FOUNDATION YEAR 2025-2026

COURSE - DATA ANALYSIS

Lesson 2

Niccolò Maffezzoli, Ca' Foscari University of Venice

January 15, 2026

Course Materials

https://drive.google.com/drive/folders/1YjLzYJ_
MtDuBa8f8ppk7Yc9qTl1fLk_B?usp=drive_link

Exercise from last time

Given the measurements of the period of a pendulum $T = \{2.8, 3.0, 3.1, 3.1\}$ seconds, calculate:

1. the mean μ
2. the variance σ^2
3. the st. deviation σ
4. the weighted mean, given the uncertainties $\sigma_i \in \{0.3,\ 0.2,\ 0.5,\ 0.1\}$ seconds.

1. Mean

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{x_1 + x_2 + x_3 + x_4}{n} = \frac{2.8 + 3.0 + 3.1 + 3.1}{4} = 3.0s$$

2. Variance

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 = \frac{1}{n}[(x_1 - \mu)^2 + (x_2 - \mu)^2 + (x_3 - \mu)^2 + (x_4 - \mu)^2]$$
$$= \frac{1}{4}[(2.8 - 3.0)^2 + (3.0 - 3.0)^2 + (3.1 - 3.0)^2 + (3.1 - 3.0)^2] \sim 0.015\ s^2$$

3. Standard deviation (or RMS)

$$\sigma = \sqrt{\sigma^2} = \sqrt{0.015} \sim 0.12\ s$$

4. Weighted mean

$$\begin{aligned}
\mu_w = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} &= \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2 + x_3/\sigma_3^2 + x_4/\sigma_4^2}{1/\sigma_1^2 + 1/\sigma_2^2 + 1/\sigma_3^2 + 1/\sigma_4^2} \\
&= \frac{2.8/(0.3^2) + 3.0/(0.2^2) + 3.1/(0.5^2) + 3.1/(0.1^2)}{1/(0.3^2) + 1/(0.2^2) + 1/(0.5^2) + 1/(0.1^2)} \\
&= 3.0\ s
\end{aligned}$$

In this particular case, the mean and the weighted mean coincide. That's not the case in general.

Summary

Given a set $\{x_1, x_2, x_3, ..., x_n\}$, the standard deviation σ expresses the dispersion/spread of the data around the mean.

$$\mu = \frac{1}{n}\sum_{i=1}^{n} x_i \textbf{ MEAN} \tag{1}$$

$$\sigma^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2 \textbf{ VARIANCE} \tag{2}$$

$$\sigma = \sqrt{\sigma^2} = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i - \mu)^2} \textbf{ STDEV (or RMS)} \tag{3}$$

If x_i are affected by different uncertainties σ_i we can calculate the weighted mean:

$$\mu_w = \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} \textbf{ WEIGHTED MEAN} \tag{4}$$

The error on the mean

Given a set $\{x_1, x_2, x_3, ..., x_n\}$, we can calculate the mean μ. How about its uncertainty σ_μ ?

$$\sigma_\mu = \frac{\sigma}{\sqrt{N}} \textbf{ UNCERTAINTY ON THE MEAN} \tag{5}$$

You first calculate σ of the set and then divide by the square root of the number of measurement N. This is a very important formula.
A more general case is the one in which each measurement x_i has its own uncertainty σ_i. The uncertainty on the weighted mean becomes:

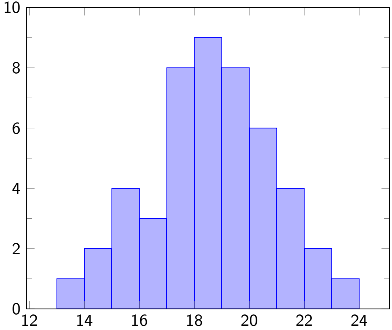
$$\sigma_{\mu,w} = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}} \tag{6}$$

In case of varying uncertainty, the σ that describes the entire population is not particularly meaningful.

How can we show our data in a compact way ? Histograms

In a movie theatre the age of 49 people watching at a movie are the following:

Age (x_i)	frequency (f_{x_i})
13	1
14	2
15	4
16	3
17	8
18	9
19	8
20	6
21	4
22	2
23	1
24	1



Exercise

Calculate mean, σ

Graphical interpretation

You should be able to indicate roughly the values of mean, σ just by eye.

Exercise

The temperature in o F on 20 days during the month of June was as follows:
70° F, 76° F, 76° F, 74° F, 70° F, 70° F, 72° F, 74° F, 78° F, 80° F, 74° F, 74° F, 78° F, 76° F, 78° F, 76° F, 74° F, 78° F, 80° F, 76° F

Calculate mean, median, mode, standard deviation and error on the mean of the temperatures for the month of June? Plot the distribution.