07-Correlation Between Water Quality Indicators

2022-11-30

1 Example 1: Correlation of water quality indicators of a stream

1.1 Introduction

Given are the following series of measurements from a polluted river (Ertel et al., 2012):

```
Station <- 1:17
COD   <- c(9.6, 12.2, 13.2, 13.3, 37.4, 21.4, 16.1, 24.8, 24.2, 26.9,
           29.2, 31.6, 18.2, 24.8, 13.7, 23.6, 24.2)
O2    <- c(9.8, 10, 8.3, 9.6, 1.5, 4.4, 6.3, 3, 3, 10, 9.4, 17.9, 9.7,
           9.7, 8.2, 9.5, 11.3)
NH4   <- c(0.15, 0.1, 0.74, 0.29, 5.04, 2.26, 0.96, 3.37, 2.44, 0.27,
           0.32, 0.68, 0.27, 0.32, 0.22, 0.58, 0.59)
Color <- c(0.59, 0.52, 0.6, 0.57, 1.34, 1.21, 1.17, 1.12, 1.1, 1.08,
           1.24, 1.25, 1.29, 1.29, 1.06, 1.25, 1.16)
```

with COD, the chemical oxygen demand (in mgL^{-1}), O_2 oxygen concentration (mgL^{-1}) and Color. The spectral absorption coefficient at 436nm (m^{-1}) is a measure of the color intensity of the filtered water. The data set is a subset from field measurements, that contained more samples from several measurement campaigns.

The question is whether Ammonium (NH4), Oxygen (O2) and Color depend on the organic load (COD).

1.2 Data analysis

For the dependency between Color and COD we can use Pearson correlation. In addition it is always a good idea to plot the data.

```
plot(COD, Color)
cor.test(COD, Color)
```

We see an almost linear dependency and get a highly significant correlation. Now, for the dependence between Ammonium on COD we can proceed with:

```
plot(COD, NH4)
cor.test(COD, NH4)
```

We find again significant correlation. However, the dependency is not very strict and it seems that there are two different data sets. This is in fact true because sampling sites 1–9 were before and the other sampling sites below a cooling reservoir of a power plant. In the following, we create first an empty `plot(...., type = "n")` and then add the station number as text labels:

```
plot(COD, NH4, type = "n")
text(COD, NH4, labels = Station)
```

We now repeat the analysis with the first 9 data pairs from the stations upstream of the cooling reservoir (50.19N, 24.40E, Link to Google Maps):

```
plot(COD[1:9], NH4[1:9])
cor.test(COD[1:9], NH4[1:9])
```

1.3 Exercises and discussion

Is oxygen concentration (O2) directly related to organic pollution (COD)? Interpret the results, and discuss the potential mechanisms how the cooling reservoir may influence water quality. Compare your conclusions with the paper of Ertel et al. (2012). Does the assumption of independent residuals hold?

Repeat the analysis with Spearman's correlation, and compare the results with the Pearson correlation coefficients.

2 Example 2: Examination results in statistics

2.1 Introduction

Given is a number of students that passed an examination in statistics. The examination was written two times, one time before and one time after an additional series of lectures. The values represent the number of points approached during the examination. Check whether there is a dependency between the results before and after the test, i.e. if there is any dependency between the results of the final test and the basic knowledge before the course.

2.2 Data and data analysis

```
x1 <- c(69, 77, 35, 34, 87, 45, 95, 83)
x2 <- c(100, 97, 67, 42, 75, 73, 92, 97)
cor.test(x1, x2)
```

If we plot this, we see a rather strange pattern, i.e. no clear linear relationship:

```
plot(x1, x2)
```

Therefore it may be a good idea to use the rank correlation:

```
cor.test(x1, x2, method="spearman")
```

Sometimes, we may get a warning that it "cannot compute exact p-values with ties", then we can use another approach and compute the Spearman correlation via the Pearson correlation of ranks:

```
cor.test(rank(x1), rank(x2))
```

This needs a little bit more effort (for the computer, not for us), but the interpretation is the same. To understand how this worked, it can be a good idea to create a scatterplot of the ranks of both variables.

2.3 Exercise and Discussion

What do the results above tell us? Compare the results with a paired t-test of the same data set. Which test tells what?

References

Ertel, A.-M., Lupo, A., Scheifhacken, N., Bodnarchuk, T., Manturova, O., Berendonk, T., & Petzoldt, T. (2012). Heavy load and high potential: Anthropogenic pressures and their impacts on the water quality along a lowland river (western bug, ukraine). *Environmental Earth Sciences, 65*, 1459–1473. https://doi.org/10.1007/s12665-011-1289-0