FOUNDATION YEAR 2025-2026

COURSE - DATA ANALYSIS

Lesson 10

Niccolò Maffezzoli, Ca' Foscari University of Venice

February 24, 2026

Exercise 1 - Data and descriptive statistics

Consider the following kilometers swam by a professional swimmer over 10 days:
18, 15, 10, 13, 17, 22, 25, 20, 19, 10

a) Calculate the mean and median.
b) Calculate the variance and standard deviation.
c) Calculate the uncertainty on the mean.

Exercise 1 - Data and descriptive statistics

Consider the following kilometers swam by a professional swimmer over 10 days:
18, 15, 10, 13, 17, 22, 25, 20, 19, 10

a) Calculate the mean and median.

b) Calculate the variance and standard deviation.

c) Calculate the uncertainty on the mean.

a) $\mu = \frac{18+15+10+13+17+22+25+20+19+10}{10} = 16.9km$. The ordered data is 10, 10, 13, 15, 17, 18, 19, 20, 22, 25. The median is the mean of the 5th and 6th, i.e. $median = (17+18)/2 = 17.5km$.
b) $\sigma^2 = \frac{1}{N}\sum_{i=1}^{10}(x_i - \mu)^2 = \frac{1}{10}[(18-16.9)^2 + (15-16.9)^2 + \ldots + (10-16.9)^2] \approx 22.1km^2$. $\sigma = \sqrt{\sigma^2} \approx 4.7km$.
c) $\sigma_\mu = \frac{\sigma}{\sqrt{N}} = \frac{4.7}{\sqrt{10}} \approx 1.49km$

Exercise 2 - Weighted mean

A Formula1 pilot has to weight more than 70kg. Three consecutive measurements report $m_1 = 70.2 \pm 0.5$; $m_2 = 71.6 \pm 0.1$; $m_3 = 69.1 \pm 0.9$. Calculate the weighted mean and its uncertainty.

Exercise 2 - Weighted mean

A Formula1 pilot has to weight more than 70kg. Three consecutive measurements report $m_1 = 70.2 \pm 0.5$; $m_2 = 71.6 \pm 0.1$; $m_3 = 69.1 \pm 0.9$. Calculate the weighted mean and its uncertainty.

$$\begin{aligned}\mu_w &= \frac{\sum_{i=1}^{n} x_i/\sigma_i^2}{\sum_{i=1}^{n} 1/\sigma_i^2} = \frac{x_1/\sigma_1^2 + x_2/\sigma_2^2 + x_3/\sigma_3^2}{1/\sigma_1^2 + 1/\sigma_2^2 + 1/\sigma_3^2} \\ &= \frac{70.2/(0.5^2) + 71.6/(0.1^2) + 69.1/(0.9^2)}{1/(0.5^2) + 1/(0.1^2) + 1/(0.9^2)} \\ &\approx 71.517\ kg\end{aligned}$$

$$\sigma_{\mu,w} = \sqrt{\frac{1}{\sum_{i=1}^{n} 1/\sigma_i^2}} = \sqrt{\frac{1}{1/\sigma_1^2 + 1/\sigma_2^2 + 1/\sigma_3^2}} = \sqrt{\frac{1}{1/(0.5^2) + 1/(0.1^2) + 1/(0.9^2)}} \approx 0.09748$$

$$\rightarrow m \approx (71.5 \pm 0.1) kg$$

Exercise 3 - Significant digits

Report the following results correctly, by stating the uncertainty to only 1 significant digit.

a) $23.999 \pm 0.51 \rightarrow$

b) $0.0011 \pm 0.0039 \rightarrow$

c) $3099 \pm 0.55 \rightarrow$

d) $1099 \pm 15 \rightarrow$

e) $0.00101 \pm 0.00059 \rightarrow$

f) $7.7777 \pm 0.0999 \rightarrow$

Exercise 3 - Significant digits

Report the following results correctly, by stating the uncertainty to only 1 significant digit.

a) $23.999 \pm 0.51 \rightarrow$ 24.0 ± 0.5

b) $0.0011 \pm 0.0039 \rightarrow$ 0.001 ± 0.004

c) $3099 \pm 0.55 \rightarrow$ 3099.0 ± 0.6

d) $1099 \pm 15 \rightarrow$ $(11.0 \pm 0.2) \cdot 10^2$

e) $0.00101 \pm 0.00059 \rightarrow$ 0.0010 ± 0.0006

f) $7.7777 \pm 0.0999 \rightarrow$ 7.8 ± 0.1

Exercise 4 - Covariance and Correlation

Consider the following data on the average cost of fruit and number of thunderstorms over 4 years. Calculate the correlation coefficient between the 2 variables. What can you say about them ?

Cost of fruit ($)	1.1	2.7	0.9	5
Number of spring thunderstorms	10	15	7	30

Exercise 4 - Covariance and Correlation

Consider the following data on the average cost of fruit and number of thunderstorms over 4 years. Calculate the correlation coefficient between the 2 variables. What can you say about them ?

Cost of fruit ($)	1.1	2.7	0.9	5
Number of spring thunderstorms	10	15	7	30

x = cost of fruit; y = No. of thunderstorms. $\mu_x = 2.43$; $\mu_y = 15.5$; $\sigma_x = 1.64$; $\sigma_y = 8.85$.

$Cov(x, y) = \frac{1}{N}\sum_{i=1}^{4}(x_i - \mu_x)(y_i - \mu_y) =$

$\frac{1}{4}[(1.1 - 2.43)(10 - 15.5) + (2.7 - 2.43)(15 - 15.5) + (0.9 - 2.43)(7 - 15.5) + (5 - 2.43)(30 - 15.5)] = 14.3625$

$\rho = \frac{cov(x,y)}{\sigma_x \sigma_y} = \frac{14.3625}{1.64 \cdot 8.85} = 0.9899$.

The two variables are very highly positively correlated, because $\rho \approx 1$.

Exercise 5 - Poisson distribution

Occasionally, a machine producing steel tools needs to be reset. The number of resets in a month and is modelled by a Poisson distribution. The mean number of resets needed per month has been found to be 6. Find the probability that:

a) 7 resets per month are needed.

b) Fewer than 3 resets per month are needed.

c) More than 4 resets per month are needed.

Exercise 5 - Poisson distribution

Occasionally, a machine producing steel tools needs to be reset. The number of resets in a month and is modelled by a Poisson distribution. The mean number of resets needed per month has been found to be 6. Find the probability that:

a) 7 resets per month are needed.

b) Fewer than 3 resets per month are needed.

c) More than 4 resets per month are needed.

a) $P(k=7, \lambda=6) = \frac{\lambda^k \cdot e^{-\lambda}}{k!} = \frac{6^7 \cdot e^{-6}}{7!} = 0.1376.$

b) $P(0)+P(1)+P(2) = \frac{6^0 \cdot e^{-6}}{0!} + \frac{6^1 \cdot e^{-6}}{1!} + \frac{6^2 \cdot e^{-6}}{2!} = 0.0025 + 0.0149 + 0.0446 = 0.062$

c)

$P(5)+P(6)+P(7)+P(\dots) = 1-[P(0)+P(1)+P(2)+P(3)+P(4)] = 1-[0.062+\frac{6^3 \cdot e^{-6}}{3!}+\frac{6^4 \cdot e^{-6}}{4!}] = 1-[0.062+0.0892+0.1338] = 0.715.$

Exercise 6 - Binomial distribution

About 8% of males are colorblind. A researcher needs three colorblind men for an experiment and begins checking potential subjects. What is the probability that she finds three or more colorblind men in the first nine she examines?

Exercise 6 - Binomial distribution

About 8% of males are colorblind. A researcher needs three colorblind men for an experiment and begins checking potential subjects. What is the probability that she finds three or more colorblind men in the first nine she examines?

$P(k \geq 3, p = 0.08, n = 9) = 1 - (P(0) + P(1) + P(2))$

$P(0) = p^k(1-p)^{n-k}\frac{n!}{k!(n-k)!} = 0.08^0(1-0.08)^{9-0}\frac{9!}{0!9!} = 0.4721$

$P(1) = p^k(1-p)^{n-k}\frac{n!}{k!(n-k)!} = 0.08^1(1-0.08)^{9-1}\frac{9!}{1!8!} = 0.3695$

$P(2) = p^k(1-p)^{n-k}\frac{n!}{k!(n-k)!} = 0.08^2(1-0.08)^{9-2}\frac{9!}{2!7!} = 0.1285$

$\rightarrow P(k \geq 3, p = 0.08, n = 9) = 1 - (P(0) + P(1) + P(2)) = 1 - 0.9701 = 0.0299$

Exercise 7 - Gaussian distribution

The average salary for first-year teachers is \$27989. Assume the distribution is Gaussian with standard deviation \$3250.

a) What is the probability that a randomly selected first-year teacher makes between \$20,000 and \$30000 each year?

b) What is the probability that a randomly selected first-year teacher has a salary less than \$20000?

Exercise 7 - Gaussian distribution

The average salary for first-year teachers is \$27989. Assume the distribution is Gaussian with standard deviation \$3250.

a) What is the probability that a randomly selected first-year teacher makes between \$20,000 and \$30000 each year?

b) What is the probability that a randomly selected first-year teacher has a salary less than \$20000?

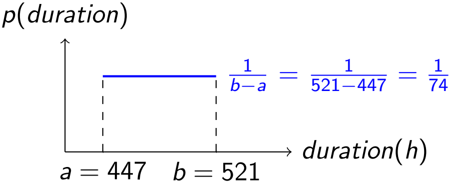
a) $z_1 = \frac{20000-27989}{3250} = -2.46$; $z_2 = \frac{30000-27989}{3250} = 0.62$;
$P(20000 < x < 30000) = P(z_1 < z < z_2) = P(z_1 < z < 0) + P(0 < z < z_2) = 0.9861/2 + 0.4647/2 = 0.7254$
b) $z_1 = \frac{20000-27989}{3250} = -2.46$. $P(x < 20000) = P(z < z_1) = \frac{1-0.9861}{2} = 0.00695$

Exercise 8 - Uniform distribution

The total duration of baseball games in the major league in the 2011 season is uniformly distributed between 447 hours and 521 hours inclusive. Draw the distribution, calculate the mean and standard deviation, calculate the probability that a random game is more than 500 hours.

Exercise 8 - Uniform distribution

The total duration of baseball games in the major league in the 2011 season is uniformly distributed between 447 hours and 521 hours inclusive. Draw the distribution, calculate the mean and standard deviation, calculate the probability that a random game is more than 500 hours.



- $\mu = \frac{a+b}{2} = \frac{447+521}{2} = 484\text{h}$. $\sigma = \frac{b-a}{\sqrt{12}} = \frac{521-447}{\sqrt{12}} = 21.4\text{h}$.
- $p(x > 500) = (521 - 500) \cdot 1/74 \approx 0.28 = 28\%$

Exercise 9 - Error propagation

A scale has an uncertainty of 2 kg. If I measure a mass $m1 = 56kg$ and a mass of $m2 = 140kg$, what is the total mass and its uncertainty ?

Exercise 9 - Error propagation

A scale has an uncertainty of 2 kg. If I measure a mass $m1 = 56kg$ and a mass of $m2 = 140kg$, what is the total mass and its uncertainty ?

$$m = 196 \pm \sqrt{\sigma_{m1}^2 + \sigma_{m2}^2} = 196 \pm \sqrt{8} \approx 196 \pm 2.82 = (196 \pm 3)kg$$

Exercise 10 - Bayes Theorem

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it does not rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

Exercise 10 - Bayes Theorem

Marie is getting married tomorrow, at an outdoor ceremony in the desert. In recent years, it has rained only 5 days each year. Unfortunately, the weatherman has predicted rain for tomorrow. When it actually rains, the weatherman correctly forecasts rain 90% of the time. When it does not rain, he incorrectly forecasts rain 10% of the time. What is the probability that it will rain on the day of Marie's wedding?

- Let F be the forecast of rain.
- $P(rain) = 5/365 = 0.0137$
- $P(F|rain) = 0.9$; $P(F|norain) = 0.1$
- $P(rain|F)$?

$$P(rain|F) = \frac{P(F|rain)P(rain)}{P(F)} = \frac{P(F|rain)P(rain)}{P(F|rain)P(rain) + P(F|norain)P(norain)}$$

$$= \frac{0.9 \cdot 0.0137}{0.9 \cdot 0.0137 + 0.1 \cdot 0.9863} \approx 0.111 \approx 11.1\%$$