

Part II

Distributions and Hypothesis Tests

This part of the book moves the focus from *Python* to statistics.

The first chapter serves to define the statistical basics, like the concepts of *populations* and *samples*, and of *probability distributions*. It also includes a short overview of *study design*. The design of statistical studies is seriously underestimated by most beginning researchers: faulty study design produces garbage data, and the best analysis cannot remedy those problems (“Garbage in—garbage out”). However, if the study design is good, but the analysis faulty, the situation can be fixed with a new analysis, which typically takes much less time than an entirely new study.

The next chapter shows how to characterize the position and the variability of a distribution, and then uses the normal distribution to describe the most important *Python* methods common to all distribution functions. After that, the most important discrete and continuous distributions are presented.

The third chapter in this part first describes a typical workflow in the analysis of statistical data. Then the concept of *hypothesis tests* is explained, as well as the different types of errors, and common concepts like *sensitivity* and *specificity*.

The remaining chapters explain the most important hypothesis tests, for continuous variables and for categorical variables. A separate chapter is dedicated to survival analysis (which also encompasses the statistical characterization of material failures and machine breakdowns), as this question requires a somewhat different approach than the other hypothesis tests presented here. Each of these chapters also includes working *Python* sample code (including the required data) for each of the tests presented. This should make it easy to implement the tests for different data sets.