- Lab 2 : Introduction to Data Analytics
- Lecturer : Prof. Youness Moukafih
- Lab Ins: Prof. Safaa Berkani

Tutorial 1 – Summarizing data

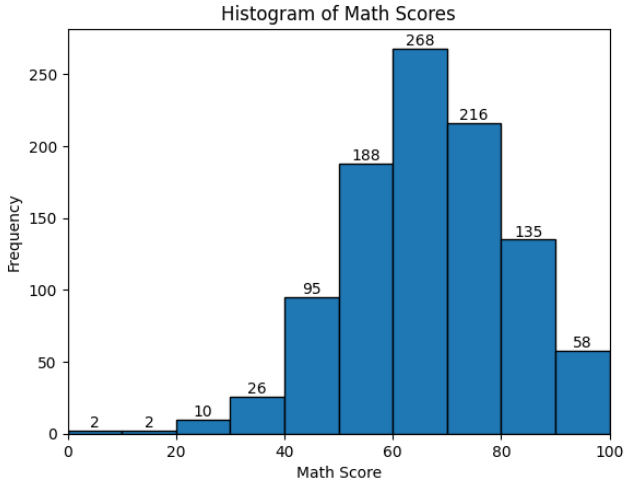
Exercise 1: Exploring the students performance dataset:

Q1: The figure below shows the histogram of the math scores of 1000 students. Using this histogram:

a- Estimate the median of X

b- Estimate Q1, Q3 and IQR of X

c- Estimate the mean of X.

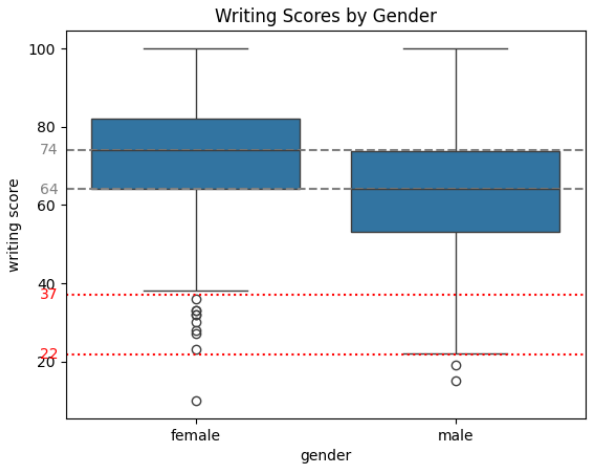


Q2: The figure below shows the boxplot of the math scores of 1000 students. Using this histogram:

a- For both genders, extract the statistics of the boxplots: the median, quartiles, and whiskers.

b- Are there any outliers in the distribution of writing scores? If so, which gender shows more outliers?

c- Which gender seems to have a higher overall math score distribution?

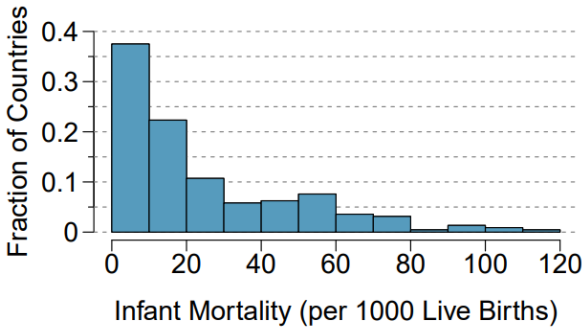


Q3 - What features of the distribution are apparent in the histogram and not the box plot? What features are apparent in the box plots but not in the histogram?

- Lab 2 : Introduction to Data Analytics
- Lecturer : Prof. Youness Moukafih
- Lab Ins: Prof. Safaa Berkani

Exercise 2: Infant Mortality:

The infant mortality rate is defined as the number of infant deaths per 1,000 live births. This rate is often used as an indicator of the level of health in a country. The relative frequency histogram below shows the distribution of estimated infant death rates for 224 countries for which such data were available in 2014.



Q1- What is the difference between the frequency histogram, and the relative frequency histogram.

Q2- Would you expect the mean of this data set to be smaller or larger than the median? Explain your reasoning.

Exercice 3: Stats scores :

Below are the final exam scores of twenty introductory statistics students.

78, 81, 94, 81, 73, 72, 69, 66, 57, 71, 89, 88, 82, 83, 83, 77, 78,74,79,79

Q1 – Calculate the mean, and estimate the median, quartiles, and whiskers.

Q2- Create a box-plot of the distribution of these scores.

Q3 – A new student with a score of 28 is added to the existing dataset of final exam scores. How does this addition affect the mean and median of the scores?

End of the lab.