

APriot: A computer program for Monte-Carlo
simulations of cumulated error probabilities in ANOVA

A O R

← I T P →

User manual

Albert-Georg Lang
Heinrich-Heine-Universität Düsseldorf
Düsseldorf, Germany

Introduction

With APriot Monte-Carlo simulations of the analysis of variance (ANOVA) can be conducted including the accumulation of the alpha error probability when intermediately inspecting the data. In order to make simulations as easy as possible there is no need to enter the simulation parameters like matrices of means, standard deviations, and correlations of residuals. Rather, you enter the data of an earlier study as you would do with other statistical software. APriot computes all parameters needed for the simulation from those raw data.

The steps for conducting a Monte-Carlo simulation are:

1. Enter the (raw) data, for example taken from a pilot study
2. Conduct an ANOVA with these data
3. Specify the effect for which the simulation will be conducted
4. Start the simulation

Entering raw data into APriot

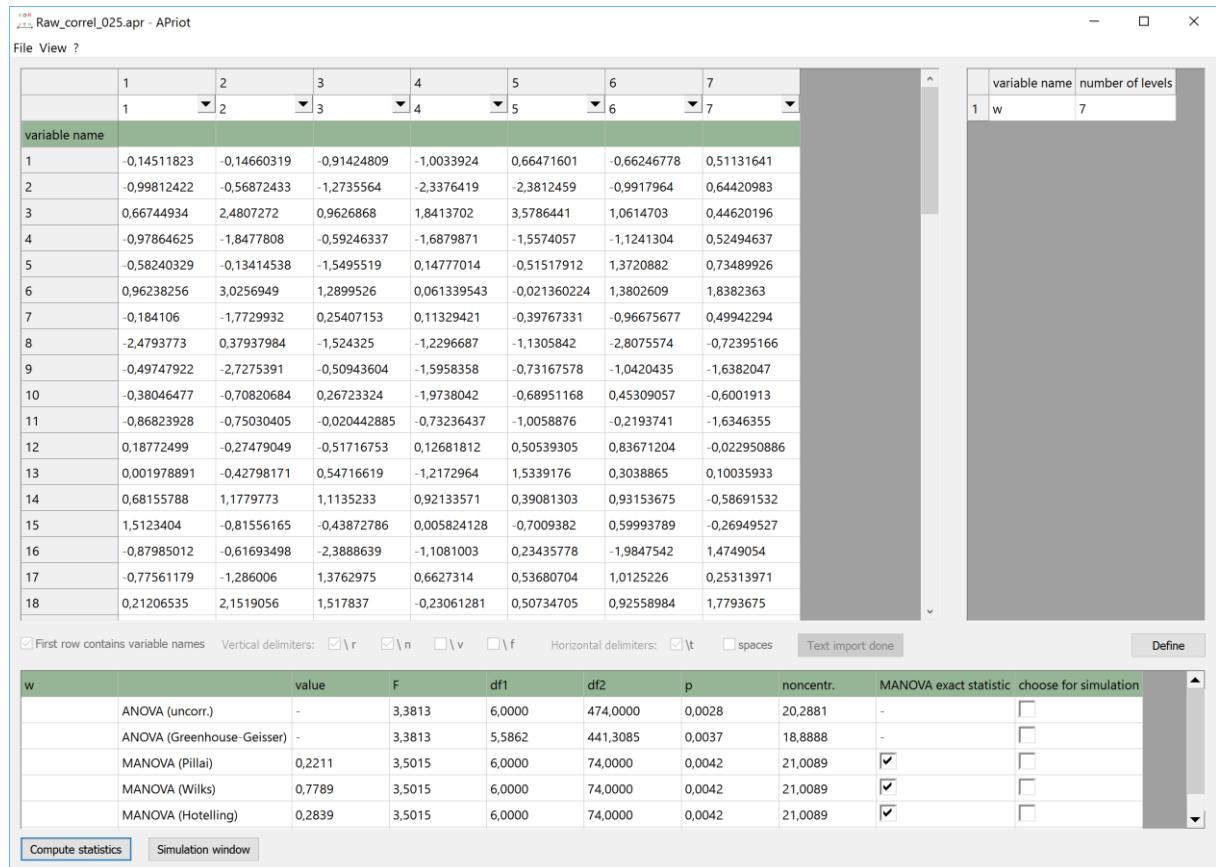


Figure 1: The main window of APriot. On the left side variables are entered, on the right side within-subject variables are defined, and on the bottom the ANOVA results are shown.

The main window of APriot contains all that is necessary to enter your (raw) data and to conduct an ANOVA with them. The first step is defining your within-subject variables on the right side of the window. Each within-subject variable is assigned a name and a number of levels. If your experimental design only contains between-subjects variables it is nevertheless necessary to define at least one within-subject variable containing one level. When all within variables are entered a click on the 'Define' button enables you to enter your data on the left side of the main window. Each column has a pull-down menu on its top where you can choose whether this column is dedicated to a between-subjects variable or to a particular combination of within-subject variables. For example, "3, 4" means that this column contains level 3 of the first within-subject variable in combination with level 4 of the second within-subject variable. If a column is declared as containing a between variable it is used like in other statistical software, this means, the levels of the between variable are coded with arbitrary different integer values.

The second row is dedicated to variable names. Please note that only entries for the between-subjects variables take effect here since the names for the within-subject variables have already been defined before in the right area of the window.

When entering variables APriot automatically increases the number of rows if the “cursor” comes near to the bottom of the window. Analogously, the number of columns is increased if the cursor comes near to the right border of the window. Additionally, you can add or delete rows and columns by hand by choosing the ‘View’ menu.

Alternatively to entering your data by hand, you may want to import them from a text file. APriot can import text files with various horizontal and vertical delimiters, for example spaces, blanks, carriage returns, etc. To import a text file, choose ‘Import text file’ in the ‘File’ menu. With the check boxes below the main area you can choose whether the first row of the file contains variable names and which delimiters are used in the file. If you check or uncheck a box APriot will immediately renew the view of the file as it is imported with the current settings. This way—if in doubt about the delimiters used in a text file—you can try until the view in APriot is correct. When you have finished the import click on the button “Text import done”.

A third alternative to enter new data is copy/pasting them from another program. If a table to be pasted into APriot has more rows/columns than are displayed in APriot by default APriot automatically increases the number of rows/columns such that the table pasted fits into APriot.

After the data have been entered clicking on the button “Compute statistics” shows the results of the ANOVA at the bottom of the main window. Here you find the results for each variable and each interaction, including *F*-values, numerator- and denominator degrees of freedom, *p*-(alpha) values, and the noncentrality parameters. For every effect, APriot computes the uncorrected and the Greenhouse-Geisser corrected ANOVA results as well as the results of the multivariate approach to repeated-measures ANOVA using Pillai’s V, Wilks’ Lambda, and Hoteling’s T. For between-subjects variables the same statistics are shown, but in this case they all are identical. For the statistics dedicated to the multivariate approach to the repeated-measures ANOVA a separate column informs you whether the statistic is exact in this case.

Changing preferences

If you select ‘Preferences’ in the ‘File’ menu a little window appears that lets you change two basic settings of APriot. By default, APriot uses all (virtual) processor cores available on your computer in order to reach optimum performance. If you want APriot to use less processor cores—e.g., because another program running on the computer needs sufficient computing capacity—you can reduce the

number of cores used by APriot. Please note that you have to restart APriot in order the change to take effect.

APriot generates random numbers internally with the Mersenne Twister algorithm. If you want to use another method for generating random numbers (e.g., use true random numbers) you can import random numbers from a text file. Please be careful that the text file contains enough random numbers for the simulation you plan to conduct.

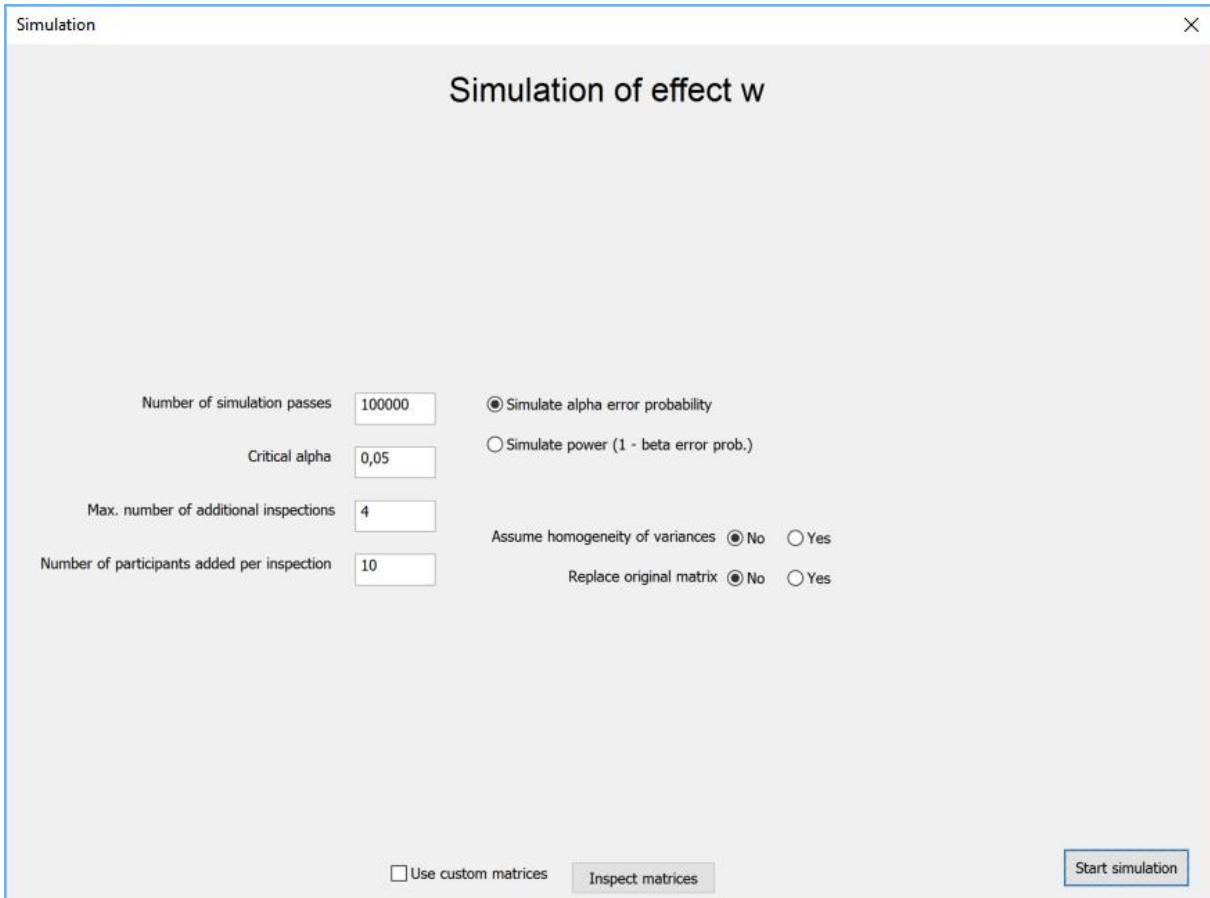


Figure 2. The simulation window of APriot.

Conducting a simulation for a chosen effect

The last column of the results view in the main window contains check boxes. By selecting a check box, you choose an effect for the further simulation. Only one effect may be chosen at a time. After clicking on the button 'Simulation window' a new window opens as shown in Figure 2. The simulation window is the core of APriot. Here, all parameters needed for conducting a simulation are chosen.

Number of simulation passes

Here you choose how many virtual experiments are conducted during the simulation. For example, if you choose 100,000 simulation passes APriot will simulate conducting 100,000 experiments.

Critical alpha

This is the alpha value for each statistical test. If you plan to conduct multiple inspections of your data (next parameter) it has to be adjusted such that total alpha does not inflate.

Max. number of additional inspections

If '0' is chosen a single ANOVA is simulated in which all participants are tested as a whole, this means, data are not inspected intermediately. If a number > 0 is chosen this number denotes how often data are inspected intermediately. The value entered here is a maximum number since the (simulated) experiment is ended as soon as an intermediate inspection shows statistical significance.

Number of participants added per inspection

If an intermediate inspection does not reveal statistical significance the study is either ended (if the "Max. number of additional inspections" is reached) or a further sample of participants is tested. "Number of participants added per inspection" specifies how large this sample is.

Simulate alpha error probability/Simulate power (1 – beta error prob.)

With these radio buttons you select which probability you want to be estimated by the simulation. Both probabilities are a result of the complete sequential test, this means they are cumulated probabilities which take into account the specified number of intermediate inspections.

Assume homogeneity of variances

Homogeneity of variances is a prerequisite for the ANOVA. It is assumed that the variances of all levels of all variables are equal.

With the two radio buttons you can choose whether APriot computes one variance for all factor levels (and factorial combinations) of the raw data and uses this variance to render the random numbers for simulation or whether APriot computes the variances separately for each factor level and combination and uses these variances as the basis for simulation.

Replace original matrix

The original matrix consists of the raw data you have entered into the main window. You can choose whether the original matrix is used for the first inspection of data or whether random numbers shall be used for all intermediate inspections (including the first inspection). If the raw data you entered are a first sample of participants of the current experiment (“internal pilot study”) you might want to use them in the simulation. If the raw data are from a (formerly conducted) pilot study the raw data only serve for computing means, standard deviations and correlations. In this case you should replace them with random numbers.

Number of participants per group

APriot offers this option only if “Replace original matrix” is set to “yes”. In this case you can specify how many participants are the basis for the first inspection of data.

The matrices window

As mentioned above, APriot uses the raw data entered in the main window to compute the parameters needed for a simulation, that is the matrices of means, standard deviations, and—if repeated-measures are involved—correlations of residuals. By clicking on ‘Inspect matrices’ you can look at these matrices (Figure 3).

Matrices							
	W1	W2	W3	W4	W5	W6	W7
W1	1	0,159	0,271	0,313	0,35	0,325	0,208
W2	0,159	1	0,258	0,146	0,281	0,187	0,232
W3	0,271	0,258	1	0,126	0,264	0,25	0,0903
W4	0,313	0,146	0,126	1	0,42	0,276	0,26
W5	0,35	0,281	0,264	0,42	1	0,154	0,298
W6	0,325	0,187	0,25	0,276	0,154	1	0,218
W7	0,208	0,232	0,0903	0,26	0,298	0,218	1

Figure 3: The matrices window. With the tabs on the top the matrix of means, standard deviations, and correlations of residuals can be chosen.

If you want to edit the matrices the option ‘Use custom matrices’ must be selected in the simulation dialog. Please note that you can go back to the matrices computed from the raw data at any time by unchecking ‘Use custom matrices’ in the simulation window.

Description of the simulation process

After the user has entered the raw data in the main window and has clicked on the “Compute statistics” button APrior starts to compute analyses for all variable combinations. The user chooses one analysis as the basis for the subsequent simulation.

As mentioned above APrior uses the raw data entered by the user to calculate all parameters needed for the simulation. As a first step a matrix of the means of all combinations of variable levels is computed; the same is done for the standard deviations. Depending on the user’s choice (“Assume homogeneity of variances”) the standard deviations are computed for each factorial combination separately or one standard deviation is computed over all factorial combinations. If there are within-

subject variables the correlations between the factorial combinations of the within-subject variables are computed as well. Standard deviations and correlations are computed as unbiased estimators. Depending on the means, standard deviations and correlations the random values for the subsequent simulation passes are generated; this means that the “virtual population” which is the basis for generating the random values has the same means, standard deviations, and correlations as specified in the respective matrices.

The alpha error probability is the probability of detecting an effect in a sample while the effect does not exist in the underlying population. Thus, if the user chooses “Simulate alpha error probability” the matrix of means is modified so that the effect associated with the chosen analysis is zero. If the user chooses “Replace original matrix” the matrix of the data entered by the user is replaced by random values with the same distributional properties as those of the original matrix. In this case the user can further choose the number of subjects per group the virtual experiment will start with. If “Replace original matrix” is not chosen the matrix entered by the user is left unchanged and is used as the basis for the further subjects-adding rounds.

One simulation pass

A simulation pass starts with a statistical computation of the selected analysis with the initial data (original matrix or replaced). For example, if the user has selected a Greenhouse-Geisser corrected ANOVA for variable “W2” this analysis is computed for the initial matrix. If the analysis result is “significant” (i.e., the resulting alpha error probability is smaller or equal to the value specified as critical alpha) this simulation pass is marked as “hit”. Otherwise the procedure continues as follows: The matrix is enlarged by further random values as specified in the field “Number of subjects added if not significant” and the analysis is computed again. If the result is now “significant” the loop pass is marked as “hit.” If the result is “not significant” further samples are added. This is repeated until either the result of the analysis is “significant” or the number of repetitions as specified in the field “Max. number of subjects-adding rounds” is reached. If the last analysis does not reach “significance” this loop pass is marked as “miss”. After all loop passes have been completed the number of “hits” (alpha error probability) or “misses” (beta error probability) is divided by the total number of loop passes.

Computations

Matrix operations are conducted using the Clapack library (Anderson et al., 1999) and the Blas library (Dongarra, Duff, Croz, & Hammarling, 1989).

By default, random values are generated using the Mersenne twister (MT 19937) algorithm (Matsumoto & Nishimura, 1998). Alternatively, the user can select to use random numbers stored in text files (“Preferences”). In each row of a text file one number has to be stored; the numbers must be normally distributed with a mean of 0 and a standard deviation of 1. This way random numbers from any source can be used, for example, from a true random source like radioactive decay.

If there are within-subject variables the random variables generated for simulations should have the same correlations as the variables of the original data. In order to obtain random numbers with a predefined correlation pattern the correlation matrix calculated from the original data is transformed using a Cholesky decomposition, then the matrix containing the random numbers is multiplied with the Cholesky-transformed correlation matrix (e.g., see Haugh, 2004).

References

- Anderson, E., Bai, Z., Bischof, C., Blackford, L. S., Demmel, J., Dongarra, J., Croz, J. D., Greenbaum, A., Hammarling, S., McKenney, A., & Sorensen, D. (1999). *LAPACK Users' Guide - Third Edition*. <http://www.netlib.org/lapack/lug/>:
- Dongarra, J., Duff, I., Croz, J. D., & Hammarling, S. (1989). *BLAS (Basic Linear Algebra Subprograms)* program]. <http://www.netlib.org/blas/>:
- Haugh, M. (2004). *The Monte Carlo Framework, Examples from Finance and Generating Correlated Random Variables*. http://www.columbia.edu/~mh2078/MCS_Framework_FEegs.pdf: Columbia University,
- Matsumoto, M., & Nishimura, T. (1998). Mersenne twister: a 623-dimensionally equidistributed uniform pseudo-random number generator. *ACM Transactions on Modeling and Computer Simulation*, 8, 3-30.