# Descriptive statistics

Frequency distributions
Percentiles

Prof. Giuseppe Verlato

Unit of Epidemiology & Medical Statistics
Department of Diagnostics & Public Health
University of Verona

## Frequency distribution

With large databases, it is very difficult to pick out the information needed at a glance. Instead, it is more convenient to summarize variables into tables called “**frequency distributions**.”

The **frequency** (n, f) of a particular observation is the number of times the observation occurs in the data.

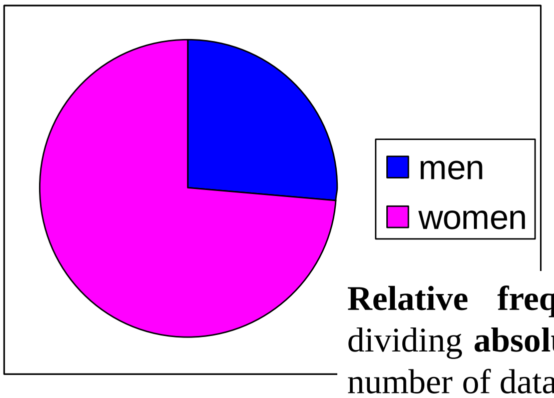
A **frequency distribution** is a table reporting the levels of a variable in the 1° column and the corresponding frequencies in the 2° column.

A frequency distribution shows the values a variable can take, and the number of people or records with each value.

- Frequency distribution tables can be used for both categorical and numeric variables.
- No data transformation is necessary to create a frequency distribution for **categorical variables** (either nominal or ordinal) as well as for **quantitative discrete variables**. Simply each level of the variable is associated with the corresponding frequency.
- For a **continuous variable,** if we associate a frequency to each distinct value of the variable, the number of frequencies will become unduly large, as a continuous variable can assume an infinite number of values within its range of variation. Hence continuous variables are discretized, i.e. recoded in class intervals.

# Frequency distribution of a categorical variable (sex)

| Sex | Number (absolute frequency) | Percent frequency |
|---|---|---|
| Men | 33 | 26.4% |
| Women | 92 | 73.6% |
| Total | 125 | 100% |




**Relative frequency** is computed by dividing **absolute frequency** by the total number of data: 33/125 = 0.264 = 26.4%

The categories should be **mutually exclusive**, i.e. non-overlapping. One statistical unit must be assigned to only one category: for instance a gay/lesbian cannot be assigned to both sexes, a gay is a male and a lesbian is a female.

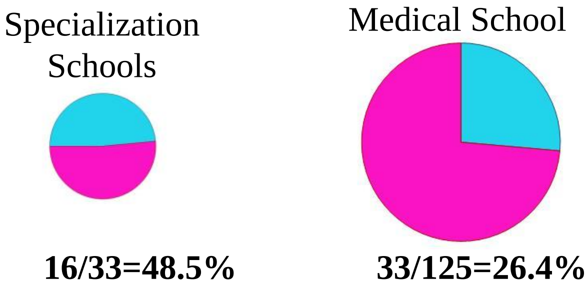
The classes should be **exhaustive**, i.e. they must cover the entire range of the data: for instance, transgender and intersex individuals should require an additional class to be classified.

## Importance of relative frequency: example

**Categorical variable = sex**

In 1995 my lessons to Specialization Schools were attended by 16 men, while my lessons to the Medical School by 33 men.

If we consider absolute frequency, men were twice as many among medical students than among specializing graduates.




**Relative frequency** **16/33=48.5%** **33/125=26.4%**

Indeed male sex is much more common among specializing medical graduates than among medical students.

## FREQUENCY DISTRIBUTION of TWO QUALITATIVE VARIABLES

Variable: Eye color

| Modality | Frequency | |
|---|---|---|
| | Absolute (n) | Relative (%) |
| **dark** | **120** | **80** |
| **light** | **30** | **20** |
| Total (∑) | 150 | 100 |

Variable: Hair colour

| Modality | Frequency | |
|---|---|---|
| | Absolute (n) | Relative (%) |
| **dark** | **110** | **73.3** |
| light | 40 | 26.7% |
| Total (∑) | 150 | 100 |

## CONTINGENCY TABLE
## 2 x 2 *(Fourfold Table)*

| | | Eye color | | |
|---|---|---|---|---|
| | | **Dark** | **Light** | |
| **Hair color** | **Dark** | **100** | **10** | **110** |
| | **Light** | **20** | **20** | **40** |
| | | **120** | **30** | **150** |

Marginal frequencies

Marginal frequencies are the sum of either a row or a column. They correspond to frequencies in univariate frequency distributions.

## CONTINGENCY TABLE
## 2 x 2 *(Fourfold Table)*

| | | Eye color | | |
|---|---|---|---|---|
| | | **Dark** | **Light** | |
| **Hair color** | **Dark** | **100** | **10** | **110** |
| | **Light** | **20** | **20** | **40** |
| | | **120** | **30** | **150** |

Joint frequencies

Joint frequencies = entries in the body of the table. 100 subjects belong both to the 1° row (dark hair) and to the 1° column (dark eyes). Hence they have both dark hair and dark eyes

## CONTINGENCY TABLE
## 2 x 2 *(Fourfold Table)*

| | | Eye color | | |
|---|---|---|---|---|
| | | **Dark** | **Light** | |
| **Hair color** | **Dark** | **100 (90.9%)** | **10 (9.1%)** | **110 (100%)** |

Row percentage

To compute row percentage, one has to focus on a single row (the 1° or the 2nd) as if it represented the entire sample.

## CONTINGENCY TABLE
## 2 x 2 *(Fourfold Table)*

| | | Eye color | | |
|---|---|---|---|---|
| | | **Dark** | **Light** | |
| **Hair color** | **Dark** | **100 (90.9%)** | **10 (9.1%)** | **110 (100%)** |
| | **Light** | **20 (50%)** | **20 (50%)** | **40 (100%)** |
| | | **120** | **30** | **150** |

Marginal frequencies

Joint frequencies

Row percentages

9.1% of subjects with dark hair have light eyes, 50% of subjects with light hair have light eyes.

---

EXERCISE: Building a 2*2 contingency table

DATA: There are 1000 elderly people, 100 have diabetes mellitus and 300 have hypertension. 70 subjects are affected by both diabetes and hypertension.

| | Hypertension | No Hyperten. | |
|---|---|---|---|
| Diabetes | 70 | 30 | 100 |
| No diabetes | 230 | 670 | 900 |
| | 300 | 700 | 1000 |

% of hypertension in the diabetic group = 70/100 = 0.70 = 70%
% of hypertension in the non-diabetic group = 230/900 = 0.256 = 25.6%

CONCLUSION: Diabetes and hypertension are highly related diseases.

EXERCISE: Building a 2*2 contingency table

DATA: There are 1000 elderly people, 100 have diabetes mellitus and 300 have hypertension. 70 subjects are affected by both diabetes and hypertension.

| | | Hypertension yes | Hypertension no | |
|---|---|---|---|---|
| diabetes | yes | 70 (70%) | 30 | 100 |
| | no | 230 (25.5%) | 670 | 900 |
| | | 300 | 700 | 1000 |

**Mendel experiment:**

**Mendel bred together smooth yellow peas (dominant traits) and wrinkled green peas (recessive traits), and further inbred the 1° generation of hybrids.**

| | Yellow | green | |
|---|---|---|---|
| Smooth | 315 | 108 | 423 |
| Wrinkled | 101 | 32 | 133 |
| | 416 | 140 | 556 |

% of green peas among smooth peas = 108/423 = 0.255 = 25.5%
% of green peas among wrinkled peas = 32/133 = 0.241 = 24.1%

CONCLUSION: The trait "surface characteristic" segregates independently of the trait "color" (**Mendel's third law = Principle of independent assortment**).

# Frequency distribution of a **discrete quantitative** variable

We want to describe the parity of a group of women, i.e. the number of children each woman has given birth to.

To construct a frequency distribution showing these data, we first list, from the lowest observed value to the highest, all the values that the variable parity can take.

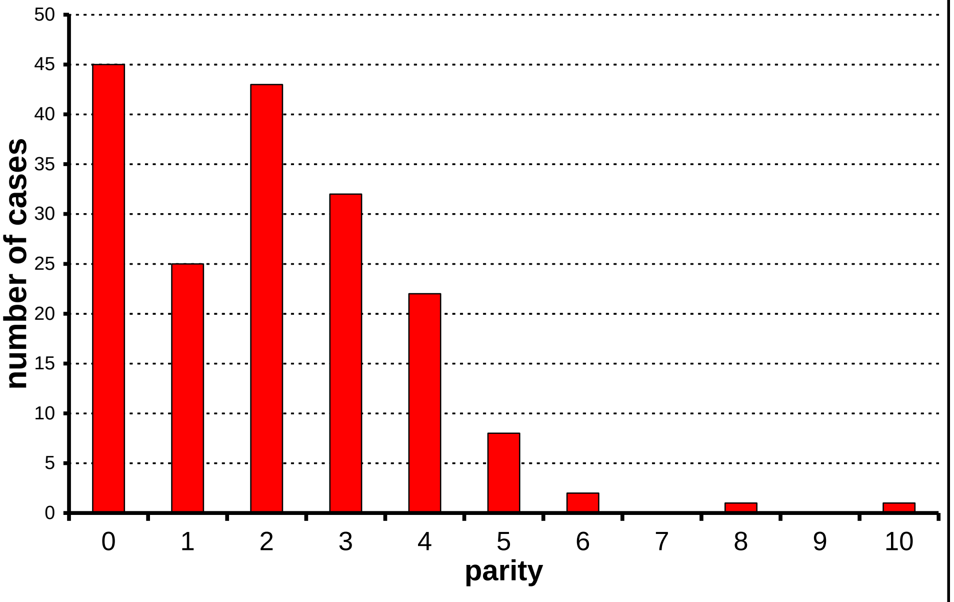
For each parity value, we then enter the number of women who had given birth to that number of children.

## Frequency distribution of a quantitative variable (parity)

The table shows the resulting frequency distribution. Notice that we listed *all* values of parity between the lowest and highest observed, even though there were no cases for some values. Notice also that each column is properly labeled, and that the total is given in the bottom row.

| parity | n° of cases | % frequency | cumulative freq. | cum. % freq. |
|---|---|---|---|---|
| 0 | 45 | 25,1% | 45 | 25,1% |
| 1 | 25 | 14,0% | 70 | 39,1% |
| 2 | 43 | 24,0% | 113 | 63,1% |
| 3 | 32 | 17,9% | 145 | 81,0% |
| 4 | 22 | 12,3% | 167 | 93,3% |
| 5 | 8 | 4,5% | 175 | 97,8% |
| 6 | 2 | 1,1% | 177 | 98,9% |
| 7 | 0 | 0,0% | 177 | 98,9% |
| 8 | 1 | 0,6% | 178 | 99,4% |
| 9 | 0 | 0,0% | 178 | 99,4% |
| 10 | 1 | 0,6% | 179 | 100,0% |
| total | 179 | 100,0% | | |

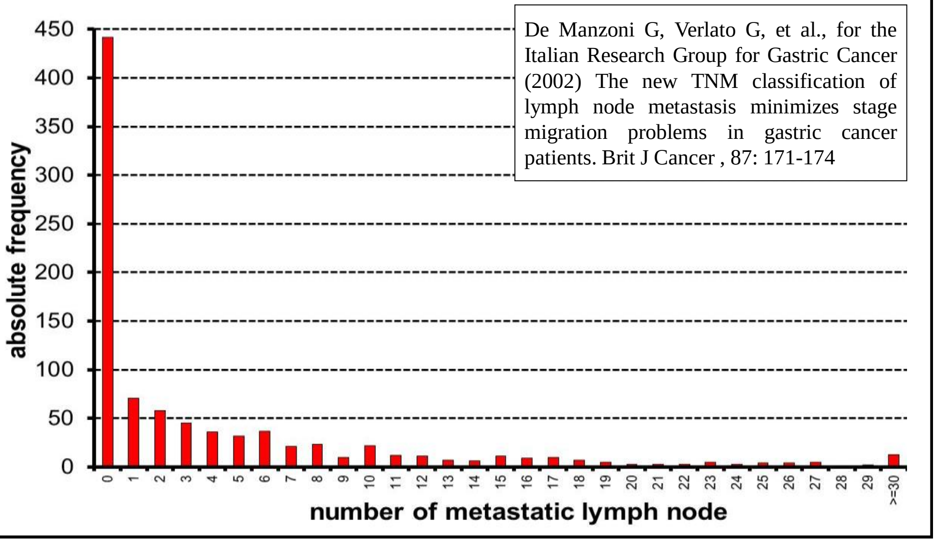
## Bar histogram




Metastatic lymph nodes in 921 patients with gastric cancer

*(graphic representation of the frequency distribution of a discrete quantitative variable – bar diagram)*

De Manzoni G, Verlato G, et al., for the Italian Research Group for Gastric Cancer (2002) The new TNM classification of lymph node metastasis minimizes stage migration problems in gastric cancer patients. Brit J Cancer , 87: 171-174



## WEIGHT, HEIGHT and SEX of 1st year MEDICAL students (FRESHERS) at VERONA UNIVERSITY in October 1995

| WEI. Kg | HEI. cm | SEX | WEI. Kg | HEI. cm | SEX | WEI. Kg | HEI. cm | SEX |
|---|---|---|---|---|---|---|---|---|
| 56 | 159 | F | 77 | 192 | M | 51 | 171 | F |
| 66 | 169 | F | 60 | 173 | F | 48 | 156 | F |
| 50 | 160 | F | 78 | 182 | M | 55 | 167 | F |
| 53 | 170 | F | 52 | 167 | F | 60 | 177 | M |
| 54 | 168 | F | 47.5 | 164 | F | 58 | 170 | F |
| 53 | 161 | F | 64 | 166 | F | 67 | 167 | F |
| 63 | 172 | M | 52 | 160 | F | 50 | 172 | F |
| 53 | 170 | F | 72 | 184 | M | 58 | 169 | F |
| 62 | 161 | F | 48 | 169 | F | 77 | 179 | M |
| 56 | 163 | F | 66 | 170 | M | 52 | 162 | M |
| 50 | 160 | F | 55 | 172 | F | 49 | 160 | F |
| 52 | 170 | F | 67 | 177 | M | 49 | 165 | F |
| 58 | 173 | F | 66 | 170 | M | 62 | 178 | M |
| 52 | 167 | F | 50 | 160 | F | 68 | 174 | M |
| 73 | 178 | M | 51 | 167 | F | 75 | 181 | M |
| 57 | 166 | F | 95 | 193 | M | 48 | 167 | F |
| 52 | 165 | F | 58 | 160 | F | 53 | 160 | F |
| 56 | 171 | F | 67 | 178 | F | 49 | 167 | F |
| 67 | 175 | M | 67 | 175 | M | 52 | 165 | F |
| 63 | 182 | F | 60 | 160 | F | 55 | 155 | F |
| 55 | 169 | F | 56 | 165 | F | 84 | 188 | M |
| 58 | 165 | F | 50 | 165 | F | 56 | 170 | F |
| 55 | 175 | M | 52 | 170 | F | 60 | 171 | F |
| 66 | 176 | M | 58 | 172 | F | 52 | 176 | M |
| 55 | 164 | F | 60 | 170 | F | 62 | 180 | F |
| 47 | 160 | F | 54 | 166 | F | | | |
| 47 | 155 | F | 60 | 165 | F | | | |
| 63 | 169 | M | 74 | 172 | M | | | |
| 61 | 177 | F | 53 | 173 | F | | | |
| 53 | 170 | F | 72 | 183 | M | | | |
| 55 | 168 | M | 52 | 168 | F | | | |
| 53 | 162 | F | 51 | 164 | F | | | |
| 62 | 162 | F | 81 | 176 | M | | | |
| 45 | 160 | F | 50 | 160 | F | | | |
| 57 | 167 | F | 51 | 171 | F | | | |
| 45 | 158 | F | 64 | 180 | F | | | |
| 53 | 168 | F | 82 | 183 | M | | | |
| 50 | 160 | F | 47 | 156 | F | | | |
| 55 | 162 | F | 70 | 175 | M | | | |
| 70 | 177 | M | 58 | 168 | F | | | |
| 64 | 178 | F | 59 | 173 | F | | | |
| 52 | 164 | F | 68 | 165 | F | | | |
| 75 | 175 | M | 63 | 177 | F | | | |
| 75 | 178 | M | 50 | 159 | F | | | |
| 70 | 165 | F | 65 | 150 | F | | | |
| 58 | 167 | F | 60 | 170 | F | | | |
| 45 | 160 | F | 51 | 167 | F | | | |
| 50 | 167 | F | 75 | 182 | M | | | |
| 56 | 156 | F | 62 | 170 | M | | | |
| 59 | 165 | F | 85 | 174 | M | | | |

# FREQUENCY DISTRIBUTION of HEIGHT

```
fre var=height.
```

| HEIGHT Value | Frequency | Percent | Valid Percent | Cum Percent |
|---|---|---|---|---|
| 150 | 1 | .8 | .8 | .8 |
| 155 | 2 | 1.6 | 1.6 | 2.4 |
| 156 | 3 | 2.4 | 2.4 | 4.8 |
| 158 | 1 | .8 | .8 | 5.6 |
| 159 | 2 | 1.6 | 1.6 | 7.2 |
| 160 | 13 | 10.4 | 10.4 | 17.6 |
| 161 | 2 | 1.6 | 1.6 | 19.2 |
| 162 | 4 | 3.2 | 3.2 | 22.4 |
| 163 | 1 | .8 | .8 | 23.2 |
| 164 | 4 | 3.2 | 3.2 | 26.4 |
| 165 | 10 | 8.0 | 8.0 | 34.4 |
| 166 | 3 | 2.4 | 2.4 | 36.8 |
| 167 | 11 | 8.8 | 8.8 | 45.6 |
| 168 | 5 | 4.0 | 4.0 | 49.6 |
| 169 | 5 | 4.0 | 4.0 | 53.6 |
| 170 | 12 | 9.6 | 9.6 | 63.2 |
| 171 | 4 | 3.2 | 3.2 | 66.4 |
| 172 | 5 | 4.0 | 4.0 | 70.4 |
| 173 | 4 | 3.2 | 3.2 | 73.6 |
| 174 | 2 | 1.6 | 1.6 | 75.2 |
| 175 | 5 | 4.0 | 4.0 | 79.2 |
| 176 | 3 | 2.4 | 2.4 | 81.6 |
| 177 | 5 | 4.0 | 4.0 | 85.6 |
| 178 | 5 | 4.0 | 4.0 | 89.6 |
| 179 | 1 | .8 | .8 | 90.4 |
| 180 | 2 | 1.6 | 1.6 | 92.0 |
| 181 | 1 | .8 | .8 | 92.8 |
| 182 | 3 | 2.4 | 2.4 | 95.2 |
| 183 | 2 | 1.6 | 1.6 | 96.8 |
| 184 | 1 | .8 | .8 | 97.6 |
| 188 | 1 | .8 | .8 | 98.4 |
| 192 | 1 | .8 | .8 | 99.2 |
| 193 | 1 | .8 | .8 | 100.0 |
| Total | 125 | 100.0 | 100.0 | |

CONSTRUCTING a FREQUENCY DISTRIBUTION for a CONTINUOUS QUANTITATIVE VARIABLE

| | |
|---|---|
| 1.Find the smallest and the largest values. | Minimum = 150 cm<br>Maximum = 193 cm |
| 2. Compute the range, i.e. the difference between the largest and the smallest value | 193 – 150 = 43 cm |
| 3.Fix the number of class intervals: between 5 (few statistical units) and 20 (several units) | 9 class intervals |
| 4. The classes should, preferably, be of equal width. | |
| 5. Fix the width of class intervals. | 43/9 = 4.78 cm ≈ 5 cm |
| 6. Construct the class intervals, which must be mutually exclusive and exhaustive | 1st interval: [150-155)<br>2nd interval: [155-160)<br>3rd interval: [160-165) |
| 7. Count the number of statistical units in each interval. | 1st interval: 1<br>2nd interval: 8<br>3rd interval: 24 |

The classes should be mutually exclusive, i.e., non-overlapping. No two classes should contain the same interval of values of the variable.

The classes should be exhaustive, i.e., they must cover the entire range of the data.

The number of classes and the width of each class should neither be too small nor too large. In other words, there should be relatively fewer classes if there are few statistical units and relatively more classes if there are many.

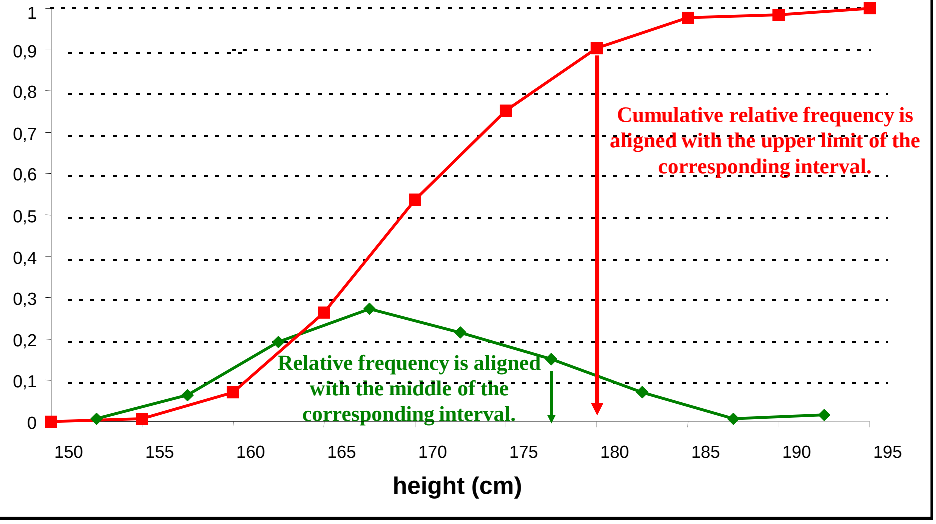
The classes should, preferably, be of equal width.

```
compute heightCLAS=trunc((height-145)/5).
fre var=heightCLAS.
```

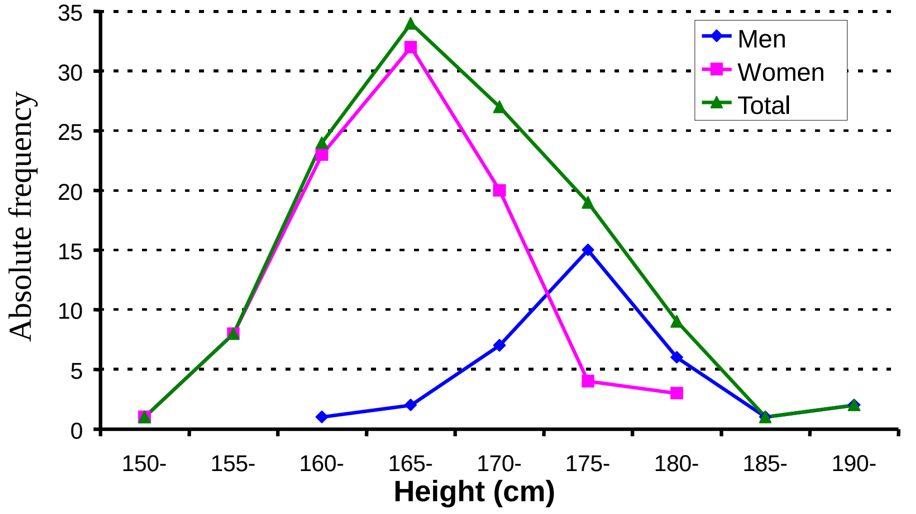
| CLASS | FREQUENCY | | CUMULATIVE FREQUENCY | |
|---|---|---|---|---|
| | ABSOLUTE | RELATIVE % | ABSOLUTE | RELATIVE % |
| 150-154,9 | 1 | 1/125= 0,8 | 1 | 1/125= 0,8 |
| 155-159,9 | 8 | 8/125= 6,4 | 1+8= 9 | 9/125= 7,2 |
| 160-164,9 | 24 | 24/125=19,2 | 1+8+24=33 | 33/125=26,4 |
| 165-169,9 | 34 | 34/125=27,2 | 1+8+24+34=67 | 67/125=53,6 |
| 170-174,9 | 27 | 21,6 | 94 | 75,2 |
| 175-179,9 | 19 | 15,2 | 113 | 90,4 |
| 180-184,9 | 9 | 7,2 | 122 | 97,6 |
| 185-189,9 | 1 | 0,8 | 123 | 98,4 |
| 190-194,9 | 2 | 1,6 | 125 | 100,0 |
| | ---- | ------- | | |
| Total | 125 | 100,0 | | |

**Cumulative frequency** = the sum of absolute frequencies of all the classes equal to or less than the considered class.

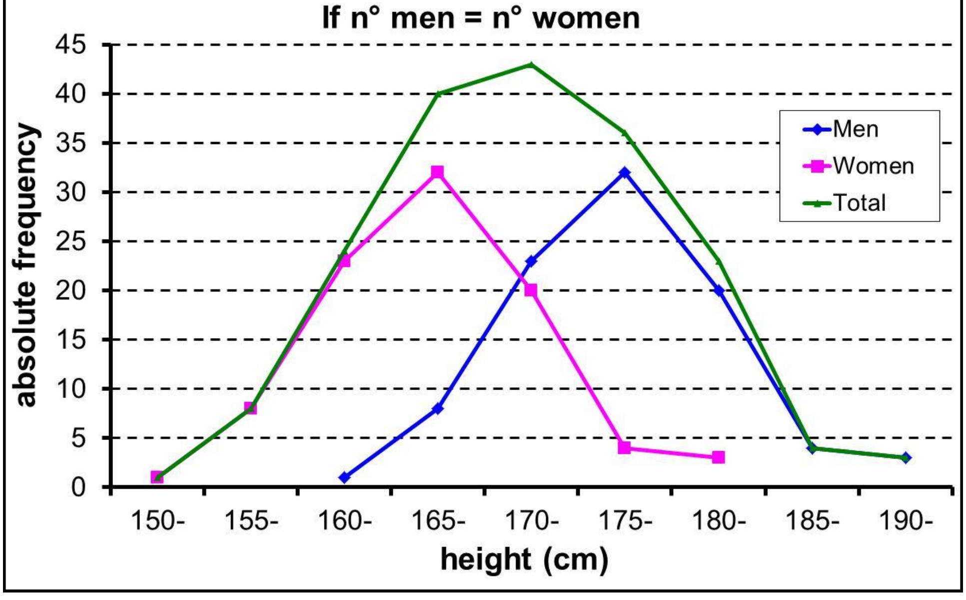
## Height of medical freshers at Verona University in 1995
## *(graphic representation by line charts)*



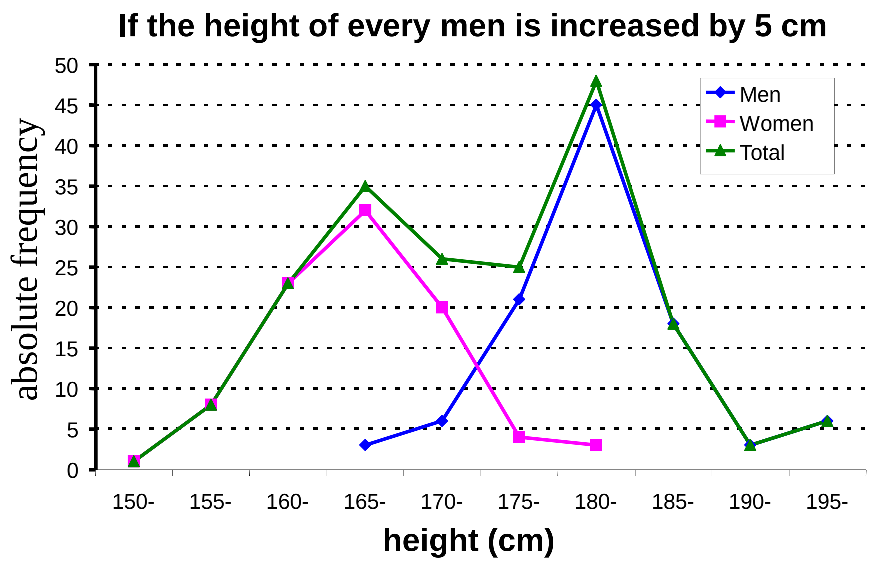
## Height of medical freshers at Verona University in 1995 as a function of gender




## Height of medical freshers at Verona University in 1995 as a function of gender



## Height of medical freshers at Verona University in 1995 as a function of gender




## CONSTRUCTING a FREQUENCY DISTRIBUTION

**Constructing class intervals**

$$\delta_i = \textbf{Range} / k$$

$\delta_i$ = interval width

$k$ = number of intervals

range: 160-192 cm

Number of class intervals = 5

Width of class intervals = (192-160)/5=32/5=6.4 ÷ 7

Height (cm)
166
164
170
192
160
174
186
176
160
165
165
173
179
168
168

class intervals
160-166.9 cm
167-173.9 cm
174-180.9 cm
181-187.9 cm
188-194.9 cm

| Height (cm) | n | p | N | P |
|---|---|---|---|---|
| 160-166.9 | 6 | 0.40 | 6 | 0.40 |
| 167-173.9 | 4 | 0.26 | 10 | 0.67 |
| 174-180.9 | 3 | 0.20 | 13 | 0.86 |
| 181-187.9 | 1 | 0.07 | 14 | 0.93 |
| 188-194.9 | 1 | 0.07 | 15 | 1.00 |

## Algorithms to choose the number of intervals / interval width

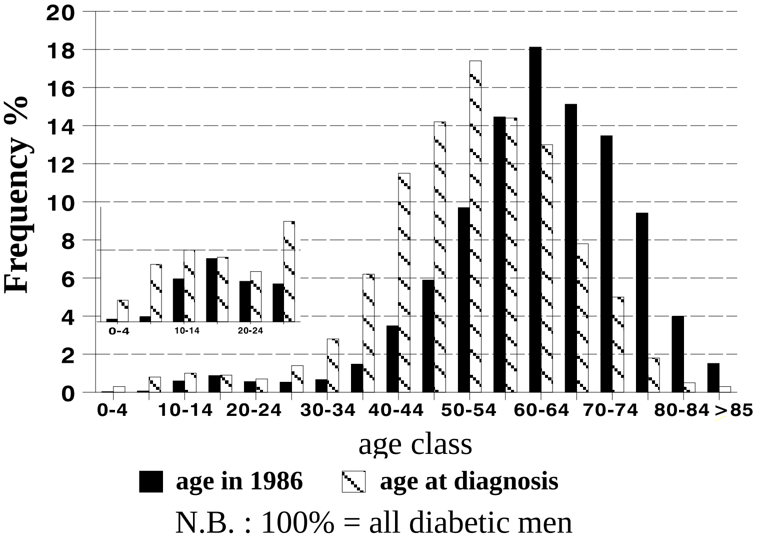
A) According to H. Sturges (1926) the optimal number of class intervals (C) can be mathematically derived from the number of observations (N):

$$C = 1 + \frac{10}{3} \cdot \log_{10}(N)$$

B) According D. Scott (1979) the optimal width (h) of class intervals, which directly determines also the number of class intervals, can be derived from the standard deviation (S) as follows:

$$h = \frac{3{,}5 \cdot S}{\sqrt{N}}$$

## DIABETIC MEN in VERONA on the 31.12.1986




N.B. : 100% = all diabetic men

Muggeo M, Verlato G, …, de Marco R (1995) The Verona Diabetes Study: a population-based survey on known diabetes mellitus prevalence and 5-year all-cause mortality. *Diabetologia*, 38: 318-325

**Absolute rank =** number specifying position in an numerically ordered series.

An **ascending order** is usually adopted in medical statistics**.**

If two or more statistical units (individuals) have the **same value,** they are assigned the **average rank** of the positions held.

| | | | | | |
|---|---|---|---|---|---|
| **RANK** | **1** | **2** | **3** | **4** | **5** |
| **VALUE** | **3** | **4** | **4** | **5** | **6** |
| | | **2,5** | **2,5** | | |

| | | | | | |
|---|---|---|---|---|---|
| **RANK** | 1 | 2 | 3 | 4 | 5 |
| **VALUE** | 3 | **4** | **4** | **4** | 5 |
| | | 3 | 3 | 3 | |

# Percentile Rank

Percentile rank of a given score is the proportion of scores which are equal to or lower than that score.

For instance, a student gets a bad mark at school. If this mark is lower than the marks obtained by 90% of his/her schoolmates, parents usually get anxious and nervous.

However, if this mark is lower than the marks obtained by 10% of the other students, parents usually relax a little.

In the first case the percentile rank is 10%, while in the second case is 90%.

**EXAMPLE**

A schoolboy has a glycaemia of 90 mg/dl.

There are 700 students in his school.

If glycaemia is sorted in ascending order, his absolute rank (position) is 500.

Which is the percentile rank (%)?

`PercentileRank = AbsoluteRank / (n+1)`

500/(700+1) = 500/701 = 0,713 = 71,3 %

Reverse equation:

`AbsoluteRank = (n+1) * PercentileRank`

# COMPUTING the PERCENTILE RANK

Let's consider two subjects whose absolute rank is 50, respectively in a group of 99 subjects or in a group of 100 subjects.

| | N=99 | N=100 |
|---|---|---|
| Subjects with higher rank | 49 | 50 |
| | 50 | 50 |
| Subjects with lower rank | 49 | 49 |

**Percentile rank =** **50/(99+1)=50%** **50/(100+1)=49.5 %**

**WRONG % rank= 50/99=50.5%** **50/100=50%**

**To compute percentile rank, we have to divide by N+1 not by N !**

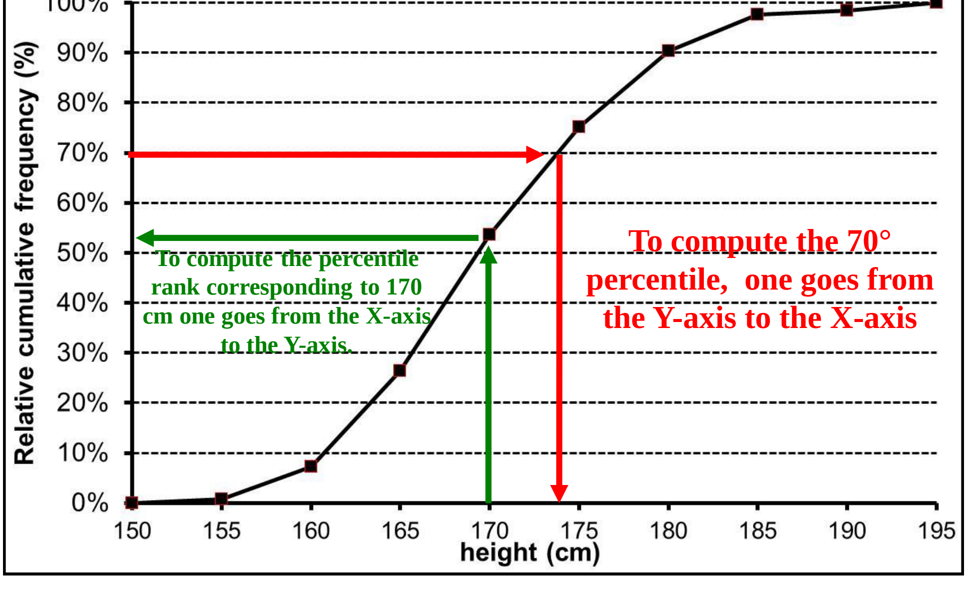
# Percentile

**Percentiles** are 99 values of a variable that divide the distribution of the variable in 100 subgroups having equal frequency.

**N.B. Quartiles** are 3 values that divide a distribution in 4 subgroups having equal frequency:

1° quartile =25° percentile
2° quartile =50° percentile
3° quartile =75° percentile

Height of medical freshers at Verona University in 1995
*(graphic representation by line charts)*



**PERCENTILE RANK = FEATURE of an INDIVIDUAL**
**PERCENTILE = FEATURE of a POPULATION**

**EXAMPLE:**
**An individual weighs 100 Kg. His percentile rank is 96%, i.e. 96% of other individuals have an equal or lower weight.**

**Which is the 96° percentile in the same population ? 100 Kg.**

**An individual with a percentile rank of 96% has a weight equal to the $96^{th}$ percentile of that population (100 Kg).**

## Computing the *k-th* percentile - 1

*(Individual data are available)*

- First of all, one should find the absolute rank corresponding to the ***k-th* percentile**

`Absolute rank = (N+1) * k / 100`

- Then one should find the value of the observation with that particular rank.

## Example (individual data)

Which is the 40° percentile of height in 1° class medical students at Verona University in 1995 ?

1) Which absolute rank corresponds to the **40° percentile ?**

**AbsoluteRank = (N+1) * k/100 = (125+1) * 40/100 = 126*0.4 = 50.4**

2) Observations with absolute ranks 50 and 51, both have a height of 167 cm.

$$\mathbf{X_{40} = 167\ cm}$$

## Computing the *k-th* percentile - 2

*(Original data not available, only a frequency table)*

- The class interval containing the ***k-th*** **percentile** should be identified, i.e. the class interval where relative cumulative frequency exceeds or equals $k$ percent
- Then a **linear interpolation** is performed

$$x_k = u_{i-1} + \frac{k - F(u_{i-1})}{F(u_i) - F(u_{i-1})} * \delta_i$$

It is assumed that values are **uniformly** distributed within each class

$k$ = percentile rank
$\chi_k$ = *k-th* percentile of the distribution
$u_{i-1}$ = lower limit of *i-th* interval
$u_i$ = upper limit of *i-th* interval
$F(u_{i-1})$= cumulative frequency of previous interval
$F(u_i)$ = cumulative frequency of *i-th* interval
$\delta_i$ = width of *i-th* interval

## Example (frequency table)

Which is the 40th percentile of height in 1st class medical students at Verona University in 1995 ?

The 40th percentile belongs to the 4th classe: [165-170) cm

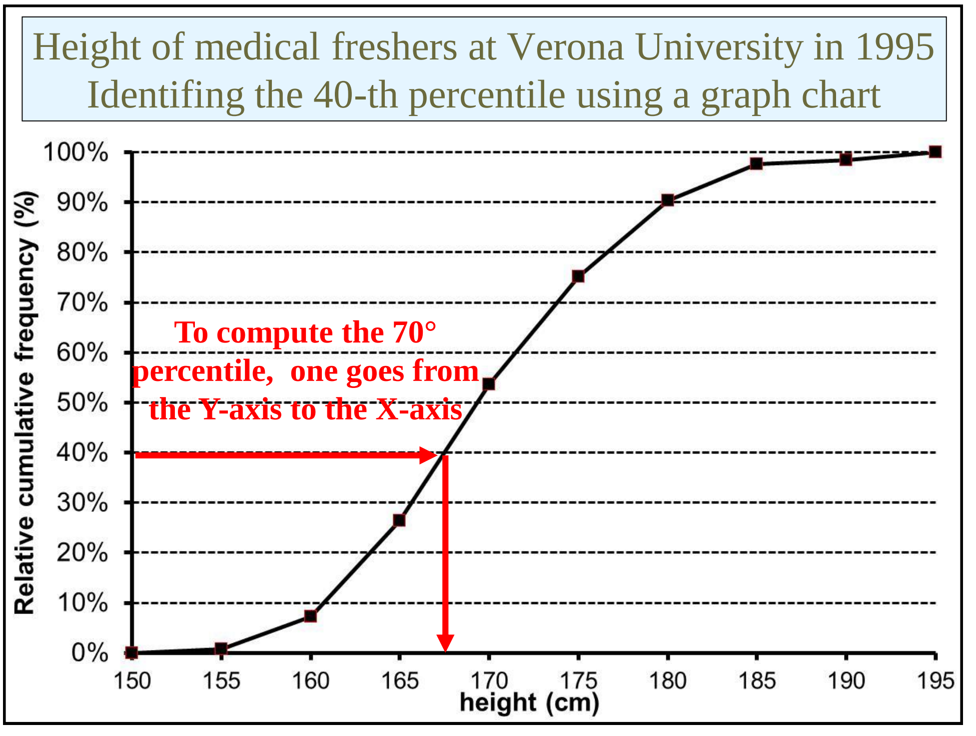
$$X_{40} = 165 + 5 * \frac{40\% - 26.4\%}{53.6\% - 26.4\%} = 165 + 5 * \frac{13.6\%}{27.2\%} =$$

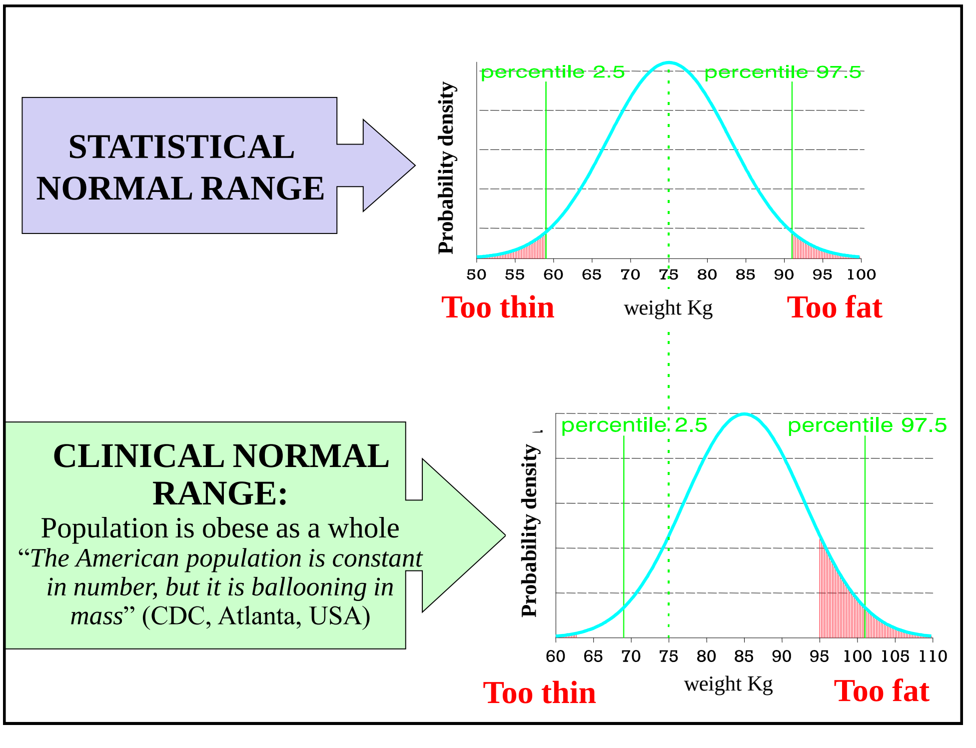
$$= 165 + 5 * 0.5 = 165 + 2.5 = 167.5 \text{ cm}$$

## Computing *k-th* percentile – 3

*(Individual data are not available, only a graphical representation of relative cumulative frequency is available)*

- The point corresponding to ***k-th* percentile rank** is located on the Y-axis
- An horizontal line is drawn from this point until it crosses the ***chart line***, showing the pattern of relative cumulative frequency
- A vertical line is drawn from the intersection point until it crosses the X-axis, reporting the values of the variable under study
- The value of the variable in the latter intersection point corresponds to the ***k-th* percentile**


Height of medical freshers at Verona University in 1995
Identifing the 40-th percentile using a graph chart
Relative cumulative frequency (%)
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
150 155 160 165 170 175 180 185 190 195
height (cm)
To compute the 70° percentile, one goes from the Y-axis to the X-axis

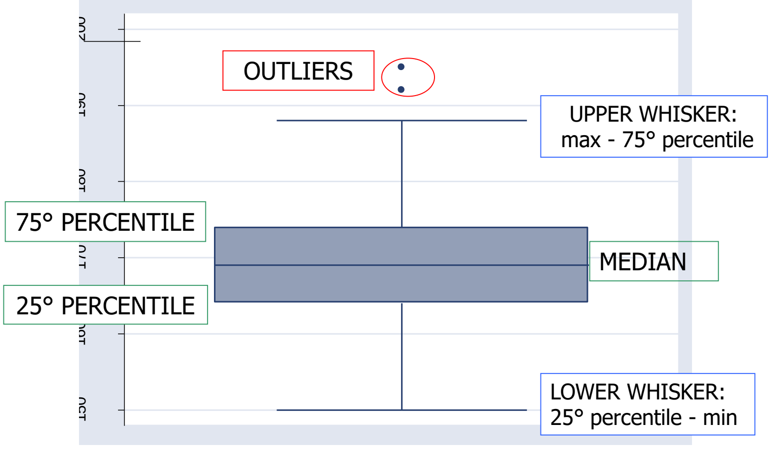

STATISTICAL NORMAL RANGE
percentile 2.5
percentile 97.5
Probability density
50 55 60 65 70 75 80 85 90 95 100
weight Kg
Too thin
Too fat
CLINICAL NORMAL RANGE:
Population is obese as a whole
"The American population is constant in number, but it is ballooning in mass" (CDC, Atlanta, USA)
percentile 2.5
percentile 97.5
Probability density
60 65 70 75 80 85 90 95 100 105 110
weight Kg
Too thin
Too fat

## NUMERICAL or GRAPHICAL SUMMARIES of DATA

| Type of variables | Numerical summary | | Graphical summary |
|---|---|---|---|
| Categorical (nominal or ordinal) | Frequency table | | pie<br>bar chart |
| Quantitative discrete | Frequency table | | bar chart |
| Quantitative continuous | Frequency table | Stem-and leaf plot | histogram<br>line chart<br>box-and-whisker plot |

```
examine statura/percentiles (2.5 25 50 75 97.5).
--------------------------------------------------------------------
          STEM-AND-LEAF DIAGRAM (DIAGRAMMA TRONCO E FOGLIE)
--------------------------------------------------------------------
  n  STEM LEAVES                   CORRESPONDING NUMBERS
--------------------------------------------------------------------
  1  15  0                                        150
  8  15  55666899                   155,155,156,156,156,158,159,159
 24  16  000000000000011222234444
 34  16  5555555555566677777777778888899999
 27  17  000000000000011112222233344
 19  17  5555566677777888889
  9  18  001222334
  1  18  8                                        188
  2  19  23                                     192,193
--------------------------------------------------------------------
 Stem width:   10
 Each leaf:     1 case(s)
--------------------------------------------------------------------
```

**HEIGHT DISTRIBUTION AMONG 1° CLASS MEDICAL STUDENTS**
**BOX-and-WHISKERS PLOT** (GRAFICO SCATOLA E BAFFI)




Whisker maximum length = 1.5 * interquartile range (box height)