Abdelhafid Boussouf University Center – Mila **Academic Year: 2025/ 2026**

Institute of Natural and Life Sciences **Module: Biostatistics**

Department of Environment and Ecology **Instructor: R. Boustila**

Practical Sessions: Series 1

📓 Exercise 1:

We are interested in studying the possible relationship between the **number of annual sunshine hours (Y)** and the **average temperature in July (X)** in several French cities.

City	Average Temperature in July (X)	Annual Sunshine (Y)
Ajaccio	22.2	2726
Bordeaux	20.8	1992
Clermont-Ferrand	19.7	1898
Brest	16.6	1492
Lille	17.9	1617
Lyon	21.3	2010
Millau	19.3	2121
Nice	23.1	2668
Paris	19.5	1630
Strasbourg	20.3	2010
Toulouse	21.6	2437
Fort-de-France	27.5	2685
Papeete	25.0	2685

Questions

a) Calculate the **mean**, **variance**, and **standard deviation** for each variable X and Y.
→ To be done using Excel's predefined functions.

b) Calculate the **covariance** and the **linear correlation coefficient**.
→ To be done both with the course formulas and with Excel's predefined functions.

c) Determine the **equation of the regression line** $Y=aX + b$ using the **least squares method**.

d) Using Excel, plot the **scatter diagram** and the **regression line** on the same graph.

e) Interpret the obtained results.

Numerical Results

Variable	Value
Mean of X	20.6
Mean of Y	2083
Variance of X	5.85
Variance of Y	158,928
Covariance Cov(X,Y)	173.71
Correlation r(X,Y)	0.93
Regression line	Y = 1505 + 39.2X

Solution :

1 Enter the Data

In Excel, enter the data in three columns as follows:

City	Temperature (X)	Sunshine (Y)
Ajaccio	22.2	2726
Bordeaux	20.8	1992
Clermont-Ferrand	19.7	1898
Brest	16.6	1492
Lille	17.9	1617
Lyon	21.3	2010
Millau	19.3	2121
Nice	23.1	2668
Paris	19.5	1630
Strasbourg	20.3	2010
Toulouse	21.6	2437
Fort-de-France	27.5	2685
Papeete	25.0	2685

2 Calculate the Mean

Use Excel formulas:

• Mean of X: =AVERAGE(B2:B14)

• Mean of Y: =AVERAGE(C2:C14)

Expected results: Mean(X) ≈ 20.6, Mean(Y) ≈ 2083

3 Variance and Standard Deviation

Use Excel formulas:

• Variance of X: =VAR.S(B2:B14)

• Variance of Y: =VAR.S(C2:C14)

• Std deviation of X: =STDEV.S(B2:B14)

• Std deviation of Y: =STDEV.S(C2:C14)

Expected results: Var(X) ≈ 5.85, Var(Y) ≈ 158,928, SD(X) ≈ 2.42, SD(Y) ≈ 399

4 Covariance

Formula: =COVARIANCE.S(B2:B14;C2:C14)

Expected result: Cov(X,Y) ≈ 173.71

5 Correlation Coefficient

Formula: =CORREL(B2:B14;C2:C14)

Expected result: r ≈ 0.93 (very strong positive correlation)

6 Regression Line

You can get regression coefficients in two ways:

Method A (formulas):

• Slope (b): =SLOPE(C2:C14;B2:B14)

• Intercept (a): =INTERCEPT(C2:C14;B2:B14)

Regression line: Y = 1505 + 39.2X

Method B (Data Analysis Toolpak):

1. Go to Data → Data Analysis → Regression.

2. Input Y Range: C2:C14, Input X Range: B2:B14.

3. Check Labels and Line Fit Plots, then click OK.

7 Scatter Plot and Regression Line

1. Select columns B and C (Temperature and Sunshine).

2. Go to Insert → Scatter (XY) → Scatter with Straight Line.

3. Click on a data point → Add Trendline → Display Equation and R^2 on chart.

Excel will display something like: Y = 39.2X + 1505 (R^2 = 0.86)

Results and Interpretation

- The mean temperature X is **20.6 °C**, and the mean sunshine duration Yis **2083 hours**.
- The variance of X is **5.85**, and that of Y is **158,928**, showing a wide dispersion of sunshine values.
- The covariance Cov(X,Y)=173. indicates a **positive joint variation** between temperature and sunshine.
- The correlation coefficient r=0.93 shows a **very strong and positive linear relationship**: as the average temperature increases, the annual sunshine duration also increases.
- The estimated regression line is:

  Y=1505+39.2X

  which can be used to **estimate the annual sunshine hours from the average July temperature**.

📓 Exercise 2 :

Researchers want to study the relationship between the **average annual rainfall (X, in mm)** and the **annual agricultural yield of wheat (Y, in quintals per hectare)** in several regions.

Data Table

Region	Rainfall X (mm)	Wheat Yield Y (q/ha)
A	600	35
B	720	42
C	850	45
D	400	25
E	500	30
F	950	48
G	800	43
H	650	38
I	700	40
J	1000	50

---

1. Questions

a) Calculate the **mean**, **variance**, and **standard deviation** of each variable X and Y.
☞ Use Excel functions (AVERAGE, VAR.S, STDEV.S).

b) Compute the **covariance** and **correlation coefficient** between X and Y.
☞ Use formulas from the course and Excel functions (COVARIANCE.S, CORREL).

c) Determine the **regression line equation** $Y=a\,X+b$ using the least squares method.
☞ Use Excel functions: SLOPE, INTERCEPT.

d) Using Excel, **plot the scatter diagram** and **add the regression line** on the same graph.

e) **Interpret the results**:
What can you say about the relationship between rainfall and wheat yield?

Solution

Variable	Description	Formula (Excel)	Result
Mean of X	Average rainfall	=AVERAGE(B2:B11)	**717.0 mm**

Variable	Description	Formula (Excel)	Result
Mean of Y	Average yield	=AVERAGE(C2:C11)	**39.6 q/ha**
Variance of X	Spread of rainfall	=VAR.S(B2:B11)	**36,112.22**
Variance of Y	Spread of yield	=VAR.S(C2:C11)	**61.6**
Covariance	Joint variation	=COVARIANCE.S(B2:B11,C2:C11)	**1473.11**
Correlation (r)	Linear relation	=CORREL(B2:B11,C2:C11)	**0.99**
Slope (b)	b=Cov(X,Y) /Var(X)	=SLOPE(C2:C11,B2:B11)	**0.041**
Intercept (a)	$a=\bar{Y}-b\bar{X}$	=INTERCEPT(C2:C11,B2:B11)	**10.35**
Regression Line		**Y = 10.35 + 0.041X**	
R^2	Explained variation	=RSQ(C2:C11,B2:B11)	**0.98**

📈 Interpretation

- The correlation **r = 0.99** shows a **very strong positive relationship** between rainfall and wheat yield.
- The regression line Y=10.35+0.041X means:
  ☞ For each **1 mm** of extra rainfall, yield increases by **0.041 q/ha** on average.
- The coefficient of determination R2=0.98 means that **98% of the variation in yield** is explained by rainfall.

□ Excel Chart

1. Enter data in two columns:
   - Column B → Rainfall (X)
   - Column C → Yield (Y)
2. Select both columns → **Insert → Scatter (XY)**.
3. Add **Trendline → Linear → Show Equation + R^2**.
   You'll see a line close to:

   Y=10.35+0.041X, R2=0.98