1.9 Exercises

1. 1 In the following situations, describe the population/sample, name and determine the type of the variable.

1) To determine the socio-economic profile of the households living in Neuchâtel, we note the number of children per household in a sample of a thousand households.
2) In Switzerland and according to the Federal Statistical Office for the year 2013, German is the language spoken most frequently among the population with a proportion of 63.5%, then with 22.5% French, with 8.2% Italian, with 0.5% Romansh and finally with 5.3% another language.
3) In a school, we ask each student about their dropout rate, that is the number of low grades divided by the number of grades. We have :

   a) Less than 1%
   b) From 1% to 15.9%
   c) From 16% to 49.9%
   d) 4.50% and more

1. 2 In a survey, 820 Swiss citizens are asked about the bilateral treaties between Switzerland and the European Union. The answers are :

Answers	**Absolute frequency**	**Relative frequency**
Very useful	95	
Useful	342	
Harmful	210	
Very harmful	46	
No opinion	127	
Total		

1) Complete the distribution table and represent, by using an appropriate chart, the distribution.
2) Calculate the confidence level of these treaties, that is the percentage of citizens that estimate them useful or very useful.

1. 3 On a road, the speed limit is 80km/h. One measures the speed of 50 vehicles :

84	81	76	71	80	81	83	84	80	83
74	75	92	76	80	82	94	73	83	83
75	81	79	**97**	78	82	76	78	82	82
78	81	91	**68**	82	73	82	79	75	77
83	80	77	81	69	78	81	83	87	87

1) Divide the data into classes of equal size and establish a distribution table.
2) Represent the organized data by using a histogram and a frequency polygon.
3) Complete : «Almost .... of the vehicles respects the speed limitation of 80km/h and ..... % drive at a speed between 80km/h and 85km/h. By taking into account a tolerance of 5km/h, ... % of the vehicles are finable.»

1.4 In a factory, during a quality control, one measures the diameter, in mm, of 50 bolts randomly selected. The results are :

Diameter [mm]	**Frequency**
$[21.5; 21.8[$	4
$[21.8; 21.9[$	6
$[21.9; 22.0[$	6
$[22.0; 22.1[$	13
$[22.1; 22.2[$	8
$[22.2; 22.3[$	7
$[22.3; 22.5[$	6
Total	50

1) Represent these data by using a histogram as well as a cumulative frequency polygon.

2) If the nominal value is 22mm, calculate the percentage of bolts that deviates of more than 0.3mm. Check the coherence of the result on the frequency polygon.

1.5 Answer the following :

1) Rewrite without using the sum notation : $\sum\limits_{i=1}^{6} (x_i - a)^2$

2) Rewrite using the sum notation : $3ax_1y_1z_1 + 3ax_2y_2z_2 + \cdots + 3ax_ky_kz_k$

3) Given that $\sum\limits_{i=1}^{6} x_i = -4$ and that $\sum\limits_{i=1}^{6} x_i^2 = 100$, calculate :

a) $\sum\limits_{i=1}^{6} (2x_i + 3)$ b) $\sum\limits_{i=1}^{6} x_i (x_i - 1)$ c) $\sum\limits_{i=1}^{6} (x_i - 5)^2$

4) The two variables x and y take the values x_i and y_i here below :

x	2	-5	4	-8
y	-3	-8	10	6

Calculate :

a) $\sum\limits_{i=1}^{4} x_i y_i$

b) $\sum\limits_{i=1}^{4} x_i \cdot \sum\limits_{i=1}^{4} x_i$

c) $\sum\limits_{i=1}^{4} x_i^2 + y_i^2$

d) $\sum\limits_{i=1}^{4} x_i^2 + \sum\limits_{i=1}^{4} y_i^2$

1.6 For that exercise, we use the wording of the exercise 2.4.

1) Calculate the average diameter of the bolts and represent that value by a triangle under the x-axis of the histogram.

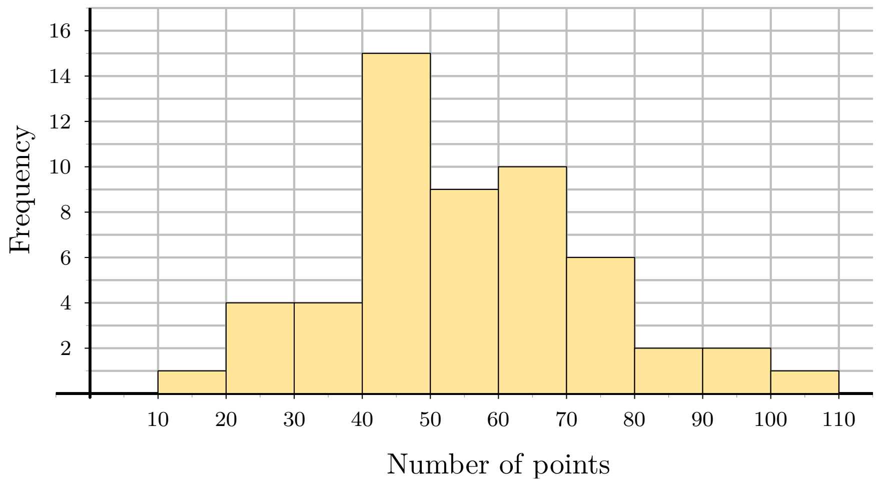
2) Determine the median diameter and mark that value by a vertical bar on the histogram

3) Determine the modal class. Is it representative in that situation ? Justify your answer.

4) Is it possible to have an idea of the shape of the distribution by using the mean, the median and the modal class ?

1.7 The sets of data $S_1 = \{3, 4, 4, 4, 4, 4, 5\}$, $S_2 = \{1, 3, 4, 4, 5, 5, 6\}$ and $S_3 = \{1, 1, 4, 4, 6, 6, 6\}$ represent the grades of seven students at three different tests. Calculate the standard deviation of each set of data and interpret the results.

1.8 The number of points obtained by the 54 Swiss schools at the competition *Mathématiques sans Frontières* is represented by the histogram here below :



1) Calculate the mean $\bar{x}$ and the standard deviation σ of these data and give an interpretation of these results. Place clearly these values on the histogram.

2) What is the standardized variable or standard score of a school that obtains 110 points ?

3) What is the number of points of school whose standardized variable equals -2 ?

1.9 The table here below represents the evolution of the price of a loaf of bread in a bakery.

Year	2010	2011	2012	2013	2014	2015
x	1	2	3	4	5	6
p [CHF]	1.75	2	2.1	2.25	2.4	2.55

1) Place these data on a set of axes. Is it possible to use an adjustment method ?

2) Let's call G_1 the average point of the first three values, G_2 the average point of the last three values. Calculate G_1, G_2 and give the equation of the line of adjustment l_1 of Mayer.

3) Determine the equation of the adjustment line l_2 by using the least squares method.

4) Compare these two methods and give the probable price for 2018.

1.10 Let's consider the table :

x	1	2	3	4
y	1	3	6	7

1) By solving a 2x2 system, find the regression line $y = ax + b$ for the given data.
2) Check your result thanks to the formulas in your theory.
3) One can prove that a regression line obtained with the least squares method passes always through the point $P(\overline{x};\overline{y})$. Check that result with the numerical values here above.

1.11 Given the following data :

x	1	3	4	6	8	9	11	14
y	1	2	4	4	5	7	8	9

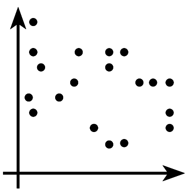
1) By using the least squares method, determine the regression line $y = ax + b$.
2) Find a new regression line by switching roles of x and y.
3) Verify that the intersection of these two lines is the point $P(\overline{x};\overline{y})$.

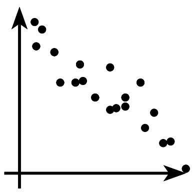
1.12 The number of bacteria y per unit of volume in a cultured broth after x hours is given by the table :

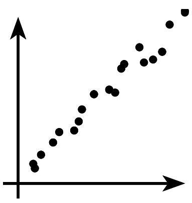
x	0	1	2	3	4	5	6
y	32	47	65	92	132	190	275

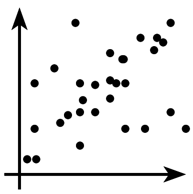
1) Thanks to a computer, draw the scatter plot and take note that there's no linear correlation between these two variables.
2) By stating that $y' = \ln(y)$, draw thanks to a computer, a new scatter plot with axes x and y' take note that this time there's a linear correlation.
3) By using a computer, determine the equation of the regression line $y' = ax + b$.
4) Determine and then draw the regression curve $y = \dots$ given that its form is $y = a \cdot b^x$.

1.13 Associate the correlation coefficients $r_1 = 0.44$, $r_2 = -0.91$, $r_3 = -0.36$ and $r_4 = 0.98$ to the scatter plots here below :

a) 

b) 

c) 

d) 

1. 14 The table here below represents the braking distance of a car on a dry road depending on the car's speed.

v [km/h]	40	50	60	70	80	90	100	110	120
d [m]	20.29	28.42	35.57	45.75	58.94	70.12	95.15	98.17	113.19

1) Determine the correlation coefficient between v and d and the one between v and $z = \sqrt{d}$.

2) Estimate the braking distance of a car driving at 200km/h.

1. 15 *With a computer*
One wants to analyse the progression of the records in the men's 100 metres. For that we have the data here below :

Year	1900	1912	1921	1930	1964	1983	1991	1999
x	0	12	21	30	64	83	91	99
t	10.80	10.60	10.40	10.30	10.06	9.93	9.86	9.79

After a first analyse, we state that $a = \mathrm{e}^{-0.00924x}$ and $b = \ln(t)$ to model our problem. Thus, we obtain the table :

a	1.000	0.895	0.824	0.758	0.554	0.464	0.431	0.401
b	2.380	2.361	2.342	2.332	2.309	2.296	2.288	2.281

1) Calculate the equation of the regression line $b = ma + h$ thanks to the least squares method.

2) Deduce that it is possible to model an expression of the form $t = \mathrm{e}^{a \cdot e^{-0.00924x}+b}$ where a and b are two real numbers that must be found.

3) Thanks to that adjustment, what record one can expect in 2010 ?

4) Calculate $\lim\limits_{x\to\infty} \mathrm{e}^{0.154 \cdot \mathrm{e}^{-0.00924x}+2.221}$.

5) What can we conclude, by using that model, about the men's 100 metres records over the very long term ?

1. 16 Answer the following :

1) By using the same model as in the theory, establish the equations in order to find by the least squares method the parabola regression $y = ax^2$.

2) Use you model and find the parabola regression for the data :

x	0.2	-0.7	0.5	0.6	-0.4
y	0.1	1	0.5	0.7	0.3

1. 17 How many different ways of sitting 7 people are there when

1) they sit on a bench ?

2) they sit on chairs around a round table ?

1. 18 Determine the number of words (anagrams) that can be written with the letters of

1) THEIR 2) UNUSUAL 3) SOCIOLOGICAL