

Statistics 243: *class notes*

William J. De Meo

October 27, 1997

1 Statistical Packages

1.1 Procedure Oriented

The highest level are Procedure Oriented Statistical Packages; e.g. SAS, BMDP, SPSS, systat. The downside of these programs is that they don't offer much flexibility, so if there is a method that they don't do, it is often very inefficient to implement it. The upside is that, if they happen to do what you want, they probably do it in the best way possible.

1.2 Programming Environment

Programming environments are interactive programs which allow you to manipulate your data on the fly, since it stores data in fast memory. Examples are Splus, gauss, blss, matlab, stata. When you have a very large data set, these programs are almost useless. If your job is taking very long, run the `top` command. If it says that your job is only using 2.4% of the cpu, it means that your job is taking long because there is too much data, and it is swapping from disk. One technique you could use to circumvent this is to try to get your programming environment to read one observation at a time.

1.3 Subroutine Libraries

The next best alternative is to use subroutine libraries.

1.4 C, Fortran, C++, java

2 S

2.1 Extensibility

Users can write first class functions, which are functions which look identical to those that come with the language. You can also link your own C, C++ or FORTRAN programs.

2.2 Books

Statistical Models in S (The White book), Chambers & Hastie.

Modern Applied Statistics using Splus (The Yellow Book), Venables & Ripley.