

## USE OF VIRTUAL EXPERIMENTS IN TEACHING DESIGN AND ANALYSIS OF EXPERIMENTS

Paul Darius, Eddie Schrevens

K. Univ. Leuven, Belgium

Kenneth Portier

University of Florida - Gainesville, United States

paul.darius@biw.kuleuven.be

*The ability to design experiments in an appropriate and efficient way is an important skill, but students typically have little opportunity to get experience. Most textbooks introduce standard general-purpose designs, and then proceed with the analysis of data already collected. In this paper we explore a tool for gaining design experience: computer based virtual experiments. These are software environments which mimic a real situation of interest, and invite the user to collect data to answer a research question. The following prototype environments will be described: an industrial process that must be optimized, a greenhouse experiment to compare the effect of different treatments on plant growth, and an arcade style applet that illustrates the use of t-tests, regression and analysis of variance. These environments are parts of a collection called env2exp, and can be freely used over the web. They have been used in several courses over the last two years.*

### INTRODUCTION

All programs leading to a degree in statistics contain an extensive course in experimental design. Many other courses, especially service courses for students in other disciplines also teach experimental design. It is quite difficult to include material in those courses that actually trains the student in setting up experiments. Most of the courses will introduce the basic concepts of design and then go through standard general purpose designs. Often there is more emphasis on analysis of data already collected than on the actual design process. Exercises are typically of a form where the data are supplied, with little information about the reasons that led to this particular way of obtaining data. Classical textbook exercises seem to be a poor medium to train students' skills in setting up an experiment.

Several textbooks (e.g., Cobb, 2002; Dean and Voss, 1998) recognized this, and have stressed the importance of including projects into the course, in which students have to prepare, perform and analyze a real experiment. Such projects are extremely valuable, but they are very time and resource consuming, for the student as well as for the teacher. For classes with a large number of students, they are practically infeasible.

We have been looking at alternative ways to give students the opportunity to gain experience in designing an experiment. We developed "virtual experiments": software environments which mimic a real situation of interest, pose a research question, then invite the user to collect associated data which, when statistically analyzed, will shed light on the research question. The data are generated by an underlying stochastic model that is invisible to the user and that incorporates the complexities of the real situation. Once the data are generated, the student can transfer it to a standard statistical package for the analysis. Faced with increasingly more complex scenarios, the user develops his/her design skills and improvement can be assessed by relating the actual design strategies and statistical analysis used to the true underlying model and the study objectives.

We have experimented with a number of virtual experiments and a collection of sorts, currently referred to as *ENV2EXP* (for Environments to Experiment, the new name for the collection previously called *VIRTEX*) is available on the web at <http://www.kuleuven.be/ucs/env2exp>. In this paper we will discuss three of the applets in detail, but the interested reader may wish to examine the complete collection (see also Darius *et al.*, 2003). A companion collection of programs referred to as *VESTAC* (Darius *et al.*, 2000, and at <http://www.kuleuven.be/ucs/java>) illustrate many associated statistical concepts. All elements of *ENV2EXP* and *VESTAC* are JAVA applets that can be run on any computer configured with a web browser that supports JAVA. Note that like all JAVA applets, they cannot read or write data

from or to the user's hard disk and hence any transfer of numeric information from the applet to the user's computer must be via a cut and paste operation.

### THE FACTORY APPLET

This applet introduces a typical industrial situation: a process running in a production plant must be optimized through experimentation on a scaled-down version of the process, running in a pilot plant. In Figure 1 the pilot plant is on the left, the production plant on the right. The process has 3 parameters, for which better settings have to be found than those currently in use at the production plant. To accomplish this, the user can do experiments on the pilot plant.

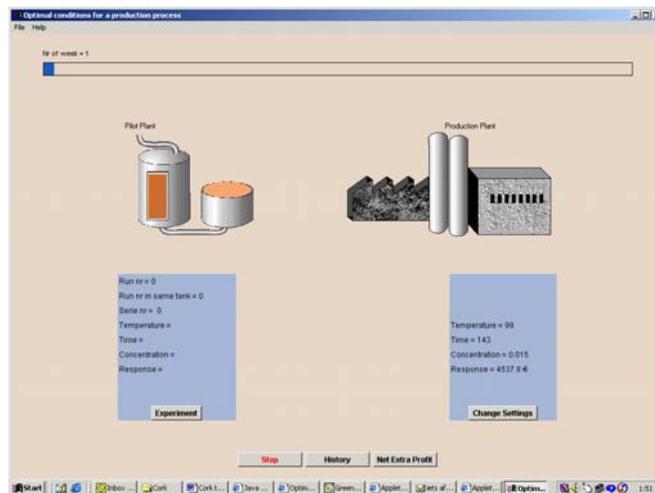


Figure 1: The FACTORY applet

The whole applet runs in “real” time: the user has 39 “weeks” (one “week” being about 3 minutes of real time) to complete the experimentation. Each experimental run takes several “days.”

To set up an experiment, the user opens an Experiment window on the pilot plant. He/she can then specify the settings of the 3 parameters to be used in any number of runs (to make things a bit more complex, the pilot plant works with raw material delivered in batches, one batch allowing 10 runs). After some time, the results can be consulted in the History window. The windows' contents can be copy/pasted into any statistical package for analysis.

The user can then continue with a new experiment or decide that he/she knows enough to reset the parameters in the production plant. This results in a cost for halting the plant, but hopefully also in an increased yield in the future. The result of the new settings is only available after 6 “weeks.”

The purpose of the whole exercise is to obtain a maximal profit over a period of 39 weeks. The costs are an initial cost, a fixed cost per experimental run with the pilot plant, and a fixed cost for each production plant process change. The benefit is the increase in yield in the production plant, multiplied by the number of weeks it was obtained.

The underlying model for this applet is a model for a particular fat-processing operation that has been in actual use in an industrial context. Additional variation (random noise and other types) have been added, resulting in different results if the same settings are run twice or on different occasions.

This applet invites the user to gain experience with many of the experimental designs covered in a Design of Experiments course. Fractional factorial or other screening designs, full factorial designs with or without center points, second order designs such as Box-Behnken or central composite designs can all be used. In addition, the batch constraint calls for the use of a blocking variable. But the user can also experience what happens with intuitive experimentation, a one factor at a time approach, or non-standard self-invented designs. He/she also will have to deal with an issue not covered in most courses: should one set up one (or a few) large designs, or

go for many designs with fewer runs and possibly a smaller experimental region. A final problem is the amount of information needed before deciding to change the production plant settings.

### THE GREENHOUSE APPLET

Here a greenhouse is used as the environment within which the user has to grow tomato plants. The greenhouse environment contains many sources of diversity in raw materials and experimental circumstances which make setting up an appropriate experiment far from trivial.

The purpose of the experiment is to find the optimal dose of a new fertilizer for use with tomato plants. Therefore the user has to set up an experiment, choosing plants and applying several fertilizer treatments, let the plants grow and examine the final result.

An initial view is shown in Figure 2. The area in the middle is the table where the plants are to be put to let them grow. The table is bordered on the left and the right with heating devices (the thick black vertical lines). The table is also lighted in a non-uniform way. There are two lighting schemes available: either a long horizontal lamp along the middle of the table, or a set of four bulbs the latter is shown in Figure 2). There is room on the table to place at most 144 plants.

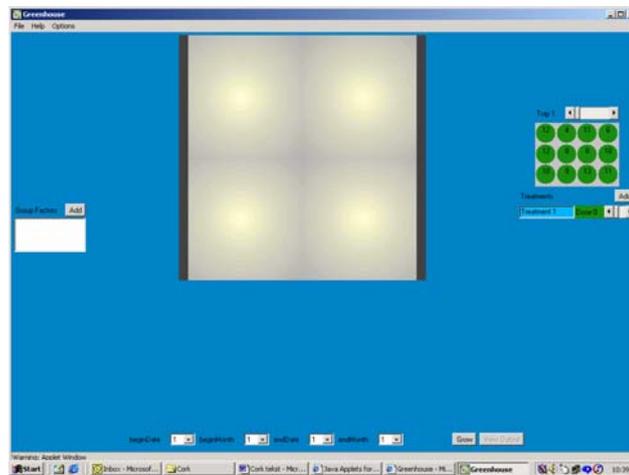


Figure 2: The GREENHOUSE applet - initial view

The available material consists of 144 young tomato plants (12 trays of 3 × 4 plants). They are shown as the circles on the right part of Figure 2. The young plants are not all the same: the initial weight is shown on the screen. The plants can be transferred to a position on the table with a simple mouse drag.

With the buttons below the young plants on Figure 2 the user can define the treatments he/she wants to apply. When this is done, each plant participating in the experiment should receive one of the treatments. This is again done with a simple mouse operation, and the color of the plant then reflects the treatment it has received.

It is also possible to define “grouping” factors to account for the differences in locations and initial weights. This is done with the buttons on the left side in Figure 2, and results in pie-parts being placed on the circle.

Figure 3 shows an example of a simple experiment. Nine plants were selected and placed on the table in 3 rows of 3. Here, only plants with initial weight 10 were selected and the light effect was ignored. Three doses (control, 4 and 8) were selected, and the plants grouped in 3 blocks (the columns) to account for the heat gradient. The color of the circle shows which treatment was applied, the color of the pie-part to which group (column) it belongs.

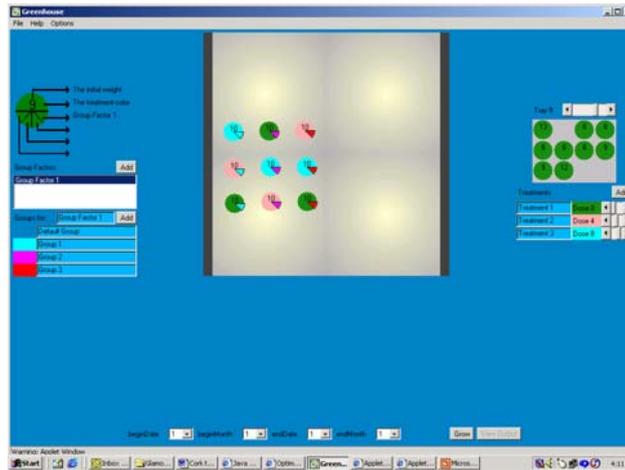


Figure 3: The GREENHOUSE applet - a simple experimental setup

When the plants have been properly placed, the user should select the time period for the plants to grow with the buttons at the bottom. When the “Grow” button is hit, the growth of each plant is simulated on an hour-to-hour basis, and the final weight (along with all the other variables) is available in the “Output” window, from where it can be copied and pasted into an external program for statistical analysis.

The underlying model for the growth simulation is an adapted version of TOMGRO, a well-known growth model for tomatoes (Jones *et al.*, 1991), and uses standard climatic data for the time period selected.

Since setting up a large experiment can be time consuming, a feature is provided that allows to save the current setup to disk (through copy/paste) and to restore a saved setup in a future session.

Note that in this applet the user has to make a substantial number of design decisions before he/she can obtain data:

- How many of the available plants should be used, and which ones? (select randomly (how?) or base the selection on the initial weight, ignore weight, block for weight or use weight as a covariate,...)
- Where should the plants be positioned on the table? (block for or ignore light, heating or both, how many blocks, how to block in the light bulbs lighting scheme,...)
- Which treatments should be used?
- How to assign the treatments to the plants?

These are the type of decisions an experimenter has to make, but they are completely ignored when the student is only given textbook-type exercises where the data are already available and all the decisions were already made.

This applet allows the user to get comparative experience with almost all classical designs (Box *et al.*, 2005; Montgomery, 2004): completely randomized, complete or incomplete block, Latin Square, etc. There is also ample opportunity to invent and use new setups, made to accommodate the specific features of the greenhouse situation.

The greenhouse applet has been used by the authors to teach experimental design to a broad audience of graduate students over the last two years. In particular, we have assessed the utility of this tool in an experimental design course for agricultural graduate students and also in courses for students in food science and bioinformatics. In an exercise students were required to generate an experiment with the applet, export the data to a statistical package, in this case *R* (*R* Development Core Team, 2005), perform the analysis and write a short report. Following the exercise, students were asked to rate the following statements:

1. With these exercises, I learned things of which I didn't realize the importance after the classroom teaching and written exercises.
2. Obtaining the data with the applets was more difficult than analyzing it with the statistical software.

3. The applet shows a reasonable image of a real situation as it occurs in practice.
4. The applet is sufficiently user friendly: it was not very difficult to specify the experiments precisely as I wanted.
5. The exercise as a whole (including the report) required too much of my time, compared to the things I learned from it.
6. The experimental situation shown in the applet was needlessly complex.

Overall, students agreed that the applet provided a realistic situation that allowed them to gain insight beyond the classroom lectures and classical exercises, and in a user friendly way. The answers to questions 2 and 5 were more neutral with wide ranges. This was probably caused by the fact that some students encountered more problems with the analysis software than others. Students who spent more time on the exercise felt that the overall gain in knowledge did not match the effort invested.

### THE SHOOTING APPLETS

Where the previous examples assumed substantial knowledge about experimental design concepts, these applets are very simple and can be used in the context of an introductory statistics course. Their aim is to let the user gain additional insight into the output of simple analyses like  $t$ -tests, simple and multiple regression and analysis of variance. They do so by allowing the user to generate his/her own data, and to “guess” how the analysis output will look like.

To generate the data, the applets use an arcade-style environment: the user is invited to “shoot” (by aiming and clicking the mouse) at a disk flying on the screen. The response is the precision of the shot: the distance between the actual position of the mouse when clicked and the center of the disk at that same moment.

The different applets then use different factors to explain the response. The first applet varies the speed of the disk. Then we can apply linear regression, with speed as the independent and precision as the dependent variable.

The second applet varies the color (or rather the shade of grey) of the disk. This also leads to linear regression. The third applet varies both speed and color, leading to multiple regression. The fourth applet invites the user to sometimes use the left hand for shooting, sometimes the right hand. A  $t$ -test can then be used, comparing the left to the right hand. The fifth applet uses red and green colored disks, again leading to a  $t$ -test.

After the shooting is completed, a scatter plot of the data is shown, as well as the standard computer output for the appropriate analysis, but with the numerical fields covered with black. When the mouse is clicked on a black field, the black is removed and the underlying number becomes visible. The idea is that the user tries to “guess” what is under the different black fields, using the information on the scatter plot, the information in the fields already uncovered, and general statistical knowledge (e.g., about the relation between  $t$ -values and  $p$ -values). It is hoped that this will lead to a better insight into the working of the analysis procedure, the interpretation of the analysis output and its relation to the data.

An additional feature is the possibility to edit the data during the analysis phase. When, for example, something has gone wrong with the shooting and an outlier was obtained, one can first look at the analysis results with the outlier, the edit the data to remove or correct the outlier, and see how this changed the analysis results.

The different applets were originally designed to illustrate different aspects. Changing the speed should normally result in a regression with a significantly possible slope. Changing the shade of grey usually results in a flat regression line, except for a few points at one extreme: most of the shades are visible enough against the background, except the darkest ones which can hardly be seen. This illustrates the point that a linear relation is not always appropriate. The left-right comparison usually leads to a significant difference, while the green-red comparison is usually non-significant. However, in practice we found large person-to-person differences.

The applets were used in the classroom. After explaining the analysis and its output, one student was invited to come in front and do the shooting on the computer connected to the beamer. Then the teacher strategically uncovered the black fields one by one, each time asking the class what they thought would be the underlying number and why. Students experienced this

as a much more lively and interesting experience than simply going through another example in the classical way.

## CONCLUSION

These applets show just a few examples, and there are countless possibilities for extension and improvement. But they clearly demonstrate that current computer technology allows the creation of special purpose, accessible and rich environments that have, in the area of data collection and analysis, the potential to give learning experiences well beyond those of traditional textbook exercises.

## REFERENCES

- Box, G. E. P., Hunter, W. G. and Hunter, J. S. (2005). *Statistics for Experimenters*. New York: Wiley.
- Cobb, G. (2002). *Introduction to Design and Analysis of Experiments*. Emeryville, CA: Key College Publishing.
- Darius, P. L., Ottoy, J-P., Solomin, A., Thas, O., Raeymaekers, B., and Michiels, S. (2000). A collection of applets for visualizing statistical concepts. In J. G. Bethlehem and P. G. M. van der Heijden (Eds.), *Proceedings in Computational Statistics 2000*, Utrecht, the Netherlands. Heidelberg: Physica Verlag.
- Darius, P. L., Ottoy, J-P., Thas, O., Michiels, S., and Raeymaekers, B. (2002). Applets for experimenting with statistical concepts. In B. Phillips (Ed.), *Proceedings of the Sixth International Conference on Teaching of Statistics*, Cape Town. Voorburg, The Netherlands: International Statistical Institute.
- Darius, P.L., van der Knaap, H. C. M., Schrevens, E., Portier, K. M., Massonnet, G., Lievens, L., and Duchateau, L. (2003). Virtual experiments and their role in teaching design and analysis of experiments. Paper presented at the 54<sup>th</sup> International Statistical Institute, Berlin, Germany, <http://www.stat.auckland.ac.nz/~iase/publications.php?show=3>.
- Dean, A. and Voss, D. (1998). *Design and Analysis of Experiments*. New York: Springer.
- Jones, J. W., Dayan, E., Allen, L. H., Van Keulen, H., and Challa, H. (1991). A dynamic tomato growth and yield model (TOMGRO). *Transactions of the ASAE* 34, No. 2, 663-672.
- Montgomery, D. C. (1997). *Design and Analysis of Experiments*. New York: Wiley.
- R Development Core Team. (2005). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.