

CHAPTER 4

STATISTICS

Sections

4.1. Sample Statistics (program <code>uvs</code>)	132
4.2. Discrete-Data Histograms (program <code>ddh</code>)	148
4.3. Continuous-Data Histograms (program <code>cdh</code>)	159
4.4. Correlation (programs <code>bvs</code> and <code>acs</code>)	172

The first three sections in this chapter consider the computation and interpretation of univariate-data sample statistics. If the size of the sample is small then essentially all that can be done is compute the sample mean and standard deviation. Welford’s algorithm for computing these quantities is presented in Section 4.1. In addition, Chebyshev’s inequality is derived and used to illustrate how the sample mean and standard deviation are related to the distribution of the data in the sample.

If the size of the sample is not small, then a sample data histogram can be computed and used to analyze the distribution of the data in the sample. Two algorithms are presented in Section 4.2 for computing a discrete-data histogram. A variety of examples are presented as well. Similarly, a continuous-data histogram algorithm and a variety of examples are presented in Section 4.3. In both of these sections concerning histograms, the empirical cumulative distribution function is presented as an alternative way to graphically present a data set. The fourth section deals with the computation and interpretation of paired-data sample statistics, in both paired-correlation and auto-correlation applications. Correlation is apparent in many discrete-event simulations. The wait times of adjacent jobs in a single-server node, for example, tend to be above their means and below their means together, that is, they are “positively” correlated.