| **Group** | **Students** | **Signatures** |
| --- | --- | --- |
|  | * Alonso Herreros Copete |  |

**IMPORTANT:** *The teachers of this course apply a ‘zero tolerance’ policy regarding academic dishonesty. Students that sign up this document agree to deliver an original work. The breach of this commitment will result in academic punishment.*

**Observations:**

Solve the exercises in the **Assignment1.pdf** file. **Note:** It is advisable to consult the manual for basic operation of MATLAB / Octave available on the website of the course.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

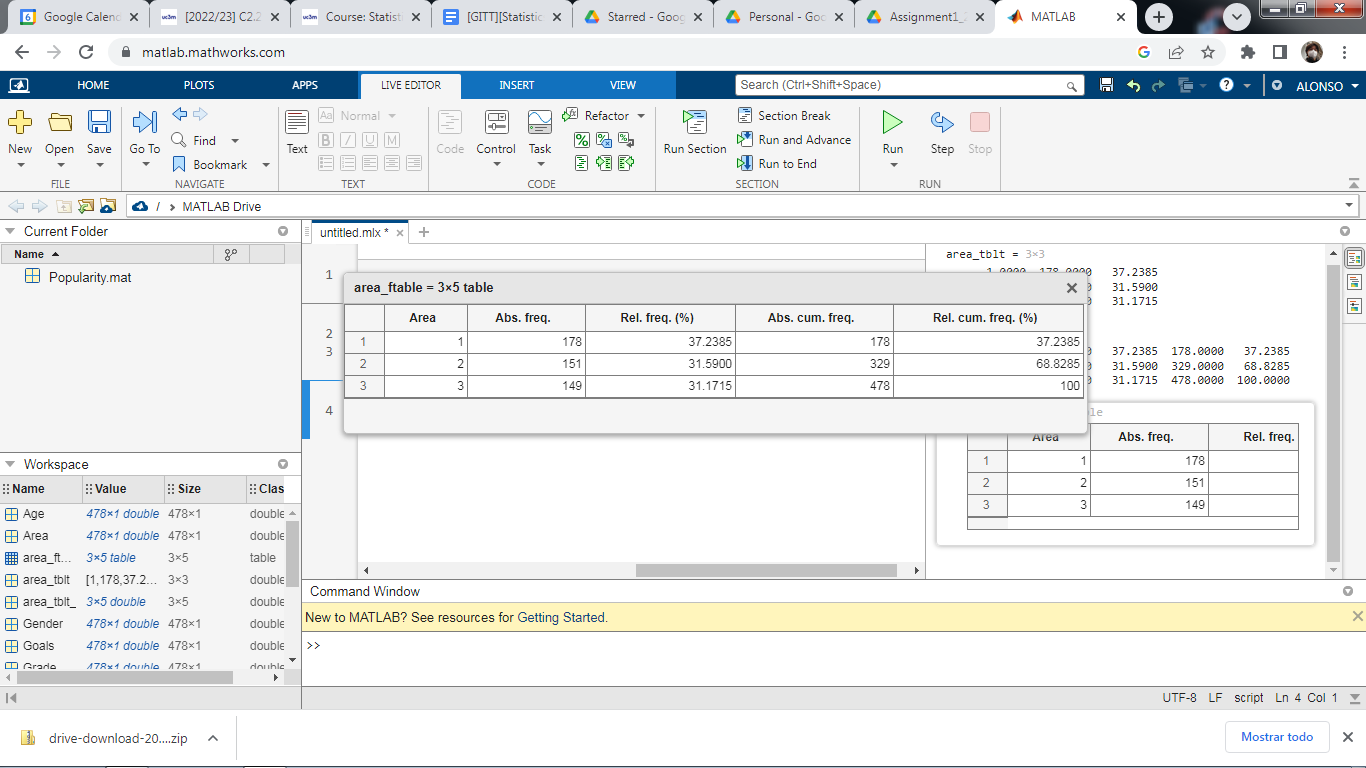
\*All code for each of the 3 problems is in a separate file called Section<n>.pdf

## Analysis of a data set

**1.** Calculate the frequency table of variable Area. The table must include the absolute, relative, cumulative absolute and cumulative relative frequencies. In which of the three types of areas most students are concentrated?

The following is the requested frequency table. There is a screenshot of the generated variable, too.

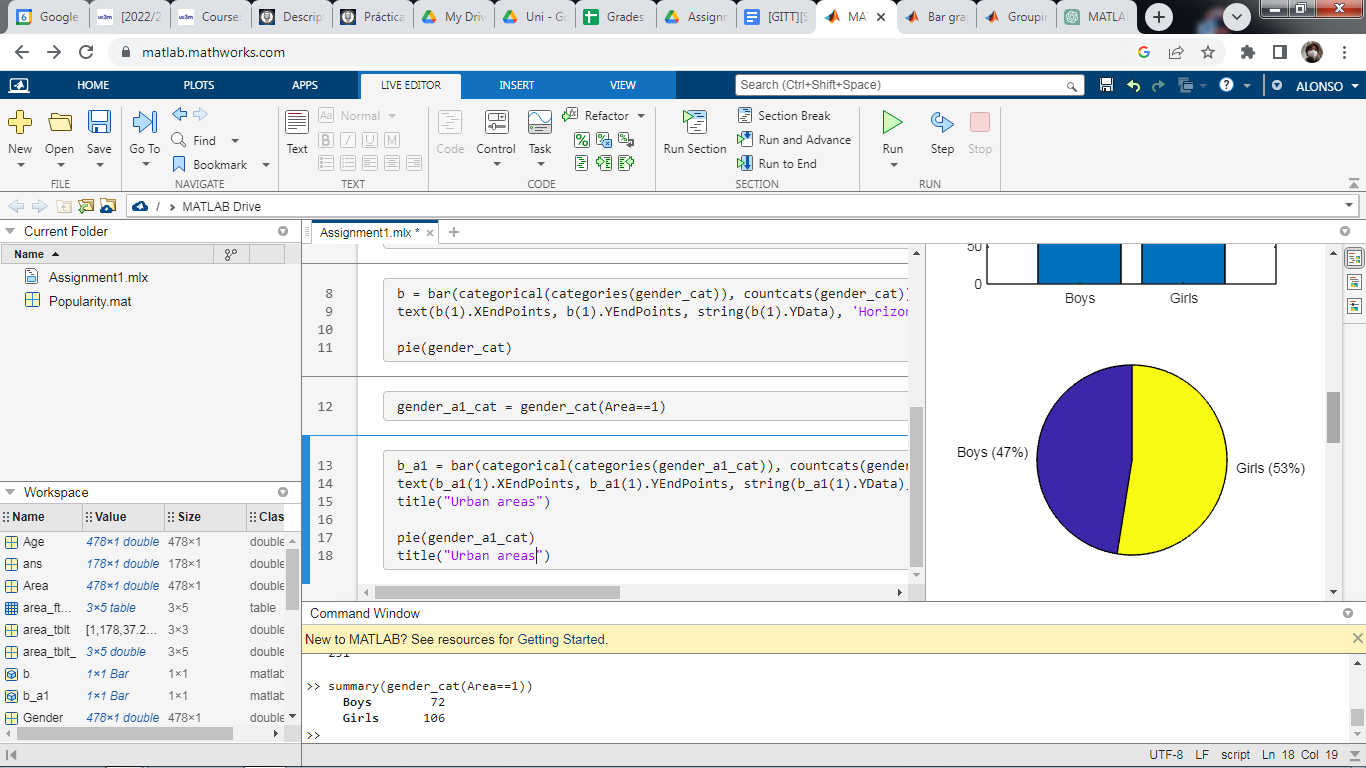
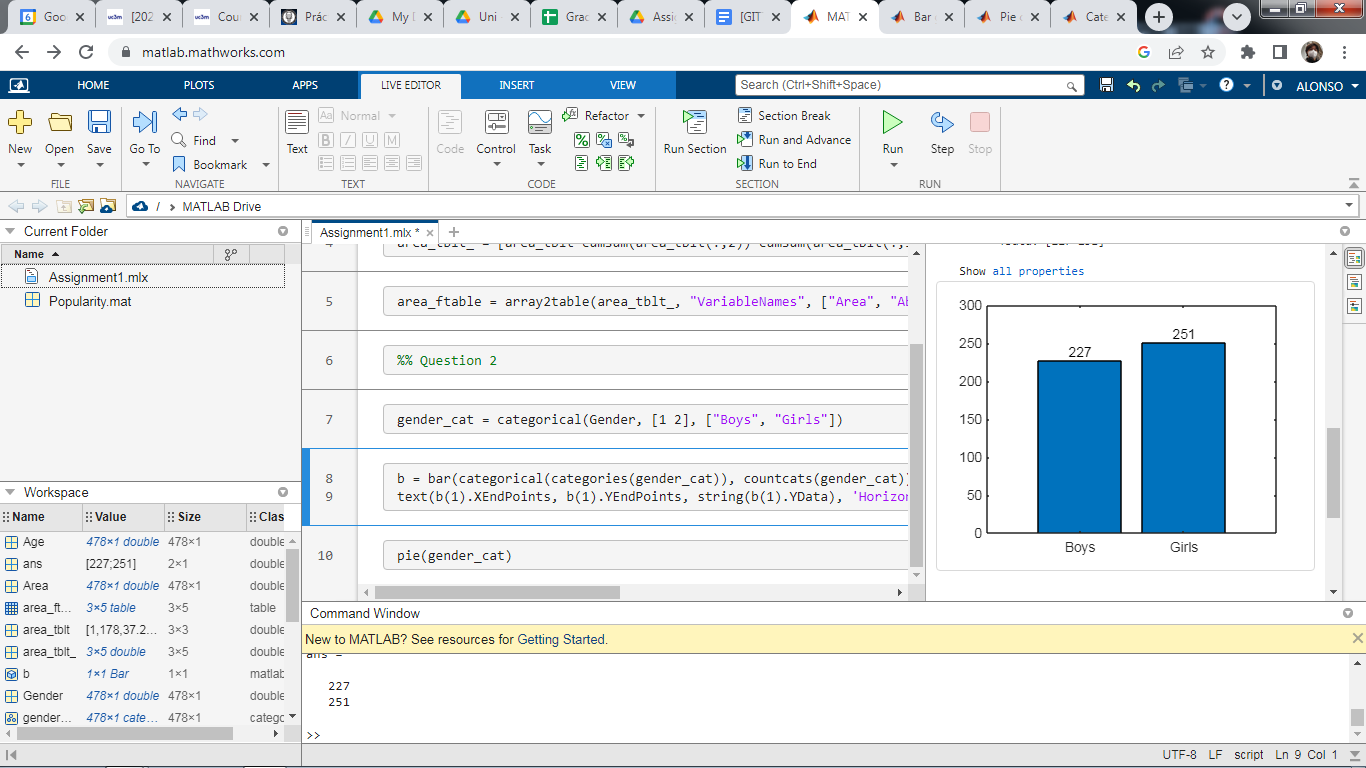
| Area | Abs. frequency | Rel. frequency (%) | Abs. cum. freq. | Rel. cum. freq (%) |
| --- | --- | --- | --- | --- |
| 1 | 178 | 37.238 | 178 | 37.2385 |
| 2 | 151 | 31.59 | 329 | 68.8285 |
| 3 | 149 | 31.1715 | 478 | 100 |



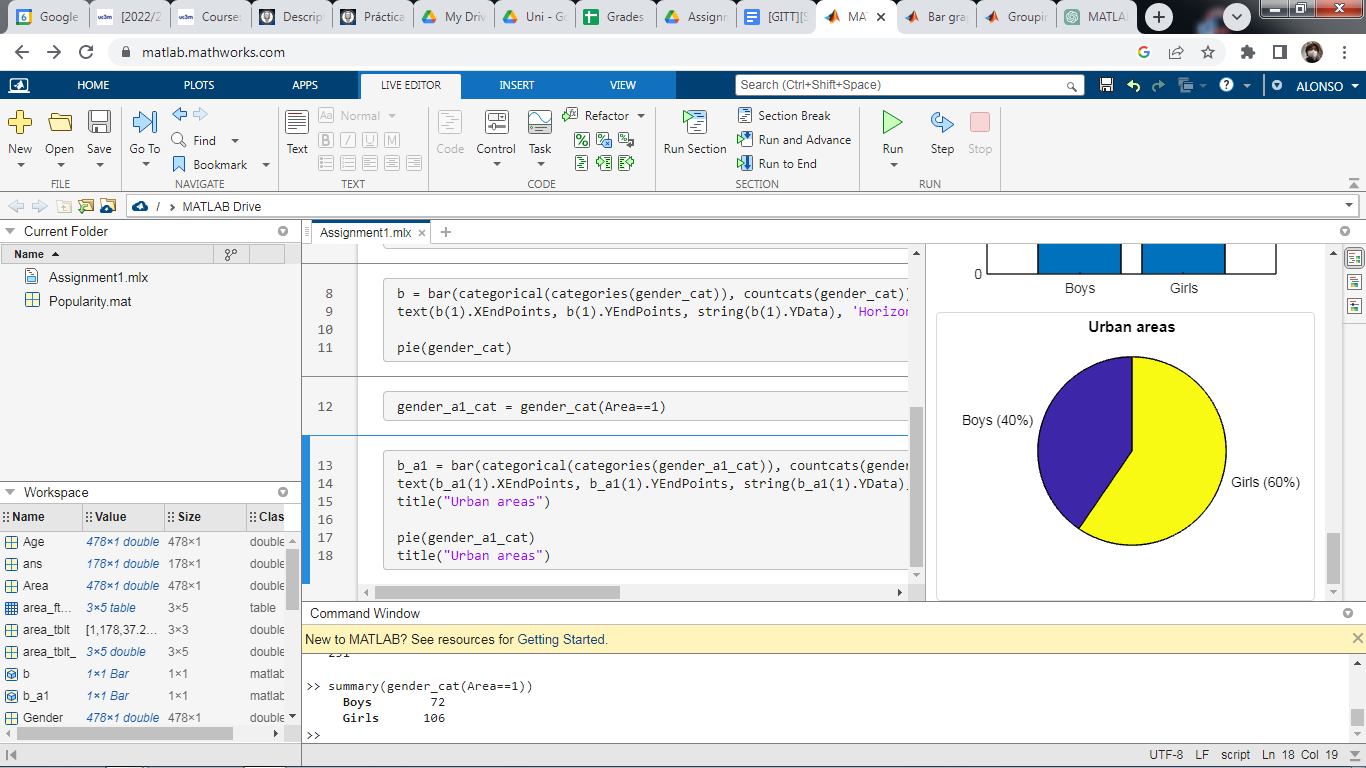
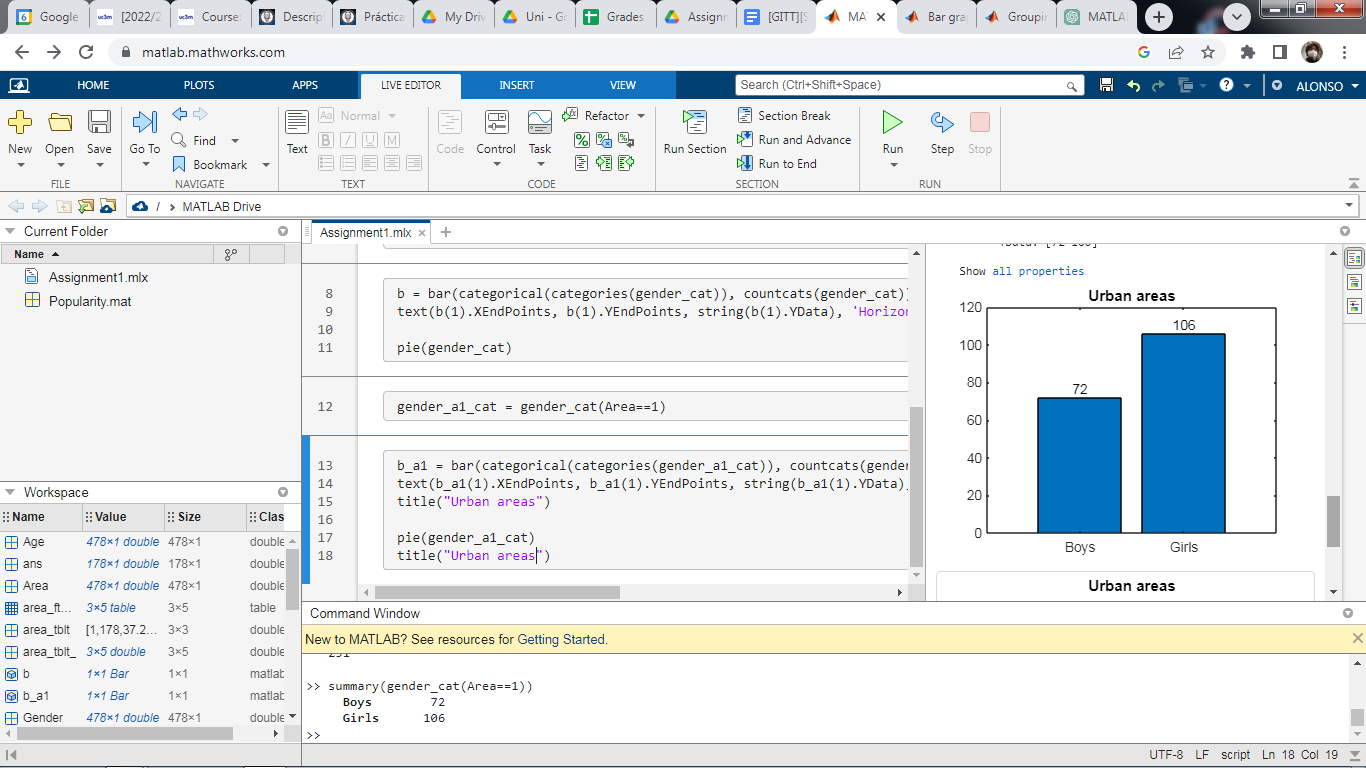
Most students are concentrated in area 1 (urban).

**2.** What is the proportion of boys and girls? Represent graphically that proportion with a bar and a pie chart. What is the proportion of boys and girls whose schools are established in urban areas?

There are 227 boys and 251 girls. That’s 47.5% boys and 52.5% girls.

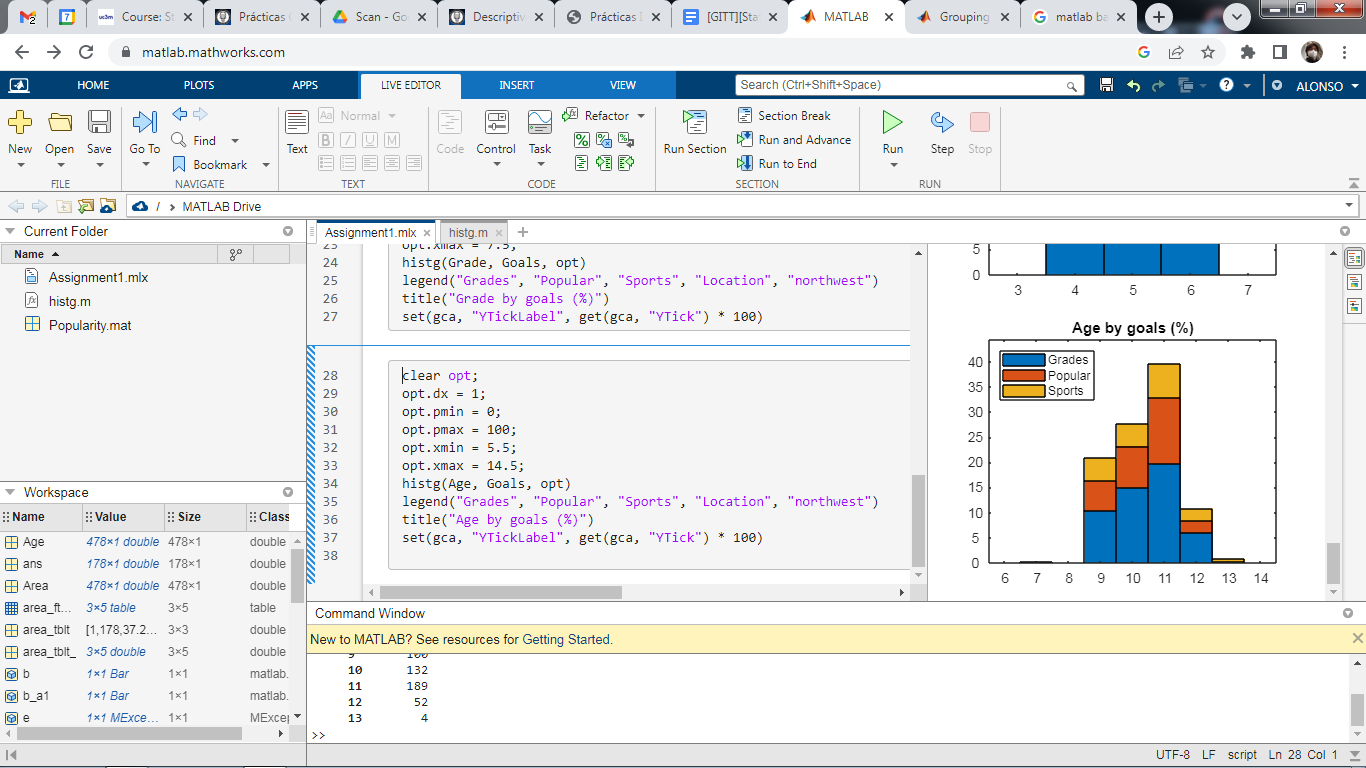
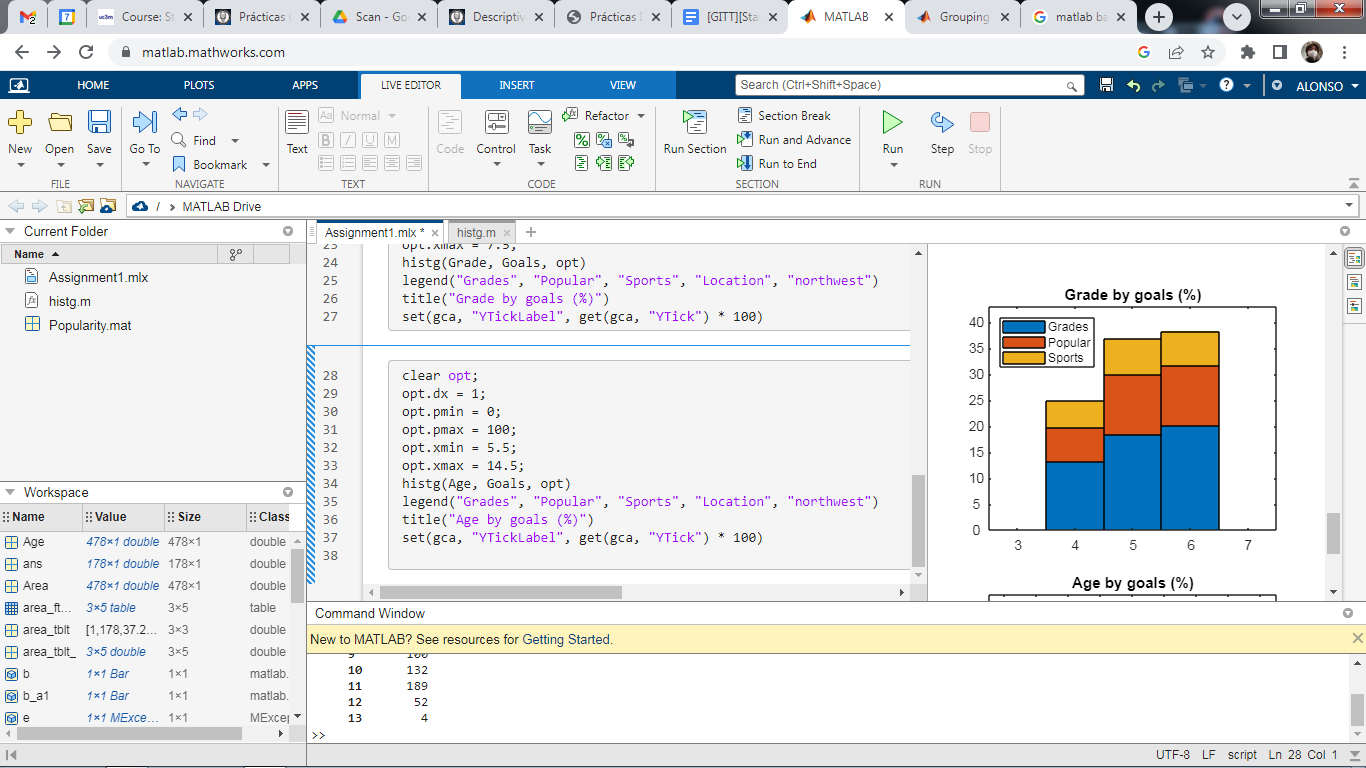
****

If we focus on urban areas (Area = 1), we can find 72 boys and 106 girls, corresponding to 60% and 40% of students in urban areas, respectively.



**3.** Do histograms of the variables Grade and Age by the variable Goals. Calculate the mean and the standard deviation of the variables Grades and Sports by Age groups.

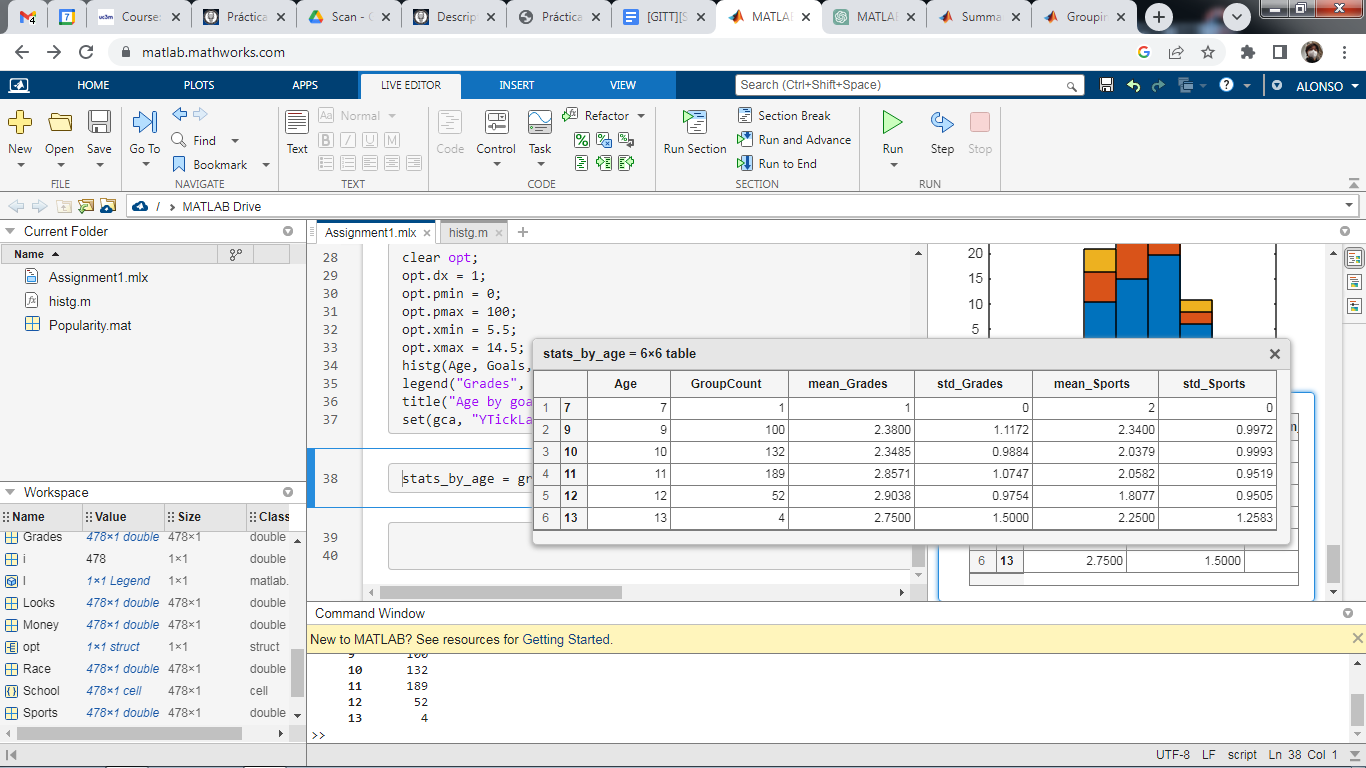
The following are screenshots of the histograms created.



The mean and standard deviation of Grades and Sports by Age groups can be seen in the following table:

| Age | Grades mean | Grades std. dev. | Sports mean | Sports std. dev. |
| --- | --- | --- | --- | --- |
| 7 | 1.00 | 0.0000 | 2.00 | 0.0000 |
| 9 | 2.38 | 1.1172 | 2.34 | 0.9972 |
| 10 | 2.35 | 0.9884 | 2.04 | 0.9993 |
| 11 | 2.86 | 1.0747 | 0.06 | 0.9519 |
| 12 | 2.90 | 0.9754 | 1.81 | 0.9505 |
| 13 | 2.75 | 1.5000 | 2.25 | 1.2583 |

This is a screenshot of the variable used to obtain that data

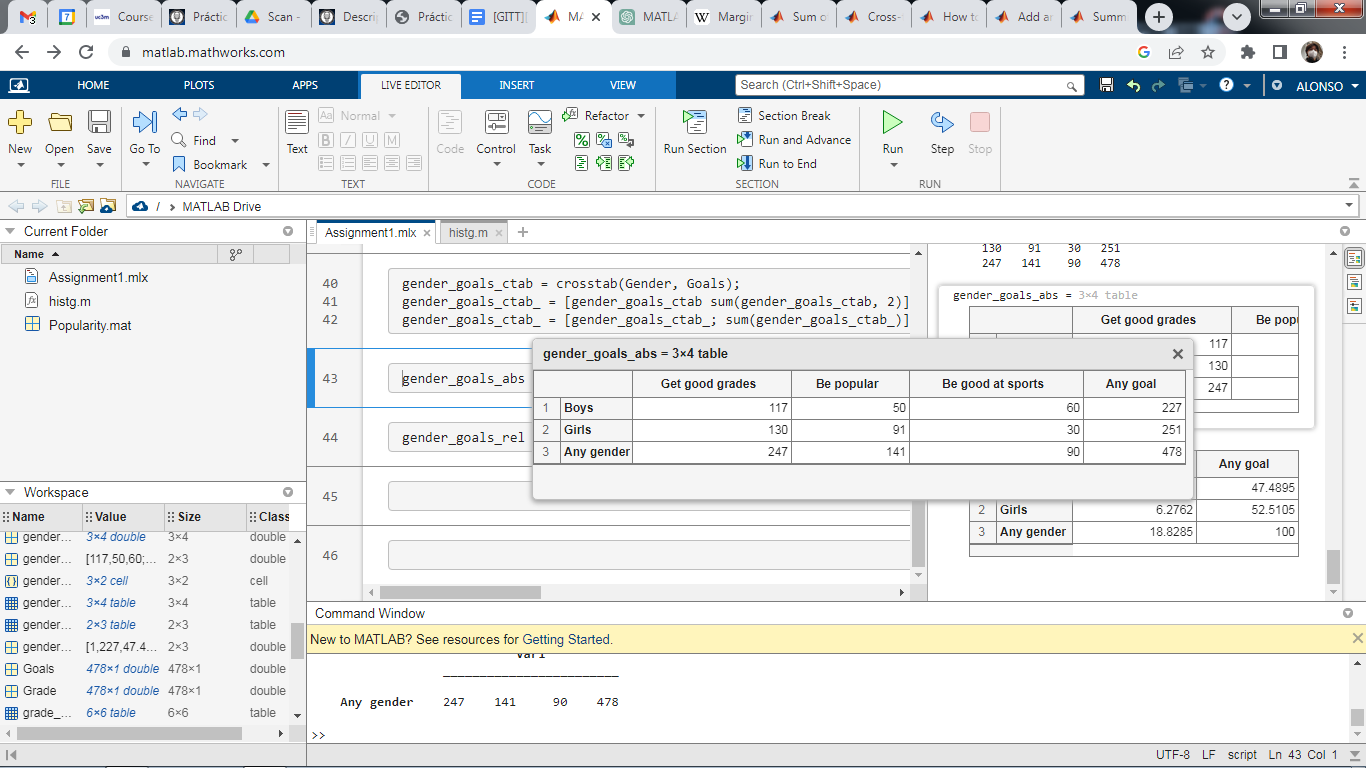


**4.** Analyse the variables Gender and Goals in a double entry table. Calculate the absolute frequency table with its marginal distributions and the relative frequency table with its marginal distributions.

The following are the generated tables, along with screenshots of the variables used. The “Any” rows and columns represent the marginal distributions.

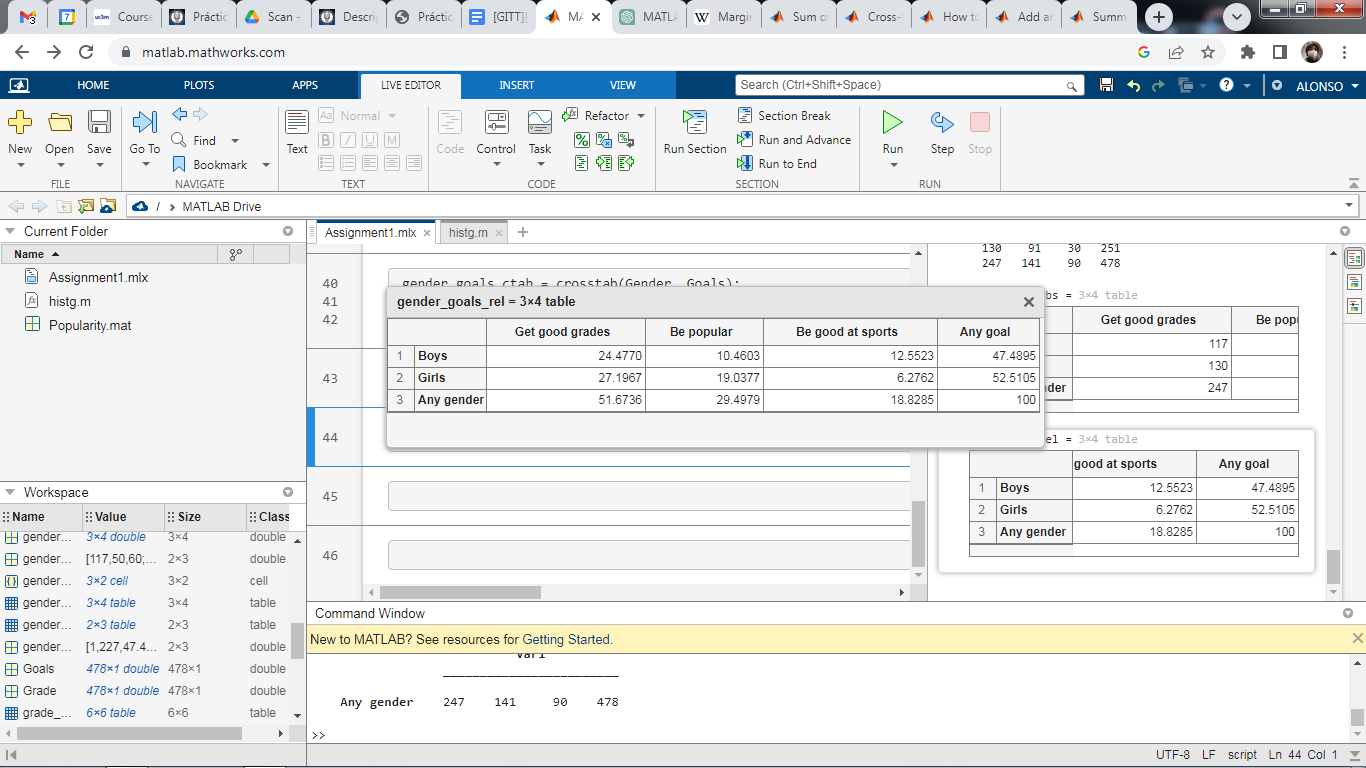
Gender and Goals, absolute frequency

|  | Get good grades | Be popular | Be good at sports | Any goal |
| --- | --- | --- | --- | --- |
| Boys | 117 | 50 | 60 | 227 |
| Girls | 130 | 91 | 30 | 251 |
| Any gender | 247 | 141 | 90 | 478 |



Gender and Goals, relative frequency (%)

|  | Get good grades | Be popular | Be good at sports | Any goal |
| --- | --- | --- | --- | --- |
| Boys | 24.48 | 10.46 | 12.55 | 47.49 |
| Girls | 27.20 | 19.04 | 6.28 | 52.51 |
| Any gender | 51.67 | 29.50 | 18.83 | 100 |



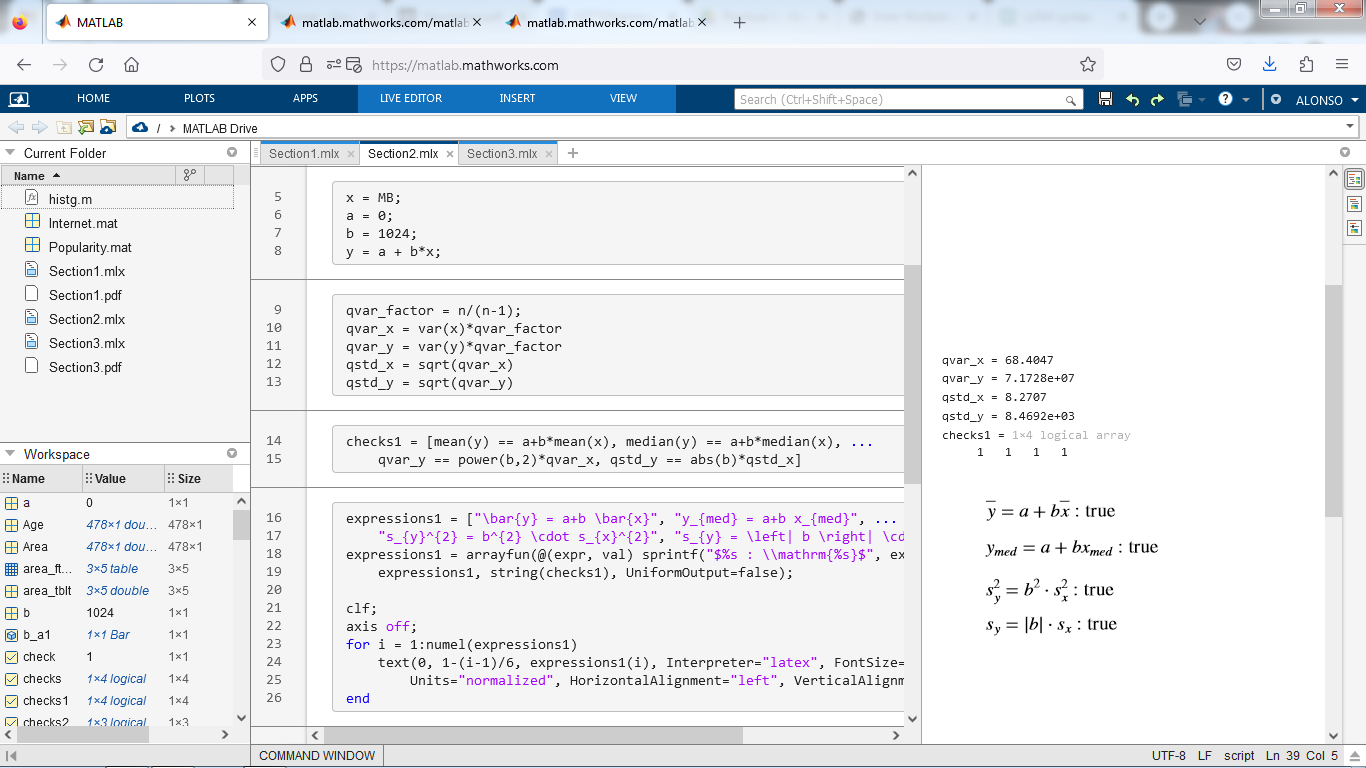
## Linear Transformations

**1.** **Change of units**. Consider the matrix internet in the file internet.mat, and consider the variable MB (“downloaded Mb”). Define a new variable, KB, as the nº of downloaded Kb, recall “1Mb = 1024Kb”. The new variable is the result of a linear transformation of the form . From this transformation, check with MATLAB/Octave the next theoretical relations:

1. .
2. , where med is the median.
3. , where is the sample quasi-variance.
4. , where is the sample quasi-standard deviation.

All relations were checked using MATLAB, and they turned out to be true.

The following is a screenshot of the code used to check these relations, and its output.



**2.** **Standardization of variables**. Consider the variable MB, and denote it as . Define a new variable as the result of the standardization of . The standardization consists of applying a linear transformation such that subtracts the mean value and divides by its standard deviation. The resulting variable has zero mean, and standard deviation and variance equal to one.

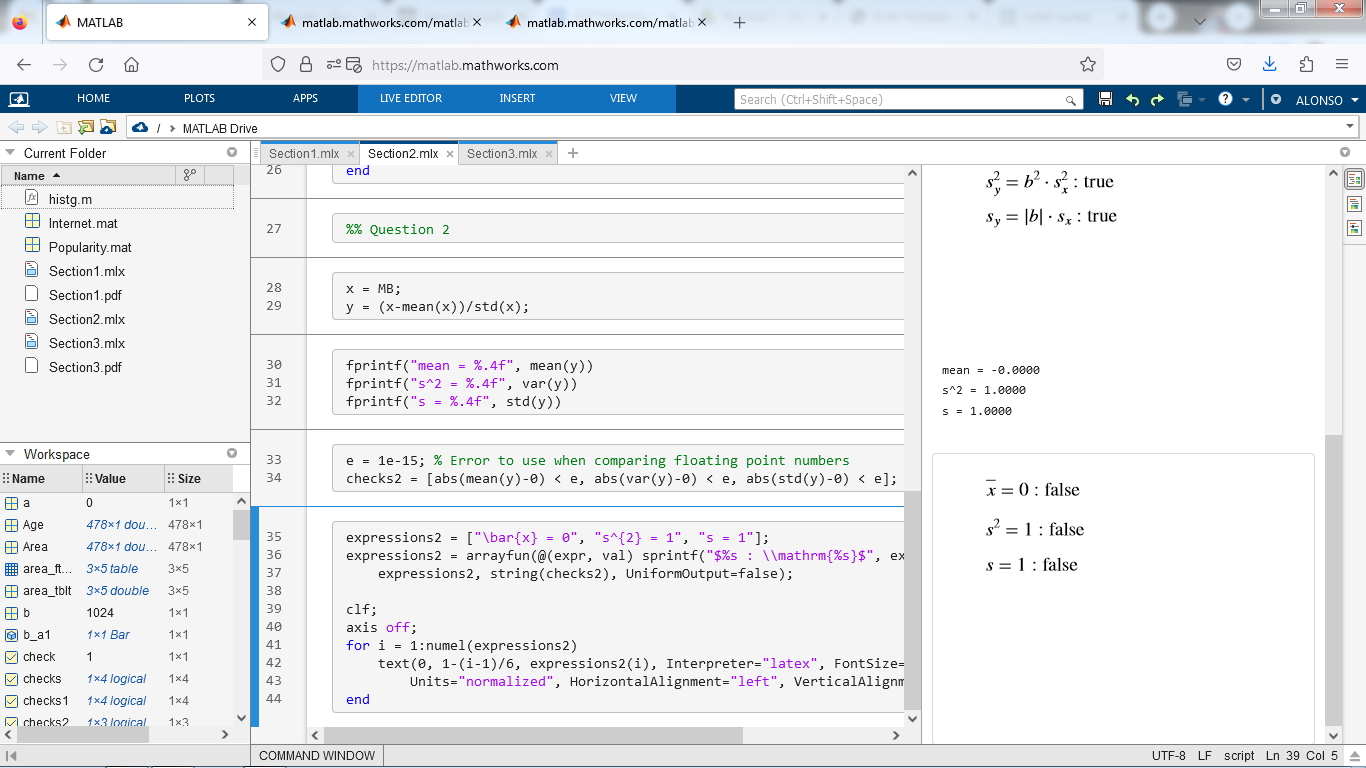
1. Determine the values of and of the corresponding linear transformation .
2. Obtain the new standardized variable and check in MATLAB/Octave, the next results:  
   , , and .

a) The normalization can be defined by the expression , being the standard deviation. We can use this expression to identify and such that in the following way:

|  |  |  |
| --- | --- | --- |
|  |

b) The expressions were checked using MATLAB, and they turned out to be true, although some adjustments had to be made to account for precision errors.

This is a screenshot of the code used to check the expressions, and its output. The values can be checked directly by looking at them. The last blocks of code perform the numerical comparisons, with a small error margin to account for precision errors, and show the results on screen.



## Correlation Between Linearly Transformed Variables

**Change of units.** Consider the matrix internet, and variables MB (“downloaded Mb”) and connection (“connection time in hours”). From them, create two new variables: “nº of downloaded KB” and “connection time in seconds”. Note that you are applying a linear transformation of the type: and , or simply a change of units: and .

Check in MATLAB/Octave, the next result:

which indicates that the correlation coefficient between two variables does not change if a change of units is applied.

The expression was checked in MATLAB, and it turned out to be true.

This is a screenshot of the code used to check the expression, as well as its output. Here, the check can be done visually by comparing the values of R\_x\_v and R\_y\_u. In the last block of code, the values are actually compared, and the result of the comparison is shown on screen.

