

Practical Session 3

Statistical graphics for quantitative variables

In this session, we can analyse a discrete and a continuous variable from our class questionnaire. The necessary techniques are illustrated using some of the data from the theory class examples

1. Discrete data example

In the theory class we have used the following data set of times voted by a group of Madrileños.

Times	3	3	3	4	1	2	4	5	2	3	1	1	3	8	4	1	3	4	2	5	0	0	5	4	2	1	2	3	3	2
voted	1	4	3	2	3	5	0	6	3	1	3	5	4	1	4	1	2	4	4	3	3	0	7	2	2	1	3	4	2	2

Before trying to set up a frequency table, it is useful to first calculate minimum and maximum values of the data (to know which values to use for the table). This can be done using the **MIN** and **MAX** commands.

Then we can set up a column of the different values of the table. (The easiest way is typing the first two values and dragging down). Then, as in the previous practical, we can use the **COUNTIF** command to calculate the frequencies of each value of the data. Remember to use the dollar trick to fix the data!

Times voted	Absolute frequency (n)
0	4
1	10
2	12
3	15
4	11
5	5
6	1
7	1
8	1
>8	0

As in the previous practical, we can calculate the sum of the data and the relative frequencies. This gives the following frequency table.

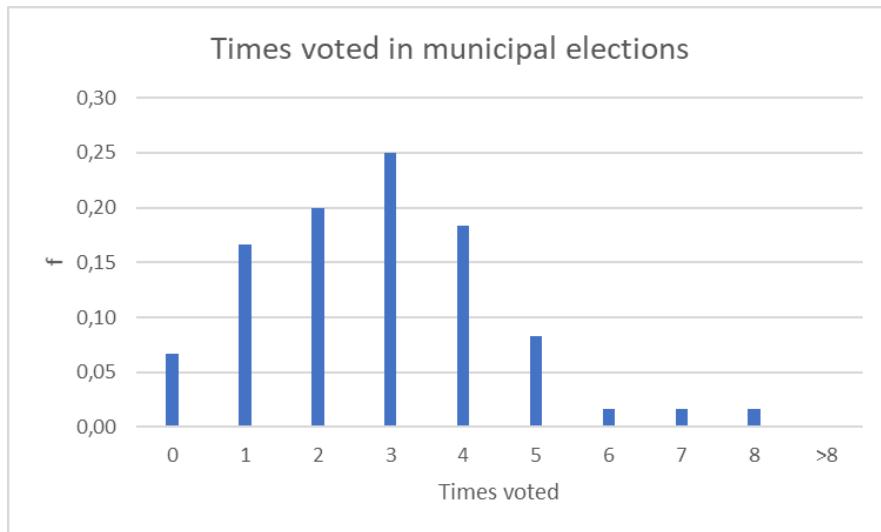
Times voted	Absolute frequency (n)	Relative frequency (f)
0	4	0.07
1	10	0.17
2	12	0.20
3	15	0.25
4	11	0.18
5	5	0.08
6	1	0.02
7	1	0.02
8	1	0.02
>8	0	0.00
TOTAL	60	1.00

As we have discrete data, we can augment the table with cumulative frequencies. Assuming that the values 4 and 0.07 above are in cells **B2** and **C2** respectively, then we can set **D2 = B2** and **E2 = C2** and then **D3 = D2+B3** and **E3 = E2+C3** and drag down. Note that it does not make sense to include values in the TOTAL file for cumulative frequencies.

Times voted	Absolute frequency (n)	Relative frequency (f)	Cumulative frequency (N)	Cumulative relative frequency (F)
0	4	0,07	4	0,07
1	10	0,17	14	0,23
2	12	0,20	26	0,43
3	15	0,25	41	0,68
4	11	0,18	52	0,87
5	5	0,08	57	0,95
6	1	0,02	58	0,97
7	1	0,02	59	0,98
8	1	0,02	60	1,00
>8	0	0,00	60	1,00
TOTAL	60	1,00		

2. Drawing a bar chart for discrete data.

As in the previous practical, we can use the Bar Chart Graphics option to construct the chart. Here however, we should adjust the bar width to its thinnest value, to show that the bar represents the number of data at an individual value. We can construct the bar chart with individual (or cumulative) frequencies. To construct the graph, it is easiest to just select the frequency values and then correct the labels later.



What features can you observe in your constructed graph? What are typical values of the data? Are the data symmetric or asymmetric? Do any data look unusual or atypical?

3. Constructing a frequency table for continuous data.

Yearly spending of different town halls (€)

```
114579 73896 59003 86165 53428 93844 61536 90628 49501
56767 78063 87750 82409 107664 60479 88872 66325 78268
38360 82436 83531 81364 63210 112842 56206 59052 52660
45000 91562 66308 50397 79964 65369 71803 60108 49264
```

When we decide on the bar intervals to use, in order to construct the frequency table, we can use the **COUNTIFS (CONTAR.SI.CONJUNTO)** function.

For example, if our data are in cells **A1: I4** and we wish to calculate the number of data greater than 30 and less than or equal to 45, we can use the command

COUNTIFS(A1:I4;">>30";A1:I4;"<=45").

Even faster, if the interval limits are in cells **K1** and **L1**, we could write:

COUNTIFS(A1:I4;">">"&K1;A1:I4;"<="&L1).

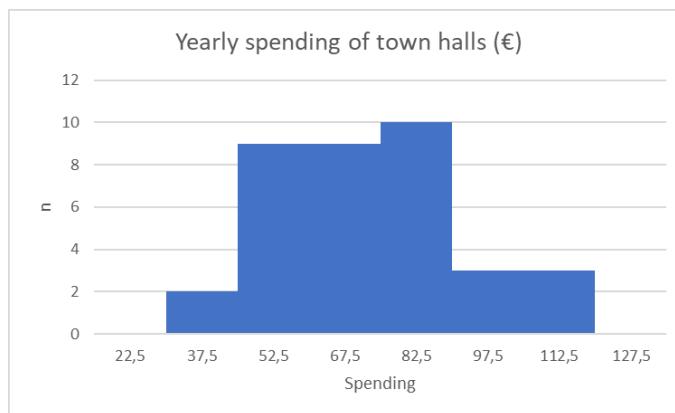
The above formula could easily be dragged down.

Yearly spending (€)	Absolute frequency
≤ 30	0
(30,45]	2
(45,60]	9
(60,75]	9
(75,90]	10
(90,105]	3
(105,120]	3
> 120	0
Total	36

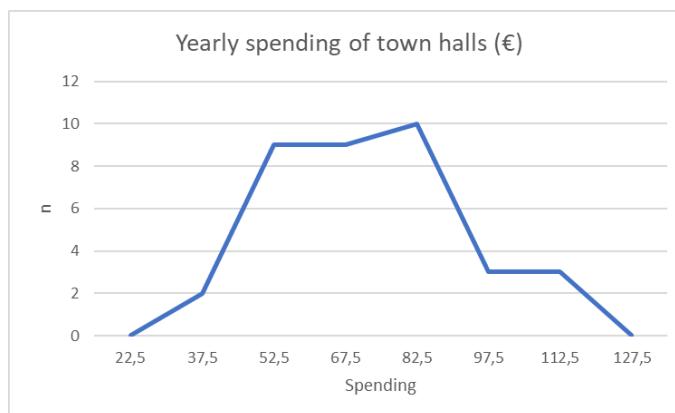
It is useful to show columns for the left and right-hand side and centre of each bar for constructing the full frequency table and corresponding graphics.

Lower limit	Centre	Upper limit	Absolute frequency (n)	Relative frequency (f)	Cumulative frequency (N)	Cumulative relative frequency (F)
15	22,5	30	0	0,00	0	0,00
30	37,5	45	2	0,06	2	0,06
45	52,5	60	9	0,25	11	0,31
60	67,5	75	9	0,25	20	0,56
75	82,5	90	10	0,28	30	0,83
90	97,5	105	3	0,08	33	0,92
105	112,5	120	3	0,08	36	1,00
120	127,5	135	0	0,00	36	1,00
Total			36	1,00		

To construct the frequency histogram, we mark the frequency column and then use the bar chart graphics option as for discrete data. We then need to adjust the gaps between bars to 0 and relabel the x axis appropriately.



To construct a frequency polygon, we can use the line graph instead of the bar chart graphic.



4. Constructing a variable width histogram.

We will not deal with this problem in the practical class as it is non-trivial in Excel. A webpage explaining how to do this (in Excel 2016) is available [here](#).

Note that most statistical packages always construct a relative frequency histogram defining height as frequency/width.