

Understanding Data Step Processing

Data step processing consists of 2 phases, the compilation phase and the execution phase. With simple examples it is shown, how the Program Data Vector (PDV) looks like during these phases when running a set or merge statement. Understanding the way SAS works helps to prevent logical errors. Especially vertical dataset concatenations, multiple sets and many to many relationships are explained, as SAS results seem sometimes not reproducible in the first moment. In addition, the lag and diff functions will be discussed against the background of conditions (if or else).