

TECHNICAL REPORT 2002/125

---

Model Classes

*L Allison*

ABSTRACT

Statistical models are used, explicitly or implicitly, in artificial intelligence, data mining, machine learning and statistics (of course), and in specific application areas. Much associated research is about devising a new kind of model, finding a better optimization algorithm for a model, or applying a model to a problem area. This paper instead examines the semantics of statistical models - what they are precisely, how they behave in *general*, and how they can be combined.

The programming language Haskell-98, with its polymorphic type system, is the tool for the exercise so one of the products is a running program. Data types and type classes, that allow models to be manipulated in a type-safe yet flexible way are developed for probability distributions, function-models (regressions), time-series, estimators, unsupervised classification, mixture models, supervised classification, and classification trees. It is shown that there are short and convenient conversion functions between various kinds of models and between their estimators. The result is a collection of useful tools in Haskell and, perhaps more importantly, the start of a theory of programming with models.