

Strategic roadmap for the IT support of the analysis and interpretation of data in drug discovery

Franky De Cooman, Business & Decision, Brussels, Belgium
Rudi Verbeeck, Johnson & Johnson Pharmaceutical Research & Development,
Beerse, Belgium

ABSTRACT

In many fields of pharmaceutical research statistics is only used to provide p-values. A fully data analytical approach, however, goes much further. Statistics can be integrated in the entire procedure to investigate scientific questions. When statisticians and other data analysts are involved in the planning of experiments, the data analytical part often becomes less complicated. In this paper we present an example of an automated statistical analysis resulting from this tight integration of discovery research, statistics and IT support. It gives the scientists the opportunity to perform their own data analysis because, after all, they know their data best.

This vision of integration of statistics and discovery research, facilitated by IT support, has strong impact on the researcher and the statistician. (1) We should make sure that the researcher's knowledge of statistics is sufficient to do basic analyses by offering a training path. These statistical courses highlight the pitfalls and explain the assumptions of the algorithms. We provide powerful yet intuitive tools like SAS/JMP and help-desk support so that scientists can analyse their own experimental data. (2) If statisticians can be freed up of routine analyses for the researcher, they could concentrate on designing new, more complex statistical solutions using advanced SAS tools (3) The statistical algorithms should be generalised and made robust, so that it can be applied to whatever datasets/variables on which the assumptions of the statistical method hold. It should also have an automatic model-building algorithm in order to provide the best model for the data.

The principle is illustrated on an ANOVA algorithm example. This algorithm is grafted into a web interface that is used by the scientist to fill in the parameters. The algorithm returns the result into a PDF file or exports the results to data visualisation applications. Alternatively, the algorithm could be made available through a web API to be included in different applications.

INTRODUCTION

Scientists know their data best, and it seems to be the best that they can analyze their own data. Instead of asking each time again to the statistician to analyze their data, it is more efficient if the researchers can do the analysis and interpretation themselves, without demanding too heavy statistical knowledge.

ROUTES TO HAPPINESS

Statisticians are rare specimens, and asking them each time again to do the same analysis is not a good idea... Asking the researchers to develop their own algorithms is not a good idea neither... What follows are three possibilities to help everybody out...

RESEARCHERS DO IT ALL BY THEMSELVES

Each time a researcher has gathered data, he runs his own analysis using SAS, S-plus or R... This frees the statistician from cumbersome work, and the researcher doesn't have to wait until the statistician can free up some time. Off course, in this solution, the researcher will have to learn statistics in depth, and has to learn the statistical software tools. This also has huge risks, because there might be some statistical pitfalls the researcher is not aware of...

STATISTICIANS DO THE ANALYSIS EACH TIME AGAIN

Each time the researcher has gathered data, he gives them to the statistician, who will analyse the data (again & again) and gives the results to the researcher.

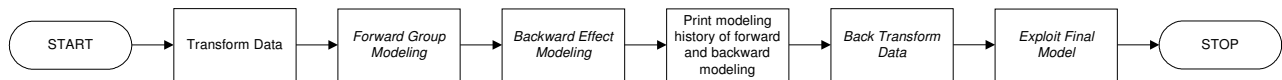
STATISTICAL PROGRAMMER AUTOMATES STATISTICIANS' WORK

The core statistical program is made by a statistician. This program is handed over to a statistical programmer, automating the program. This program is then grafted into a web-interface, enabling the researcher to call his statistical routine by sampling filling in some parameters.

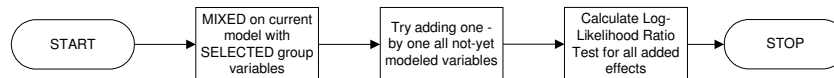
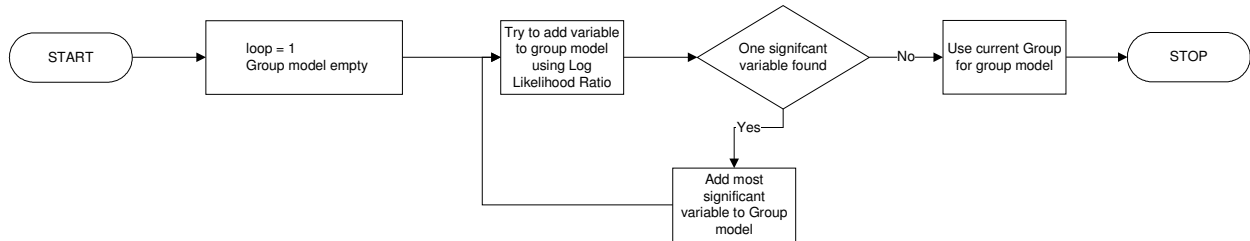
EXAMPLE OF A WEB-INTERFACED STATISTICAL PROGRAM

The principle is shown live by the means of an Anova example. In SAS, one can do automatic model building for a regression problem using the built-in feature. However, this does not exist for an Analysis Of Variance. Using some common statistical knowledge and some nifty Macro programming, this can be accomplished.

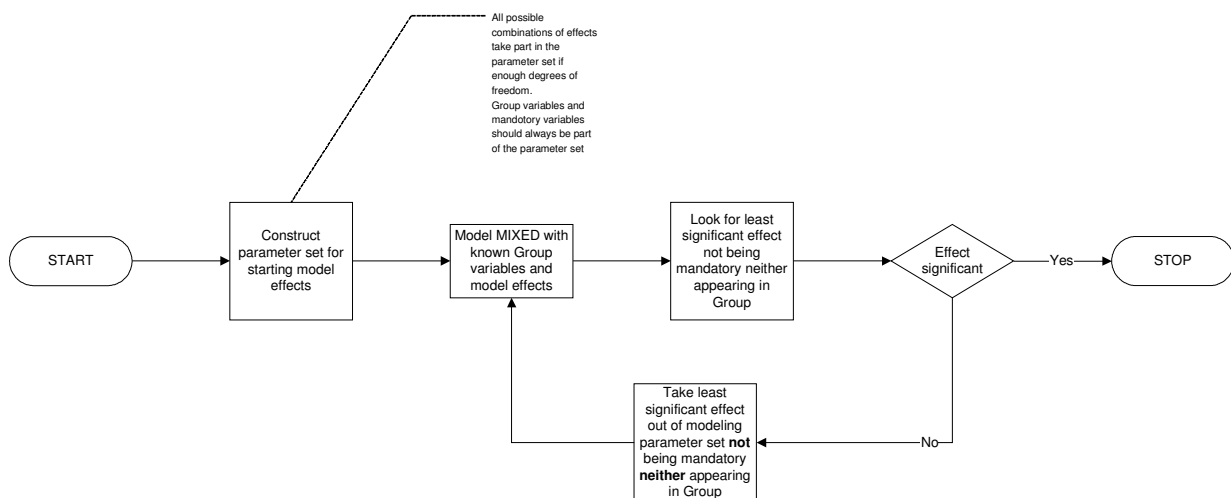
OVERVIEW OF PROGRAM



GROUP MODELING



EFFECT MODELING



CONCLUSION

Automating statistical programs is not an easy job, but it is worth the effort, especially when the routine is to be used on a regular time base. It avoids the statistician to do repetitive work, so he can concentrate on new statistical novelties.

ACKNOWLEDGMENTS

- Luc Bijmens (J&J PRD)
- Luc Wouters (Barrier Therapeutics)
- Luc Bernolet (LACO NV)
- LUC (Limburgs Universitair Centrum, now Universiteit Hasselt)

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Franky De Cooman
Business & Decision Benelux
Omwentelingsstraat 8
1000 Brussels
02/510 05 40
Franky.decooman@businessdecision.com
www.businessdecision.be

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.