# Theoretical principle

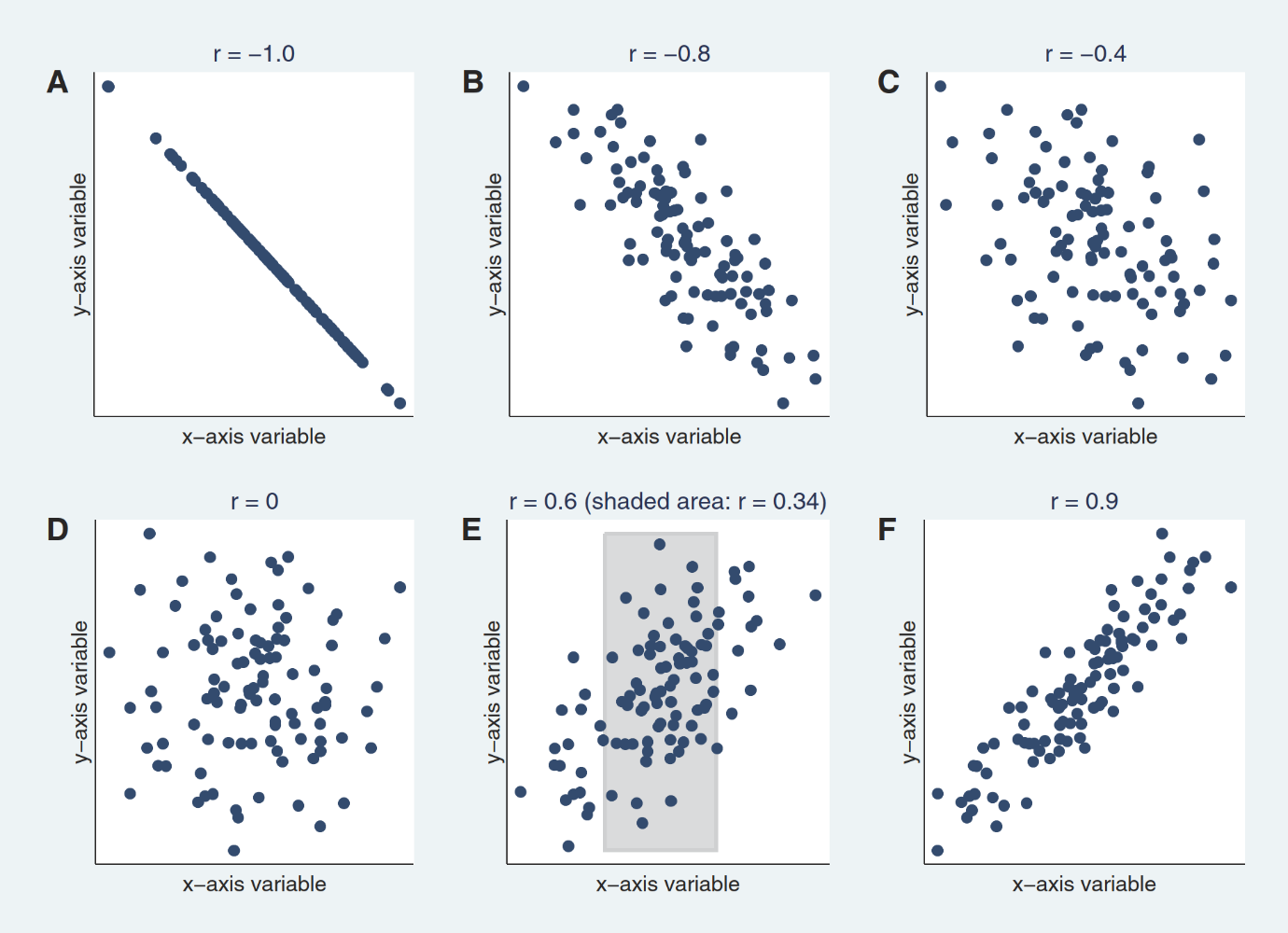
This chapter introduces the basic knowledge of the two most commonly used correlation coefficients – the Pearson coefficient and the Spearman coefficient – required to comprehend the correlation analysis performed in **chapter 4**. In the following, we focus on how they should and should not be used and interpreted.

## Correlation coefficients

Correlation is a measure of a monotonic relationship between two variables in a correlated data, where the increase of the value of one variable tend to result in either an increase (positive correlation) or a decrease (negative correlation) of the value of the other one, and vice versa. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Peason product-moment correlation

A special case of a monotonic association is a linear relationship between two variables. Most often, the term correlation is used in conjunction with such a linear relationship, known as Pearson product-moment correlation, commonly abbreviated as *r*. This coefficient is a dimensionless measure and ranges from -1 to 1. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

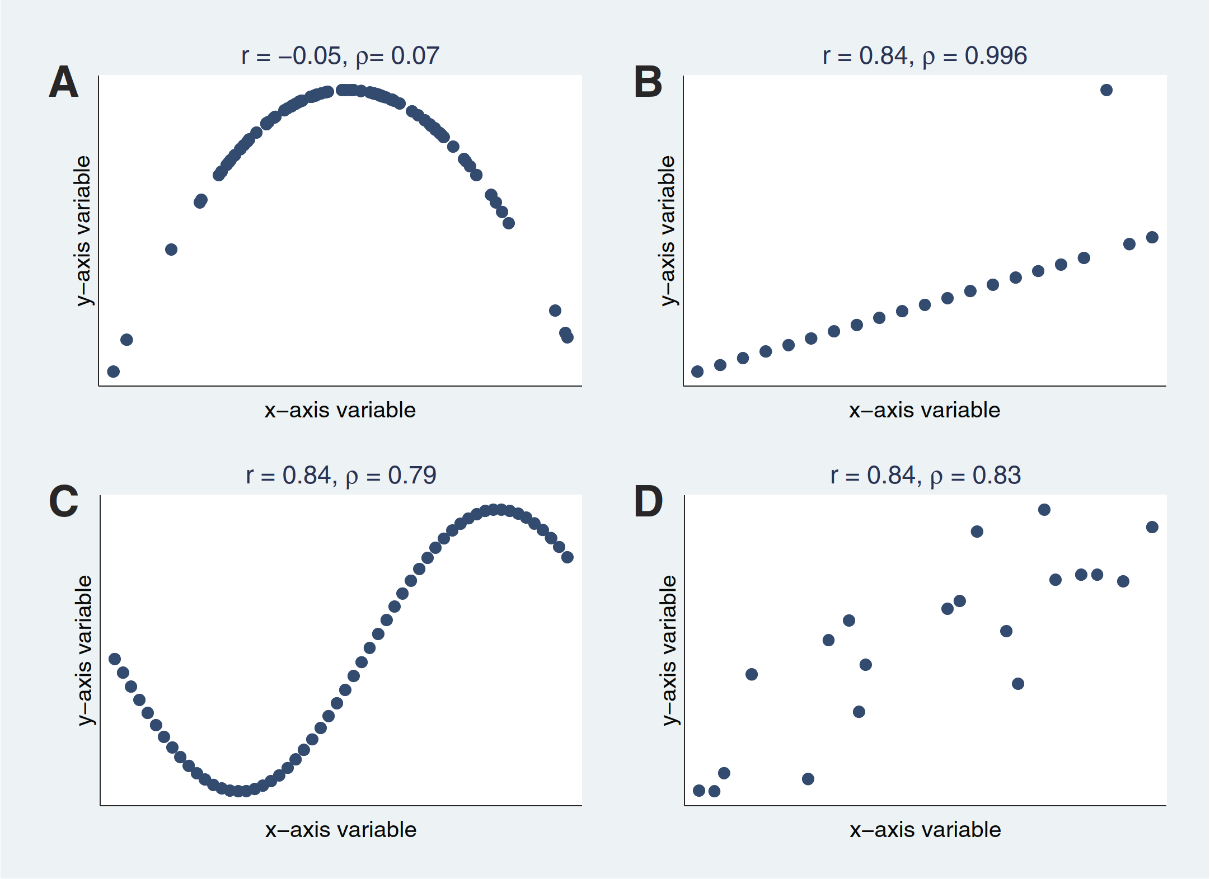
The **figure** below depicts scatterplots of sample data with different Pearson correlation coefficients.

**Figure XY** A illustrates a perfect correlation of -1. A perfect correlation of -1 or 1 implies that all the data points lie exactly on a straight line. In **Figure XY** B and F, the scatterplot approaches a straight line as the coefficient tends towards -1 or 1, whereas in **Figure XY** D there is no linear relationship, as the coefficient is 0. **Figure XY** E displays that the correlation depends on the range of the assessed value, a wider range leans towards higher correlation than the smaller range in the shaded area. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Spearman rank correlation

In contrast to a Pearson correlation, a Spearman correlation – generally abbreviated as *ρ* (rho) or *rs* – can be used to analyse nonlinear monotonic relationships. Furthermore, it is relatively robust against outliers. The Spearman correlation also ranges from -1 to 1, whereas *ρ* = 0 implies that there is no association, while *ρ* = -1 or 1 implicate a perfect correlation. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**

### Interpretation of the correlation coefficients

The scatterplots in the following **figure XY** illustrate the two correlation methods – Pearson and Spearman – on a sample dataset. Note, that the correlation coefficient should always be assessed by a visual representation of the data. For example, in **figure XY** A, both coefficients are close to 0, which connotes that there is no association between the x-axis and y-axis variables, when in fact, the plot suggests a strong quadratic relationship. Another interesting observation is, that despite the same Pearson correlation coefficient values *r* in **figures XY B through D**, the data is quite different in each of the panels. **Figure XY B** reveals, on the one hand, the robustness of the Spearman coefficient against outliers and on the other hand, its notable influence on the Pearson coefficient. In **figure XY** C, a sinusoid relationship – neither linear nor monotonic – is depicted, both correlation methods are unable to capture it. This can be further observed in **figure XY D**.

Over the course of years, several threshold values to translate a correlation coefficient into descriptors such as “weak”, “moderate” or “strong” relationship – which are arbitrary and inconsistent – have been proposed. While most researchers would agree that a correlation less than 0.1 indicates a negligible and one greater than 0.9 a strong relationship, values in between are disputable and therefore should be interpreted within the context of the posed research question. **(source: CorrelationCoefficients-AppropriateUseandInterpretation)**