Tasks for class 6

Quantiative Economics, Fall 2025

October 23, 2025

TASK 1: *Working with vectors*

You are a PE teacher in a school. Imagine that you received a vector containing the number of goals scored during PE classes by each of your n students. You want to get a sense of what the average number of goals scored in each semester is, and what is the standard deviation of those. When calculating the standard deviation for a sample, you want to use this unbiased estimator:

$$\texttt{standard\_deviation} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \texttt{average})^2}$$

Here, `average` is the estimated sample mean, n is the number of students, and g_i is the i-th element of a vector `n_goals`, which contains the number of goals scored by each student.

Let this vector be defined as:

$$\texttt{n\_goals} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix} = \begin{bmatrix} \text{goals scored by student 1} \\ \text{goals scored by student 2} \\ \vdots \\ \text{goals scored by student n} \end{bmatrix}$$

In `Julia`, the function `length()` returns the number of elements in a vector.

1. **Calculate the mean:**
   - Calculate the mean by summing the array (function `sum`) and dividing by the number of elements (function `length`).
2. **Calculate the `demeaned` vector:**
   - Use broadcasting to subtract the mean from each element of the array. You want `Julia` to perform the following operation and define the **d** vector. **Remember to use the dot operator "." for broadcasting!**

$$\textbf{demeaned} = \begin{bmatrix} g_1 \\ g_2 \\ \vdots \\ g_{n-1} \\ g_n \end{bmatrix} - \begin{bmatrix} \text{average} \\ \text{average} \\ \vdots \\ \text{average} \\ \text{average} \end{bmatrix}$$

3. **Calculate the vector of squared differences (`squared_d` vector):**

$$\textbf{squared\_d} = \begin{bmatrix} (g_1 - \text{average})^2 \\ (g_2 - \text{average})^2 \\ \vdots \\ (g_{n-1} - \text{average})^2 \\ (g_n - \text{average})^2 \end{bmatrix}$$

4. **Calculate the variance:**
   - Compute the variance. Note that the variance is the sum of `squared_d` divided by $n - 1$:

$$\texttt{variance} = \frac{\text{sum}(\textbf{squared\_d})}{n-1}$$

5. **Calculate the standard deviation:**
   - Finally, calculate the standard deviation by taking the square root of the variance:

$$\texttt{standard\_deviation} = \sqrt{\texttt{variance}}$$

TASK 2: Conditional Extraction

You (again!) are a PE teacher in a school. You have a vector containing the number of absences of each student.

Let this vector be:

$$\texttt{n\_absences} = \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_{n-1} \\ a_n \end{bmatrix} = \begin{bmatrix} \text{absences of student 1} \\ \text{absences of student 2} \\ \vdots \\ \text{absences of student n} \end{bmatrix}$$

You want to calculate how many students have a greater number of absences than 3. You want to do this using **conditional extraction**. Then, you want to calculate the average number of absences **among students who have more than 3 absences.**

Your tasks:

1. First, write a mask that will contain the value `1` whenever a particular student has more than 3 absences.

For example, if:

$$\texttt{n\_absences} = \begin{bmatrix} 2 \\ 5 \\ 1 \\ 6 \\ 3 \end{bmatrix} \quad \Rightarrow \quad \texttt{mask} = \begin{bmatrix} \texttt{0} \\ \texttt{1} \\ \texttt{0} \\ \texttt{1} \\ \texttt{0} \end{bmatrix}$$

2. Calculate the number of 1s in such a vector to find out how many students satisfy the condition.

3. Use this mask on the vector to extract number of absences only of those students who satisfy the condition.

4. Calculate the mean number of absences among students who satisfy the condition.