ADAPTATION OF ‘EXPRESS’ TO THE IBM PC: 
A TOOL FOR BUILDING KNOWLEDGE-BASED STATISTICAL SYSTEMS USING EXISTING PACKAGES

I. Heuch, J. H. Aarseth, G. Ottersen, F. Carlsen
Department of Mathematics, University of Bergen
Allégt. 55, N-5007 Bergen, Norway

The program system EXPRESS constitutes a tool for constructing knowledge-based statistical systems requiring repeated cycles of statistical analysis on given data sets. It makes it possible to specify chains of rules in such a way that intermediate results determined by execution of standard packages may affect the type of analysis to be carried out in subsequent steps. These intermediate results are assumed to occur in the ordinary output file produced by the packages involved, and they are stored in a special knowledge base maintained by EXPRESS. If the rules refer to statistical quantities which have not yet been found for the data sets in question, the system will automatically generate control language for the appropriate external package, initiate execution of this package, scan the output, and store the relevant results in the knowledge base.

EXPRESS was originally constructed for use on a main-frame computer, with rules built into the system program code itself (Carlsen & Heuch, 1986). The present version is modified substantially and allows for rules being specified in separate subprograms by the knowledge engineer. In addition, the system has been adapted to run under PC-DOS. It incorporates a number of features facilitating a more interactive usage.

At entry to EXPRESS, the user is presented with a menu providing a choice between different sets of rules applicable to various general statistical situations. For each rule set, the user may decide to attack different questions relating to the data, corresponding to separate levels of complexity in the analyses involved. The system then uses backward chaining in order to answer the question selected. Depending on system options, which may be changed during the session, brief or comprehensive explanations are given of the sequence of rules invoked. If any rule refers to a quantity that should be found from a statistical package, the user may watch how the system sets up the suitable control language, and how the output is subsequently searched for the information desired, after the external package has been run by EXPRESS. At
any time, the user may inquire about the values of parameters stored in the knowledge base, or set the values of such parameters.

New sets of rules may be introduced by the knowledge engineer by writing particular subprograms returning parameter values to EXPRESS, conforming to simple specifications. In this way additional rules may be compiled and incorporated into the main program. In the applications considered until now, rules have been specified using Fortran, but in principle any language can be used which permits subprograms to be written that can be linked with the compiled version of EXPRESS. This way of formulating rules contrasts somewhat with solutions used in other knowledge-based systems utilizing external statistical packages. Thus GLIMPSE, relying on the statistical system GLIM, uses PROLOG for specification of rules (Wolstenholme & Nelder, 1986). In TAXSY (Darius, 1986), rules are formulated in the language of SAS, the actual package applied in the statistical computations. Other systems, such as PRINCE (Duijsens et al., 1988), require rules to be written in a special file under a particular coding system. We have found that writing rules as ordinary subprograms provides the knowledge engineer with a great freedom in handling numerical quantities. At the same time rules may be specified in a very flexible way, taking advantage of techniques used in standard programming.

In the applications studied thus far, the statistical packages utilized have included SAS, BMDP, SPSS and Minitab. The application area considered most extensively has been two-sample tests, but it is also our intention to explore applications in connection with estimation of relative risk in various situations in epidemiology.

**References**


