take any values in a given interval, then we are talking about continuous quantity row.

- **Types of quantity rows:**

Frequency series: shows the frequency of occurrence of the criterion (they are the variants). Frequency (f_i) shows that how much time the individual variants occur in the observed population. If the individual frequencies ar related to the total frequency, we receive the relative frequency (g_i) of the given variant:

$$g_i = \frac{f_i}{\Sigma f_i}$$

where: g_i: relative frequency of the *i*-th variant

f_i: frequency of the *i*-th variant

$\Sigma f_i = n$: number of elements of the population

- If the number of variants is large, then the data are ranked that makes easier to classify the variable. Classification compresses the information. The interval given by the maximum and minimum values of the variants is divided into classes so that within each interval the frequencies are of almost equal distribution, so that the medium value of the class is able to characterize the class itself.

  - Classes are groups of data, where quantity changes between the different classes cover a qualitative change.
  - The number of class intervals is the smallest *k* for which : $2^k > N$
  - The length of the class interval:

$$h = \frac{x_{max} - x_{min}}{k}$$

  - If the group criterion is of continuous nature, the quantity row is called class interval frequency row (or series).

Example

For determining nitrogen-content of carrot, a 24-element sample was taken and the nitrogen-content was measured. The results of the measurement (mg/100g):

104,2	100,2	93,4	98,6	117,4	100,8	88,2	100,1
101,9	116,8	112,9	97,8	99,3	111,5	109,3	100,7
118,0	96,4	89,4	98,9	105,3	112,4	99,5	102,7

Example:
Prepare a rank from the data!
How much is the optimal number of the intervals?
Solution:

88,2	*89,4*	*93,4*	*96,4*	*97,8*	*98,6*	*98,9*	*99,3*
99,5	*100,1*	*100,2*	*100,7*	*100,8*	*101,9*	*102,7*	*104,2*
105,3	*109,3*	*111,5*	*112,4*	*112,9*	*116,8*	*117,4*	*118,0*

n=24 $2^5=32 > 24$, hence k=5 $\rightarrow$ h=(118,0-88,2)/5=5,96.
$\Rightarrow$ The length of the interval is around 6

- **Value amount row**: If multiplication of variants are shown at frequency rows then we receive a value-amount row.

  Value-amount is indicated by: s_i.

$$s_i = f_i * x_i$$

Example

Table 1

Distribution of families according to the number of children, at a given settlement

Number of children	Number of families	Number of all children (s_i)
0	992	0
1	954	954
2	761	1522
3	148	444
4	40	160
5	15	75
6	5	30
Total	2915	3185

- **Cumulative frequency row**: We can gain special information by calculating cumulative frequencies of the frequency row. Cumulation means cumulative summation, and informs us that how many times occur a value smaller than a limit (cumulative from below), or a value higher than a limit (cumulated from above). Cumulative frequency is indicated by: f_i'.

- ***Statistical table***

A comprehensive system of statistical rows. It is received when arranging data on several characteristics.

❖ **Delivery may be by:**

basic or collection table,
procession, or a work table,
communication or scoreboard.

❖ **Based on the distribution of the type of the summarized rows:**

✓**Simple table:** a comprehensive system of data rows comprising no grouping. Usually it contains descriptive and comparative rows.

Name	2007	2008	2009
Number of businesses			
Equity			
Foreign capital			

- ✓ **Grouping table:** a comprehensive system of statistical rows comprising a grouping based on one criterion.

Age	2007	2008	2009
0-6			
7-14			
15-22			
.			
.			
Total			

- **Combination table:** a comprehensive system of combinative classification based on more than one criterion.

Names	Budapest	Other city	Village	Total
With comfort				
With half-comfort				
Without comfort				
Total				

- ***Preparing statistical tables***

  Formal and substantive requirements must prevail to meet the purpose of the investigation.

- ***Formal requirements:***
  - title of the tables,
  - names in overhead and side box,
  - indication unit,
  - source.

.

Example

Table 3

The number of people receiving child protection care, 2005

1st column: age category (year); 2nd column: number of children (person)

Age category (year)	No. of children (person)
0-3	1848
4-5	1221
6-9	3075
10-11	2164
12-14	3827
15-17	5010
Total	17145

Source: Magyar statisztikai zsebkönyv (Hungarian Statistical Pocket Book), 2006.

Content requirement:

Every cell in the table should be used to communicate information, so the following standard abbreviations are used:

- specific data: numerical information,
- pulled out box (-): no information
- …: data exists but is not available
- +: data is the result of a preliminary estimation
- *: indicates a comment at the bottom of the table
- 0.0: there is data, but it is too small compared to other data.

Statistical rows (series)

- **<u>Grouping statistical row</u>:**

  It shows the internal consistency of the population, it is made for grouping, its data can be summarized

  (temporal, areal, qualitative, quantitative)

Criterion	Number of units
C1	f1
C2	f2
.	.
.	.
Ci	fi
.	.
Ck	fk
Total	N

Statistical rows (series)

- For example:

  Students sitting in the class room according to their hair colour.

  Qualitative grouping statistical row (series)

Criterion (hair colour)	No. of students
Brown	23
Blond	12
Black	4
Red	2
White	2
Other	1
Total	44

Statistical rows (series)

- <u>Comparative statistical row:</u>

  Comparative data arranged in row with the aim of comparison, its data cannot be summated.

  (time series, areal)

Criterion	Value / unit
C1	data
C2	data
.	
.	
Ci	data
.	
Ck	data

Statistical rows (series)

- For example:

  Mean monthly scholarship of full-time students of a higher educational institution, 2004-2010

  Comparative temporal row (series)

Criterion (year)	Monthly mean scholarship (HUF/ student)
2004	12.600 Ft
2005	13.200 Ft
2006	13.800 Ft
2007	14.100 Ft
2008	14.000 Ft
2009	14.200 Ft
2010	15.000 Ft

Statistical rows (series)

Components of statistical rows:

- Title (exact nomination of the population, listing of the common criteria)
- Listing of characteristics and variants
- Listing frequencies of variants
- Column „Total” (only in case of grouping row)
- Nominating of the source

Statistical tables

Comprehensive system of statistical rows

- Simple table (comparative and/or descriptive rows)
  It has no grouping row, one of its data is an element of a statistical row.
- Grouping table (grouping and/or comparative or descripitive rows)
  It contains a one-way grouping, one of its data is an element of a statistical row.
- Combination table (grouping rows)
  It contains only grouping rows, one of its data is an element of more than one statistical row at the same time.

Statistical tables

- Simple statistical table

Medical supply in a city

Year	No. of physicians	No. of inhabitants	**No. of physicians / inhabitants**
1990	240	80 000	333,3
1999	360	100 000	277,8

Statistical tables

- Grouping statistical table

Data of wheat production, 1991

Region	Production, thousands tons	Mean production, tons / ha
Over the Danube area	2000	5,2
Great Plain	3000	5,31
North	705	4,71
Total	**5705**	...

Statistical tables

- Combination statistical table

Marks of full-time students of a higher educational institution on statistics, 1991/1992, seminar 2

Mark	A	B	C	Total
	Distribution of the students according to faculties			
5	19	23	19	61
4	32	49	40	121
3	24	36	56	116
2	20	36	82	138
1	1	2	18	21
Total	96	146	215	457

Statistical tables

Parts of statistical tables

- Column (a column of the table)
- Row (a row of the table)
- Cell (the section of a row and a column)
- Box head (1st row of the table comprising the variants of a criterion as text)
- Column heading (1st column of the table comprising the variants of a criterion as text)
- Box amount (comprising the total amount of rows and columns)

Statistical tables

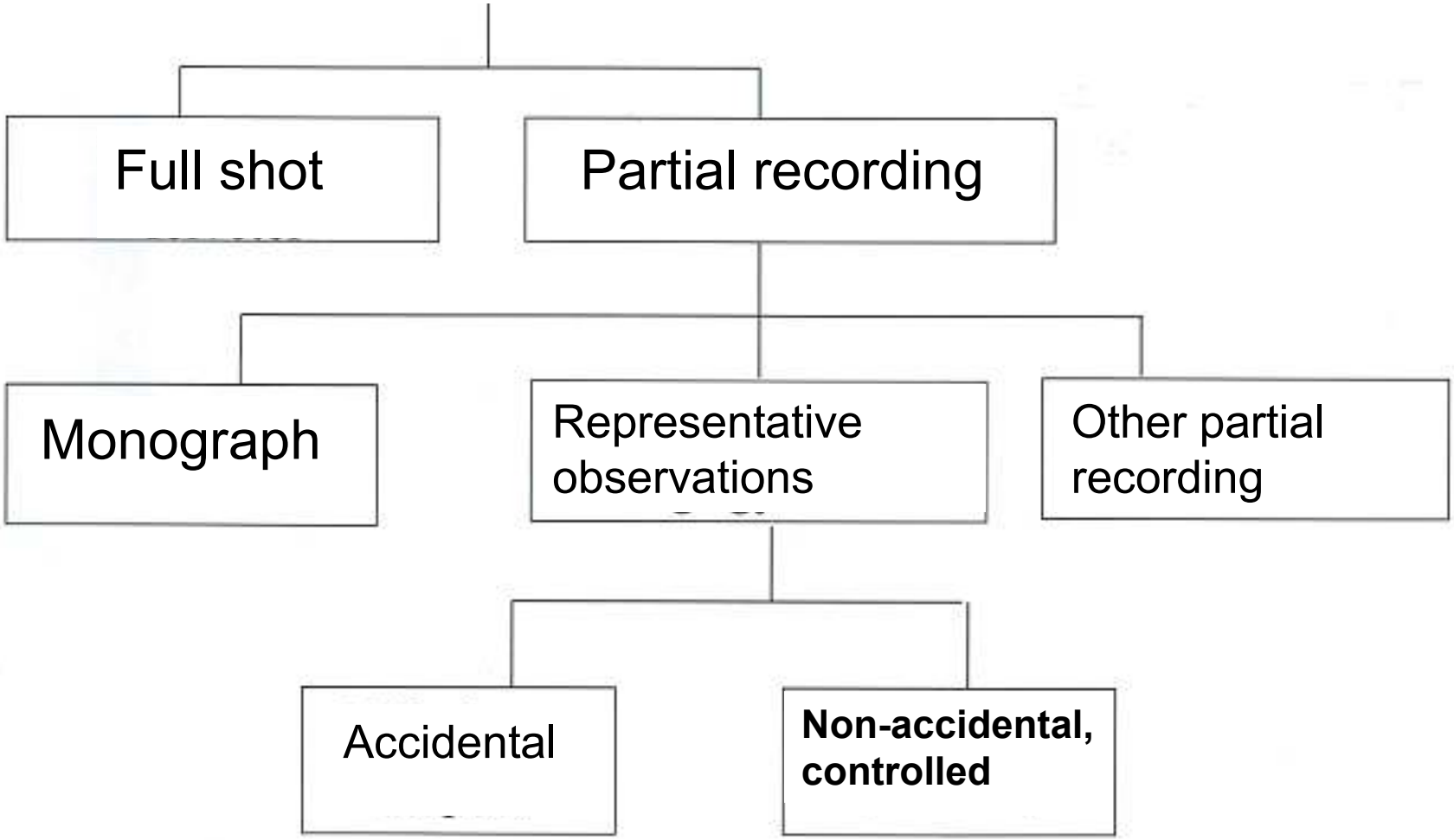
Number of dimension:

shows that a statistical data of the table is an element of how many statistical rows at the same time;

Rules of making tables

- Title (with identifiers! Time, location, etc.)
- Nomination of column pages (box head, column heading)
- No any empty box ; (--, •(•); … , 0,0)
- Comment (if a data in any column is of different unit)
- Source

Data acquisition modes



Constructing a questionnaire

- Thorough professionalism;
- Concise, clear, easy questions answered;
- Mostly multiple choice (empty circle, x, low exerted answers);
- Should not be too long;
- Recommended: anonymous survey;
- Compromise: just ask the most important things; before finalizing a test query;
- If prizes offered, response rate can be increased;

Questionnaire, László Makra, survey, for a research analyzing possible parameters associated with genetic and health related parameters, habit, household and environment in risk assessment of respiratory diseases

Nem: férfi 1 nő 2

Születési dátum négyjegyű év–hó–nap:
Születési hely (irányítószámmal):
Jelenlegi/utolsó foglalkozás:
Korábbi foglakozás:

Iskola (aláhúzással jelölve):
Kevesebb, mint 8 általános
8 általános
Középiskola (szakiskola vagy egy része)
Főiskola, vagy egy része
Egyetem, vagy egy része

Apa betegségei (max. 5 db betegség):

Anya betegségei (max. 5 db betegség):

Testvérek betegségei (minden egyes testvér esetében külön-külön, max. 5 db betegség):

Ha gyermekei vannak, a gyermek(ek) betegségei (minden egyes gyermek esetében külön-külön, max. 5 db betegség):

Saját betegségei (max. 5 db betegség):

Anyatejjel táplálták: igen – nem (aláhúzással jelölve):
Hány hónapig?:

Nem allergiás betegségei (aláhúzással jelölve):
Magas vérnyomás: igen – nem
Érbetegség: igen – nem
Szívbetegség: igen – nem
Tüdőbetegség: igen – nem
Cukorbetegség: igen – nem
Elhízás: igen – nem
Daganatos betegség: igen – nem

Allergiás megbetegedései (aláhúzással jelölve) saját megítélése szerint. (Ha az alábbiaktól eltérő(ek), akkor írja ide szövegesen a betegség(ek)et!)

Porallergia
Virágpor allergia
Ételallergia
Bőrallergia
Háziállat allergia
Gyógyszerallergia
Rovarcsípés allergia

Dohányzás (aláhúzással jelölve):
Nem dohányzik
Leszokott
Dohányzik, ... éve, ...db cigaretta/nap

Alkoholfogyasztás (aláhúzással jelölve):
Alkoholt nem fogyaszt
Hetente fogyaszt .....miből, mennyit (szöveges)
Naponta fogyaszt .....miből, mennyit (szöveges)

Lakókörülmények (aláhúzással jelölve):
Városi lakásban lakik: igen – nem
Ezen belül lakótelepi lakásban lakik: igen – nem
Családi/falusi házban lakik: igen – nem

Lakás paraméterei (aláhúzással jelölve):
Ház anyaga: Beton – Tégla – Vályog
Falak: szárazak – nyirkosak
Padló: parketta (laminált padló) – szőnyegpadló – kő – egyéb
Ágynemű: toll – nem toll
Háziállatok: kutya – macska – baromfi – sertés – tehén

Vannak-e olyan tünetei, amelyek allergiára utalnak? Igen – nem (aláhúzás)

Van-e orvos által diagnosztizált allergiája? Igen – nem (aláhúzás)

Milyen gyógyszereket szed rendszeresen? (felsorolás)

Always look on the bright side of things!

We finished for today, goodbye!

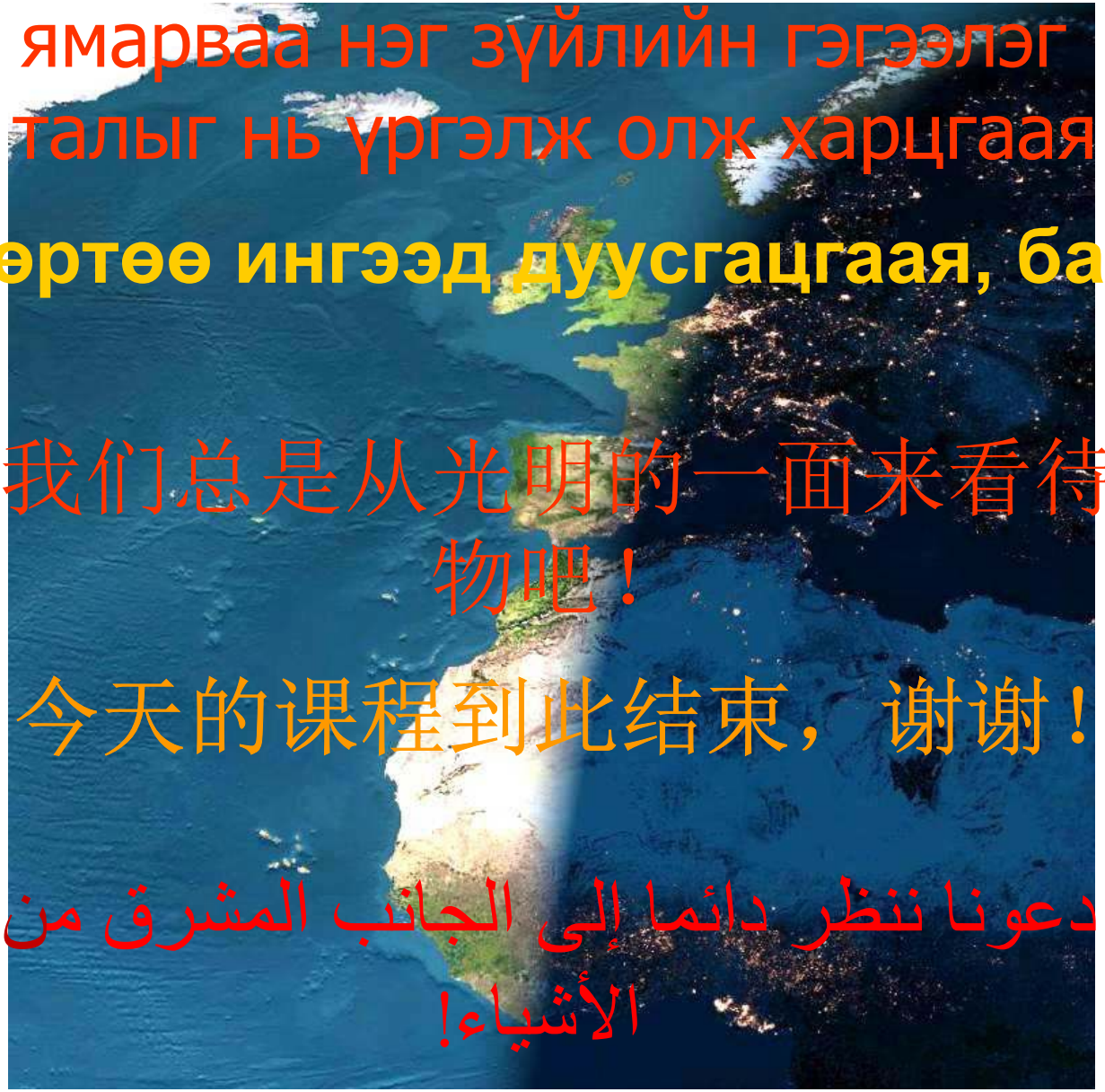
ямарваа нэг зүйлийн гэгээлэг
талыг нь үргэлж олж харцгаая

өнөөдөртөө ингээд дуусгацгаая, баяртай

让我们总是从光明的一面来看待事物吧！

今天的课程到此结束，谢谢！

دعونا ننظر دائما إلى الجانب المشرق من الأشياء!



انتهينا لهذا اليوم، وداعا!