**Correlation & causation (practice in R)**

**Problem 1:** Testing for marginal correlation

The first thing we need is to load the data into RStudio:

Изображение выглядит как текст

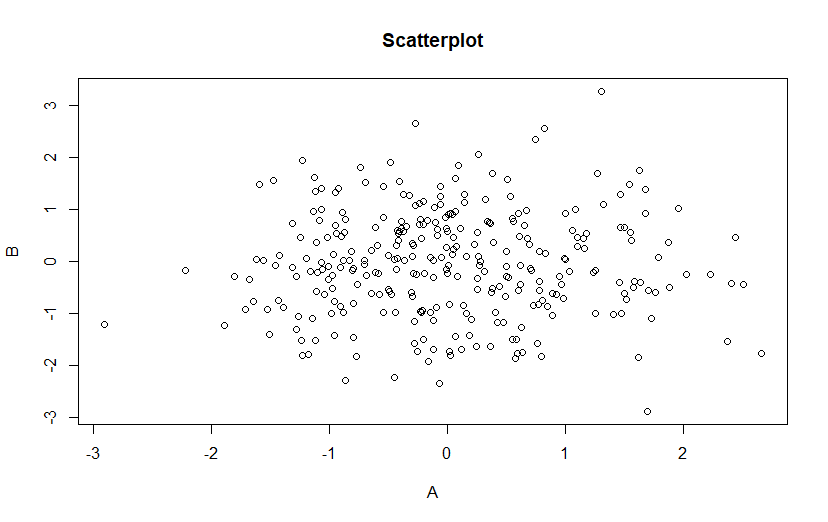
Автоматически созданное описание

The next step is to apply the attach function to the data so that A and B can be called without accessing data$. Then display observations A and B as a scatterplot:

Изображение выглядит как текст

Автоматически созданное описание

Output:



So we don't see correlation. This is consistent with the fact that in Figure 1 they are not connected.

Now we need to check the correlation between A and B:

Изображение выглядит как текст

Автоматически созданное описание

Correlation = 0.01169715, it is close to zero.

P-value = 0.840103, it is very high (usually we use 0,05).

**Conclusion:** there is no correlation.

**Problem 2: Testing for partial correlation**

To check the partial correlation of A and B we use linear regression and find the residuals:

Изображение выглядит как текст

Автоматически созданное описание

Then display residuals A ~ C and B ~ C as a scatterplot:

Изображение выглядит как текст

Автоматически созданное описание

Output:

**Изображение выглядит как диаграмма

Автоматически созданное описание**

By the look of the cloud of points, we can assume that it is slightly negatively correlated. Let's also check the correlation:

Изображение выглядит как текст

Автоматически созданное описание

Correlation = -0.3992521, it is what we expected

P-value = 6.6e-13, so we can be sure in our results

**Conclusion:** As a result, there is a slight negative correlation here. This is consistent with the fact that in Figure 1 A - C and B - C are connected.

**Problem 3: Running the PC algorithm**

To perform this task, install and download the pcalg package, as well as all that was required for further visualization:

****

Convert the data into the required format:

****

We run the algorithm for different alpha values and visualize the resulting graphs:

**Изображение выглядит как стол

Автоматически созданное описание**

Output:

Изображение выглядит как диаграмма

Автоматически созданное описание

We see that for large alpha the graph is fully connected.

With decreasing level of significance more and more edges are removed.

We also see that at alpha up to 1e-4 the connection between C and D is removed while the connection between B and E is still two-sided.

**Conclusion:** accordingly, it is not possible to achieve the same dependence as in Figure 1.