FOUNDATION YEAR 2025-2026
COURSE - DATA ANALYSIS

Lesson 3

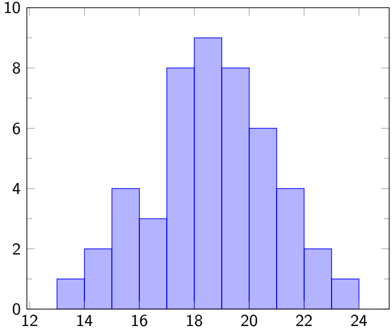
Niccolò Maffezzoli, Ca' Foscari University of Venice

January 19, 2026

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



Graphical interpretation

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

Calculate mean, σ
$\mu = 18.24489$ $\sigma = 2.146849773993232$
Does it make sense to use all these digits ? How many do we need ?

Significant digits 1/2

Rules

1. **All nonzero digits in a measurement are significant.** E.g. 237 has three significant digits. 1.897 has four significant digits.
2. **Zeros between two non-zero digits ARE significant.** E.g 2051 has four significant digits. 5.02 has three significant digits
3. **Zeros that appear in front of all of the nonzero digits are called leading zeros. Leading zeros are never significant.** E.g. 0.54 has only two significant digits. 0.0032 also has two significant digits.
4. **Zeros that appear after all nonzero digits are called trailing zeros**. **Trailing zeros in a number with a decimal point are significant.** This is true whether the zeros occur before or after the decimal point. E.g. 620.0 has four sd; 19.000 has five sd.
5. **A number with trailing zeros that lacks a decimal point may or may not be significant.** Use scientific notation to indicate the appropriate number of significant digits. E.g. 1400 is ambiguous. $1.4 \cdot 10^3$ has two sd; $1.40 \cdot 10^3$ has three sd; $1.400 \cdot 10^3$ has four sd.

Significant digits 2/2

Example

Give the number of significant digits in each. Identify the rule for each:
a) 5.87
b) 0.031
c) 52.90
d) 0.2001
e) 500
f) 0.606000

Significant digits 2/2

Example

Give the number of significant digits in each. Identify the rule for each:
a) 5.87
b) 0.031
c) 52.90
d) 0.2001
e) 500
f) 0.606000

Solution

a) Answer: 3, rule 1
b) Answer: 2, rule 3
c) Answer: 4, rule 4
d) Answer: 4, rules 2,3
e) Answer: 1,2 or 3. Ambiguous, use scientific notation, e.g. $5.0 \cdot 10^2$ with 2 sd.
f) Answer: 6, rules 2,3,4

Rounding and scientific notation

Rounding

If the digit is five or more, round up. Example: 14.5, 14.6, 14.7, 14.8, 14.9 are rounded to 15. **If the digit is less than 5, you round down.** Example: 14.0, 14.1, 14.2, 14.3, 14.4 are rounded to 14.

Scientific notation

When reporting measurements, the scientific notation is:

$$x = (0.241 \pm 0.005)m$$

Typically, when giving results only quote the error with one significant digit ! If the example above, the error on the measurement is 5 mm. If, in another experiment, the error $=$ 10mm, I should use one digit (and round x):

$$x = (0.24 \pm 0.01)m$$

Exercise

Write the mean and its error of the movie theatre exercise (slide 2) using the correct number of digits and the scientific notation.

Rounding and scientific notation

Rounding

If the digit is five or more, round up. Example: 14.5, 14.6, 14.7, 14.8, 14.9 are rounded to 15. **If the digit is less than 5, you round down.** Example: 14.0, 14.1, 14.2, 14.3, 14.4 are rounded to 14.

Scientific notation

When reporting measurements, the scientific notation is:

$$x = (0.241 \pm 0.005)m$$

Typically, when giving results only quote the error with one significant digit ! If the example above, the error on the measurement is 5 mm. If, in another experiment, the error = 10mm, I should use one digit (and round x):

$$x = (0.24 \pm 0.01)m$$

Exercise

Write the mean and its error of the movie theatre exercise (slide 2) using the correct number of digits and the scientific notation. $\mu = (18.2 \pm 0.3) = (1.82 \pm 0.03) \cdot 10^1$

Covariance

Suppose your data sample consists of (x, y) pairs:
$\{(x_1, y_1), (x_2, y_2), (x_3, y_3), ..., (x_n, y_n)\}$. You can calculate μ_x, μ_y, σ_x, σ_y.
However, there is more information - are x and y independent or do they depend on one another ?

Covariance

$$\text{cov}(x, y) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y) \tag{1}$$

- A positive covariance means that 'big' x corresponds to 'big' y, 'small' x corresponds to 'small' y.
- A negative covariance means that 'big' x corresponds to 'small' y, 'small' x corresponds to 'big' y.
- The covariance is zero if there is no obvious relationship between x and y.
- The covariance has units: $[x] \cdot [y]$

Example

There is probably positive covariance between height and weight of adults.

Pearson's correlation coefficient

Correlation

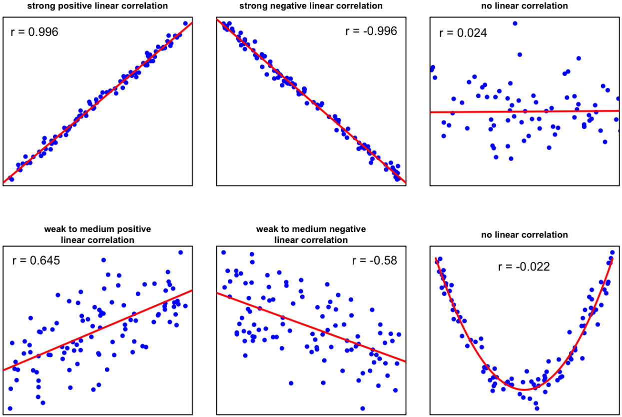
$$\rho_{xy} = \frac{\text{cov}(x,y)}{\sigma_X \sigma_Y} = \frac{\sum_{i=1}^{n}(x_i - \mu_x)(y_i - \mu_y)}{\sqrt{\sum_{i=1}^{n}(x_i - \mu_x)^2}\sqrt{\sum_{i=1}^{n}(y_i - \mu_y)^2}} \tag{2}$$

- ρ_{xy} dimensionless
- $-1 < \rho_{xy} < 1$
- $\rho_{xy} \simeq 1$ strong positive correlation
- $\rho_{xy} \simeq -1$ strong negative correlation
- $\rho_{xy} \simeq 0$ no (linear) correlation

What is the meaning ?

This formula measures the strength and direction of the linear relationship between two variables, and it takes on values between -1 and 1, with values closer to -1 or 1 indicating a stronger linear relationship, and values closer to 0 indicating a weaker linear relationship.

Pearson's correlation coefficient



Exercise - covariance and correlation (ex 2.4 pag 18 Barlow)

The marks of 12 students in classical mechanics and quantum mechanics are as follows:

Classical	22	48	76	10	22	4	68	44	10	76	14	56
Quantum	63	39	61	30	51	44	74	78	55	58	41	69

Calculate the two means and standard deviations, the covariance and the correlation.

Exercise - covariance and correlation (ex 2.4 pag 18 Barlow)

The marks of 12 students in classical mechanics and quantum mechanics are as follows:

Classical	22	48	76	10	22	4	68	44	10	76	14	56
Quantum	63	39	61	30	51	44	74	78	55	58	41	69

Calculate the two means and standard deviations, the covariance and the correlation.

SOLUTION: The averages are 37.5 and 55.25 with standard deviation 25.9 and 14.2. The covariance is 207.5 and the correlation is 0.57. TODO at home: plot the data on a scatter plot.

Exercise - histogram (ex. 2.6, pag 19 Barlow)

Here are 80 numbers - histogram them using a suitable bin size.

90	90	79	84	78	91	88	90	85	80
88	75	73	79	78	79	67	83	68	60
73	79	69	74	76	68	72	72	75	60
61	66	66	54	71	67	75	49	51	57
62	64	68	58	56	79	63	68	64	51
58	53	65	57	59	65	48	54	55	40
49	42	36	46	40	37	53	48	44	43
35	39	30	41	41	22	28	36	39	51

Also: calculate the mean, σ, σ_μ and report them using the scientific notation.