Topic 1 Introduction to statistics, Qualitative and Quantitative data sorting

Statistics – štatistika

Definition:
Statistics is the science of:

- collecting
- organizing
- summarizing
- analyzing
- and making inferences from data.

The subject of statistics is divided into two broad areas:

- descriptive statistics
- inferential statistics.

Data – údaje, dáta

Definition:
Data are the values or measurements that variables describing an event can assume.

Data set – množina údajov

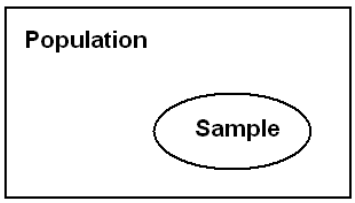
Definition:
Data set is a collection of values and each value is called a data value.

Population – základný súbor

Definition:
A population consists of all elements that are being studied.

Sample – výberový súbor

Definition:
A sample is a subset of the population.



Statistical elements, units – štatistické jednotky

Statistical attribute – štatistický znak

Definition:
There are two types of statistical attributes (variables):

- qualitative – data values that can be categorized, f.e. gender, residence, hair colour
- quantitative – numeric values, f.e. height, weight.

Elaborated by: Ing. Martina Majorová, Dept. of Statistics and Operations Research, FEM SUA in Nitra
Reference: JAISINGH, L.: Statistics for the Utterly Confused

Qualitative statistical attributes (variables) are **nonnumeric** in nature and can be classified into two groups:

- dichotomic – data values can assume 2 possibilities exactly (gender – either male or female)
- polynomic – data values can assume more than 2 possibilities (education – primary school, secondary school and university).

Quantitative statistical attributes (variables) can assume **numeric values** and can be classified into two groups:

- discrete – numeric values that can be counted (integer values), f.e. N$^{o.}$ of children
- continuous - numeric values that can be measured and so can assume also decimal values, f.e. monthly salary

---

Qualitative data sorting (sorting of qualitative statistical attributes) – triedenie kvalitatívnych štatistických znakov

---

Quantitative data sorting (sorting of quantitative statistical attributes) – triedenie kvantitatívnych štatistických znakov

---

Pivot table (frequency distribution) – asociačná tabuľka

Definition:
A frequency distribution is an organization of raw data in a tabular form, using classes and frequencies. We construct pivot table when all statistical attributes are *dichotomic*.

Count of ID	DOMICILE		
GENDER	city	village	Grand Total
F	18	10	28
M	26	18	44
Grand Total	44	28	72

Contingency table – kontingenčná tabuľka

Definition:
A contingency table describes the distribution of *two or more* statistical attributes (variables) simultaneously from which at least one of them is *polynomic*.

Count of ID		EDUCATION			
GENDER	DOMICILE	PS	SS	U	Grand Total
F	city	2	6	10	18
	village	2	4	4	10
F Total		4	10	14	28
M	city	8	6	12	26
	village	4	6	8	18
M Total		12	12	20	44
Grand Total		16	22	34	72

Frequency (frequency count; n_i) – absolútne (bežné) početnosti

Definition:
The frequency or the frequency count for a data value is the number of times the value occurs in the data set.

Relative frequency (f_i) – relatívne početnosti

Definition:
The relative frequency for any class is obtained by dividing the frequency for that class by the total number of observations.

$$f_i = \frac{n_i}{n}$$

Number of observations (n) – počet pozorovaní

Cumulative frequency (N_i) – kumulatívne absolútne početnosti

Definition:
The cumulative frequency for a specific value in a frequency table is the sum of the frequencies for all values at or below the given value.

Cumulative relative frequency (F_i)– kumulatívne relatívne početnosti

Definition:
The cumulative relative frequency for a specific value in a frequency table is the sum of the relative frequencies for all values at or below the given value.

UL	*ni*	*fi*	*Ni*	*Fi*
159	4	5,56	4	5,56
164	8	11,11	12	16,67
169	8	11,11	20	27,78
174	12	16,67	32	44,44
179	8	11,11	40	55,56
184	18	25,00	58	80,56
189	10	13,89	68	94,44
194	2	2,78	70	97,22
More	2	2,78	72	100,00
	72	100,00		

Grouped frequency distribution – intervalové (skupinové) rozdelenie početností

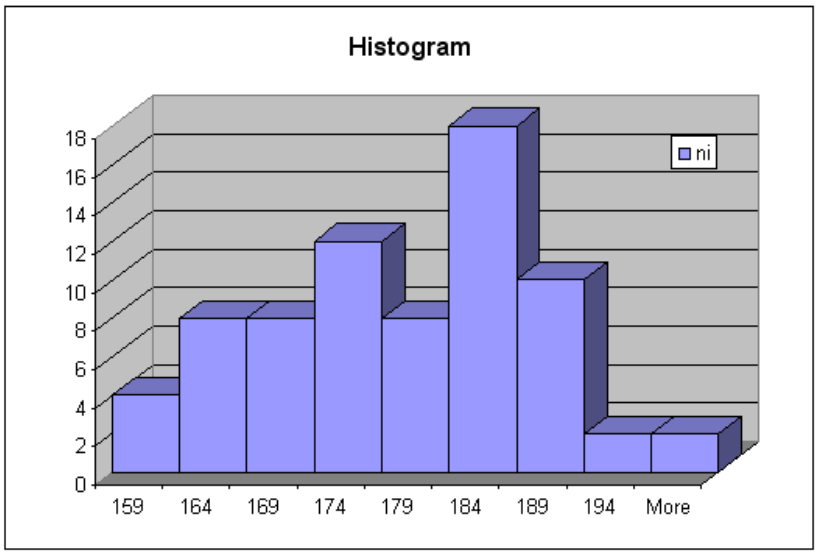
Definition:
A grouped frequency distribution is obtained by constructing classes (or intervals) for the data, and then listing the corresponding number of values (frequency count) in each interval.

Ungrouped frequency distribution – rad rozdelenia početností

Definition:
An ungrouped frequency distribution simply lists the data values with the corresponding number of times or frequency count with which each value occurs.

Histogram (bar graph) – histogram

Definition:
A histogram is a graphical display of a frequency or a relative frequency distribution that uses classes and vertical bars (rectangles) of various heights to represent the frequencies. It is a graphical representation of *grouped frequency distribution*.



Frequency polygon (line plot) – frekvenčný polygón

Definition:
A frequency polygon is a graph that displays the data using lines to connect points plotted for the frequencies. The frequencies represent the heights of the vertical bars in the histograms. It is a graphical representation of *ungrouped frequency distribution.*