Assignment 1

Hans Harder, s3754766

January 2019

Task 4. The covariance measures the variation between two variables. If the correspondence of values from both variables is high, the covariance is positive. If both variables seem to scale inversely to each other, a negative covariance is the result [1]. If the covariance is 0, both variables are independent of each other, meaning that a change of one variable has no impact on the other. It is defined as:

$$\text{cov}(\mathbf{X}, \mathbf{Y}) = \text{E}[(\mathbf{X} - \eta_x)(\mathbf{Y} - \eta_y)] \tag{1}$$

However, in order to obtain a measure on how much both variables change linearly in respect to to each other, it is necessary to normalize the covariance by using the Pearson Correlation Coefficient $\frac{1}{\sigma_x \sigma_y}$. The result is the correlation, which is defined as:

$$\text{cor}(\mathbf{X}, \mathbf{Y}) = \frac{1}{\sigma_x \sigma_y} \text{cov}(\mathbf{X}, \mathbf{Y}) \tag{2}$$

Task 5. The standardized matrix is calculated as follows:

$$\text{std}(\mathbf{X}) = \frac{\mathbf{X} - \eta_x}{\sigma_x} \tag{3}$$

It can be calculated using R by calling `scale(X)`.

Task 7a Results for plotting the 5 highest CPU average utilizations:

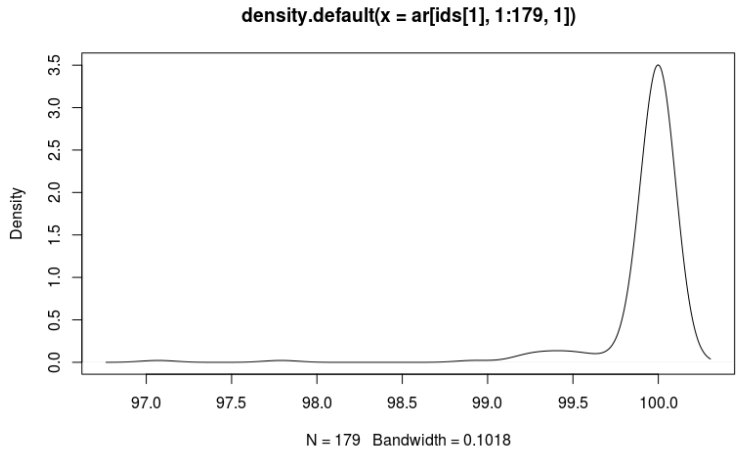

Figure 1: VM with highest average CPU utilization

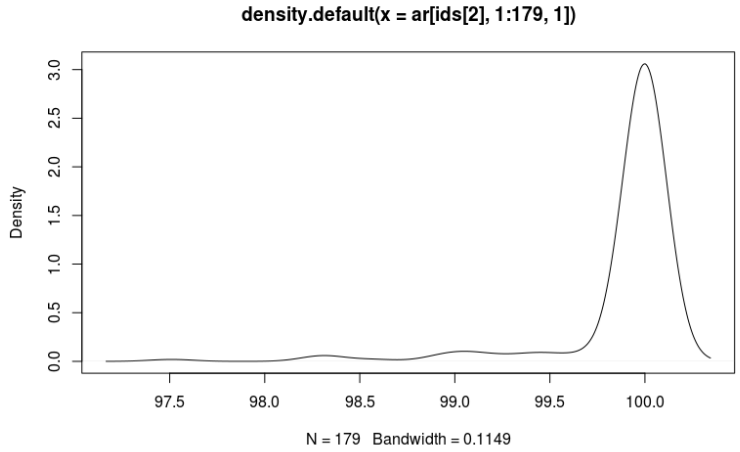

Figure 2: VM with 2nd highest average CPU utilization

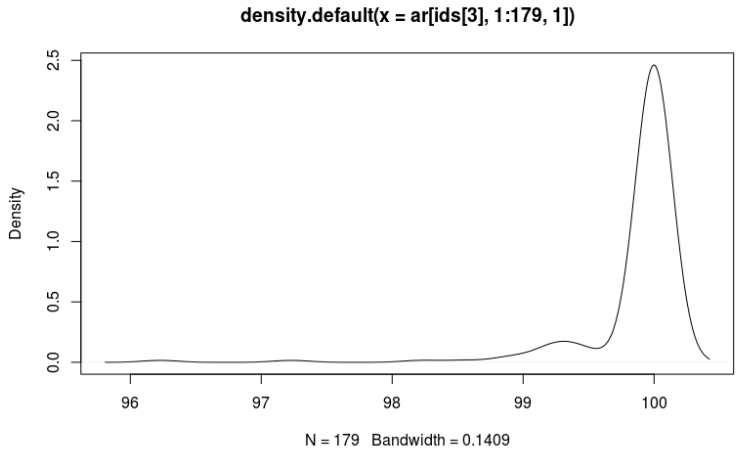

Figure 3: VM with 3rd highest average CPU utilization

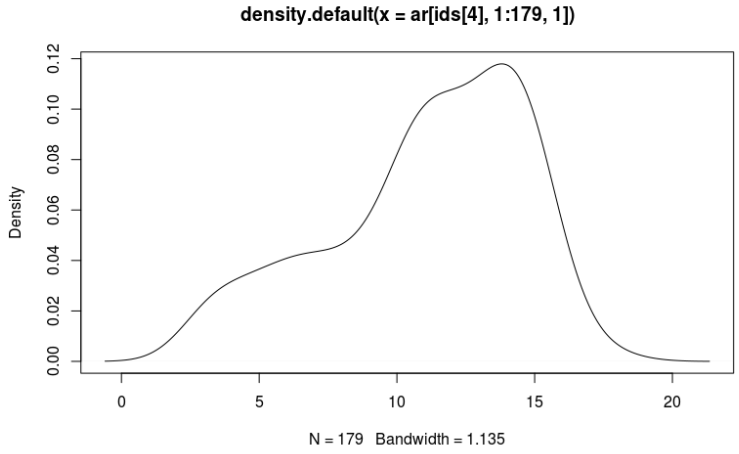

Figure 4: VM with 4th highest average CPU utilization

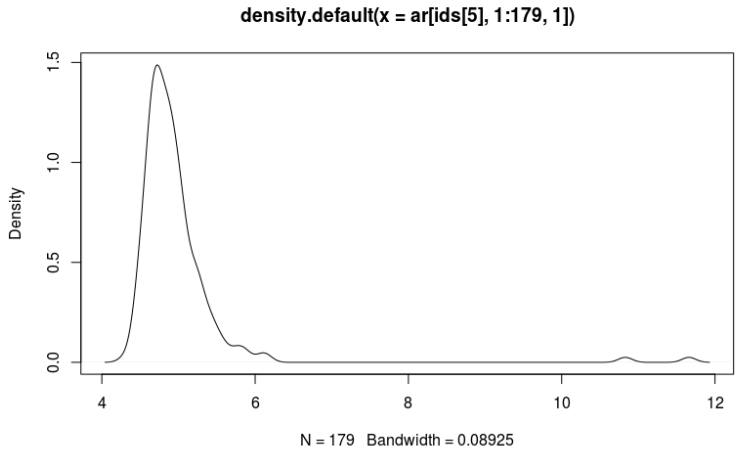

Figure 5: VM with 5th highest average CPU utilization

Task 7b. Result of plotting the 5 highest CPU utilizations variances:

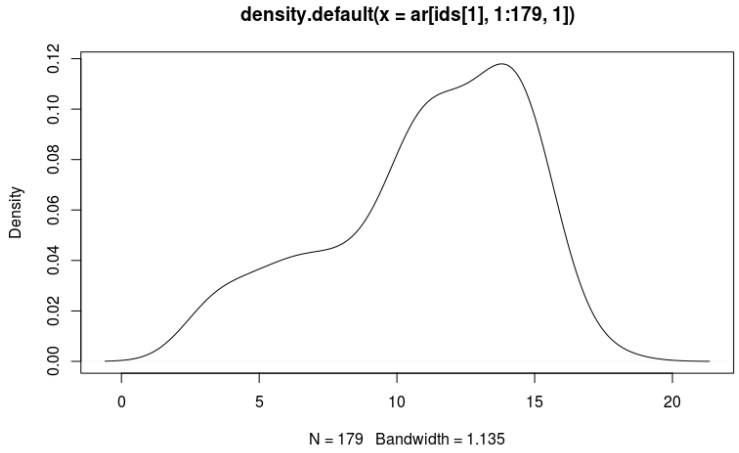

Figure 6: VM with highest CPU utilization variance

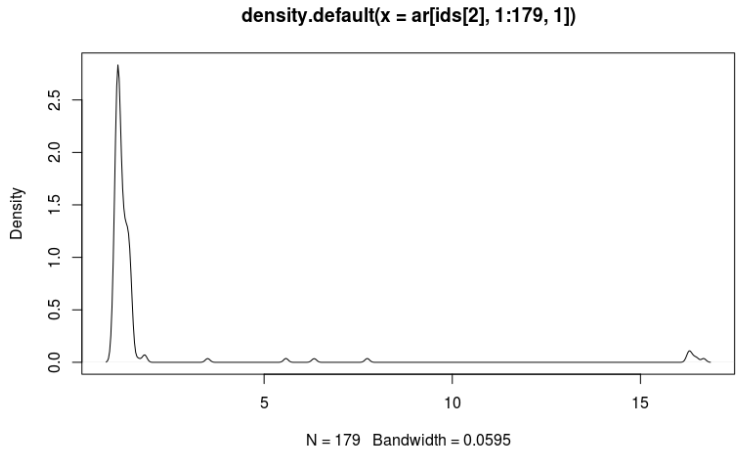

Figure 7: VM with 2nd highest CPU utilization variance

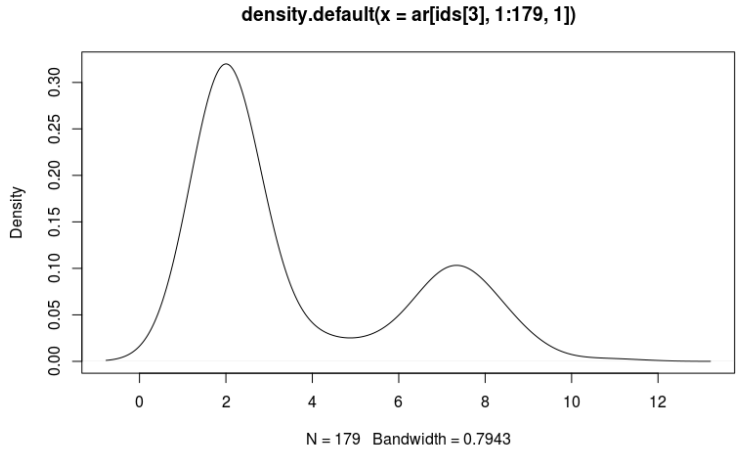

Figure 8: VM with 3rd highest CPU utilization variance

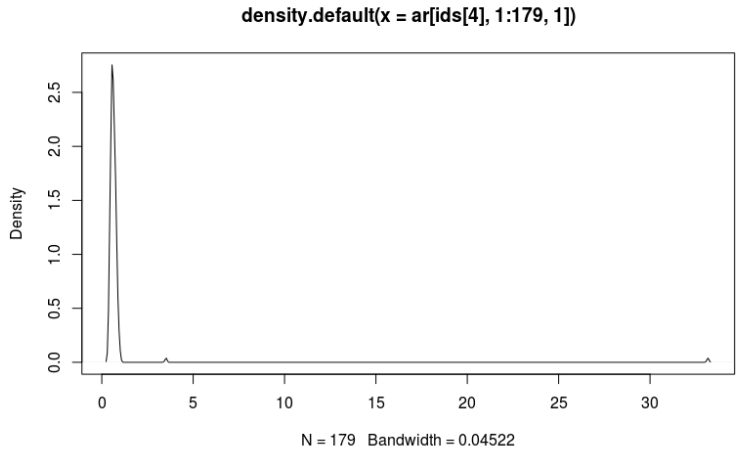

Figure 9: VM with 4th highest CPU utilization variance

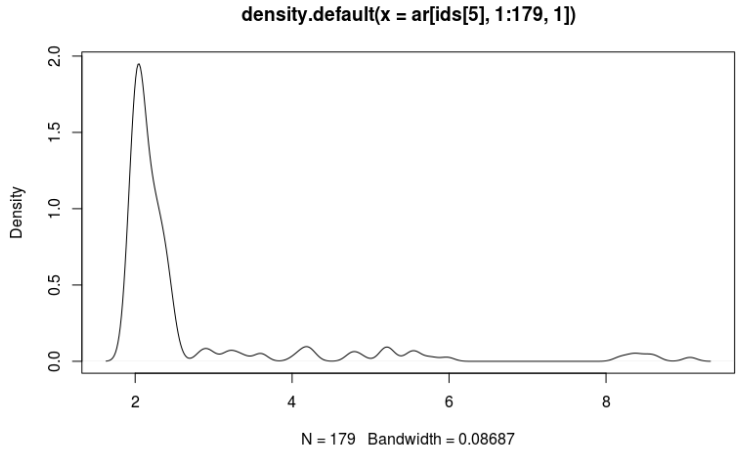

Figure 10: VM with 5th highest CPU utilization variance

Task 8. The correlation indicates strongest linear dependency of two R.V. for values near -1 or 1. 0 indicates independence. The mean and variance can be derived from the correlation of each VMs CPU and MEM usage. For the calculation NAs are ignored.

Average correlation CPU / MEM:

$$\eta_{\text{cor}} = -0.01004689 \tag{4}$$

Variance correlation CPU / MEM:

$$\sigma^2_{\text{cor}} = 0.004311276 \tag{5}$$

The data seems to show very little correlation between CPU and MEM, as most of the values are near 0. Also the variance is very low, indicating that most of the samples can be found in this region. However, the mean is slightly negative, making it seem like that a higher CPU usage correlates to a lower memory consumption and vice versa.

Nr. VM	Correlation CPU / MEM
1	0.003952292
2	-0.01962783
3	-0.1093626
4	-0.02433467
5	0.0350389
6	-0.01998985
7	-0.1350182
8	-0.03992745
9	0.03335977
10	-0.08739993
11	NA
12	-0.0825247
13	0.0005087898
14	NA
15	0.00557757
16	0.0752409
17	0.06399201
18	NA
19	-0.0147721
20	-0.01328674
21	-0.02168817
22	-0.0839365
23	NA
24	0.03863146
25	0.02552046
26	0.04091271
27	-0.02860617
28	-0.06541306
29	-0.05310957
30	NA
31	0.1354208
32	-0.09167298
33	-0.04853519
34	-0.04085905
35	0.09645776
36	0.1139977

References

[1] Eric W. Weissenstein. Covariance.