# A statistics exercise	Season	01
	Episode	05
	Time frame	1 period

Prerequisites : Statistics. This is the last episode before the test.

Objectives :
- Check the understanding of the main statistical notions and methods.
- Peer-to-peer teaching to ascertain and consolidate the understanding.

Materials :
- *Test papers.*
- *Slideshow.*

1 – Cooperative work — 45 mins

Step 1 (10 minutes) : Each student is handed out a test paper and starts answering to the questions on the paper.

Step 2 (15 minutes) : Each student gives his paper to the student behind. This second student has to check the answers of the first one, make any corrections that may be appropriate, than continue answering the questions.

Step 2 (20 minutes) : Each student gives his paper to the student behind. This third student has to check the answers of the first and second, make any corrections that may be appropriate, than continue answering the questions.

2 – Correction — 10 mins

Papers are finally given to a fourth student who will mark it and grade it according to the answers given by teacher with a slideshow. Papers are collected at the end of the hour and every students is credited of the average marks of the three papers he or she contributed to.

A statistics exercise

Season	01
Episode	05
Document	Test paper

Student 1 _____

Student 2 _____

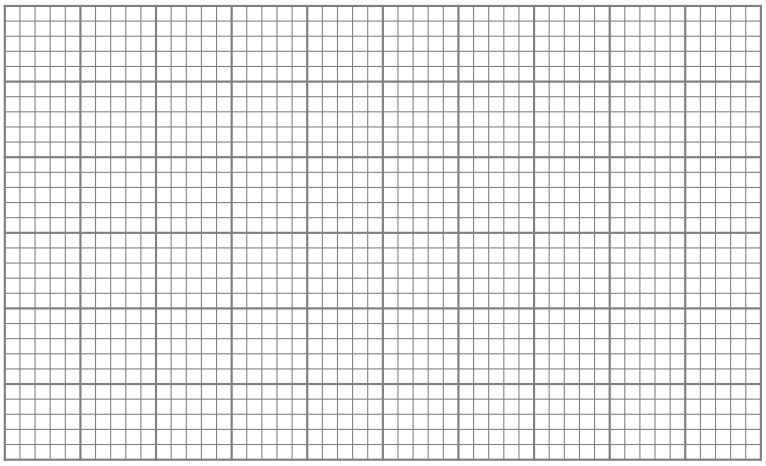
Student 3 _____

In a casino game, people can win between 0 and 20 pounds. The fee to play this game is 10 pounds. A study has been carried over 200 people, each playing 1 game. Their winnings after the games are shown in the table below. The first line gives the possible winnings, gathered in intervals. The second line shows the number of people whose winnings at the end of the game was in each interval.

Winnings	$[0;4[$	$[4;6[$	$[6;8[$	$[8;10[$	$[10;14[$	$[14;20[$
Absolute frequency	38	50	54	46	10	2

For all computations, whenever necessary, the middle of each interval will be used.

1. Represent this data as a histogram.



2. Compute the average of these data and represent it as a blue vertical line on the histogram.

3. Compute the median and the quartiles of these data and represent it as a red vertical line on the histogram.

Another study has been carried over 500 other people. The results are shown in the table below.

Winning	0	1	2	3	4	5	6	7	8	9
Frequency	27	34	37	34	42	41	39	41	35	29
Winning	10	11	12	13	14	15	16	17	18	19
Frequency	25	23	24	17	12	9	11	8	5	7

4. Find out the minimum, maximum and range (or spreading) of these 500 values.

5. Compute the average winnings of these 500 values.

6. Find out the median of these values and show all the steps in your method.

7. According to these results, would you play this game ?

We now have 700 values : 200 in the first table, gathered in intervals, and 500 in the second table.

8. Use the two averages to compute the average of the 700 values. Show your computation.

9. Can you deduce in the same way the median of the 700 values ? If so, do it. If not, explain why.

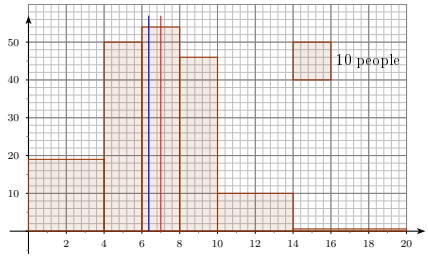
A statistics exercise	Season	01
	Episode	05
	Document	Correction

In a casino game, people can win between 0 and 20 pounds. The fee to play this game is 10 pounds. A study has been carried over 200 people, each playing 1 game. Their winnings after the games are shown in the table below. The first line gives the possible winnings, gathered in intervals. The second line shows the number of people whose winnings at the end of the game was in each interval.

Winnings	$[0;4[$	$[4;6[$	$[6;8[$	$[8;10[$	$[10;14[$	$[14;20[$
Absolute frequency	38	50	54	46	10	2

For all computations, whenever necessary, the middle of each interval will be used.

1. Represent this data as a histogram.



2. Compute the average of these data and represent it as a blue vertical line on the histogram.

$$\overline{x} = \frac{2 \times 38 + 5 \times 50 + 7 \times 54 + 9 \times 46 + 12 \times 10 + 17 \times 2}{200} = 6.36$$

3. Compute the median and the quartiles of these data and represent it as a red vertical line on the histogram.

There are 200 values.
It's an even number so the median is between the 100th and the 101st value.
The first quartile is the $\frac{200}{4} = 50$th value.
The third quartile is the $\frac{3 \times 200}{4} = 150$th value.
To find the actual values, we need the cumulative absolute frequencies.

W.	$[0;4[$	$[4;6[$	$[6;8[$	$[8;10[$	$[10;14[$	$[14;20[$
A.F.	38	50	54	46	10	2
C.A.F.	38	88	142	188	198	200

So, finally,

$$Med = [6;8[\,;\ Q_1 = [4;6[\text{ and } Q_3 = [8;10[.$$

or

$$Med = 7\,;\ Q_1 = 5 \text{ and } Q_3 = 9.$$

Another study has been carried over 500 other people. The results are shown in the table below.

Winning	0	1	2	3	4	5	6	7	8	9
Frequency	27	34	37	34	42	41	39	41	35	29
Winning	10	11	12	13	14	15	16	17	18	19
Frequency	25	23	24	17	12	9	11	8	5	7

4. Find out the minimum, maximum and range (or spreading) of these 500 values.

   The minimum is 0. The maximum is 19. The range is $19 - 0 = 19$.

5. Compute the average winnings of these 500 values.

   $$\frac{0 \times 27 + 1 \times 34 + \cdots + 19 \times 7}{500} = 6.99$$

   The average is 6.99.

6. Find out the median of these values and show all the steps in your method.

   There are 500 values. It's an even number so the median is between the 250th and the 251st value. Using the C.A.F. (cumulative absolute frequencies) we see that the median is 6.

7. According to these results, would you play this game ?

   Why not, but it's more likely to lose money than to win, as the fee is 10 pounds and the average winnings after 500 games only amounts to $6,99$ pounds.

We now have 700 values : 200 in the first table, gathered in intervals, and 500 in the second table.

8. Use the two averages to compute the average of the 700 values. Show your computation.

   In this computation, we have to consider the fact that there are 200 values in the first set of data and 500 in the second. So the global mean is

   $$\frac{200 \times 6.36 + 500 \times 6.99}{700} = 6.81.$$

9. Can you deduce in the same way the median of the 700 values ? If so, do it. If not, explain why.

   We can't find the value of the median, as in one case the data is gathered in intervals, and in the second it's just listed.
   We could gather the 700 values in classes, but it wouldn't be very accurate.