Part III
Statistical Modeling

While hypothesis tests can decide if two or more sets of data samples come from the same population or from different ones, they cannot quantify the strength of a relationship between two or more variables. This question, which also includes the quantitative prediction of variables, is addressed in the third part of this book. The basic algebraic tools that come with *Python* may suffice for simple problems like line-fits or the determination of correlation coefficients. But a number of packages significantly extend the power of *Python* for statistical data analysis and modeling. This part will show applications of the following packages:

- statsmodels
- PyMC
- scikit-learn
- scikits.bootstrap

In addition, a very short introduction to generalized linear models is included. The section on logistic regression has also been placed in this part, as logistic regression is a generalized linear model. An introduction to Bayesian statistics, including a practical example of a running Markov-chain–Monte-Carlo simulation, rounds off the chapter.