University Aboubekr Belkaid,Tlemcen	L1 Computer Science
Sciences Faculty	Tutorial 2
Computer Science Departement	Proba-Stat

Exercice 1 *A statistical study conducted with 50 high school students focused on the variable X, which represents a student's height (in cm), and the variable Y, which represents their weight (in kg). The values taken by the pair of random variables (X, Y) along with their respective partial frequencies are provided in the following table.*

X (in cm)/Y (in kg)	y_1=52	y_2=60	y_3=68
x_1=165	n_{11}= 10	n_{12}=11	n_{13}=3
x_2=170	n_{21}=4	n_{22} =0	n_{23}=8
x_3=175	n_{31}=7	n_{32}=2	n_{33}=5

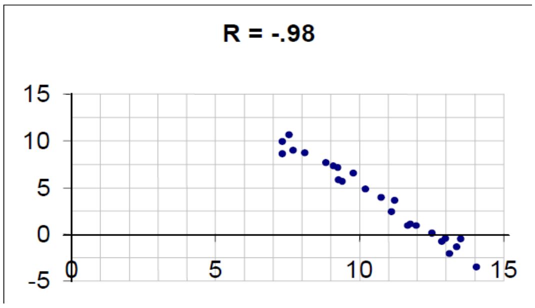
1) interpret n_{23} et n_{32}.
2) How many students have a weight less than 60 kg and a height greater than 165 cm?
3) Among the students who have a height of 170 cm, what is the percentage of those who weigh more than 60 kg?
4) Determine the distribution of partial frequencies for the pairs (X, Y).
5) Calculate the marginal distributions of the partial frequencies of X and Y.
6) deduce $\bar{x}$, $Var(X)$, σ_X, $\bar{y}$, $Var(Y)$, σ_Y.
7) Among students who are 170 cm of tall, what is the percentage of those that weight is 68kg?
8) Calculate $cov(X, Y)$.
9) Calculate the correlation coefficient r for the couple (X, Y).

Exercice 2 *The table below presents grades of individual students on the first and second hourly exams in a certain course.*

Exam 1	Exam 2	Exam 1	Exam 2
55	54	75	85
95	85	45	56
63	58	70	23
47	78	40	44
74	79	72	75
41	47	76	23
71	61	81	88

1. Using the table above, prepare a scattter plot (the first exam grade for the horizontal axis).
2. Calculate marginal sample means of Exam 1 and Exam 2.
3. Find the covariance of the exam grades.
4. Calculate the correlation coefficient r .
5. Calculate the regression equation of Y=Exam 2 and X=Exam 1.

Exercice 3 *The figure below shows a scatter plot for a set 25 data values. The correlation coefficient is $r = -0.98$.*



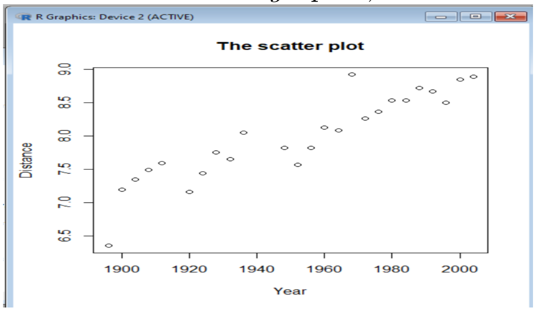
1. *The value of* $\bar{x} = 10.6$, $\bar{y} = 3.9$.*Mark the approximation location of the point* $(\bar{x}, \bar{y})$ *on the graph, how many data values have both* x *and* y *greater than their respective means or both less than these means? how this is related to the value of* r.
2. *Estimate the range of* x *and* y *from the graph.*

Exercice 4 *A biologist wants to model the growth of a bacterial population over time. The following data were recorded during an experiment:*

Time (hours)	*1*	*2*	*3*	*4*	*6*
Bacterial poulation	*120*	*200*	*320*	*500*	*920*

It is suspected that the growth follows an exponential function of the form $y = a\exp(bx)$, *where* y *is the bacterial population,* x *is the time in hours, and* a *and* b *are constants to be determined.*
1. *Plot the data on a scatter plot.*
2. *Transform the exponential model into a linear form by taking the logarithm of the population* y *and calculate the coefficients* a *and* b *of the exponential regression.*
3. *Write the final equation of the model.*
4. *Predict the bacterial population after 5 hours.*

Exercice 5 *Data concern the gold-medal winning distance, in metres, for the men's long jump for the Olympic games for the years 1896 to 2004. (Some years were missing owing to the two world wars.)*
1. *Which is the independent variable and which is the dependent variable?*
2. *Given the below graphic, describe the association between the distance and year.*



3. *Given that* $n = 25$, $\sum(Year) = 48800$, $\sum(Year^2) = 95285440$, $\sum(Year * Distance) = 390239.3$, $\sum(Distance) = 199.67$, *determine the correlation coefficient* r.
4. *Determine the linear equation regression.*
5. *Use your equation to predict the winning distance in the year 2008.*