

Experiment No -1

Title of Experiment: Covariance Testing

Abstract

Covariance testing is implemented in this experiment using the following:

- Numerical Analysis
- Graphical Analysis

Introduction

The **covariance** of two variables x and y in a data set measures how the two are linearly related.

Prerequisite for experiment

In mathematics and statistics, covariance is a measure of the relationship between two random variables. The metric evaluates how much – to what extent – the variables change together. A positive covariance would indicate a positive linear relationship between the variables, and a negative covariance would indicate the opposite. The sample covariance is defined in terms of the sample means as:

$$\text{Cov}(X, Y) = \frac{\sum (X_i - \bar{X})(Y_j - \bar{Y})}{n - 1}$$

- X_i – the values of the X-variable
- Y_j – the values of the Y-variable
- \bar{X} – the mean (average) of the X-variable
- \bar{Y} – the mean (average) of the Y-variable
- n – the number of the data points

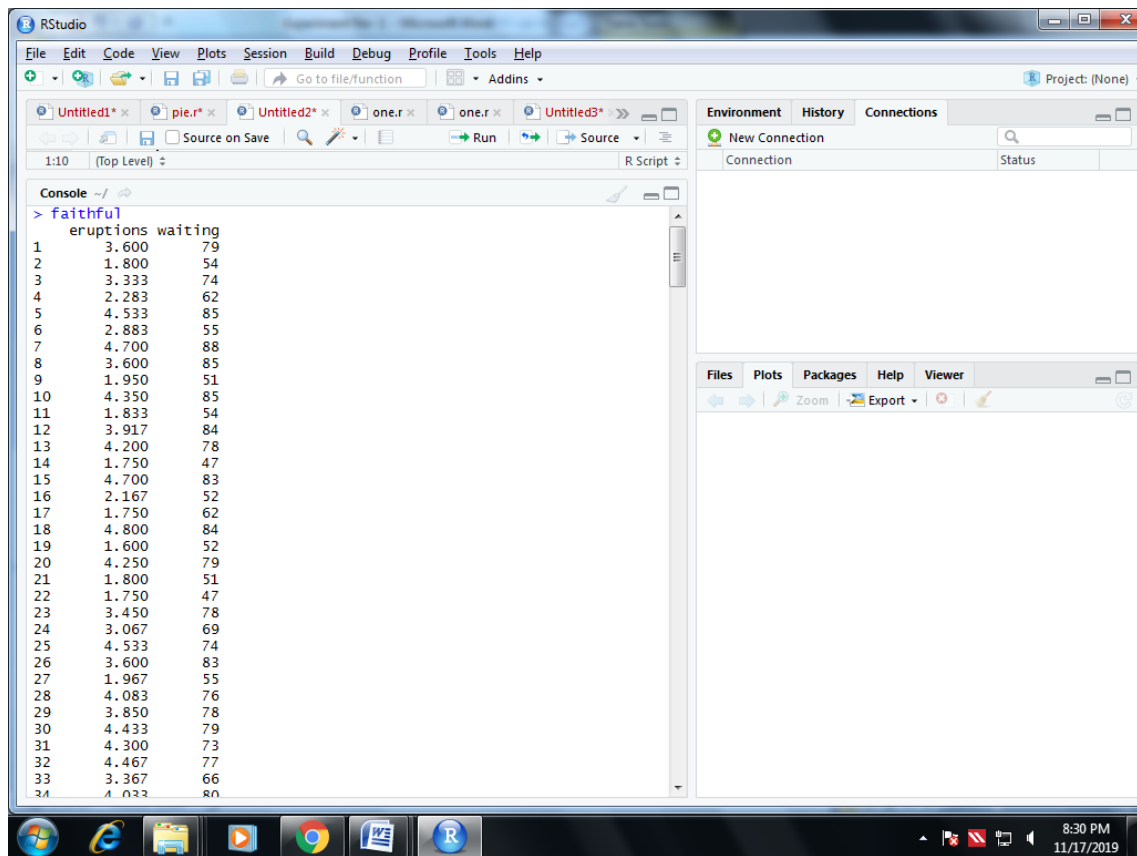
Similarly, the **population covariance** is defined in terms of the population mean μ_x, μ_y as:

$$\sigma_{xy} = \frac{1}{N} \sum_{i=1}^N (x_i - \mu_x)(y_i - \mu_y)$$

Experiment No -1

Find the covariance of eruption duration and waiting time in the data set named faithful.

Observe if there is any linear relationship between the two variables.



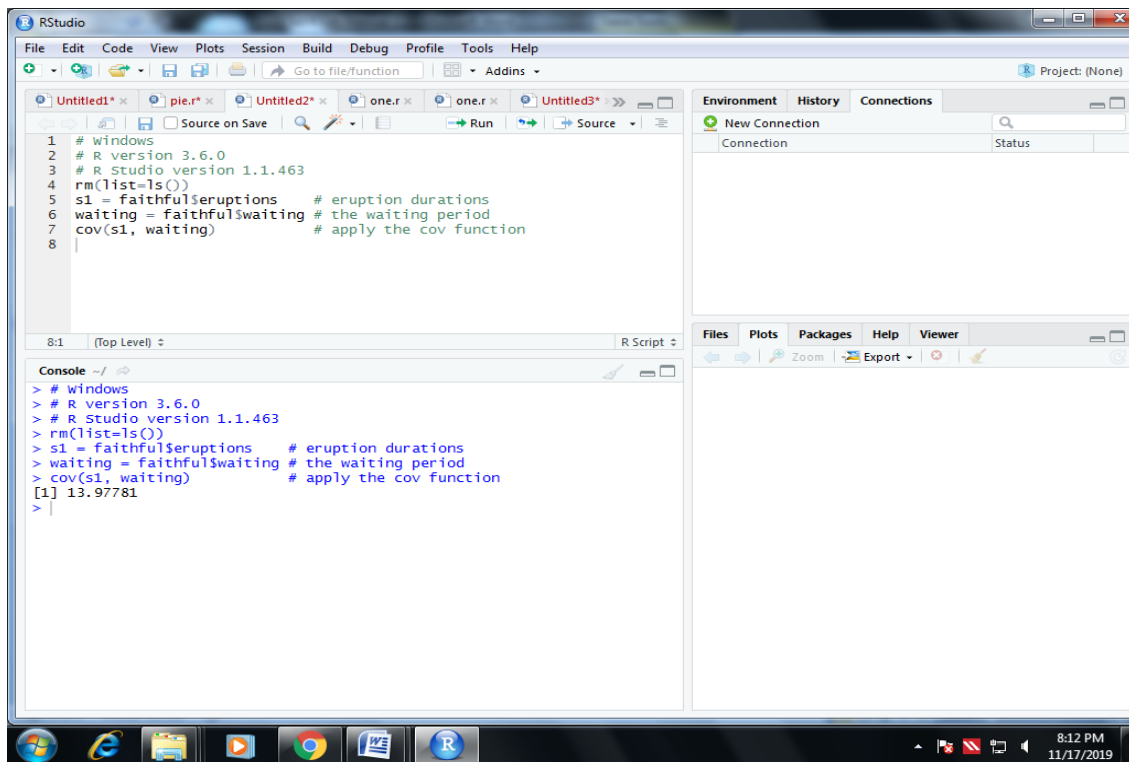
Solution

We apply the covariance function to compute the covariance of eruptions and waiting.

```
>duration = faithful$eruptions # eruption durations
> waiting = faithful$waiting   # the waiting period
>cov(duration, waiting)       # apply the cov function
```

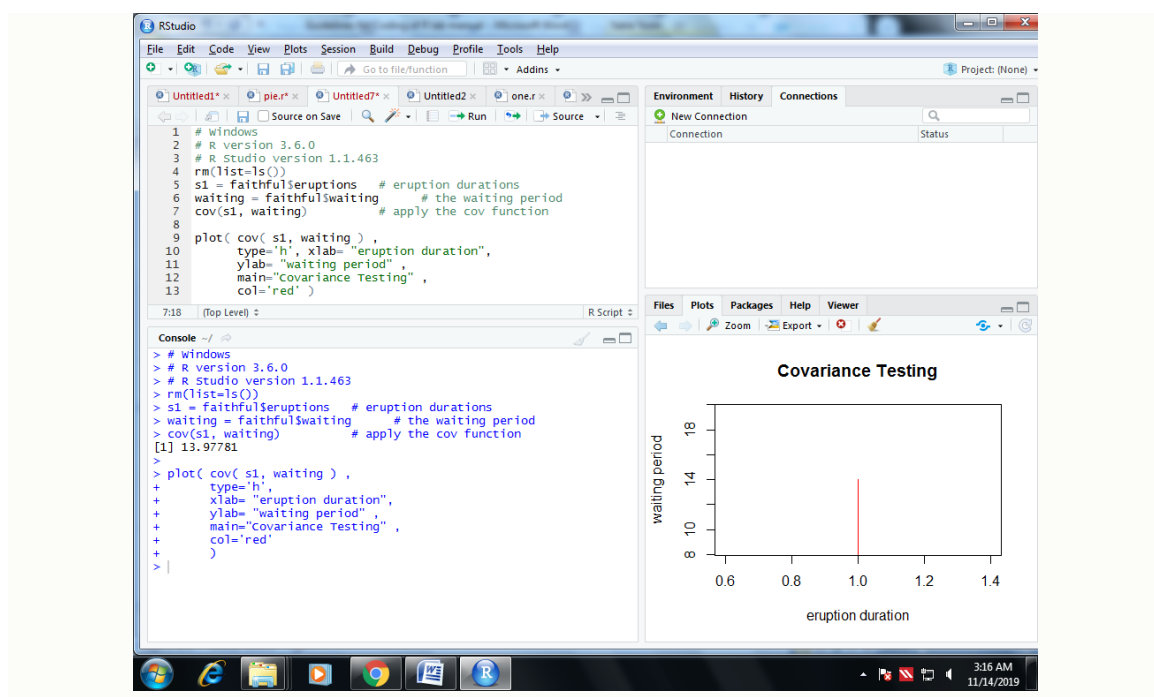
[1] 13.978

1) **Numerical Analysis with R to identify covariance between eruption and waiting time, Environment - (#windows 7 32 bit R version 3.6.1 and RStudio-1.1.463)**



Conclusion: The covariance of eruption duration and waiting time is about $13.97781 \approx 14$. It indicates a positive linear relationship between the two variables.

2) Graphical Analysis using R for eruption and waiting as two variables to identify covariance



Conclusion: In the graph above the covariance of eruption duration and waiting time is about 14. It indicates a positive linear relationship between the two variables and also proves that the analysis is consistent with the numerical value derived from the first program.