Fitting Statistical Models in Julia

Douglas Bates

06/26/2014

Characteristics of Statistical Computing

- It is widely used. Long ago university mainframes spent most of their CPU time in programs like SPSS, SAS, and BMDP.
- Hot topics like Bioinformatics, Analytics (whatever that is) and Big Data have spurred even more interest.
- There is a wide range of abilities, both mathematical and computational, in those who use such software.
- It is not uncommon for mathphobes in social and life sciences to find themselves taking a required statistics course.
- Researchers in many disciplines are expected to provide a statistical analysis of their results, which far too often means "a p-value $< 5\%$".

The "there is only one formula" phenomenon

- Statistics is often taught as rote application of formulas from some text.
- The concept that there is a model behind the formula is often never mentioned.
- This leads to the conviction that there is only one possible way of evaluating the result.
- It is well-known that the only possible way to evaluate regression coefficients is

$$\widehat{\beta} = (X'X)^{-1} X'y$$