Graphical representation of the frequency distribution

Graphs for Frequency Distributions

Graphs allow one

◘ to visualize the main characteristics of the frequency distribution

◘ to display anomalous values or relevant patterns in the data

◘ to compare the distribution of different subgroups of the sample

Hence graphs are an effective way to summarise and communicate the information in the data

Graphs for numerical variables: Vertical lines

Put the X values $x_1, \ldots, x_k$ on the abscissa axis and the corresponding frequencies $n_1, \ldots, n_k$ on the y axis.

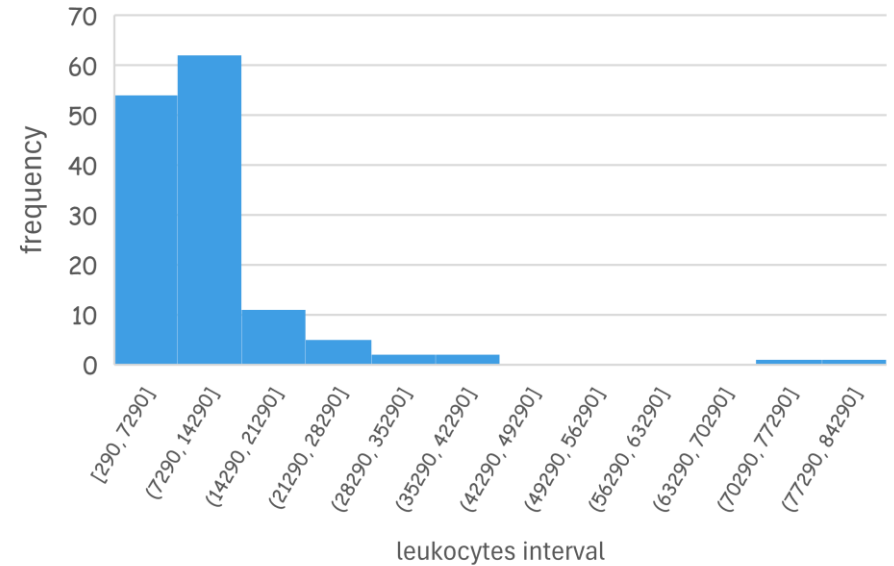Draw for each x_i a vertical line up to (x_i, n_i)
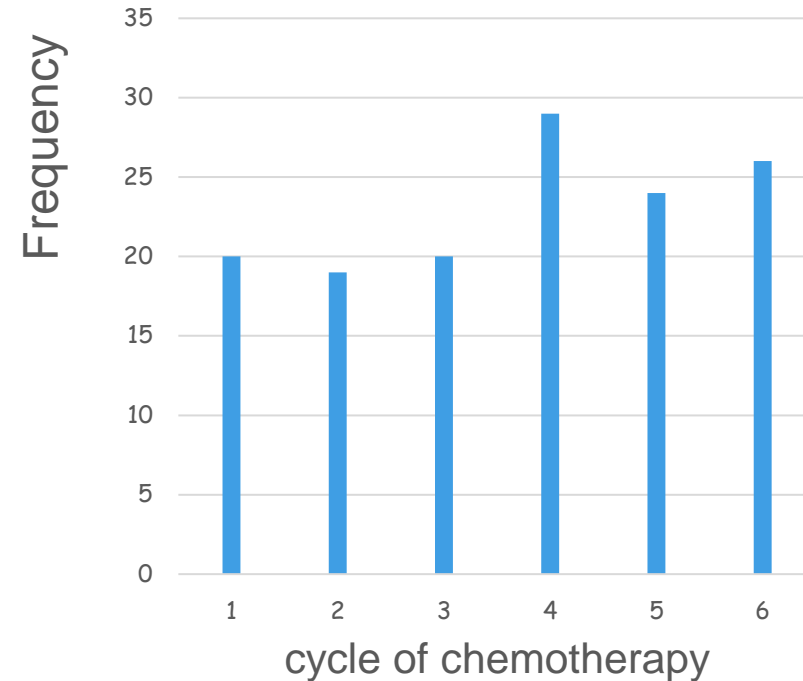
A similar graph can be made using relative frequencies (or percentages)

Frequency and relative frequency distributions have the same shape, only the y-axis scale changes

Graphs for numerical variables: histograms

A histogram is a graph suitable for representing
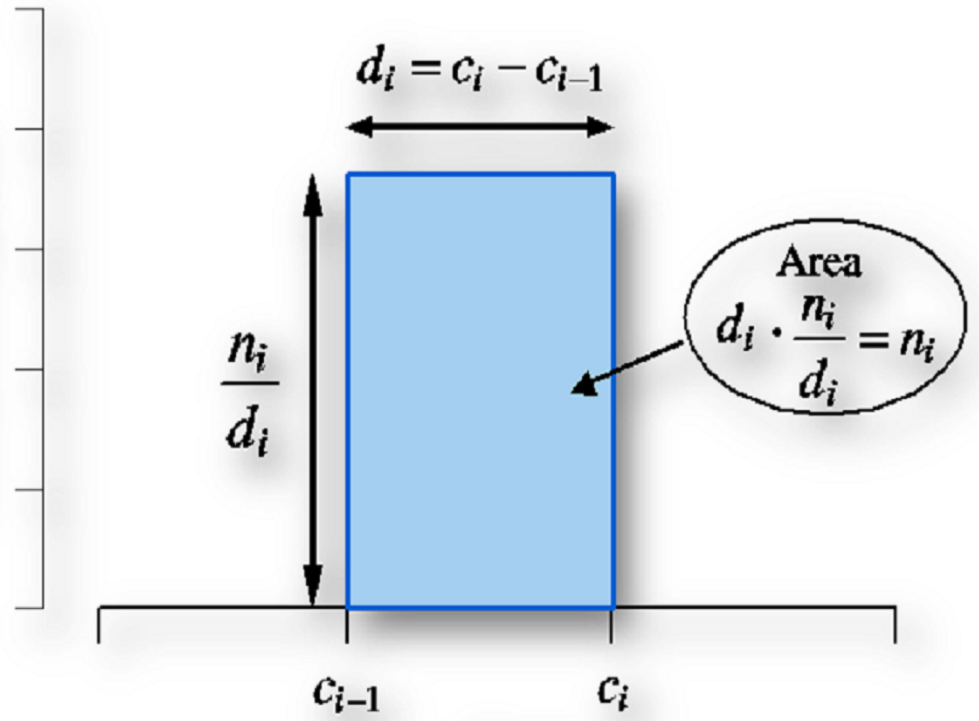
a grouped frequency distribution.

It is obtained by reporting the

class endpoints on the x-axis and the *frequency densities* on the y-axis and drawing a

rectangular bar over each class.

The height of the bar is equal to the *class frequency density*.

Therefore, the area of each bar equals the class frequency.

Histogram bars

The ratio

$$h_i = \frac{n_i}{d_i}$$

is called the frequency density

Plotting the frequencies instead of frequency density on the vertical axis produces misleading effects when the classes have different widths

Graphs for categorical variables: bar charts

In the bar chart categories are displayed in the y-axis (or x-axis). Each category is represented by a bar whose length (or height) equals the corresponding sample frequency
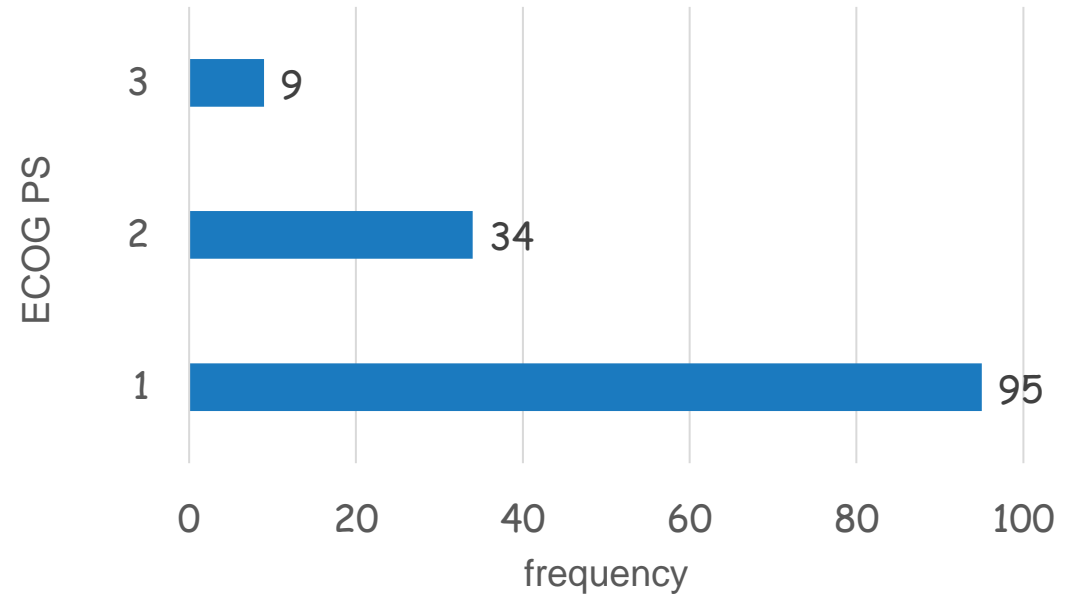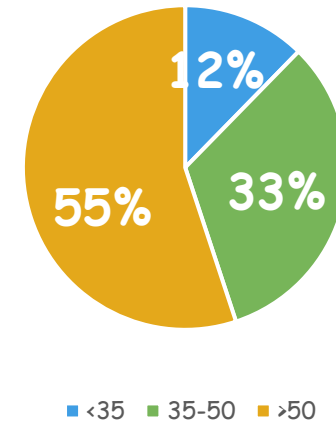
Graphs for categorical variables: pie charts

A circle is divided into sections.

Each section represents a category of the variable.

The size of each section is proportional to the frequency of the corresponding variable's class.

Sectors are coloured using different colours.

Classes of age frequency distribution

- <35
- 35-50
- >50

Graphs for comparing frequency distributions

Consider the relative frequency distribution (rfd) of a given variable (say Y) within a particular subgroup (say $X = \tilde{x}_i$)

$$p_{j|i} = \frac{n_{ij}}{n_{i\cdot}}$$

(*conditional frequency distribution*)

where $n_{i\cdot} = \sum_{j=1}^{t} n_{ij}$ is the overall number of units in category i.

Bar charts corresponding to different rfd allow one to compare different distribution shapes across different groups.

	$\tilde{y}_1$		$\tilde{y}_j$		$\tilde{y}_t$	
$\tilde{x}_1$	$p_{1|1}$	$\cdots$	$p_{j|1}$	$\cdots$	$p_{j|1}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\tilde{x}_i$	$p_{1|i}$	$\cdots$	$p_{j|i}$	$\cdots$	$p_{t|i}$	1
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$\tilde{x}_s$	$p_{1|s}$	$\cdots$	$p_{j|s}$	$\cdots$	$p_{t|s}$	1