

INTRODUCTION TO MEDICAL STATISTICS: FREQUENCY DISTRIBUTIONS AND GRAPHIC PRESENTATION

Mirjana Kujundžić Tiljak

FIRST STEP

→ summarize data to get a “feel” for the data

DIAGRAMS

TABLES

SUMMARY STATISTICS

ONE VARIABLE

EMPIRICAL FREQUENCY DISTRIBUTION of variable relates each possible observation, class of observations (i.e. range of values) or category, as appropriate, to its observed **frequency** of occurrence

replacement of each frequency by a **relative frequency** (the percentage of the total frequency)
→ to compare frequency distributions in two or more groups of individuals

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

3

CATEGORICAL OR DISCRETE NUMERICAL DATA

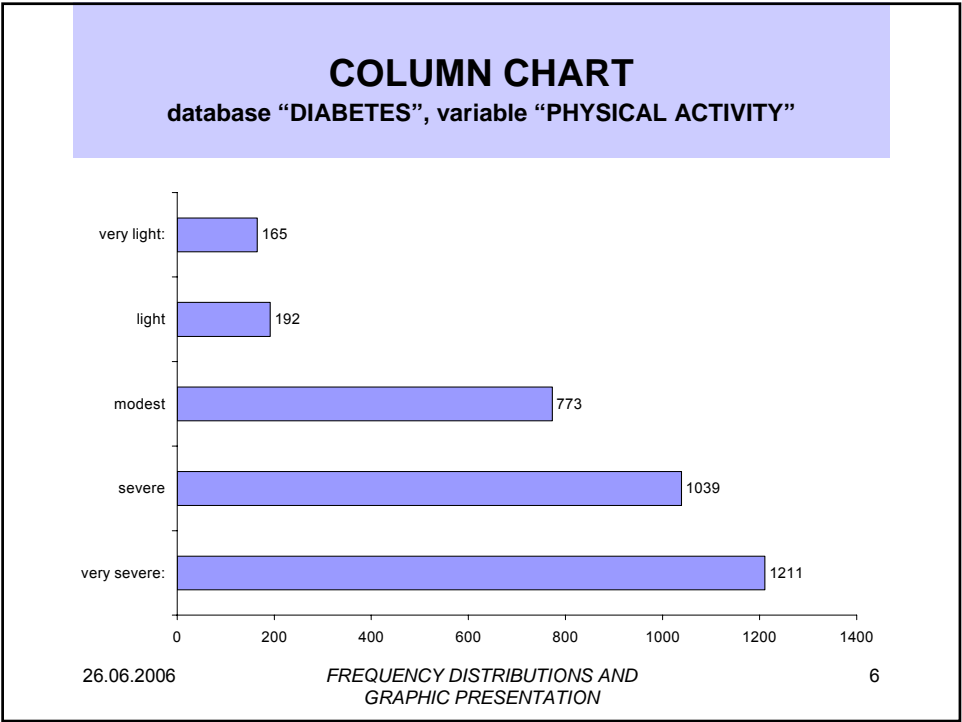
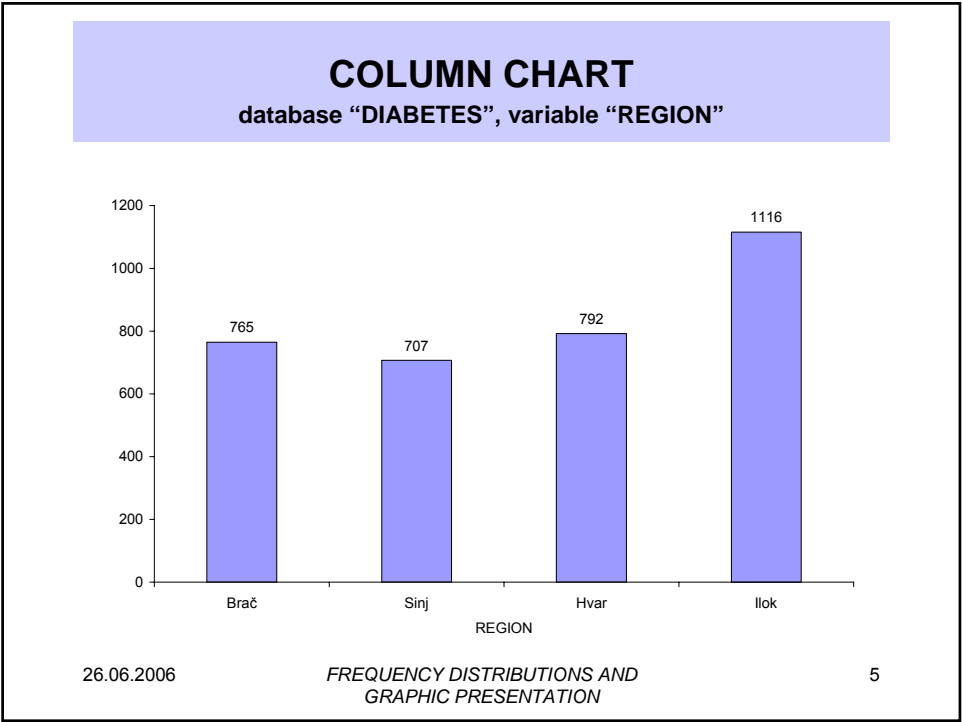
BAR or COLUMN CHART

- separate horizontal or vertical bar is drawn for each category,
- bar's length being proportional to the frequency in that category
- bars are separated by small gaps to indicate that the data are categorical or discrete

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

4



PIE CHART

- circular “pie” is split into sections
- one section for each category
- area of each section is proportional to the frequency in that category

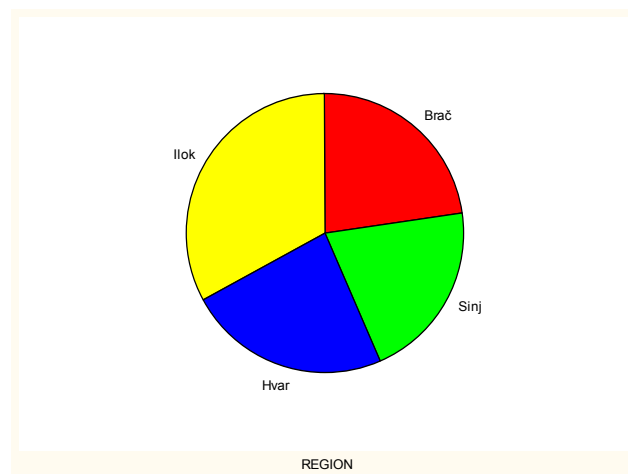
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

7

PIE CHART

database “DIABETES”, variable “REGION



26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

8

CONTINUOUS NUMERICAL DATA

more difficult to display

→ the data may need to be summarized before being drawn

HISTOGRAM

- similar to a bar chart - no gaps between the bars as the data are continuous
- width of each bar of the histogram relates to a range of values for the variable

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

9

- the area of the bar is proportional to the frequency in that range
- if one of the groups covers a wider range than the others – base will be wider and height shorter to compensate
- usually between five and 20 groups are chosen
- ranges should be so narrow to illustrate patterns in the data, but not so narrow that they are the raw data
- histogram should be labeled carefully – to make it clear where the boundaries lie

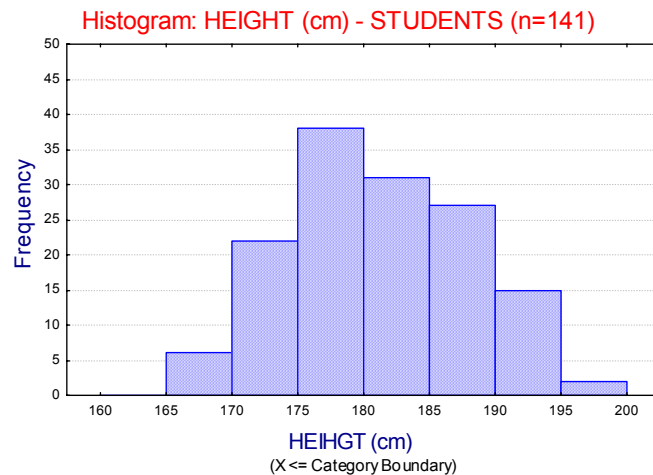
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

10

HISTOGRAM

database "STUDENT", variable "HEIGH", male students



26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

11

DOT PLOT

- each observation is represented by one dot on a horizontal or vertical line
- very simple to draw
- summary measure of the data (i.e. mean or median) is shown on the diagram
- may be used for discrete data

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

12

DOT PLOT

database "DIABETES", variable "AGE", male examinees



26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

13

STEM-AND-LEAF PLOT

- mixture of a diagram and a table
- looks similar to histograms turned on its side
- Is effectively the data values – written in increasing order of size
- vertical **stem** – first few digits of the values, arranged in order
- Protruding from the stem = **leaves** (i.e. final digit of each of the ordered values, which are written horizontally in increasing numerical order)

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

14

STEM AND LEAF PLOT

database "STUDENT", variable "HEIGHT", female examinees

Stem and Leaf Plot: HEIGHT, FEMALES

one leaf=1 case

	Class n	Percentiles
15 3	2	
15 67889	8	
16 00000001222222333344444	44	25%
16 5555566677777778888888889999	58	median
17 00000000000111112222222233334444	66	75%
17 555555667889	21	
18 00134	7	
18	0	
min = 153,0000 max = 184,0000 Total N:	206	

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

15

BOX PLOT or BOX-AND-WHISKER PLOT

- vertical or horizontal rectangle
- the ends of the rectangle = upper and lower quartiles of the data values; standard deviation
- line drawn through the rectangle = median value; mean value
- whiskers, starting at the ends of the rectangle = minimum and maximum values; particular percentiles (e.g. 5th and 95th percentiles)
- outliers may be marked

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

16

FREQUENCY DISTRIBUTION – “SHAPE”

- The choice of the most appropriate statistical method will often depend on the shape of distribution
- **unimodal** – single “peak”
- **bimodal** – two peaks
- **uniform** – each value is equally likely, no peaks
- Unimodal distribution – main aim = to see where the majority of the data values lie, relative to the maximum and minimum values

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

17

FREQUENCY DISTRIBUTION – “SHAPE”

- **symmetrical** – centered around some mid-point, with one side being a mirror-image of the other
- **skewed to the right (positively skewed)** – long tail to the right with one of a few high values
- **skewed to the left (negatively skewed)** – long tail to the left with one or a few low values

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

18

TWO VARIABLES

- one variable is categorical → separate diagrams showing the distribution of the second variable can be drawn for each of the categories
- clustered or segmented bar or column charts

26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

19

TWO VARIABLES

- both continuous or ordinal variable – the relationship between the two
→ SCATTER DIAGRAM
- plots one variable against the other in a two way diagram
- X variable – on horizontal axis
- Y variable – on vertical axis

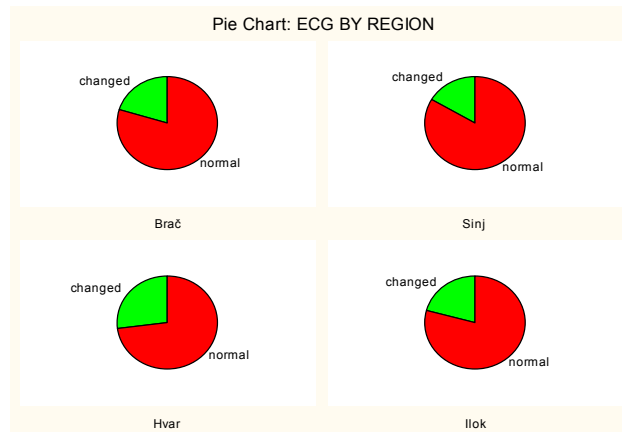
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

20

TWO CATEGORICAL VARIABLES

database "DIABETES", variables "REGION" and "ECG"



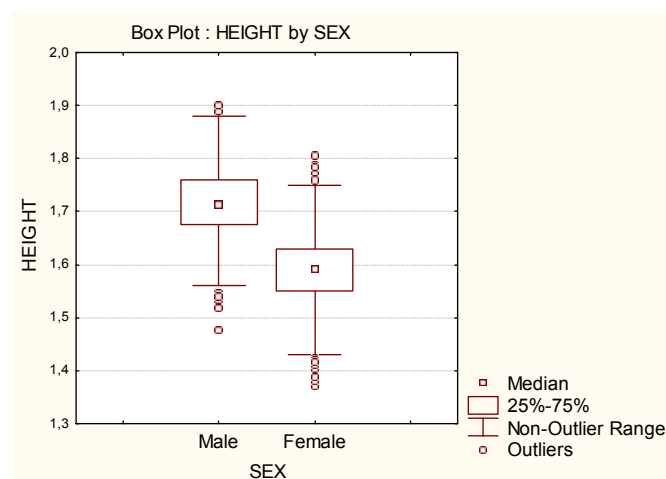
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

21

CATEGORICAL and NUMERICAL VARIABLE

database "DIABETES", variables "HEIGHT" and "SEX"



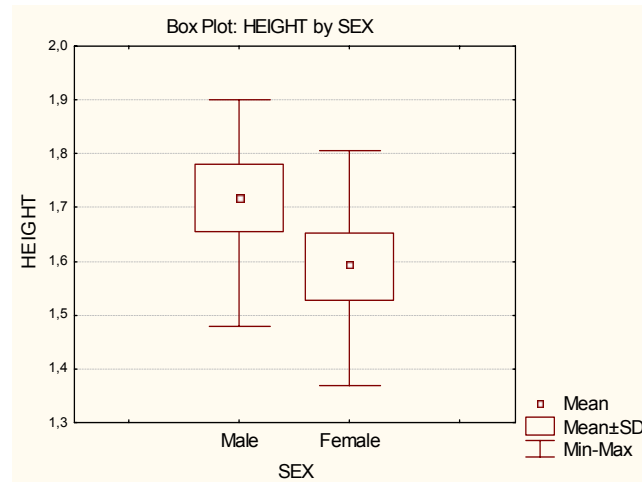
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

22

CATEGORICAL and NUMERICAL VARIABLE

database "DIABETES", variables "HEIGHT" and "SEX"



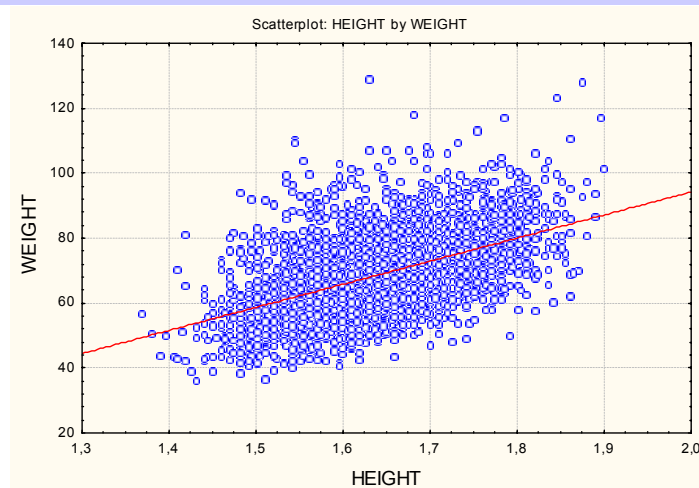
26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

23

TWO CONTINUOUS VARIABLES

database "DIABETES", variables "WEIGHT" and "HEIGHT"



26.06.2006

FREQUENCY DISTRIBUTIONS AND
GRAPHIC PRESENTATION

24