School of Psychology
Dpt. Experimental Psychology

Design and Data Analysis in Psychology I
English group (A)

Salvador Chacón Moscoso
Susana Sanduvete Chaves
Milagrosa Sánchez Martín

Lesson 3

Group and individual indexes

Group and individual indexes

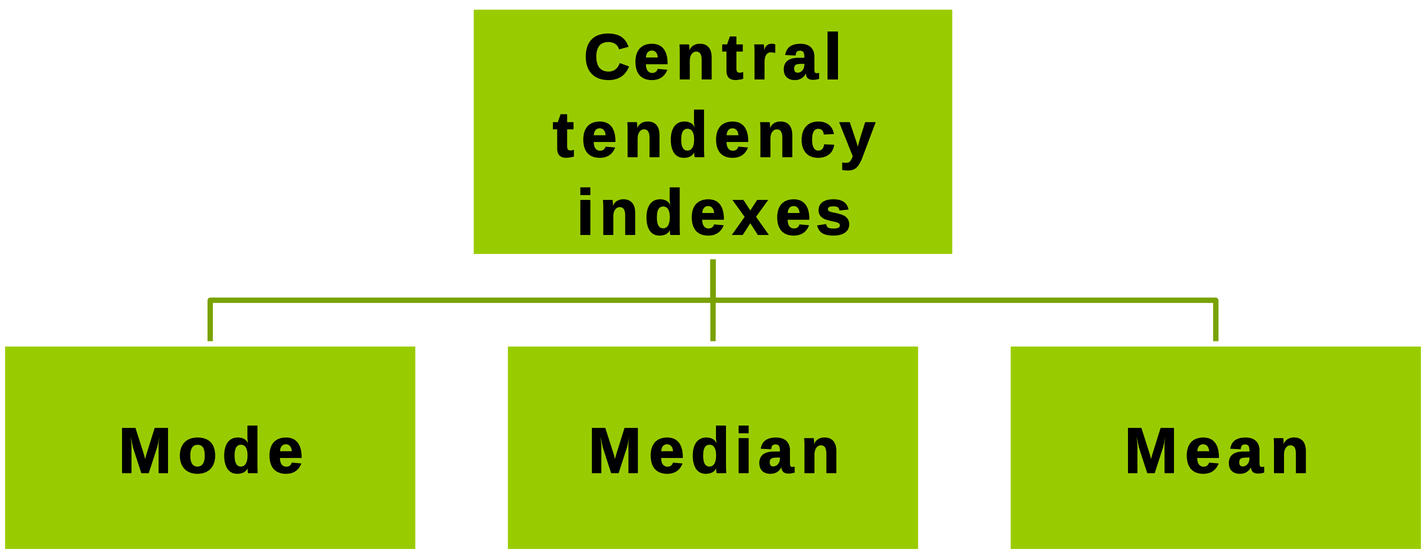
Group indexes	Individual indexes
↓	↓
Central tendency	Position: Deciles (D_i), Percentiles (P_i), Quartiles (Q_i)
Variability (Dispersion)	Raw scores (X_i)
Bias or Skewness (Asymmetry)	Differential scores (x_i)
Kurtosis	Standard scores (Z_i)

To describe a data distribution we need at least two statistics:

1. One that reflects the central tendency: value which represents the group.

2. Another that reflects the dispersion around this center. It determines how far or together the data are from each other.

Which value does represent the group?



Central tendency indexes

Definition

- They are a brief description of a mass of data, usually obtained from a sample.

- They serve to describe, indirectly, the population from which the sample was extracted.

1. Which is the most repeated value?

Mode (Mo)

1. Mode (Mo)

Definition:

- The most repeated value.
- The value most frequently observed in a sample or population.
- The value of a variable with the highest absolute frequency.
- It is symbolized by *Mo* (Fechner and Pearson)

1. Mode (Mo)

1.1. Type I distributions: small data set

a) Unimodal distribution: there is only one mode.

- Example:

8 – 8 – 11 – 11 – 15 – 15 – 15 – 15 – 15 – 17 – 17 – 17 – 19 - 19

1. Mode (Mo)

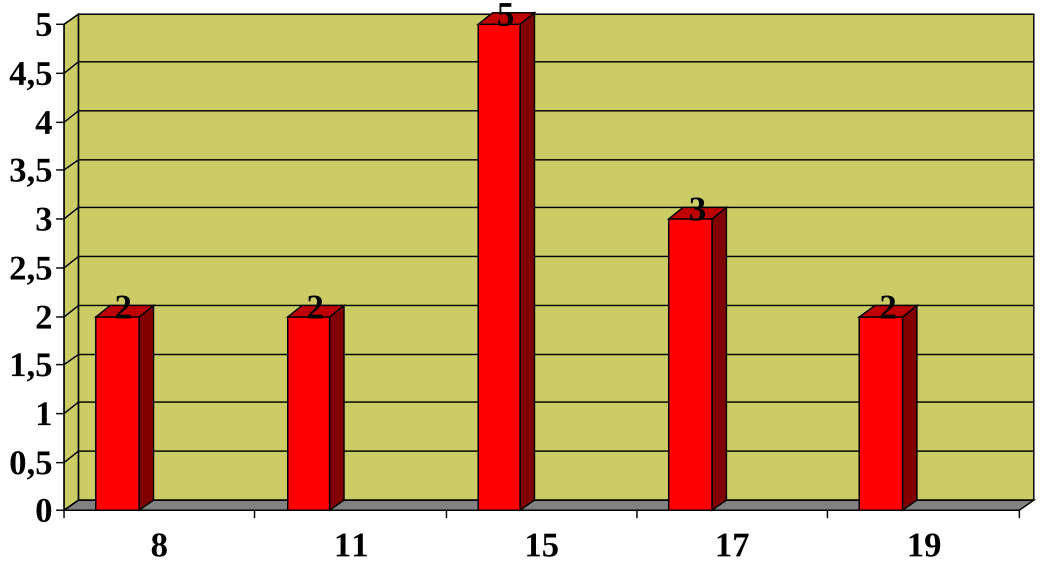
1.1. Type I distributions: small data set

a) Unimodal distribution:

- Example:

8 – 8 – 11 – 11 – 15 – 15 – 15 – 15 – 15 – 17 – 17 – 17 – 19 - 19

Mo = 15



1. Mode (Mo)

1.1. Type I distributions: small data set

b) Amodal distribution: more than 80% of the values of X present the highest absolute frequency.

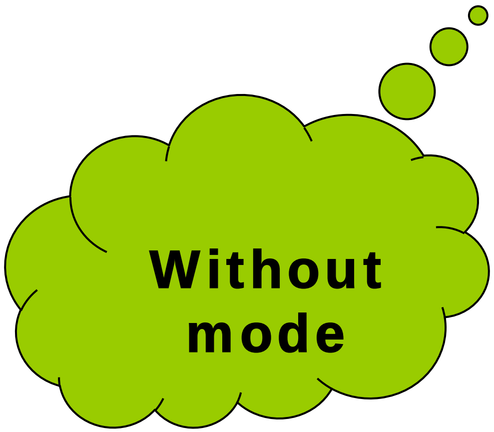
- Example: 8 – 8 – 8 – 11 – 11 – 11 – 15 – 15 – 15 – 17 – 17 – 17 – 19 – 19 –19

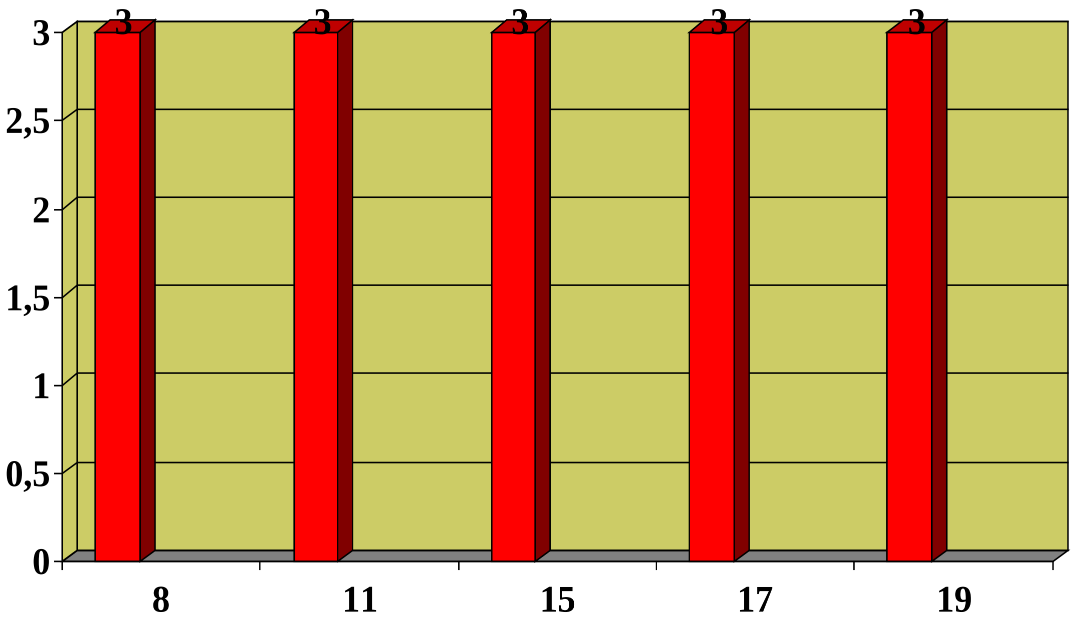
1. Mode (Mo)

1.1. Type I distributions: small data set

b) Amodal distribution:

- Example: 8 – 8 – 8 – 11 – 11 – 11 – 15 – 15 – 15 – 17 – 17 – 17 – 19 – 19 –19





1. Mode (Mo)

1.1. Type I distributions: small data set

c) Bimodal distribution: There are two modes.

- Example:

8 – 9 – 9 – 10 – 10 – 10 – 10 – 11 – 11 – 13 – 13 – 13 – 13 – 15

1. Mode (Mo)

1.1. Type I distributions: small data set

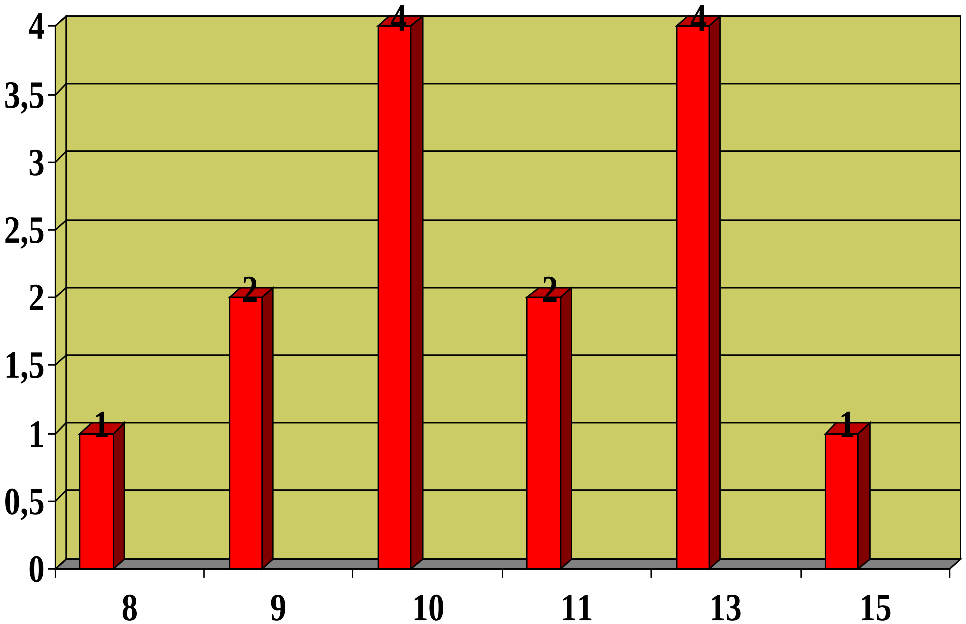
c) Bimodal distribution:

- Example:

8 – 9 – 9 – 10 – 10 – 10 – 10 – 11 – 11 – 13 – 13 – 13 – 13 – 15

Mo_1 = 10

Mo_2 = 13



1. Mode (Mo)

1.1. Type I distributions: small data set

d) Multimodal distribution: there are more than two modes.

- Example:

8 – 8 – 9 – 9 – 9 – 10 – 11 – 11 – 11 – 12 – 12 – 13 – 13 – 13 – 14 – 15 - 15

1. Mode (Mo)

1.1. Type I distributions: small data set

d) Multimodal distribution:

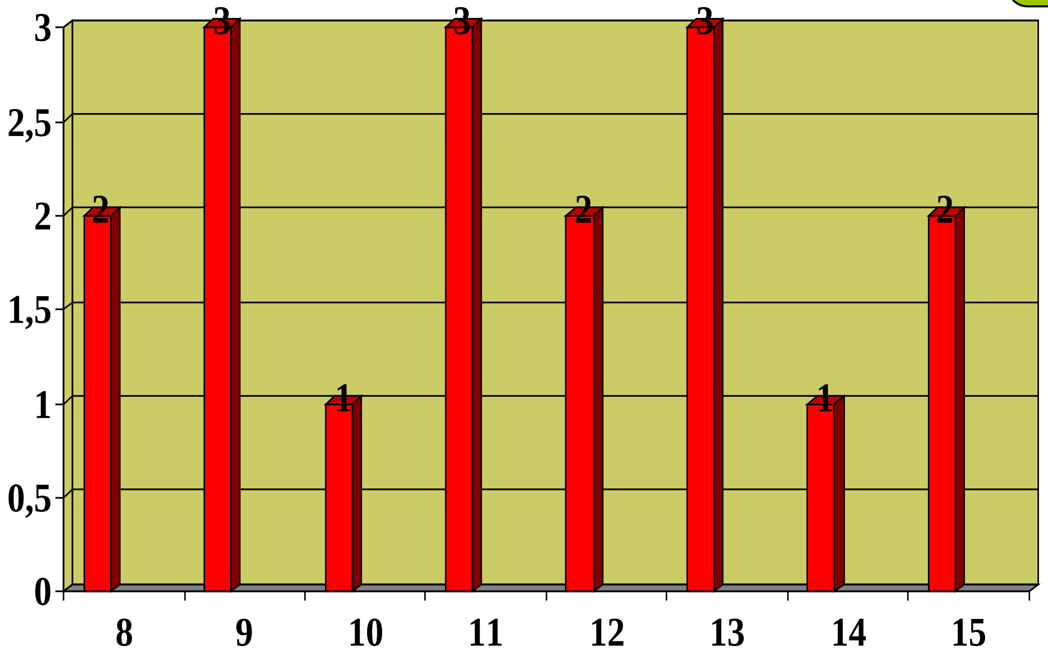
- Example:

8 – 8 – 9 – 9 – 9 – 10 – 11 – 11 – 11 – 12 – 12 – 13 – 13 – 13 – 14 – 15 - 15

Mo_2 = 11

Mo_1 = 9

Mo_3 = 13



1. Mode (Mo)

1.2. Type II distributions: big data set

Frequency table

a) Unimodal distribution: example:

X_i	f_i
12	3
14	25
16	10
18	5

1. Mode (Mo)

1.2. Type II distributions: big data set

Frequency table

a) Unimodal distribution: example:

Mo = 14

X_i	f_i
12	3
14	25
16	10
18	5

MOST REPEATED VALUE

1. Mode (Mo)

1.2. Type II distributions: big data set

b) Bimodal distribution: example

X_i	f_i
2	10
4	5
6	10
8	7

1. Mode (Mo)

1.2. Type II distributions: big data set

b) Bimodal distribution: example:

$Mo_1 = 2$ and

$Mo_2 = 6$

MOST REPEATED VALUES

X_i	f_i
2	10
4	5
6	10
8	7

1. Mode (Mo)

Example: Complete the table knowing that the modes are: -2, -1 and 5; and $f_3 = f_4$

X_i	f_i	rf_i	$\%_i$
-2	5		
-1			
0		O.12	
2			
5			20
6			

1. Mode (Mo)

Example: solution

X_i	f_i	rf_i	$\%_i$
-2	5	0.2	20
-1	5	0.2	20
0	3	O.12	12
2	3	0.12	12
5	5	0.2	20
6	4	0.16	16
Σ	25	1	100

2.

Which is the value exceeded by the half of the participants?

Median (Mdn)

2. Median (Mdn)

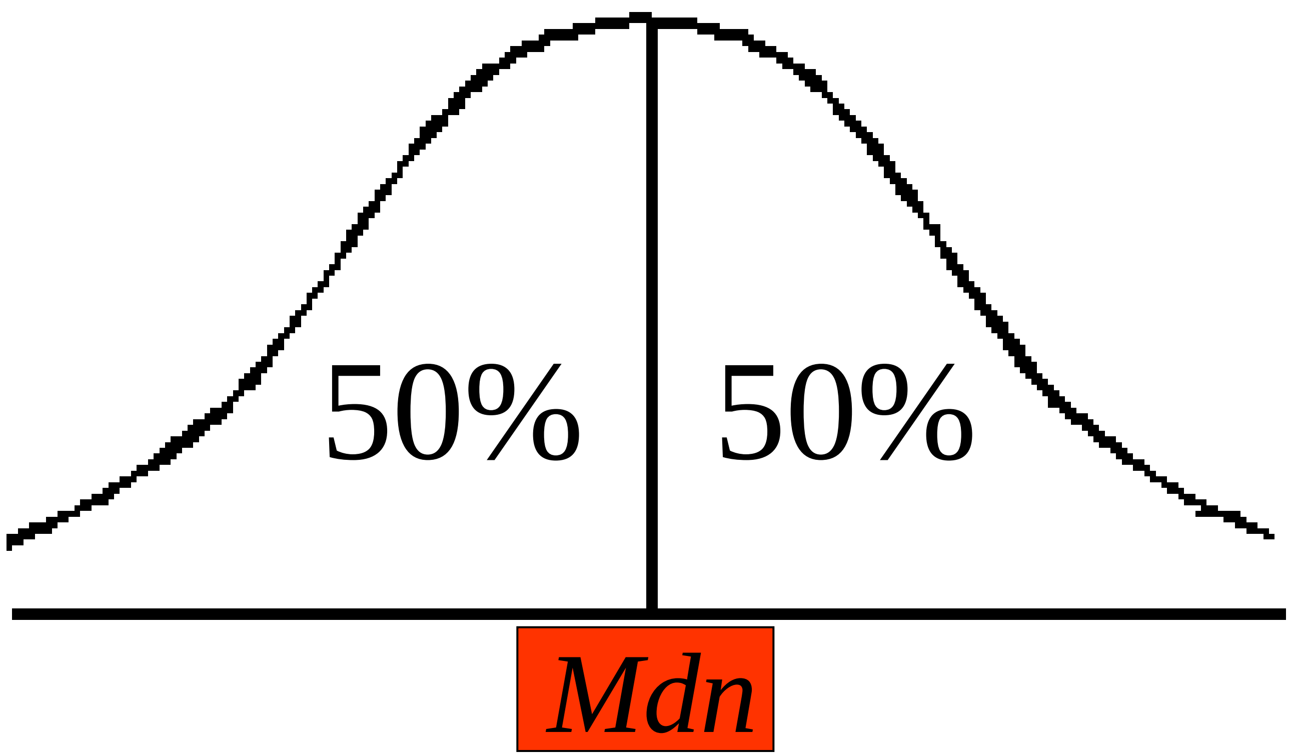
- Definitions:

  - It is the point of the distribution that divides it into 2 equal parts.
  - It is the value with the property that the number of observations smaller than itself is equal to the number of observations higher than itself.
  - It is the value that holds the central point of an ordered series of data.
  - The 50% of the values is above and the other 50% is below it.

2. Median (Mdn)

2.1. Graphic representation

- The Mdn is not defined as a data or particular measure, but a **point** (a value).
  - A point whose value does not necessarily have to match any observed value.



2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

ODD data set:

Example:

7 – 11 – 6 – 5 – 7 – 12 – 9 – 8 – 10 – 6 – 9

1st) Data is sorted from the lowest to the highest.

2nd) Central value is obtained:

$$\frac{n+1}{2}$$

2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

ODD data set:

1st) Data is sorted from the lowest to the highest:

5 – 6 – 6 – 7 – 7 – 8 – 9 – 9 – 10 – 11 – 12

2nd) Central value is obtained:

$$\frac{n+1}{2} = \frac{11+1}{2} = 6$$

2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

1º	**5**
2º	**6**
3º	**6**
4º	**7**
5º	**7**
6º	**8**
7º	**9**
8º	**9**
9º	**10**
10º	**11**
11º	**12**

Mdn = 8

2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

EVEN data set:

Example

23 – 35 – 43 – 29 – 34 – 41 – 33 – 38 – 38 – 32

1st) Data is sorted from the lowest to the highest:

2nd)

$$Mdn = \frac{CentralValue_1 + CentralValue_2}{2}$$

2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

EVEN data set:

1st) Data is sorted from the lowest to the highest:

23 – 29 – 32 – 33 – 34 – 35 – 38 – 38 – 41 – 43

2nd)

$$Mdn = \frac{34 + 35}{2} = 34.5$$

2. Median (Mdn)

2.2. Non-grouped data

Type I distributions: small data set

Mdn = 34.5

1º	2º	3º	4º	5º	6º	7º	8º	9º	10º
23	29	32	33	34	35	38	38	41	43

2. Median (Mdn)

2.3. Grouped data

Type II distributions: big data set

Procedure:

1. Calculate F_i: cumulative frequencies
2. Calculate n/2
3. Determine L_i: the lower exact limit from the interval that includes n/2 (the 50% of the observed data)

Exact limits = value ± 0.5 x measurement unit

4. Determine the f_i in that interval
5. Determine the F_i before that interval
6. Calculate the interval amplitude:

I = max – min (exact limits of the interval)

7. Calculate the formula:

$$Mdn = L_i + \frac{I}{f_i}\left(\frac{n}{2} - F_i\right)$$

X_i	f_i
16-21	4
22-27	9
28-33	21
34-39	22
40-45	16
46-51	11
52-57	7
58-63	8
64-69	2

2. Median (Mdn)

2.3. Grouped data

Type II distributions: big data set

1. F_i
2. $n/2 = 50$
3. L_i: $34 - 0.5 \times 1 = 33.5$
4. $f_4 = 22$
5. $F_3 = 34$
6. $I = 39.5 - 33.5 = 6$
7. $$Mdn = 33.5 + \frac{6}{22}\left(\frac{100}{2} - 34\right) = 37.86$$

X_i	f_i	^1F_i
16-21	4	4
22-27	9	13
28-33	21	34
34-39	22	56
40-45	16	72
46-51	11	83
52-57	7	90
58-63	8	98
64-69	2	100

3.

Which is the average score?

Arithmetic mean

$(\overline{X})$

3. Arithmetic mean

- Definition:
  - It is the central tendency index most commonly used.
  - It is the sum of all the observed values divided by the total number of them.

3. Arithmetic mean

3.1. Non-grouped data

Type I distributions: small data set

$$\overline{X} = \frac{\sum_{i=1}^{k} X_i}{n} = \frac{X_1 + X_2 + X_3 + \ldots + X_k}{n}$$

- Example: The 10 numbers below are the items remembered by 10 children in an immediate memory task

6 – 5 – 4 – 7 – 5 – 7 – 8 – 6 – 7 - 8

Calculate the arithmetic mean

3. Arithmetic mean

3.1. Non-grouped data

Type I distributions: small data set

$$\overline{X} = \frac{\sum_{i=1}^{10} X_i}{n} = \frac{6+5+4+7+5+7+8+6+7+8}{10} = \frac{63}{10} = 6.3$$

3. Arithmetic mean

3.1. Non-grouped data

Type I distributions: small data set

Example:

3 – 10 – 8 – 4 – 7 – 6 – 9 – 12 – 2 – 4

Calculate the mean

3. Arithmetic mean

3.1. Non-grouped data

Type I distributions: small data set

$$\overline{X} = \frac{\sum_{i=1}^{10} X_i}{n} = \frac{3+10+8+4+7+6+9+12+2+4}{10} = \frac{65}{10} = 6.5$$

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

Possibility 1:

MEAN

$$\bar{X} = \frac{\sum f_i X_i}{n} = \frac{f_1 X_1 + f_2 X_2 + f_3 X_3 + \ldots + f_k X_k}{n}$$

FREQUENCY TABLE

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

X_i	f_i
0	3
1	6
2	7
3	3
4	1

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

X_i	f_i		f_iX_i
0	3	3 x 0	0
1	6	6 x 1	6
2	7	7 x 2	14
3	3	3 x 3	9
4	1	1 x 4	4

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{(3x0)+(6x1)+(7x2)+(3x3)+(1x4)}{20} = \frac{33}{20} = 1.65$$

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

Possibility 2:

$$\overline{X} = \frac{\sum f_i X_i}{n} = \sum rfX_i$$

Example:

X_i	f_i
0	3
1	6
2	7
3	3
4	1

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

$$\overline{X} = \sum rfX_i = 0 + 0.30 + 0.70 + 0.45 + 0.20 = 1.65$$

X_i	f_i	rf_i	rf_iX_i
0	3	0.15	0.00
1	6	0.30	0.30
2	7	0.35	0.70
3	3	0.15	0.45
4	1	0.05	0.20
Σ	20	1	1.65

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

$$\overline{X} = \frac{\sum f_i X_i}{n} = \sum rfX_i$$

Example:

Calculate the arithmetic mean using the two formulas

Intervals	X_i	f_i
64-69		2
58-63		8
52-57		7
46-51		11
40-45		16
34-39		22
28-33		21
22-27		9
16-21		4

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

Intervals	X_i	f_i	f_iX_i	rf_i	rf_iX_i
64-69	66.5	2	133	0.02	1.33
58-63	60.5	8	484	0.08	4.84
52-57	54.5	7	381.5	0.07	3.815
46-51	48.5	11	533.5	0.11	5.335
40-45	42.5	16	680	0.16	6.8
34-39	36.5	22	803	0.22	8.03
28-33	30.5	21	640.5	0.21	6.405
22-27	24.5	9	220.5	0.09	2.205
16-21	18.5	4	74	0.04	0.74
Σ		100	3950	1	39.5

3. Arithmetic mean

3.2. Grouped data

Type II distributions: big data set

Possibility 1**:**

$$\boxed{\overline{X} = \frac{\sum f_i X_i}{n} = \frac{3950}{100} = 39.5}$$

Possibility 2**:**

$$\boxed{\overline{X} = \sum rfX_i = 39.5}$$

3. Arithmetic mean

3.3. Properties:

1. Sum of the deviation of all the values from their arithmetic mean is 0:

$$\sum (X_i - \overline{X}) = 0$$

Example:

9, 3, 6, 7, 5

3. Arithmetic mean

3.3. Properties:

Example:

$$\overline{X} = \frac{\sum X}{n} = \frac{30}{5} = 6$$

X_i	$(X_i - \overline{X})$
9	3
3	-3
6	0
7	1
5	-1

$$\sum (X_i - \overline{X}) = 3 + (-3) + 0 + 1 + (-1) = 0$$

3. Arithmetic mean

3.3. Properties:

2. Sum of the deviations square with regard to the arithmetic mean is less than with regard to any other value or average:

$$\sum (X_i - \overline{X})^2 < \sum (X_i - c)^2$$

$$(c \neq \overline{X})$$

Example:

9, 3, 6, 7, 5

c = 3

3. Arithmetic mean

3.3. Properties:

$$\sum(X_i - \overline{X})^2 = (9-6)^2 + (3-6)^2 + (6-6)^2 + (7-6)^2 + (5-6)^2 =$$
$$3^2 + (-3)^2 + 0^2 + 1^2 + (-1)^2 = 9 + 9 + 0 + 1 + 1 = 20$$

$$\sum(X_i - c)^2 = (9-3)^2 + (3-3)^2 + (6-3)^2 + (7-3)^2 + (5-3)^2 =$$
$$6^2 + (0)^2 + 3^2 + 4^2 + (2)^2 = 36 + 0 + 9 + 16 + 4 = 65$$

$$\sum(X_i - \overline{X})^2 < \sum(X_i - c)^2 \Rightarrow 20 < 65$$

3. Arithmetic mean

3.3. Properties:

3. If every value of the variable X is increased by a constant, the arithmetic mean will be increased by the same constant:

$$\frac{\sum (X_i + a)}{n} = \overline{X} + a$$

Example:

9, 3, 6, 7, 5

a=2

3. Arithmetic mean

3.3. Properties:

X_i	$X_i + a$
9	11
3	5
6	8
7	9
5	7
Σ=30	Σ=40

$$\frac{\sum (X_i + a)}{n} = \frac{40}{5} = 8$$

$$\overline{X} + a = 6 + 2 = 8$$

$$\frac{\sum (X_i + a)}{n} = \overline{X} + a \Rightarrow 8 = 8$$

3. Arithmetic mean

3.3. Properties:

4. If every value of the variable X is multiplied by a constant, the arithmetic mean will be multiplied by the same constant:

$$\frac{\sum (X_i * a)}{n} = \overline{X} * a$$

Example:

9, 3, 6, 7, 5

a=4

3. Arithmetic mean

3.3. Properties:

X_i	$X_i * a$
9	36
3	12
6	24
7	28
5	20
Σ=30	Σ=120

$$\frac{\sum (X_i * a)}{n} = \frac{120}{5} = 24$$

$$\overline{X} * a = 6 * 4 = 24$$

$$\frac{\sum (X_i * a)}{n} = \overline{X} * a \Rightarrow 24 = 24$$

3. Arithmetic mean

3.3. Properties:

5. If $\overline{X}_1$ and $\overline{X}_2$ are the means of the two groups computed from the values n_1 and n_2, then the mean is given by the formula:

$$\overline{X} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2}{n_1 + n_2}$$

3. Arithmetic mean

3.3. Properties:

Example: Three groups with 3, 4 and 5 participants respectively are conformed in order to check the effect of a drug in a perceptive task. The first group receive a placebo; the second group, 1 mg of the drug; and the third one, 2 mg. The results are presented in the table below:

Group 1	Group 2	Group 3
4	3	1
6	5	2
5	4	4
	0	2
		1

Calculate the arithmetic mean using the two formulas below:

$$\overline{X} = \frac{n_1 \overline{X}_1 + n_2 \overline{X}_2 + n_3 \overline{X}_3}{n_1 + n_2 + n_3}$$

$$\overline{X} = \frac{\sum X}{n}$$

3. Arithmetic mean

3.3. Properties:

Group 1	Group 2	Group 3
4	3	1
6	5	2
5	4	4
	0	2
		1
Σ=15	Σ=12	Σ=10

$$\overline{X}_1 = \frac{15}{3} = 5$$

$$\overline{X}_2 = \frac{12}{4} = 3$$

$$\overline{X}_3 = \frac{10}{5} = 2$$

$$\overline{X} = \frac{n_1\overline{X}_1 + n_2\overline{X}_2 + n_3\overline{X}_3}{n_1 + n_2 + n_3} = \frac{3x5 + 4x3 + 5x2}{3+4+5} = \frac{15+12+10}{12} = \frac{37}{12} = 3.083$$

$$\overline{X} = \frac{\sum X}{n} = \frac{4+6+5+3+5+4+0+1+2+4+2+1}{12} = \frac{37}{12} = 3.083$$

3. Arithmetic mean

3.3. Properties:

6. The arithmetic mean of a variable that is a lineal combination of other variables is equal to the lineal combination of the arithmetic means of those variables.

   Whether:

$$X = a_1 X_1 + a_2 X_2 + \ldots + a_k X_k$$

   Then:

$$\overline{X} = a_1 \overline{X}_1 + a_2 \overline{X}_2 + \ldots + a_k \overline{X}_k$$

3. Arithmetic mean

3.3. Properties:

Example: The data presented in the table below are the scores obtained by 4 participants in three intelligence tests (X_1, X_2 and X_3):

X_1	X_2	X_3
2	4	5
3	5	6
4	6	4
3	5	1

From these variables, a new one is conformed: $X=2X_1-X_2+3X_3$.

Calculate:

a) The value of X for each participant.

b) The arithmetic mean of the variable X using the two formulas below:

$$\overline{X} = a_1\overline{X}_1 + a_2\overline{X}_2 + \ldots + a_k\overline{X}_k \qquad \overline{X} = \frac{\sum X}{n}$$

3. Arithmetic mean

3.3. Properties:

X_1	X_2	X_3	$2X_1$	$-X_2$	$3X_3$	a) X
2	4	5	4	-4	15	15
3	5	6	6	-5	18	19
4	6	4	8	-6	12	14
3	5	1	6	-5	3	4
Σ=12	Σ=20	Σ=16				Σ=52

$$\overline{X}_1 = \frac{12}{4} = 3$$

$$\overline{X}_2 = \frac{20}{4} = 5$$

$$\overline{X}_3 = \frac{16}{4} = 4$$

$$\overline{X} = a_1\overline{X}_1 + a_2\overline{X}_2 + \ldots + a_k\overline{X}_k = 2\overline{X}_1 - \overline{X}_2 + 3\overline{X}_3 = 2x3 - 5 + 3x4 =$$
$$6 - 5 + 12 = 13$$

$$\overline{X} = \frac{\sum X}{n} = \frac{52}{4} = 13$$

Comparison between measures of central tendency

- We usually prefer the **mean**:
  - Other statistics are based on the mean.
  - It's the best estimator of its parameter.
- We only prefer the **median**:
  - When the variable is ordinal.
  - When there are very extreme data.
  - When there are open intervals.
- We only prefer the **mode**:
  - When the variable is qualitative or nominal.
  - When the open interval includes the median.

Example 1

The degree of agreement in considering "shouting" as a sign of aggression in a sample is presented in the table below:

X_i	f_i
1	3
2	6
3	5
4	6
5	12

Calculate the most appropriate central tendency index.

Example 1

X_i	f_i	F_i
1	3	3
2	6	9
3	5	14
4	6	20
5	12	32
Σ	32	

$$\frac{n+1}{2} = \frac{32+1}{2} = \frac{33}{2} = 16.5$$

Central values corresponding to positions 16th and 17th

$$Mdn = \frac{centralvalue_1 + centralvalue_2}{2} = \frac{4+4}{2} = 4$$

Example 2

The table below represents the number of rituals that students do before an exam:

X_i	f_i
2	1
3	8
4	10
5	21
6	19
7	9
8	2
9	1

1. Calculate the mode, the median and the arithmetic mean.
2. Which is the most appropriate central tendency index?

Example 2

X_i	f_i	F_i	f_iX_i
2	1	1	2
3	8	9	24
4	10	19	40
5	21	40	105
6	19	59	114
7	9	68	63
8	2	70	16
9	1	71	9
Σ	71		373

1. Mo = 5

   Mdn:

   $$\frac{n+1}{2} = \frac{71+1}{2} = \frac{72}{2} = 36$$

   Position 36th → Mdn = 5

   $$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{373}{71} = 5.253$$

2. The most appropriate central tendency index is the arithmetic mean (for quantitative variables).

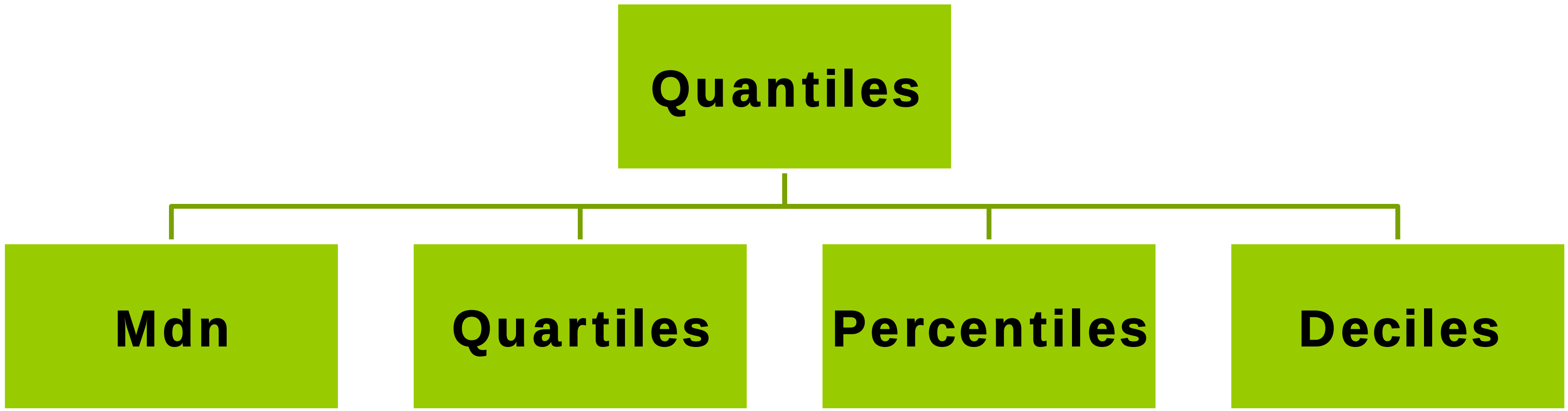
Group and individual indexes

Group indexes

- Central tendency
- Variability (Dispersion)
- Bias or Skewness (Asymmetry)
- Kurtosis

Individual indexes

- Position
  - Deciles (D_i)
  - Percentiles (P_i)
  - Quartiles (Q_i)
- Raw scores (X_i)
- Differential scores (x_i)
- Standard scores (Z_i)


Quantiles
Mdn
Quartiles
Percentiles
Deciles

Position indexes

- Position measures are used to provide information about the relative position of a **case** with respect to its data set. They are individual indexes.

Quantiles

Definition: They divide the distribution in k parts with the same amount of data.

- **<u>Mdn</u>**: divides the distribution in 2 parts:

$$\boxed{k = 2}$$

- **<u>Quartiles</u> (Q_i)**: divide the distribution in 4 parts: Q_1, Q_2, Q_3:

$$\boxed{k = 4}$$

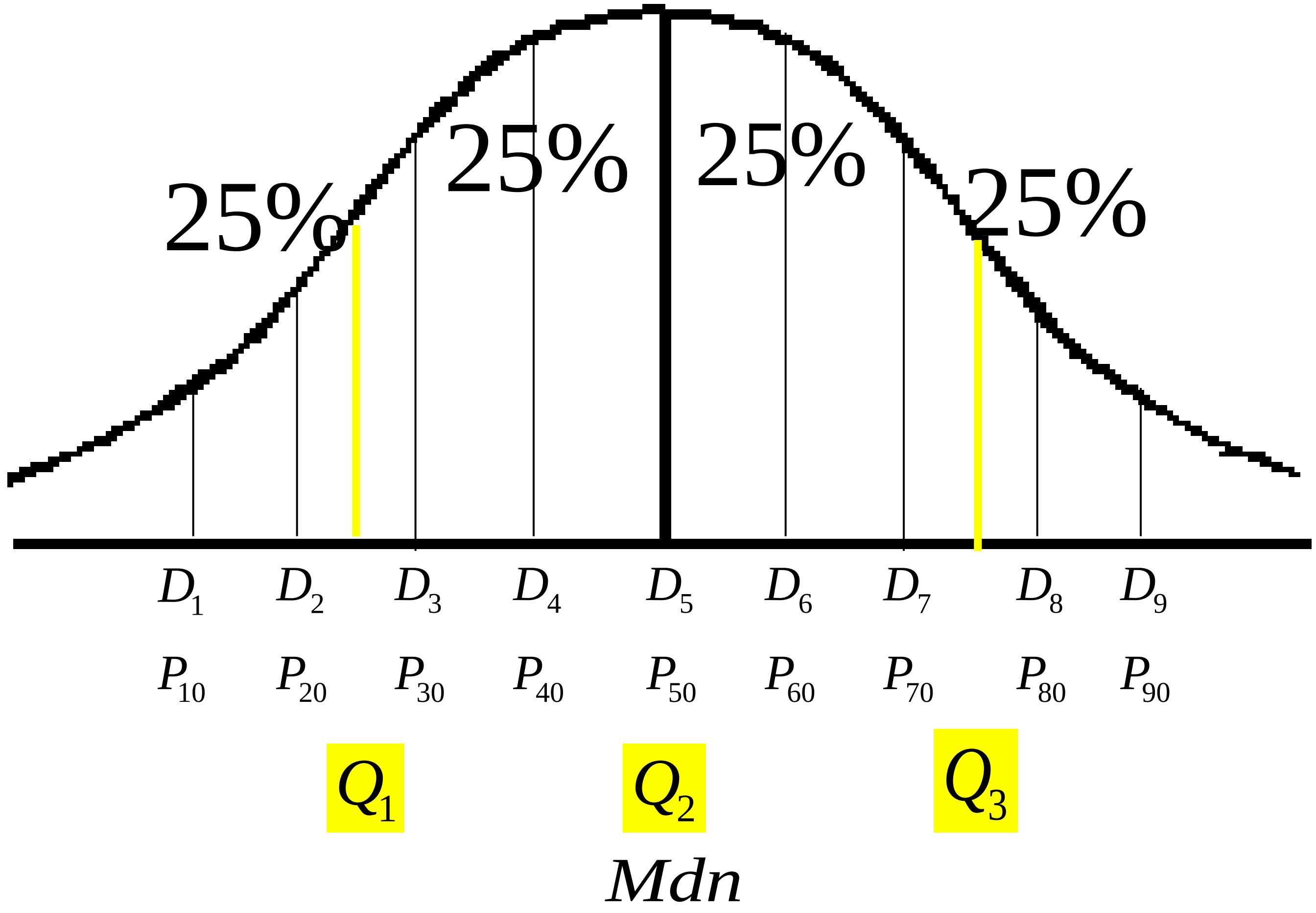
- **<u>Deciles</u> (D_i)**: divide the distribution in 10 parts : D_1, D_2, … , D_9:

$$\boxed{k = 10}$$

- **<u>Percentiles</u> or Centiles(P_i or C_i)**: divide the distribution in 100 parts: P_1, P_2, … , P_{99}:

$$\boxed{k = 100}$$

Quantiles
Graphic representation



$C_{50} = P_{50} = Q_2 = D_5 = \text{Mdn}$

Percentile. Definition

Percentile rank indicates the participant's standing relative to other participants in the norm group. For example, a student's percentile rank on a norm-referenced test tells us what proportion of students in the norm group scored the same or lower than a target student.

Procedure

1. Position:

$$\frac{i(n+1)}{k} = I + D$$

i = the quartile, decile or percentile in study.

k = number of divisions.

I = integer.

D= decimals.

2. Value:

$$X_I + D(X_{I+1} - X_I)$$

X_I = value of the position of the integer part.

X_{I+1}= value of the following position to the integer part.

Example 1: Q_3

X_i	f_i
2	3
3	11
4	15
5	30
6	23
7	12
8	4
9	1

Example 1: Q_3

1. Position:

$$\frac{i(n+1)}{k} = \frac{3(99+1)}{4} = 75$$

2. Value:

$Q_3 = 6$
(the result is an integer)

X_i	f_i	F_i	Positions
2	3	3	1 – 3
3	11	14	4 – 14
4	15	29	15 – 29
5	30	59	30 – 59
6	23	82	60 – 82
7	12	94	83 – 94
8	4	98	95 – 98
9	1	99	99

Examples 2 and 3: D_6 and P_{13}

X_i	f_i
2	1
3	8
4	10
5	21
6	19
7	9
8	2
9	1

Example 2: D_6

1. Position:

$$\frac{i(n+1)}{k} = \frac{6(71+1)}{10} = 43.2$$

X_i	f_i	F_i	Positions
2	1	1	1
3	8	9	2 – 9
4	10	19	10 – 19
5	21	40	20 – 40
6	19	59	41 – 59
7	9	68	60 – 68
8	2	70	69 – 70
9	1	71	71

2. Value:

$D_6 = 6$

Example 3: P_{13}

1. Position:

$$\frac{i(n+1)}{k} = \frac{13(71+1)}{100} = 9.36$$

X_i	f_i	F_i	Positions
2	1	1	1
3	8	9	2 – 9
4	10	19	10 – 19
5	21	40	20 – 40
6	19	59	41 – 59
7	9	68	60 – 68
8	2	70	69 – 70
9	1	71	71

2. Value:

9 = I; 0.36 = D

$X_I + D(X_{I+1} - X_I)$

$P_{13} = 3 + 0.36(4 - 3) = 3.36$

Example 4: $C_{50} = P_{50} = Q_2 = D_5 = Mdn$

X	f_i
1	30
2	15
3	25
4	6
5	4

Example 4: $C_{50} = P_{50} = Q_2 = D_5 = Mdn$

X	f_i	F_i
1	30	30
2	15	45
3	25	70
4	6	76
5	4	80

1. Position:

$C_{50} = P_{50}$

$$\frac{i(n+1)}{k} = \frac{50(80+1)}{100} = 40.5$$

Q_2

$$\frac{i(n+1)}{k} = \frac{2(80+1)}{4} = 40.5$$

D_5

$$\frac{i(n+1)}{k} = \frac{5(80+1)}{10} = 40.5$$

Mdn

$$\frac{i(n+1)}{k} = \frac{1(80+1)}{2} = 40.5$$

2. Value:

$C_{50} = P_{50} =$
$Q_2 = D_5 =$
$Mdn = 2$

Example 5: C_{25}, C_{75}, C_{38}, C_{90}, C_{95}

X	f_i
1	30
2	15
3	25
4	6
5	4

Example 5: C_{25}, C_{75}, C_{38}, C_{90}, C_{95}

X	f_i	F_i
1	30	30
2	15	45
3	25	70
4	6	76
5	4	80

$C_{25} = 1$

$$\frac{i(n+1)}{k} = \frac{25(80+1)}{100} = \frac{2025}{100} = 20.25$$

$C_{75} = 3$

$$\frac{i(n+1)}{k} = \frac{75(80+1)}{100} = \frac{6075}{100} = 60.75$$

$C_{38} = X_l + D(X_{l+1} - X_l) = 1 + 0.78(2-1) = 1.78$

$$\frac{i(n+1)}{k} = \frac{38(80+1)}{100} = \frac{3078}{100} = 30.78$$

Example 5: C_{25}, C_{75}, C_{38}, C_{90}, C_{95}

X	f_i	F_i
1	30	30
2	15	45
3	25	70
4	6	76
5	4	80

$C_{90} = 4$

$$\frac{i(n+1)}{k} = \frac{90(80+1)}{100} = \frac{7290}{100} = 72.9$$

$C_{95} = X_I + D(X_{I+1} - X_I) = 4 + 0.95(5-4) = 4.95$

$$\frac{i(n+1)}{k} = \frac{95(80+1)}{100} = \frac{7695}{100} = 76.95$$

Example 6:

The tables below present the results obtained from a sample of women and men in the province of Seville in a psychological test:

Women	
X_i	f_i
19	25
20	15
21	6
22	5
25	1
29	8

Men	
X_i	f_i
20	4
21	8
24	10
26	12
27	8
29	8

Calculate Q_3 in women sample and P_{84} in men sample

Example 6:

Women		
X_i	f_i	F_i
19	25	25
20	15	40
21	6	46
22	5	51
25	1	52
29	8	60

$$\frac{i(n+1)}{k} = \frac{3(60+1)}{4} = \frac{183}{4} = 45.75$$

$Q_3 = 21$

Example 6:

Men		
X_i	f_i	F_i
20	4	4
21	8	12
24	10	22
26	12	34
27	8	42
29	8	50

$$\frac{i(n+1)}{k} = \frac{84(50+1)}{100} = \frac{4284}{100} = 42.84$$

$P_{84} = X_l + D(X_{l+1} - X_l) = 27 + 0.84(29-27) = 27 + 0.84x2 = 27 + 1.68 = 28.68$

Procedure when there are intervals

It is analogous to the Mdn calculation. Steps:

1. Calculate F_i: cumulative frequencies
2. Calculate i*n/k
3. Determine L_i: the lower exact limit from the interval that includes i*n/k

   Exact limits = value ± 0.5 x measurement unit
4. Determine the f_i in that interval
5. Determine the F_i before that interval
6. Calculate the interval amplitude:

   I = max – min (exact limits of the interval)
7. Calculate the formula:

$$P_i, C_i, D_i, Q_i = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right)$$

Example:

Intervals	f_i
16-21	4
22-27	9
28-33	21
34-39	22
40-45	16
46-51	11
52-57	7
58-63	8
64-69	2

Calculate C_{90}

Example:

Intervals	f_i	${}^1F_i$
16-21	4	4
22-27	9	13
28-33	21	34
34-39	22	56
40-45	16	72
46-51	11	83
52-57	7	90
58-63	8	98
64-69	2	100

1. F_i
2. $i*n/k = 90 \times 100/100 = 90$
3. L_i: $52 - 0.5 \times 1 = 51.5$
4. $f_7 = 7$
5. $F_6 = 83$
6. $I = 57.5 - 51.5 = 6$
7.

$$C_{90} = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) =$$

$$51.5 + \frac{6}{7}\left(\frac{90*100}{100} - 83\right) =$$

$$51.5 + 0.857(90 - 83) = 51.5 + 0.857*7 =$$

$$51.5 + 6 = 57.5$$

The question could be the opposite:

Intervals	f_i
16-21	4
22-27	9
28-33	21
34-39	22
40-45	16
46-51	11
52-57	7
58-63	8
64-69	2

With the same set of data, which percentile corresponds to X = 57.5

$$i = \frac{\dfrac{(X - L_i) * f_i}{I} + F_i}{n} * 100$$

$$i=\frac{\frac{(57.5-51.5)*7}{6}+83}{100}*100=\frac{6*7}{6}+83=\frac{42}{6}+83=7+83=90$$

57.5 corresponds to centile 90

Group and individual indexes

Group indexes

- **Central tendency**
- **Variability (Dispersion)**
- **Bias or Skewness (Asymmetry)**
- **Kurtosis**

Individual indexes

- **Position**
  - **Deciles (D_i)**
  - **Percentiles (P_i)**
  - **Quartiles (Q_i)**
- **Raw scores (X_i)**
- **Differential scores (x_i)**
- **Standard scores (Z_i)**

How is the data arranged with respect to the distribution center?

How far or together are the data from each other?

Variability or dispersion indexes

- **Total amplitude (Ta)**
- Mean deviation (MD)
- Variance (S^2)
- **Standard deviation (S)**
- **Semiinterquartile amplitude (Q)**
- **Pearson variation coefficient (VC)**

Variability or dispersion indexes

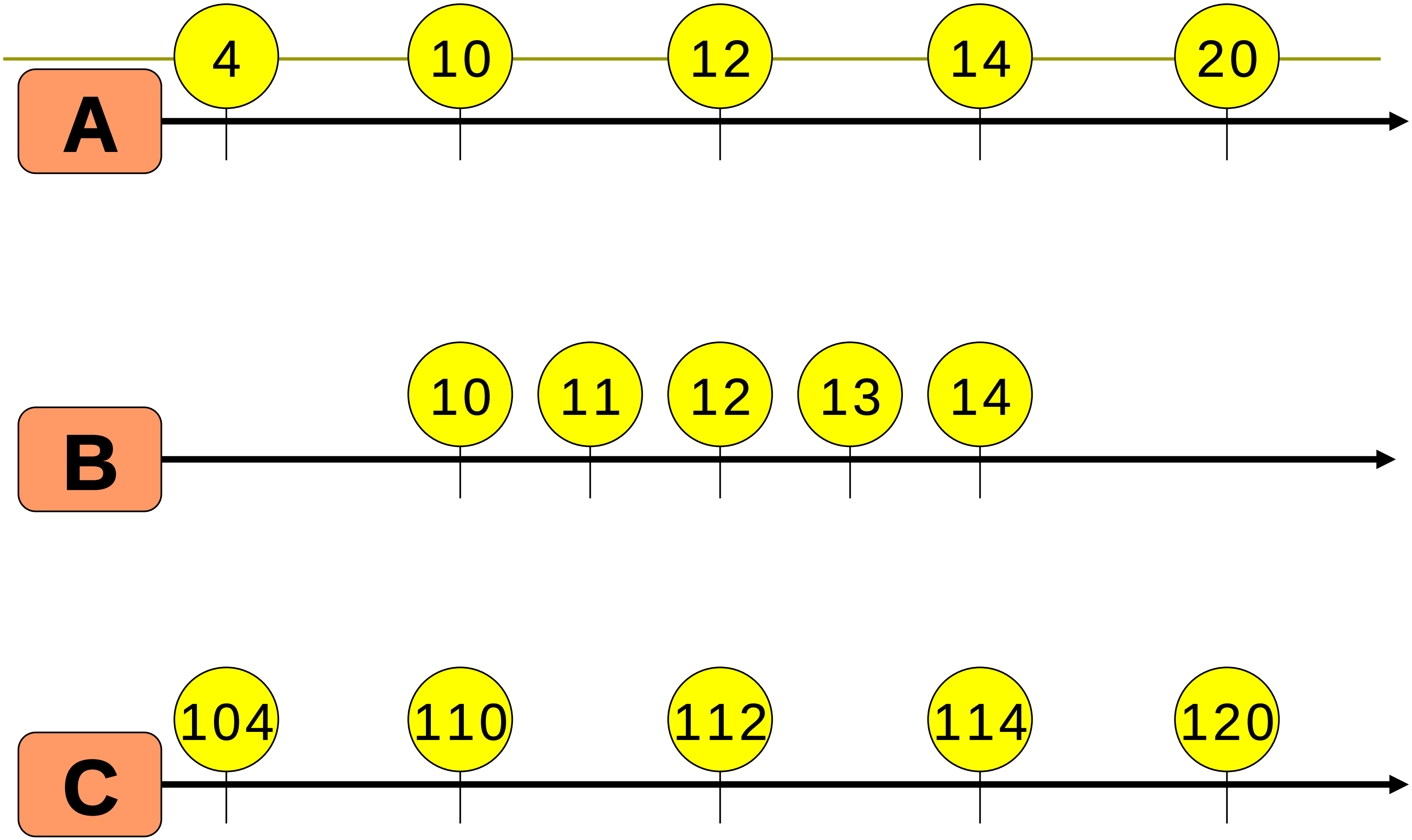
Introduction

A	**4**	**10**	**12**	**14**	**20**	$\overline{X}_A = 12$
B	**10**	**11**	**12**	**13**	**14**	$\overline{X}_B = 12$
C	**104**	**110**	**112**	**114**	**120**	$\overline{X}_C = 112$

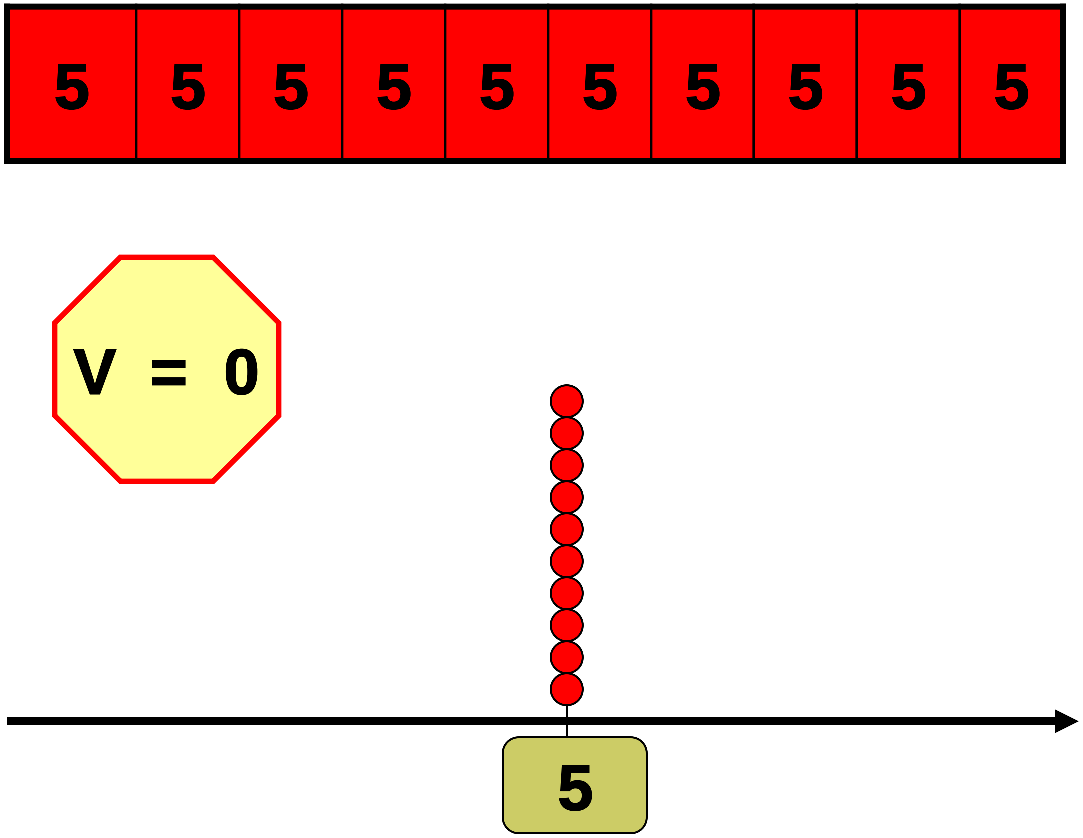
$$\overline{X}_A = \overline{X}_B \neq \overline{X}_C$$

$$V_A = V_C \neq V_B$$

Introduction



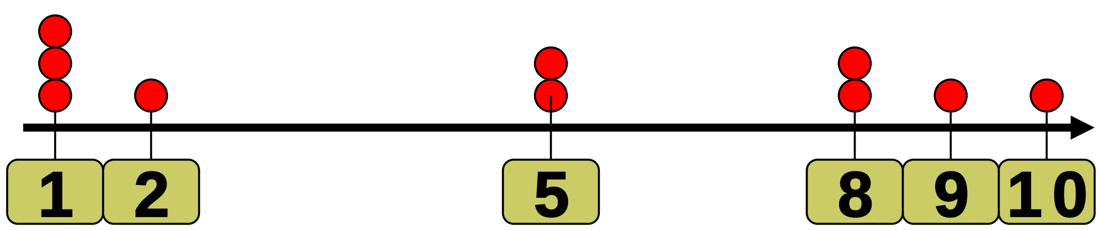
Introduction



Introduction

1	1	1	2	5	5	8	8	9	10

$V \neq 0$



Introduction

INDEXES

- T_A — 1. TOTAL AMPLITUDE
- MD — 2. MEAN DEVIATION
- S^2 — 3. VARIANCE
- S — 4. STANDARD DEVIATION
- Q — 5. SEMIINTERQUARTILE AMPLITUDE
- vc — 6. PEARSON VARIATION COEFFICIENT

1. Total amplitude (or range)

- Definition: It is the distance between the maximum and minimum value of a data set.

$$T_A = X_{Max} - X_{Min}$$

- Advantage: easy to calculate.
- Disadvantages:
  - It is unstable. It only uses two data from the sample, so it is very sensitive to extreme values and insensitive to average values.
  - It is not independent of the sample sizes (T_A obtained in samples of different sizes are not directly comparable).

1. Total amplitude (or range)

Example:

A) 3 7 8 9 10 11 12 13

B) 7 7 8 9 10 11 12 13

C) 7 10 10 10 10 10 10 13

1. Total amplitude (or range)

Example:

A) 3 7 8 9 10 11 12 13

$T_A = 13 - 3 = 10$

B = C

B) 7 7 8 9 10 11 12 13

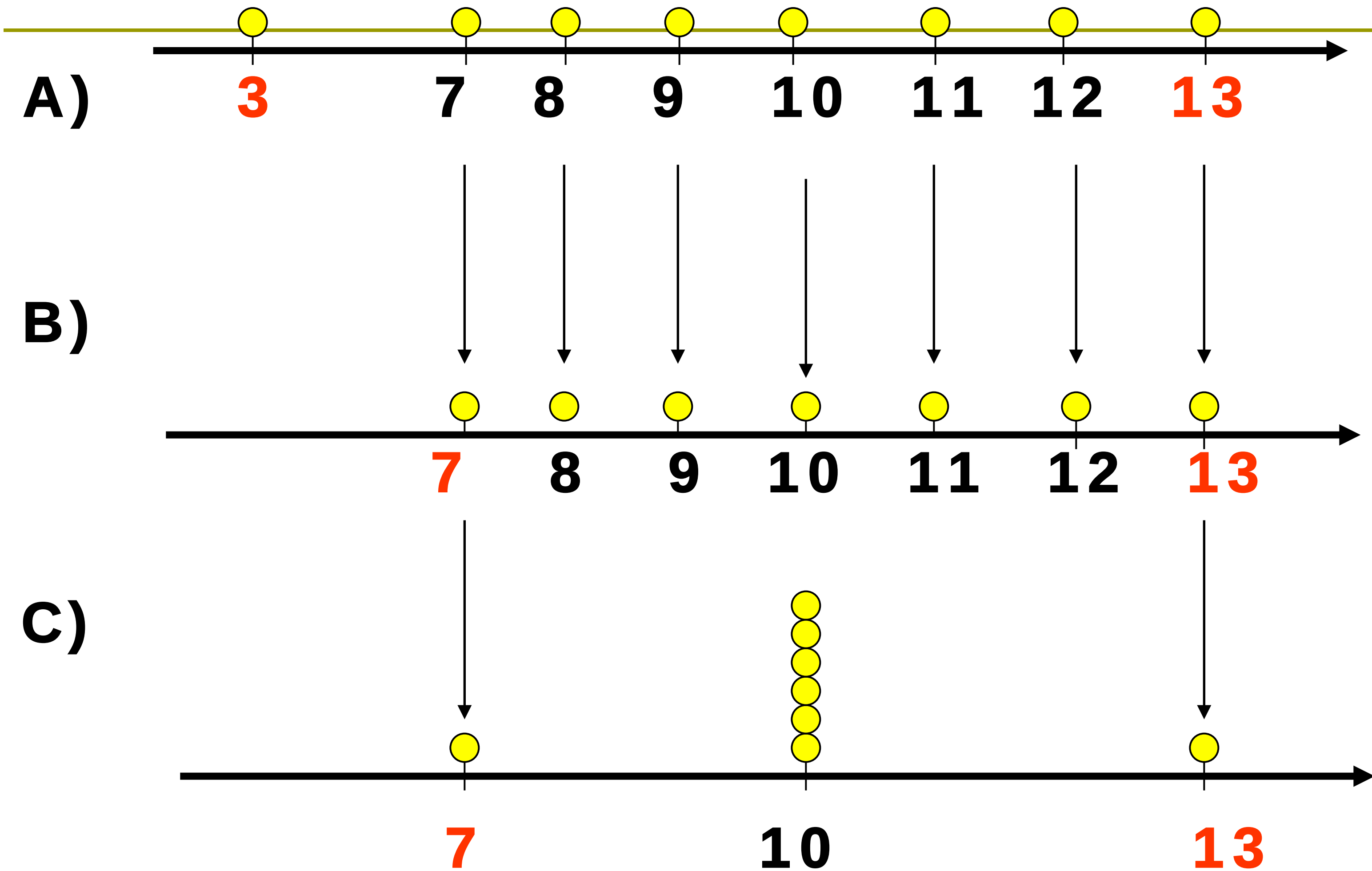
$T_A = 13 - 7 = 6$

A > B and C

C) 7 10 10 10 10 10 10 13

$T_A = 13 - 7 = 6$

1. Total amplitude (or range)



2. Mean deviation

- Definition: arithmetic mean of all the deviations taken with positive sign.

$$MD = \frac{\sum f_i \left| X_i - \overline{X} \right|}{n}$$

- Interpretation: the highest MD, the most deviation.
- Advantage: It is independent of the sample sizes.
- Disadvantage: statistic difficulty due to the absolute values.

2. Mean deviation

- Example:

X	f_i
4-6	2
7-9	8
10-12	10
13-15	8
16-18	2

Calculate the MD.

2. Mean deviation

X	X_i	f_i	f_iX_i	$X_i - \overline{X}$	$\left\|X_i - \overline{X}\right\|$	$f_i\left\|X_i - \overline{X}\right\|$
4-6	5	2	10	-6	6	12
7-9	8	8	64	-3	3	24
10-12	11	10	110	0	0	0
13-15	14	8	112	3	3	24
16-18	17	2	34	6	6	12
		n=30	Σ=330			72

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{330}{30} = 11$$

$$MD = \frac{\sum f_i \left|X_i - \overline{X}\right|}{n} = \frac{72}{30} = 2.4$$

3. Variance and 4. Standard deviation

Variance:

$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2$$

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n}$$

$$S^2 = \sum rf_i X_i^2 - \overline{X}^2$$

Standard deviation:

$$S = \sqrt{S^2}$$

Quasi-variance:

$$\hat{S}^2 = S^2 \frac{n}{n-1}$$

Quasi standard deviation:

$$\hat{S} = \sqrt{\hat{S}^2}$$

They estimate more accurately the variance and the standard deviation of the population

3. Variance and
4. Standard deviation

Example 1:

Calculate the variance using the formulas:

X_i
3
6
9
12
15

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n}$$

$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2$$

3. Variance and
4. Standard deviation

X_i	$X_i - \overline{X}$	$(X_i - \overline{X})^2$
3	-6	36
6	-3	9
9	0	0
12	3	9
15	6	36
Σ	0	90

$$\overline{X} = \frac{\sum X_i}{n} = \frac{45}{5} = 9$$

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n} = \frac{90}{5} = 18$$

3. Variance and
4. Standard deviation

X_i	X^2
3	9
6	36
9	81
12	144
15	225
45	495

$$\overline{X} = \frac{\sum X_i}{n} = \frac{45}{5} = 9$$

$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2 = \frac{495}{5} - 9^2 = 99 - 81 = 18$$

3. Variance and
4. Standard deviation

Example 2:

X_i	f_i
3	1
4	7
5	6
6	3
7	3

Calculate the variance using the formulas:

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n}$$

$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2$$

$$S^2 = \sum rf_i X_i^2 - \overline{X}^2$$

Calculate the standard deviation, the quasi-variance and the quasi standard deviation

3. Variance and
4. Standard deviation

X_i	f_i	f_iX_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$
3	1	3	-2	4	4
4	7	28	-1	1	7
5	6	30	0	0	0
6	3	18	1	1	3
7	3	21	2	4	12
Σ	20	100	0	10	26

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{100}{20} = 5$$

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n} = \frac{26}{20} = 1.3$$

3. Variance and
4. Standard deviation

X_i	f_i	$f_i X_i$	X_i^2	$f_i X^2$
3	1	3	9	9
4	7	28	16	112
5	6	30	25	150
6	3	18	36	108
7	3	21	49	147
Σ	20	100		526

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{100}{20} = 5$$

$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2 = \frac{526}{20} - 5^2 = 26.3 - 25 = 1.3$$

3. Variance and
4. Standard deviation

X_i	f_i	$f_i X_i$	rf_i	X_i^2	rf_iX^2
3	1	3	0.05	9	0.45
4	7	28	0.35	16	5.6
5	6	30	0.3	25	7.5
6	3	18	0.15	36	5.4
7	3	21	0.15	49	7.35
Σ	20	100	1		26.3

$$S^2 = \sum rf_i X_i^2 - \overline{X}^2 = 26.3 - 5^2 = 1.3$$

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{100}{20} = 5$$

$$S = \sqrt{S^2} = \sqrt{1.3} = 1.14$$

$$\hat{S}^2 = S^2 \frac{n}{n-1} = 1.3 \frac{20}{19} = 1.3x1.053 = 1.369$$

$$\hat{S} = \sqrt{\hat{S}^2} = \sqrt{1.369} = 1.17$$

3. Variance and
4. Standard deviation

Example with intervals:

X	f_i
60-64	20
55-59	30
50-54	100
45-49	30
40-44	20

Calculate the variance using the formula:

$$S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n}$$

3. Variance and
4. Standard deviation

X	f_i	X_i	f_iX_i	$(X_i - \overline{X})$	$(X_i - \overline{X})^2$	$f_i(X_i - \overline{X})^2$
60-64	20	62	1240	10	100	2000
55-59	30	57	1710	5	25	750
50-54	100	52	5200	0	0	0
45-49	30	47	1410	-5	25	750
40-44	20	42	840	-10	100	2000
Σ	200		10400	0		5500

$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{10400}{200} = 52 \qquad S^2 = \frac{\sum f_i \left(X_i - \overline{X}\right)^2}{n} = \frac{5500}{200} = 27.5$$

5. Semiinterquartile amplitude

- Definition: it is the semidistance between quartile 3 and quartile 1.

$$Q = \frac{Q_3 - Q_1}{2}$$

- It is usually calculated:
  - When we only want to consider the central scores of the distribution.
  - When the central tendency index recommended is the median.

5. Semiinterquartile amplitude

A

X_A	f_A
1	35
2	40
3	75
4	30
5	20

B

X_B	f_B
1	5
2	20
3	40
4	80
5	55

Example:

The tables present the frequency distribution to two different items (A and B).

1. Calculate Q in both items A and B.
2. Which is the item that presents more homogeneity in their answers?

5. Semiinterquartile amplitude

A

X_A	f_A	F_A
1	35	35
2	40	75
3	75	150
4	30	180
5	20	200

$$Q = \frac{Q_3 - Q_1}{2} = \frac{3.75 - 2}{2} = \frac{1.75}{2} = 0.875$$

Q_3

1. Position $\frac{i(n+1)}{k} = \frac{3(200+1)}{4} = 150.75$
2. Value

   $X_I + D(X_{I+1} - X_I)$

   $Q_3 = 3 + 0.75\ (4 - 3) = 3.75$

Q_1

1. Position $\frac{i(n+1)}{k} = \frac{1(200+1)}{4} = 50.25$
2. Value

   $Q_1 = 2$

5. Semiinterquartile amplitude

B

X_B	f_B	F_B
1	5	5
2	20	25
3	40	65
4	80	145
5	55	200

$$Q = \frac{Q_3 - Q_1}{2} = \frac{5-3}{2} = \frac{2}{2} = 1$$

Q_3

1. Position $\frac{i(n+1)}{k} = \frac{3(200+1)}{4} = 150.75$
2. Value

$Q_3 = 5$

Q_1

1. Position $\frac{i(n+1)}{k} = \frac{1(200+1)}{4} = 50.25$
2. Value

$Q_1 = 3$

$Q_A = 0.875 < Q_B = 1$

The item that presents more homogeneity in their answers is item A

5. Semiinterquartile amplitude

Example with intervals:

Intervals	**f_i**
50-59	5
60-69	6
70-79	18
80-89	31
90-99	9
100-109	7
110-119	3
120-129	1

Steps:

1. F_i
2. i*n/k
3. L_i: value - 0.5 x measurement unit
4. f_i
5. F_i
6. I = max – min (exact limits of the interval)
7. $$Q_i = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right)$$

Calculate the semiinterquartile amplitude

5. Semiinterquartile amplitude

$$Q = \frac{Q_3 - Q_1}{2} = \frac{89.513 - 74.495}{2} = \frac{15.018}{2} = 7.509$$

Intervals	f_i	${}^1F_i$
50-59	5	5
60-69	6	11
70-79	18	29
80-89	31	60
90-99	9	69
100-109	7	76
110-119	3	79
120-129	1	80

Q_3

1. F_i
2. $i*n/k = 3 \times 80/4 = 60$
3. L_i: $80 - 0.5 \times 1 = 79.5$
4. $f_4 = 31$
5. $F_3 = 29$
6. $I = 89.5 - 79.5 = 10$
7.

$$Q_3 = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) = 79.5 + \frac{10}{31}\left(\frac{3*80}{4} - 29\right) =$$

$$79.5 + 0.323(60 - 29) = 79.5 + 10.013 = 89.513$$

5. Semiinterquartile amplitude

Intervals	**f_i**	**^1F_i**
50-59	5	5
60-69	6	11
70-79	18	29
80-89	31	60
90-99	9	69
100-109	7	76
110-119	3	79
120-129	1	80

Q_1

1. F_i
2. i*n/k = 1 x 80/4 = 20
3. L_i: 70 – 0.5 x 1 = 69.5
4. f_3 = 18
5. F_2 = 11
6. I = 79.5 – 69.5 = 10
7.

$$Q_1 = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) = 69.5 + \frac{10}{18}\left(\frac{1*80}{4} - 11\right) =$$

$$69.5 + 0.555(20 - 11) = 69.5 + 0.555x9 = 69.5 + 4.995$$

$$= 74.495$$

6. Pearson variation coefficient

$$VC = \frac{S}{\overline{\overline{X}}} 100$$

- It is useful to compare two standard deviations of different samples or different variables.
- It also measures the representativeness of the arithmetic mean. The bigger the VC is, the less representative the arithmetic mean is.

6. Pearson variation coefficient

Example: We studied the reaction time to two stimuli, A and B. The results were as follows:

	A	B
$\overline{X}$	50	600
S_X	5	6

1. Which stimulus does present more variation?
2. Which arithmetic mean is more representative?

6. Pearson variation coefficient

$$VC_A = \frac{S_A}{\overline{X}_A} 100 = \frac{5}{50} 100 = 10$$

$$VC_B = \frac{S_B}{\overline{X}_B} 100 = \frac{6}{600} 100 = 1$$

$VC_A = 10 > VC_B = 1$

1. A presents more variation.
2. The arithmetic mean of B is more representative.

Group and individual indexes

Group indexes	**Individual indexes**
↓	↓
Central tendency	Position: Deciles (D_i), Percentiles (P_i), Quartiles (Q_i)
Variability (Dispersion)	Raw scores (X_i)
Bias or Skewness (Asymmetry)	Differential scores (x_i)
Kurtosis	Standard scores (Z_i)

How are the data arranged with respect to the rest?

Are data piled at one side?

Bias, Skewness or Asymmetry indexes (A_s)

- Pearson asymmetry
- Fisher asymmetry
- Interquartile asymmetry

Bias, skewness or asymmetry indexes

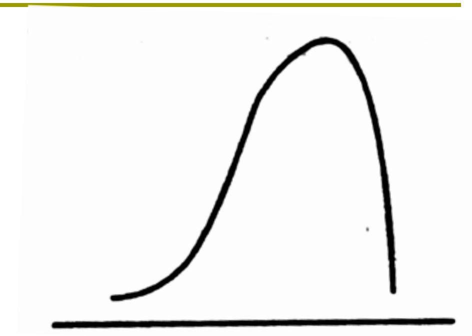
Introduction

A_s

$A_s < 0$

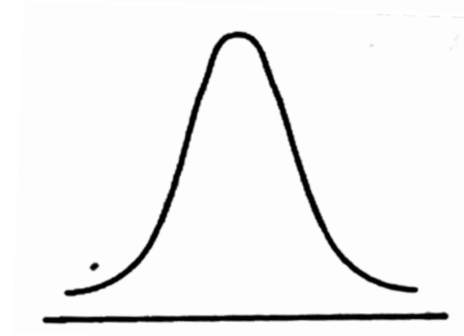
Asymmetric negative

$\overline{X} < Mdn < Mo$



$A_s = 0$ Symmetric

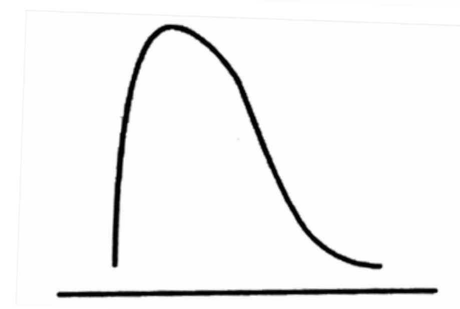
$\overline{X} = Mdn = Mo$



$A_s > 0$

Asymmetric positive

$\overline{X} > Mdn > Mo$



Introduction

INDEXES (A_s)

1. PEARSON ASYMMETRY
2. FISHER ASYMMETRY
3. INTERQUARTILE ASYMMETRY

1. Pearson asymmetry

$$A_s = \frac{\overline{X} - Mo}{S_X}$$

- It only can be calculated when the distribution is unimodal.

1. Pearson asymmetry

With the data below:

Intervals	f_i
1-3	2
4-6	7
7-9	13
10-12	18
13-15	10

1. Calculate Pearson asymmetry index.
2. Is the distribution asymmetric negative, symmetric or asymmetric positive?

1. Pearson asymmetry

Intervals	X_i	f_i	f_iX_i	X^2	f_iX^2
1-3	2	2	4	4	8
4-6	5	7	35	25	175
7-9	8	13	104	64	832
10-12	11	18	198	121	2178
13-15	14	10	140	196	1960
Σ		**50**	**481**		**5153**

$$A_s = \frac{\overline{X} - Mo}{S_X} = \frac{9.62 - 11}{3.24} = \frac{-1.38}{3.24} = -0.43$$

1. Pearson asymmetry

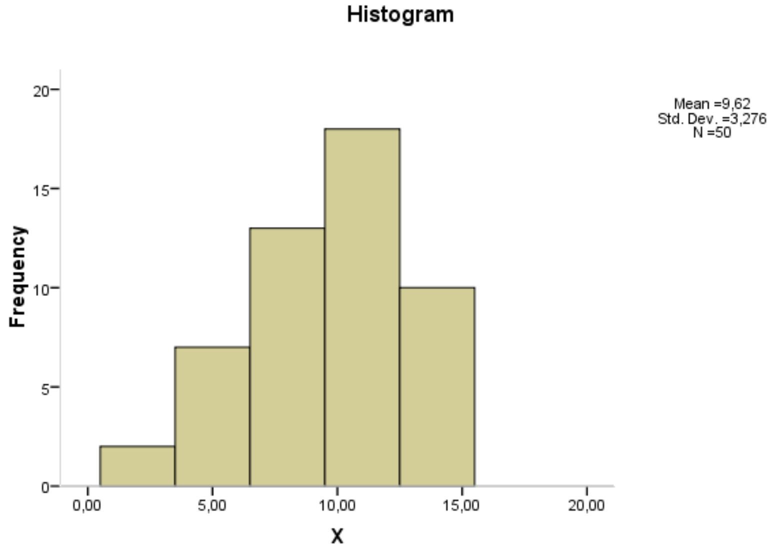
$$\overline{X} = \frac{\sum f_i X_i}{n} = \frac{481}{50} = 9.62$$

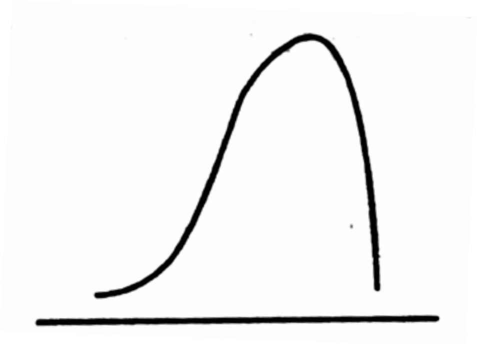
$$S^2 = \frac{\sum f_i X_i^2}{n} - \overline{X}^2 = \frac{5153}{50} - 9.62^2 = 103.06 - 92.54 = 10.52$$

$$S = \sqrt{S^2} = \sqrt{10.52} = 3.24$$

1. Pearson asymmetry

$$A_s = -0.43 \rightarrow A_s < 0 \rightarrow \text{ Asymmetric negative}$$





2. Fisher asymmetry

- It is the best of the asymmetry indexes.

$$A_s = \frac{\sum f_i (X_i - \overline{X})^3 \Big/ n}{S_X^3}$$

2. Fisher asymmetry

Example: with the same example used before:

Intervals	f_i
1-3	2
4-6	7
7-9	13
10-12	18
13-15	10

Calculate Fisher asymmetry index.

2. Fisher asymmetry

Intervals	X_i	f_i	$X_i - \overline{X}$	$(X_i - \overline{X})^3$	$f_i(X_i - \overline{X})^3$
1-3	2	2	-7.62	-442.45	-884.9
4-6	5	7	-4.62	-98.61	-690.27
7-9	8	13	-1.62	-4.25	-55.25
10-12	11	18	1.38	2.63	47.34
13-15	14	10	4.38	84.03	840.3
Σ		**50**			**-742.78**

$$A_s = \frac{\sum f_i (X_i - \overline{X})^3 \Big/ n}{S_X^3} = \frac{-742.78/50}{3.24^3} = \frac{-14.86}{34.01} = -0.44$$

$$\overline{X} = 9.62$$

$$S = 3.24$$

3. Interquartile asymmetry

- Advantage: values between -1 and +1.

$$\boxed{A_s = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1}}$$

3. Interquartile asymmetry

Example: with the same example used before:

Intervals	f_i
1-3	2
4-6	7
7-9	13
10-12	18
13-15	10

Steps:

1. F_i
2. $i*n/k$
3. L_i: value - 0.5 x measurement unit
4. f_i
5. F_i
6. I = max – min (exact limits of the interval)
7. $$Q_i = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right)$$

Calculate the interquartile asymmetry index.

3. Interquartile asymmetry

$$A_s = \frac{(Q_3 - Q_2) - (Q_2 - Q_1)}{Q_3 - Q_1} = \frac{(12.08 - 10) - (10 - 7.31)}{(12.08 - 7.31)} = \frac{2.08 - 2.69}{4.77} = \frac{-0.61}{4.77} = -0.13$$

Intervals	f_i	F_i ¹
1-3	2	2
4-6	7	9
7-9	13	22
10-12	18	40
13-15	10	50

Q_3

1. F_i
2. $i*n/k = 3 \times 50/4 = 3 \times 12.5 = 37.5$
3. L_i: $10 - 0.5 \times 1 = 9.5$
4. $f_4 = 18$
5. $F_3 = 22$
6. $I = 12.5 - 9.5 = 3$
7.

$$Q_3 = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) = 9.5 + \frac{3}{18}(37.5 - 22) = 9.5 + 2.58 = 12.08$$

3. Interquartile asymmetry

Intervals	f_i	F_i [1]
1-3	2	2
4-6	7	9
7-9	13	22
10-12	18	40
13-15	10	50

Q_2

1. F_i
2. i*n/k = 2 x 50/4 = 2 x 12.5 = 25
3. L_i: 10 – 0.5 x 1 = 9.5
4. f_4 = 18
5. F_3 = 22
6. I = 12.5 – 9.5 = 3
7.

$$Q_2 = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) = 9.5 + \frac{3}{18}(25-22) = 9.5 + 0.5 = 10$$

3. Interquartile asymmetry

Intervals	f_i	F_i [1]
1-3	2	2
4-6	7	9
7-9	13	22
10-12	18	40
13-15	10	50

Q_1

1. F_i
2. $i*n/k = 1 \times 50/4 = 1 \times 12.5 = 12.5$
3. L_i: $7 - 0.5 \times 1 = 6.5$
4. $f_3 = 13$
5. $F_2 = 9$
6. $I = 9.5 - 6.5 = 3$
7.

$$Q_1 = L_i + \frac{I}{f_i}\left(\frac{i*n}{k} - F_i\right) = 6.5 + \frac{3}{13}(12.5 - 9) = 6.5 + 0.81 = 7.31$$

Group and individual indexes

Group indexes	Individual indexes
↓	↓

Group indexes

- **Central tendency**
- **Variability (Dispersion)**
- **Bias or Skewness (Asymmetry)**
- **Kurtosis**

Individual indexes

- **Position**
  - **Deciles (D_i)**
  - **Percentiles (P_i)**
  - **Quartiles (Q_i)**
- **Raw scores (X_i)**
- **Differential scores (x_i)**
- **Standard scores (Z_i)**

Which form is the distribution?

Is it flattened or sharp?

Kurtosis index

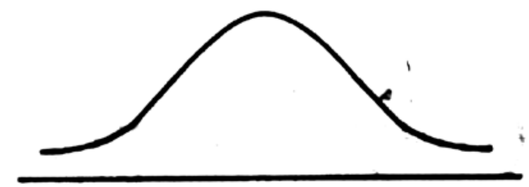
k_r

Kurtosis index

Introduction

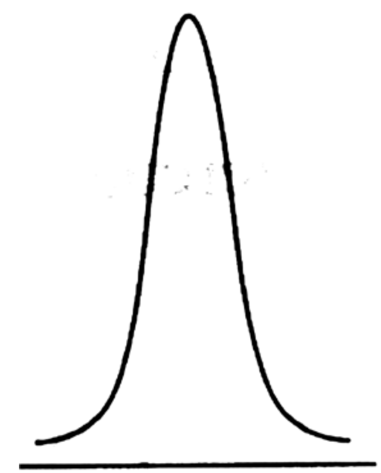
K_r

K_r < 0 Platykurtic



K_r = 0 Mesokurtic (normal distribution)

K_r > 0 Leptokurtic



Kurtosis index

$$\boxed{K_r = \frac{\sum f_i (X_i - \overline{X})^4 \Big/ n}{S_X^4} - 3}$$

Kurtosis index

Example: with the same example used before:

Intervals	f_i
1-3	2
4-6	7
7-9	13
10-12	18
13-15	10

1. Calculate the kurtosis index.
2. Is the distribution platykurtic, mesokurtic or Leptokurtic?

Kurtosis index

Intervals	X_i	f_i	$X_i - \overline{X}$	$(X_i - \overline{X})^4$	$f_i(X_i - \overline{X})^4$
1-3	2	2	-7.62	3371.47	6742.94
4-6	5	7	-4.62	455.58	3189.06
7-9	8	13	-1.62	6.89	89.57
10-12	11	18	1.38	3.63	65.34
13-15	14	10	4.38	368.04	3680.4
Σ		**50**			**13767.31**

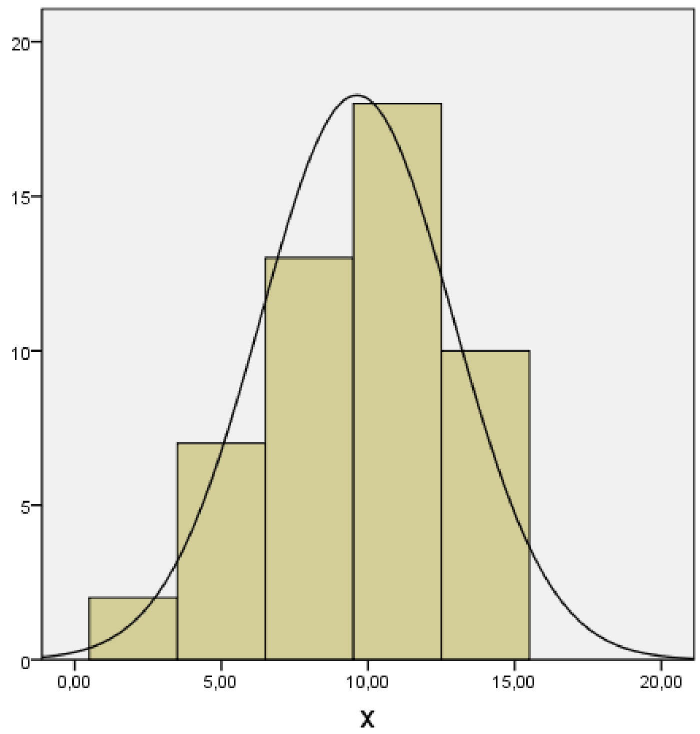
$$K_r = \frac{\sum f_i (X_i - \overline{X})^4 / n}{S_X^4} - 3 = \frac{13767.31/50}{3.24^4} - 3 = \frac{275.35}{110.2} - 3 = 2.5 - 3 = -0.5$$

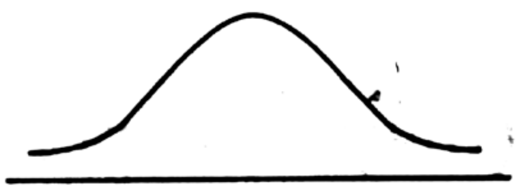
$$\overline{X} = 9.62$$

$$S = 3.24$$

Kurtosis index

$$K_r = -0.5 \rightarrow K_r < 0 \rightarrow \quad \text{Platykurtic}$$





Group and individual indexes

Group indexes	**Individual indexes**
↓	↓

- Central tendency
- Variability (Dispersion)
- Bias or Skewness (Asymmetry)
- Kurtosis
	- Position
	  - Deciles (D_i)
	  - Percentiles (P_i)
	  - Quartiles (Q_i)
	- Raw scores (X_i)
	- Differential scores (x_i)
	- Standard scores (Z_i)

1. Introduction

- Raw scores (X_i): they are the scores given directly (example: punctuation in a test).

$$\boxed{X_i}$$

- Differential scores (x_i):

$$\boxed{x_i = X_i - \overline{X}}$$

- Standard scores (Z_i):

$$\boxed{Z_i = \frac{X_i - \overline{X}}{S}}$$

1. Introduction

- The comparison using raw scores (X_i) can lead us to misleading conclusions.

- The solution based on the distances to the mean (x_i) is not an entirely satisfactory solution.

- The solution is to use standard scores.

2. Raw scores

CASE 1: a participant, a variable.

A participant obtains a score *on an intelligence questionnaire:*

$X=25$

- Interpretation:
  - A raw score (example: 25) does not give us more than a number. Is it too much? Is it enough?
  - It depends on two factors:
    - Mean
    - Variability
- Conclusion: raw scores are not enough to compare.

2. Raw scores

CASE 2: a participant, two variables.

John weighs 75 kg. and he is 1.80 m. tall.

- His weight, is it more or less than his height?
- They are not directly comparable.

2. Raw scores

<u>CASE 3</u>: a participant, two supposedly comparable variables.

A student X has recently been examined in two different subjects: Motivation and Methods.

- If the scores obtained were respectively 30 and 15, can we say that the student has done better in Motivation than in Methods?
- It depends on his group of reference.

3. Differential scores

Scores of

- Deviation
- Dispersion
- Errors or bias

$$x_i = X_i - \bar{X}$$

3. Differential scores

- They tell us if a raw score is higher, smaller or equal to the mean.

- This information is inferred from the sign of the differential score (positive, negative or zero respectively).

3. Differential scores

X_i
7
9
10
11
13

1. Calculate the mean, the variance and the standard deviation.
2. Calculate the differential scores.
3. Calculate the mean, the variance and the standard deviation of these differential scores.

3. Differential scores

X_i	X_i^2
7	49
9	81
10	100
11	121
13	169
50	520

$$\overline{X} = \frac{\sum X_i}{n} = \frac{50}{5} = 10$$

$$S^2 = \frac{\sum X_i^2}{n} - \overline{X}^2 = \frac{520}{5} - 10^2 = 104 - 100 = 4$$

$$S = \sqrt{S^2} = \sqrt{4} = 2$$

3. Differential scores

X_i
7
9
10
11
13
50

↔

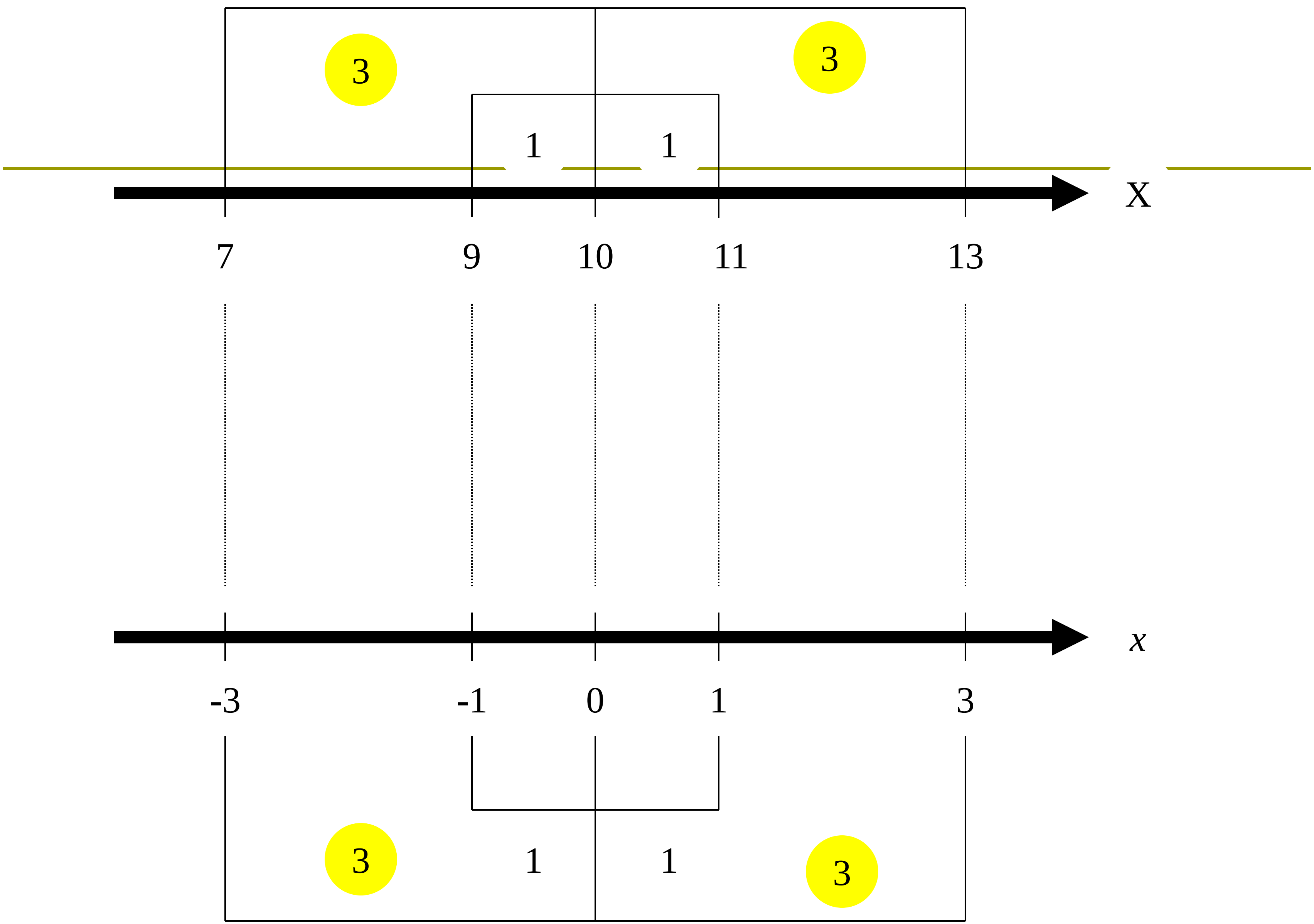
$x_i=(X_i - \overline{X})$	x^2
7-10 = -3	9
9-10 = -1	1
10-10 = 0	0
11-10 = 1	1
13-10 = 3	9
0	20

$$\overline{X} = \frac{\sum X_i}{n} = \frac{0}{5} = 0$$

$$S^2 = \frac{\sum X_i^2}{n} - \overline{X}^2 =$$

$$\frac{20}{5} - 0^2 = 4 - 0 = 4$$

$$S = \sqrt{S^2} = \sqrt{4} = 2$$


3
3
1
1
X
7
9
10
11
13
x
-3
-1
0
1
3
3
1
1
3

3. Differential scores: conclusions

RAW SCORES		DIFFERENTIAL SCORES
$\overline{X} = 10$	$\neq$	$\overline{x} = 0$
$S^2_X = 4$	$=$	$S^2_x = 4$
$S_X = 2$	$=$	$S_x = 2$

3. Differential scores: example

The intellectual level of two groups (A and B) was measured:

- A: 97 102 107 112 117
- B: 92 97 102 107 112 117 122

Supposing that 2 students, one belonging to group A and another to group B, obtained the same intellectual level of 117 points,

1. Which were the differential scores for both students?
2. Do both scores imply the same intellectual level?

3. Differential scores: example

$$\overline{X}_A = \frac{\sum X_i}{n} = \frac{97+102+107+112+117}{5} = \frac{535}{5} = 107$$

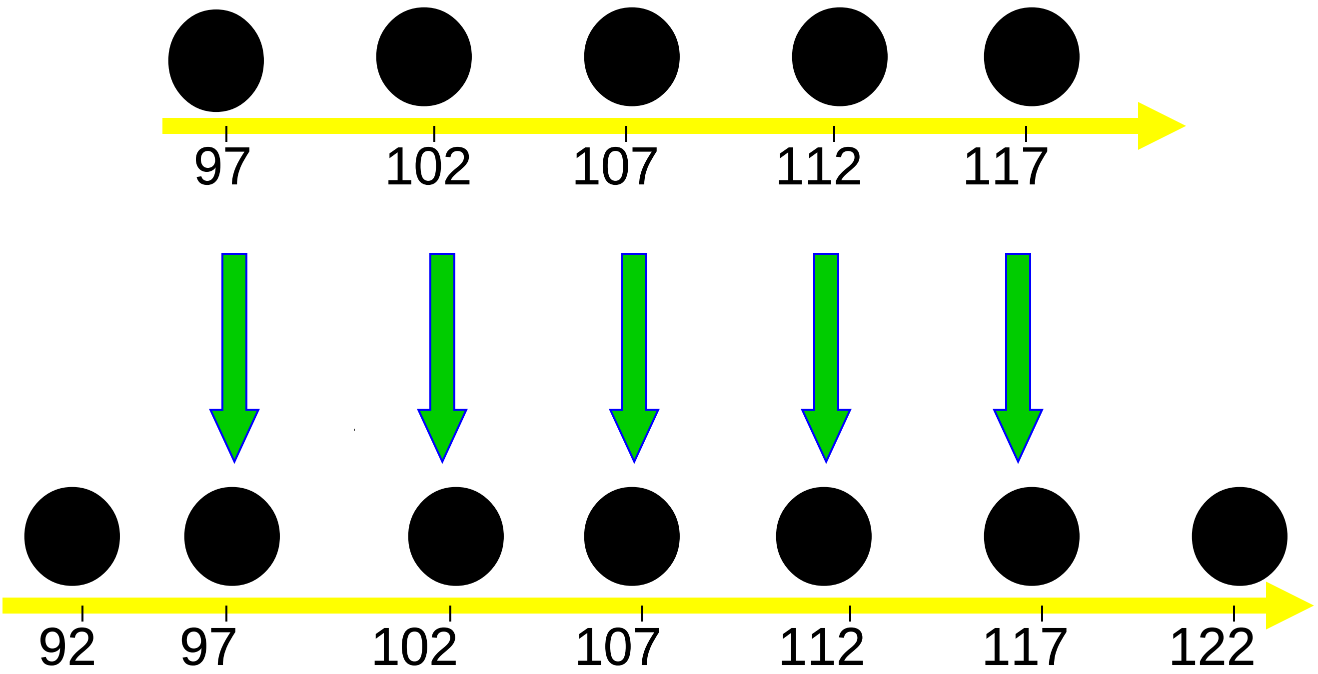
$$\overline{X}_B = \frac{\sum X_i}{n} = \frac{92+97+102+107+112+117+122}{7} = \frac{749}{7} = 107$$

$$\overline{X}_A = \overline{X}_B = 107$$

$$x = X - \overline{X} = 117 - 107 = +10$$

3. Differential scores: example

- **Group A**
- **Group B**



3. Differential scores: example

- Conclusions:
  - Equality in differential scores may be masking different situations.
  - Group A was more homogeneous than group B.
  - While the score 117 in group A represented an extreme value (because it was the maximum score), the same score in group B was not so extreme (there was a higher score: $X_i=122$).
  - As a solution to solve this problem, standard scores can be used.

4. Standard scores (Z_i)

- They are also called typified or standardized scores.
- Definition: The standard, typified or standardized score indicates the number of standard deviations that a particular raw score is separated from its mean.

$$\boxed{Z_i = \frac{X_i - \overline{X}}{S}}$$

4. Standard scores (Z_i). Example

X_i
7
9
10
11
13

Calculate the differential and the standard scores.

4. Standard scores (Z_i). Example

X_i	x	Z_i	X^2
7	-3	-1.5	49
9	-1	-0.5	81
10	0	0	100
11	1	0.5	121
13	3	1.5	169
Σ=50	0		520

$$\overline{X} = \frac{\sum X}{n} = \frac{50}{5} = 10$$

$$S = \sqrt{\frac{\sum X^2}{n} - \overline{X}^2}$$

$$= \sqrt{\frac{520}{5} - 10^2} =$$

$$\sqrt{104 - 100} = \sqrt{4} = 2$$

4. Standard scores (Zi)

- There is a shift to the left side of the x-axis when we convert raw scores into differential:

x_i X_i

[-3 -1 0 1 3] **[7 9 10 11 13]**

- And there is a concentration of scores when they become standard:

Z_i $\left\{ \textbf{-1.5 } \; \textbf{–0.5} \; \textbf{0} \; \textbf{0.5} \; \textbf{1.5} \right\}$

4. Standard scores (Z_i). Properties

1 The sum of standard scores is 0:

$$\sum Z_i = 0$$

X_i	Z_i
7	-1.5
9	-0.5
10	0
11	0.5
13	1.5
Σ=50	

4. Standard scores (Z_i). Properties

2 The mean of standard scores is 0:

$$\bar{Z} = 0$$

X_i	Z_i
7	-1.5
9	-0.5
10	0
11	0.5
13	1.5
Σ=50	

4. Standard scores (Z_i). Properties

2 The mean of standard scores is 0:

$$\overline{Z} = 0$$

X_i	Z_i
7	-1.5
9	-0.5
10	0
11	0.5
13	1.5
Σ=50	Σ=0

$$\overline{Z} = \frac{\sum Z_i}{n} = \frac{0}{5} = 0$$

4. Standard scores (Z_i). Properties

3 The sum of the standard scores squared is n:

$$\Sigma Z_i^{\,2} = n$$

X_i	Z_i
7	-1.5
9	-0.5
10	0
11	0.5
13	1.5
Σ=50	Σ=0

4. Standard scores (Z_i). Properties

3 The sum of the standard scores squared is n:

$$\Sigma Z_i^{\,2} = n$$

X_i	Z_i	Z^2
7	-1.5	2.25
9	-0.5	0.25
10	0	0
11	0.5	0.25
13	1.5	2.25
Σ=50	Σ=0	5

$n = 5$

4. Standard scores (Z_i). Properties

4 The standard deviation and variance of standard scores is equal to 1：

$$S_Z^2 = 1$$

$$S_Z = 1$$

X_i	Z_i	Z^2
7	-1.5	2.25
9	-0.5	0.25
10	0	0
11	0.5	0.25
13	1.5	2.25
Σ=50	Σ=0	5

4. Standard scores (Z_i). Properties

4 The standard deviation and variance of standard scores is equal to 1 :

$$S_Z^2 = 1$$
$$S_Z = 1$$

X_i	Z_i	Z^2
7	-1.5	2.25
9	-0.5	0.25
10	0	0
11	0.5	0.25
13	1.5	2.25
Σ=50	Σ=0	5

$$S_Z^2 = \frac{\sum Z_i^2}{n} - \overline{Z}^2 = \frac{5}{5} - 0^2 = 1$$

$$S_Z = \sqrt{S_Z^2} = \sqrt{1} = 1$$