

Take-Home Assignment for Segment 11

Write a program to create a correlation matrix for samples represented by a set of numerical variables (e.g., microbiomes represented by OTU composition).

Input:

An $N \times M$ matrix in tab-delimited format, with line and column labels. The lines represent N samples and columns M variables.

Output:

A tab-delimited $N \times N$ matrix (triangular or square) with Pearson correlation coefficients between the samples (lines in the input matrix).

Maximum reward for completing the assignment is 20 points.

Extra credit option:

Also assess the p-values for the correlation coefficients using the permutation test (same as in assignment THA08). Use the upper right triangle of the output matrix to show the values of the correlation coefficient and lower left to show the p-values.

Maximum extra credit for completing this part is 5 points.

Notes and recommendations:

- Sample input (courtesy of Rachel Dockman from Liz Ottesen's lab) and output files are included with the assignment on eLC.
- The correlation matrix is symmetrical (that is, $r[i][j] = r[j][i]$). It is common in such situation to display only half of the values (upper right triangle or lower left triangle) and leave the rest empty (sometimes placing a * or other symbol in the diagonal cells $r[i][i]$) but you can print the full matrix if you prefer (unless you do the extra credit part, in which case the two triangles would show different values).
- **Do not use the <algorithms> container (#include <algorithms>) and do not use any external function to calculate the correlation coefficient or to manipulate the matrices.** When necessary, write loops going through all elements of the matrix.
- **If anything is unclear, ask.**

Submit your code via eLC.