Automated Univariate Analysis of Variance Methods for Nested Mixed Effects Linear Models

Timothy Hall*

Abstract
A complete univariate Analysis of Variance (ANOVA) of a mixed effects linear model with nested factors may become extremely complicated when the data justify the use of some terms in the fully factorial, i.e., unconditional, model while failing to accept others at a given level of significance. This paper presents an analytical routine (written in elementary ANSI C code) that automatically calculates the applicable ANOVA table, including the Expected Mean Squares (EMS) expressions, and exact and Welch-Satterthwaite approximate F tests and statistics, that a given dataset would present as significant at a given level of Type I error. The output is the Plain TeX code that displays the ANOVA table and a listing of the intermediate steps taken to calculate the final model. Also included is a discussion of the possible valid adjustments an analyst may make to non-fully factorial datasets, or datasets with an unequal number of within-factor observations, to make them appropriate for use with this code.

Key Words: Analysis of Variance, Embedded System, Nested Mixed Effects Linear Model

1. Introduction

Embedded systems are becoming more and more critically important towards realizing an important achievement – automating the fundamental benefits of statistical analyses into the intelligence of a control system. Rather than focus on statistical analyses that may be performed once or twice by teams of people over weeks or even years, embedded systems offer the opportunity to perform thousands or even millions of a variety of modifiable statistical analyses within one second by a single microprocessor. Such a capability may provide a control system with immediate feedback on changing conditions, thus enabling more precise and quick-responding control. Inefficient and unproductive characteristics of common processes, such as vibration, friction, heat loss, and incomplete combustion may be made ultra-efficient by the use of a control system that has the ability to react to changing conditions and variable boundary values according to assessments and conclusions provided by a statistical analysis embedded system.

2. Statistical Programming Principles

To achieve this goal of embedding statistical analysis capabilities into the intelligence of a control system, a radically different approach to numerical programming and hardware implementation is needed compared to the current dominant computer architecture model. In particular, the orientation towards compiled programs must be replaced by hardware systems that implement the analytical methods required for a statistical analysis, with flexibility for alternate methods should circumstances require such diversity.

2.1 Atomic, Modular, Reusable, Hierarchical

The first required principle of statistical programming methods is that all embedded system statistical analysis functionality be oriented to atomic, modular, reusable, and hierarchical

*PQI Consulting, P. O. Box 425616, Cambridge, MA, USA, 02142-0012

structures. These concepts are independent yet cumulative within an embedded system, i.e., enhancing/degrading one aspect does not enhance/degrade any of the other aspects. Each of these concepts is equally important with respect to the integrity and effectiveness of every statistical analysis embedded system.

2.1.1 Atomic

An embedded system structure is said to be *atomic* if it defines functionality that is not a combination of other, more basic, functionality. Since the ultimate goal of any embedded system is to reside completely in a hardware implementation, the atoms of any statistical analysis embedded system will depend on the particular structures available in the hardware. However, for development purposes, the terms “atoms” will refer to the fundamental and indivisible units of an intermediate software implementation (see Sections 2.3 – 2.5).

2.1.2 Modular

A *module* is a narrowly-defined, application-specific, and practical segment of statistical analysis functionality. It may be oriented towards numerical, database, or signaling aspects within the context of the entire embedded system. An example of a module is the floating-point addition of two numbers. Other non-numeric examples would be deleting an entry from the operator queue, and changing the value of an overflow register.

2.1.3 Reusable

A module is said to be *reusable* if it may be used in combination with other modules to perform a more complication calculation, or a more intricate database change, or signal a more sophisticated status change to the embedded system. The fact that floating-point addition may be repeatedly used, along with the floating-point division module, to calculated the average value of a dataset, shows that floating-point addition (and division) are reusable modules.

2.1.4 Hierarchical

A collection of modules, sometimes called a library, is said to be *hierarchical* if each member of the collection is either atomic, i.e., only defined in terms of lower-level operations, or is defined in terms of a combination of other modules in the collection. For example, a variance module might be defined in terms of a module that calculates the mean, and modules/atoms that perform floating-point subtraction, multiplication, and division.

2.2 Theoretical Foundations

The first step in designing an atomic, modular, reusable, and hierarchical statistical analysis embedded system is to carefully define and delineate the functionality of each structure. In this way, minimal changes will be required during the implementation phases of such a system since all the pitfalls and deadends that complicated analytical systems may introduce have been anticipated and avoided. Such theoretical foundations include defining the inputs needed to the module, the analytical methods used to make the necessary calculations (including having the required modules available in the library), and the outputs provided by the module. Even though these aspects of system design appear to be obvious or trivial, careful attention to these matters goes a long way to minimize the need for “rework” when extensions are attempted to system functionality.

2.3 High-level Language Prototype

A highly effective starting place for designing a statistical analysis embedded system is to code all modules in terms of a high-level language such as `C`, `UNIX Shell`, `MAPLE`®, `Mathmatica`® , etc. Note that these examples *do not* include statistics-oriented software such as `R`, since that type of software is oriented to performing complicated statistical calculations behind-the-scenes to the user – precisely the sort of thing embedded systems programming needs to avoid. Rather, symbolic manipulation and programmable mathematics software provides a general, yet proprietary programming language that is well suited to the goals of module development. The purpose of starting with a high-level language is to focus on the functionality involved in the module, rather than the underlying details that will be explored in the next section.

2.4 Low-level Language Prototype

Once the fundamental functionality of a module has been documented in the high-level language prototype, this code may be used to translate the same functionality to a low-level language such as `MMIX`, which is a generic Reduced Instruction Set Computing (RISC) 64-bit assembly language invented by Dr. Donald E. Knuth of Stanford University. For example, the high-level language implementation for the variance module (which essentially uses floating-point addition, subtraction, multiplication, and division) may be used to code the low-level language implementation using the MMIX operations `FADD`, `FSUB`, `FMUL`, and `FDIV`. The purpose of translating the high-level language version into a low-level version is to easily and more intuitively move one step closer to the ultimate goal of hardware implementation of a statistical analysis embedded system.

2.5 Hardware-level Implementation

Finally, the low-level language implementation of any module is easily translated to a hardware description language at the register-transfer level of abstraction (versus property specifications) such as documented under the IEEE 1364 standard. It is from this kind of implementation that fabrication masks are created.

2.6 Policies

A library of module policies should be included along with every embedded system code library, as such documentation is just as critically important as any other embedded system documentation is to the ultimate success of an embedded system as is the programming code itself. It provides guidance for analysts to use, extend, and otherwise modify the functionality of modules (and their interactions within a library). These policies may also be used to anticipate special cases and particular implementation issues that might arise due to the particular nature of the statistical analyses involved. They also provide a controlling reference for enhancements and version branches as the complexity and applicability of the module library expands.

3. Minimal Instruction Set

An important consideration for defining the modules within the low-level language context involves the number of instructions available. For example, the `MMIX` language provides

® MAPLE is a registered trademark of Maplesoft, Inc., Waterloo, Ontario, Canada.
® Mathematica is a registered trademark of Wolfram Research, Inc., Champaign, Illinois, USA.

for 256 instructions (consistent with its 64-bit architecture), yet perhaps only 25 are needed to implement a particular library of statistical analysis modules. However, this creates the question: How many instructions is optimal? Are 10 instructions sufficient? Or are 19 the absolute minimum needed? It turns out that for all statistical analyses, regardless of how complicated and how intricate they might become, only 1 instruction is needed to implement any module; all other instruction would be defined in terms of repeated uses of that one instruction. This would be the ultimate “efficient” code.

A practical determination of the absolute minimum number of instructions needed to complete a module calculation involves the trade-off between the number of instructions, the number of clock cycles required to perform all of the calculations within the module, and the requirements of how often such a calculation is needed. With 25 instructions available, a module may require 300 “oops” (computer science lingo for the number of clock cycles required to complete all instructions), which, on a computer with an effective throughput rate of 3GHz, would mean the module calculation may be performed 10,000,000 times per second. However, with 3 instructions available, that same module may require 300,000 oops, which would mean it could only be completed 10,000 times per second. If the application addressed by the embedded system required an update every microsecond, i.e., 1,000,000 times per second, then the 3 instruction version, as elegant as its code might be, would be insufficient to meet this requirement, whereas the 25 instruction version would be able to do so.

3.1 Floating-point Operators

By far the most frequently occurring mathematical operations used in any statistical analysis are those than may be accomplished through floating-point calculations. These include addition and multiplication, and their inverses (subtraction and division), the (positive) square root, floor and ceiling conversions to integers, and the conversions between integer and floating-point expressions. Many of these are redundant, e.g., the square root may be implemented as a module involving only arithmetic operations, so a smaller instruction set is possible even at this fundamental level.

3.2 Stochastic Functions

Just as the floating-point operators are the most commonly used in any statistical analysis, the highly-complicated, multiple parameter stochastic functions, such as a χ^2 distribution density function or an F distribution inverse cumulative function, may still be expressed in terms of a low-level language with a minimal instruction set. The hierarchical nature of the module library that leads to such functionality is not immediate: Numerical integration and orthogonal sets of polynomials must be defined before any stochastic function may be implemented as a module. The primary point here is that such implementation are not only possible, they are required, regardless of complexity, to maintain the atomic, modular, reusable, and hierarchical structures that make up any statistical analysis embedded system.

4. Automated ANOVA For Nested Mixed Effects Linear Models

In general, the term “ANOVA” refers to a collection of statistical models used to analyze the contribution of various sources of variation (depending on the model) to the total variation found in a dataset. Note how the dataset itself is part of the parameters of an analysis of variance. In particular, when the statistical model of a univariate dependent value consists of the linear sum of a set of factors (and their interactions), where these factors may take on “fixed” values in the dataset, i.e., they are only capable of taking on certain specific values, as well as "random" values in the dataset, i.e., they are capable of taking on any of a continuous range of values (not necessarily the same for each factor), then the ANOVA refers to a mixed effects linear model. Furthermore, if a factor is "nested" in another factor, i.e., there is a restriction on the randomization of the treatments to experimental units that prevents an assessment of the variance contribution due to the interaction of those factors, then the ANOVA refers to a nested mixed effects linear model. For the purposes of this paper, the dataset involved in a nested mixed effects linear model is "fully factorial," i.e., every estimable combination of factors is represented at least once in the dataset, and occurs the same number of times as any other estimable combination of factors (see also Section 6).

The following information documents the functionality of a library of statistical analysis modules that may be used to implement a nested mixed effects linear model ANOVA as an embedded system. This library consists of low-level programming modules (not associated with any particular hardware manifestation) that implement the required mathematical and stochastic functions in terms of a subjectively minimal (25-many) instruction set, i.e., one that does not consider any timing or cycle count requirements — rather only as a programming convenience. This library will be referred to as "the automated ANOVA embedded system."

4.1 Overview Of Functionality

There are three steps in any use of the automated ANOVA embedded system: (1) Reading of the Configuration File, which contains all the parameter definitions, labels, and values; (2) The presentation of the Summary Table, which contains the results of all ANOVA calculations; and (3) The presentation of the Disposition Table, which details the step-by-step reduction of the original nested mixed effects linear model into its final form based on the significance of the factors found in the datasets.

For terms in the ANOVA model with an exact F-test statistic, those that are deemed not significant in the dataset are consolidated into other terms of the current model (starting with the highest-order interaction term in the model). This determination is based on the denominator of the F-test statistic as given by the expected mean squares calculation. A new consolidation pass is made through the model each time a term is consolidated.

For terms in the ANOVA model without an exact F-test statistic, a Welch-Satterthwaite approximate F-test statistic is calculated in a manner that balances the degrees of freedom between the numerator and denominator as closely as possible. Such terms are retained in the final model regardless of significance level, since the consolidation into other significance terms is made problematical by the approximating nature of the F-test statistic.

The two output tables (summary and disposition) are actually ASCII-text files consisting of Plain $\TeX$ code that produces the graphical presentation of those tables. These $\TeX$ files are produced through the input/output functionality of the low-level programming language that would not be necessary in the hardware version (unless such output were desirable and affordable in clock cycles).

4.2 The Configuration File

The configuration file for the automated ANOVA embedded system consists of one of five available statements, one per line, beginning with a backslash escape character (ASCII 92), terminated by a return/linefeed (ASCII 10) character. The available statements are `factordef`, `nesteddef`, `errordef`, `significance`, and `datafile`.

4.2.1 Factordef

This statement takes three arguments delineated by curly brackets (ASCII 123 and 125): (1) The name of the factor (a single character); (2) The type of factor – "F" for fixed and "R" for random; and (3) The number of levels in the dataset. The use of this statement means the factor is not nested in any other factor (whether yet defined or otherwise), and it is not the error term of the model. As currently implemented, up to seven factors may be defined in any single use of the embedded system.

4.2.2 Nesteddef

This statement takes four arguments delineated by curly brackets (ASCII 123 and 125): (1) The name of the factor (a single character); (2) The type of factor – "F" for fixed and "R" for random; (3) The number of levels in the dataset; and (4) The term in the model in which the factor is nested (a single character). The use of this statement means the defined factor is nested in the given factor (the fourth argument), and it is not the error term of the model. As currently implemented, up to six factors may be nested in any single use of the embedded system.

4.2.3 Errordef

This required statement defines the error term of the model and takes three arguments delineated by curly brackets (ASCII 123 and 125) – even though only two are at the discretion of the analyst. The first argument is the name of the factor (a single character), followed by "R" (the error term must be considered a random factor in every model), followed by the number of levels in the dataset. The use of this statement means the factor is not the name of an independent variable in the model.

Note that all terms in the ANOVA model that would otherwise be tested by the error term would become untestable if the number of levels is set to 1.

4.2.4 Significance

This statement takes one argument delineated in curly brackets (ASCII 123 and 125) and it defines the minimum level of significance a term in the model must have in the dataset for that term to be retained in the model, i.e., reported in the summary table. If a term is not deemed significant in the dataset, then the disposition table will account for its consolidation into other significant terms in the model as explained in Section 4.1. The significance level may be any number between 0 and 1 (inclusive) expressed as a percentage, to whatever precision is available in the calculation environment.

4.2.5 Datafile

This statement defines the name of the file containing the comma-delimited ASCII-text dataset. Only one such statement may be used in a configuration file, and only one file may be named, which must have a `*.dat` filename extension.

4.3 Processing Steps

The automated ANOVA embedded system operates as follows.

1. Process the contents of the configuration file.
2. Load the contents of the dataset file into memory.
3. Calculate the ...

   (a) Sum Of Squares (Orthogonal Partition)
   (b) Mean Squares
   (c) Expected Mean Squares (EMS)
   (d) Exact F-Test Statistics
   (e) Approximate F-Test Statistics (When Needed)
   (f) Significance Level Of F-Test Statistics
4. Based on the EMS terms, and starting with the highest-level interaction term in the model, sequentially consolidate the model terms evaluated as not significant.
   (a) If consolidation occurs, record the event in the Disposition Table, and restart the significance evaluation from the beginning.
   (b) If consolidation does not occur, continue with the model terms evaluation without writing anything to the Disposition Table.
5. Finish writing the Disposition Table.
6. Write the Summary Table.

4.4 The Summary Table

The summary table presents the final version of the ANOVA model that is deemed significant in the dataset. For example, for the following configuration file, where the significance

```
\factordef{C}{F}{4}
\factordef{M}{F}{2}
\factordef{T}{R}{3}
\errordef{e}{R}{5}
\significance{0.0}
\datafile{FFR.dat}
```

Figure 1: `FFR` Dataset Configuration Listing

is set to 0.0 to ensure all terms appear in the summary table, the automated ANOVA embedded system produces the following processed TEX code for the data from file `FFR.dat`. The first column in the summary table "Source" gives the name of the term in the model.

Source	df	SS	MS	F	T*	(N):D**	Signf	EMS
C_i	3	1.725180	0.575060	0.089345	E	$-:CT$	0.036732	$\sigma^2_\epsilon + 10\sigma^2_{CT} + 30\phi_C$
M_j	1	12.910080	12.910080	4.580857	E	$-:MT$	0.832127	$\sigma^2_\epsilon + 20\sigma^2_{MT} + 60\phi_M$
T_k	2	25.593945	12.796972	1.750521	E	$-:\epsilon$	0.820818	$\sigma^2_\epsilon + 40\sigma^2_T$
CM_{ij}	3	19.793940	6.597980	0.455080	E	$-:CMT$	0.276627	$\sigma^2_\epsilon + 5\sigma^2_{CMT} + 15\phi_{CM}$
CT_{ik}	6	38.618534	6.436422	0.880450	E	$-:\epsilon$	0.487628	$\sigma^2_\epsilon + 10\sigma^2_{CT}$
MT_{jk}	2	5.636535	2.818268	0.385516	E	$-:\epsilon$	0.318852	$\sigma^2_\epsilon + 20\sigma^2_{MT}$
CMT_{ijk}	6	86.991104	14.498517	1.983278	E	$-:\epsilon$	0.924515	$\sigma^2_\epsilon + 5\sigma^2_{CMT}$
$\epsilon_{l(ijk)}$	96	701.796570	7.310381					σ^2_ϵ

* (E) Exact; (A) Approximate – (df); (U) Untestable
** No Numerator Given For Exact Tests

Figure 2: `FFR` Dataset Summary Table

It is followed by the degrees of freedom "df" for that term, followed by the dataset Sum of Squares "SS" and Mean Squares "MS" statistics. The "F" column shows the calculated F-test statistic for each term in the final model that is deemed significant. This statistic is based on the EMS expressions (last column) for the numerator and denominator test values. The "T" column lists whether the F-test statistic is Exact, Approximate, or Untestable.

The following column shows the numerator and denominator terms involved in the F-test statistic, where no numerator term is given for exact tests, and "ϵ" is symbolic for the error term. Nested factors are signaled by their index symbols in parentheses.

As another more complicated example, for the following configuration file, which

```
\factordef{C}{F}{4}
\nesteddef{M}{F}{2}{C}
\factordef{T}{R}{3}
\nesteddef{V}{F}{5}{T}
\factordef{W}{R}{2}
\errordef{e}{R}{3}
\significance{90.0}
\datafile{FFRFR.dat}
```

Figure 3: `FFRFR` Dataset Configuration Listing

contains five factors, two of which are nested, the automated ANOVA embedded system produces the following processed TEX code at 90% significance for the data from file `FFRFR.dat`. Note that there is only one surviving term (and factor) in this analysis, since

Source	df	SS	MS	F	T*	(N):D**	Signf	EMS
$V_{l(k)}$	12	188.837753	15.736479	1.891855	E	$- : \epsilon$	0.967744	$\sigma_\epsilon^2 + 48\phi_V$
$\epsilon_{n(ijklm)}$	707	5880.838379	8.318018					σ_ϵ^2

* (E) Exact; (A) Approximate – (df); (U) Untestable
** No Numerator Given For Exact Tests

Figure 4: `FFRFR` Dataset Summary Table

all other terms have been deemed not significant (see the next section).

4.5 The Disposition Table

The disposition table presents the logical flow of F-tests and consolidations performed during the analysis that results in the final model presented in the summary table (see the previous section). For example, for the `FFR` dataset and configuration file (see Figure 1), with the significance set to 100% in this example (which means all terms in the model will be rejected to fully demonstrate the consolidation of terms), the automated ANOVA embedded system produces the processed TEX code for the disposition table presented in Figure 5. The first column of the disposition table "Step" is a counter for each separate F-test or consolidation calculation made by the embedded system. It is followed by the "#" column which denotes the term number in the fully factorial expansion of the factor names and their interactions. The next column explicitly lists the term referenced by the "#" column. The next two columns graphically show whether the term is retained in the model ("in"), or rejected from the model ("out"), based on the F-test statistic. Once again, the "E" term in these columns indicates that the decision was made based on an exact F-test statistic from the EMS expressions. The "EMS Term" column shows the numerator and denominator terms for the F-test statistic, where "ϵ" is symbolic for the error term. The next column shows the degrees of freedom corresponding to the previous column expression, which is followed by the explicit F-test ratio values, with the reduced value found in the

PQICSTAT EMS Table Disposition Report (Section 1 Of 1)*

Step	#	Term	Model In Out	EMS Test	Test df:df	F Ratio	Ratio Value	Signf
1	7	CMT	E	$CMT : \epsilon$	6 : 96	14.498517/7.310381	1.983278	0.924515
				Term CMT (#7) Consolidated Into ϵ [New SSE/df = 788.787659/102].				
2	6	MT	E	$MT : \epsilon$	2 : 102	2.818268/7.733212	0.364437	0.304512
				Term MT (#6) Consolidated Into ϵ [New SSE/df = 794.424194/104].				
3	5	CT	E	$CT : \epsilon$	6 : 104	6.436422/7.638694	0.842608	0.460101
				Term CT (#5) Consolidated Into ϵ [New SSE/df = 833.042725/110].				
4	4	CM	E	$CM : \epsilon$	3 : 110	6.597980/7.573116	0.871237	0.541601
				Term CM (#4) Consolidated Into ϵ [New SSE/df = 852.836670/113].				
5	3	T	E	$T : \epsilon$	2 : 113	12.796972/7.547227	1.695586	0.811874
				Term T (#3) Consolidated Into ϵ [New SSE/df = 878.430603/115].				
6	2	M	E	$M : \epsilon$	1 : 115	12.910080/7.638527	1.690127	0.802317
				Term M (#2) Consolidated Into ϵ [New SSE/df = 891.340698/116].				
7	1	C	E	$C : \epsilon$	3 : 116	0.575060/7.683972	0.074839	0.026588
				Term C (#1) Consolidated Into ϵ [New SSE/df = 893.065857/119].				

* Created Tuesday, April 02, 2013 at 21:40:32 EDT for configuration/data file prefix EMSwthSSMSwthFTwthK2b.

Figure 5: `FFR` Dataset Disposition Table

following column. The final column shows the maximum significance level found in the configuration file that would be needed for this term to be retained in the model (at that point in the analysis).

Note how each term in the model is eventually consolidated into the error term (the F-test denominator in each case) since the significance level has been set to 100%.

As another more complicated example, for the `FFRFR` dataset and configuration file (see Figure 3), with the significance level set to 90%, the automated ANOVA embedded system produces the processed TEX code for the disposition tables presented in Figures 6 and 7. Note how the three-way interaction term $\#23$, the "MTW" term, is repeatedly retained in the model during various re-evaluations after consolidations have occurred, e.g., Steps $2, 4, 6, ..., 23, 26, 29$ in both figures, yet it is eventually rejected from the model (at Step 32) since its sum of squares value is consolidated with the "MT" term ($\#10$) at Step 31, and the resulting increase in "MTW" degrees of freedom (from 12 to 20 from the addition of the 8 from "MT") no longer results in a significance greater then 90% (at Step 32 it drops under 73%). As the end of the disposition table shows (at Step 35 in Figure 7), the only term left in the model after all consolidations is term $\#4$, the "V" term, as was represented in the summary table (Figure 4).

5. Complexity Warning

It is very easy for even "simple" appearing configurations to present highly complicated summary tables and very long disposition tables. For example, the summary table found in Figure 8 corresponds to a seven-factor three-nested model configuration file with the significance level set to 0.0. However, the disposition table for this analysis grows to $1{,}063$ steps consuming 40 pages (for brevity, they are not included here) when the significance level is set to 90.0. Care must be taken to ensure that a statistical analysis embedded system does not produce such extensive results that the number of clock cycles is artificially inflated simply due to extreme reporting duties for routine results.

PQICSTAT EMS Table Disposition Report (Section 1 Of 2)*

Step	#	Term	Model In	Model Out	EMS Test	Test df:df	F Ratio	Ratio Value	Signf
1	24	MVW		E	$MVW : \epsilon$	48 : 480	8.077482/8.547336	0.945029	0.419515
Term MVW (#24) Consolidated Into ϵ [New SSE/df = 4490.440430/528].									
2	23	MTW	E		$MTW : \epsilon$	8 : 528	15.270254/8.504622	1.795524	0.924583
3	21	CVW		E	$CVW : \epsilon$	36 : 528	8.597583/8.504622	1.010931	0.545995
Term CVW (#21) Consolidated Into ϵ [New SSE/df = 4799.953613/564].									
4	23	MTW	E		$MTW : \epsilon$	8 : 564	15.270254/8.510556	1.794272	0.924516
5	20	CTW		E	$CTW : \epsilon$	6 : 564	7.716419/8.510556	0.906688	0.510435
Term CTW (#20) Consolidated Into ϵ [New SSE/df = 4846.251953/570].									
6	23	MTW	E		$MTW : \epsilon$	8 : 570	15.270254/8.502196	1.796036	0.924876
7	15	VW		E	$VW : \epsilon$	12 : 570	10.012615/8.502196	1.177650	0.704460
Term VW (#15) Consolidated Into ϵ [New SSE/df = 4966.403320/582].									
8	23	MTW	E		$MTW : \epsilon$	8 : 582	15.270254/8.533339	1.789482	0.923677
9	14	TW		E	$TW : \epsilon$	2 : 582	6.282489/8.533339	0.736229	0.520638
Term TW (#14) Consolidated Into ϵ [New SSE/df = 4978.968262/584].									
10	23	MTW	E		$MTW : \epsilon$	8 : 584	15.270254/8.525631	1.791100	0.923994
11	11	MV		E	$MV : \epsilon$	48 : 584	7.974335/8.525631	0.935337	0.400820
Term MV (#11) Consolidated Into ϵ [New SSE/df = 5361.736328/632].									
12	23	MTW	E		$MTW : \epsilon$	8 : 632	15.270254/8.483760	1.799939	0.925852
13	9	CW		E	$CW : \epsilon$	3 : 632	1.997460/8.483760	0.235445	0.128326
Term CW (#9) Consolidated Into ϵ [New SSE/df = 5367.728516/635].									
14	23	MTW	E		$MTW : \epsilon$	8 : 635	15.270254/8.453116	1.806465	0.927078
15	8	CV		E	$CV : \epsilon$	36 : 635	5.395785/8.453116	0.638319	0.048458
Term CV (#8) Consolidated Into ϵ [New SSE/df = 5561.976562/671].									
16	23	MTW	E		$MTW : \epsilon$	8 : 671	15.270254/8.289086	1.842212	0.933534
17	7	CT		E	$CT : \epsilon$	6 : 671	8.013901/8.289086	0.966802	0.553234
Term CT (#7) Consolidated Into ϵ [New SSE/df = 5610.060059/677].									
18	23	MTW	E		$MTW : \epsilon$	8 : 677	15.270254/8.286647	1.842754	0.933646
19	5	W		E	$W : \epsilon$	1 : 677	2.888000/8.286647	0.348513	0.444165
Term W (#5) Consolidated Into ϵ [New SSE/df = 5612.948242/678].									
20	23	MTW	E		$MTW : \epsilon$	8 : 678	15.270254/8.278685	1.844527	0.933946

* Created Tuesday, April 02, 2013 at 21:41:09 EDT for configuration/data file prefix EMSwthSSMSwthFTwthK4a.

Figure 6: `FFRFR` Dataset Disposition Table Part 1

6. Enhancements

The following enhancements to the functionality of an automated ANOVA embedded system might be considered by an implementing analyst.

1. *Incomplete Factorial Models*

   It is common in practical datasets for some combination of factor levels to be underrepresented and some to be over-represented. This means the dataset presents an incomplete factorial model in the ANOVA that must also be present in the analytical calculations. An embedded system that compensates for an incomplete factorial model, such as using averaged data, the elimination of factors, the unconditional consolidation of terms, subgroup sampling, etc., would have to sense the nature and extent of the deviation from a fully factorial model, and adjust the calculations accordingly, each time the statistical analysis is performed. This issue is also related to the general concept of "missing data," especially when there are more missing data cells than populated data cells.

PQICSTAT EMS Table Disposition Report (Section 2 Of 2)*

Step	#	Term	Model In	Model Out	EMS Test	Test df:df	F Ratio	Ratio Value	Signf
21	4	V	E		$V:\epsilon$	12 : 678	15.736479/8.278685	1.900843	0.968675
22	3	T		E	$T:\epsilon$	2 : 678	3.913201/8.278685	0.472684	0.376469
		Term T (#3) Consolidated Into ϵ [New SSE/df = 5620.774414/680].							
23	23	MTW	E		$MTW:\epsilon$	8 : 680	15.270254/8.265845	1.847392	0.934436
24	4	V	E		$V:\epsilon$	12 : 680	15.736479/8.265845	1.903796	0.969008
25	1	C		E	$C:\epsilon$	3 : 680	7.935525/8.265845	0.960038	0.588882
		Term C (#1) Consolidated Into ϵ [New SSE/df = 5644.581055/683].							
26	23	MTW	E		$MTW:\epsilon$	8 : 683	15.270254/8.264394	1.847716	0.934498
27	4	V	E		$V:\epsilon$	12 : 683	15.736479/8.264394	1.904130	0.969053
28	12	MW		E	$MW:MTW$	4 : 8	10.894315/15.270254	0.713434	0.394438
		Term MW (#12) Consolidated Into MTW (#23) [New SS/df = 165.739288/12].							
29	23	MTW	E		$MTW:\epsilon$	12 : 683	13.811607/8.264394	1.671218	0.931191
30	4	V	E		$V:\epsilon$	12 : 683	15.736479/8.264394	1.904130	0.969053
31	10	MT		E	$MT:MTW$	8 : 12	3.496434/13.811607	0.253152	0.029869
		Term MT (#10) Consolidated Into MTW (#23) [New SS/df = 193.710754/20].							
32	23	MTW		E	$MTW:\epsilon$	20 : 683	9.685538/8.264394	1.171960	0.727969
		Term MTW (#23) Consolidated Into ϵ [New SSE/df = 5838.291992/703].							
33	4	V	E		$V:\epsilon$	12 : 703	15.736479/8.304825	1.894860	0.968074
34	2	M		E	$M:\epsilon$	4 : 703	10.636625/8.304825	1.280777	0.723981
		Term M (#2) Consolidated Into ϵ [New SSE/df = 5880.838379/707].							
35	4	V	E		$V:\epsilon$	12 : 707	15.736479/8.318018	1.891855	0.967744

* Created Tuesday, April 02, 2013 at 21:41:09 EDT for configuration/data file prefix EMSwthSSMSwthFTwthK4a.

Figure 7: `FFRFR` Dataset Disposition Table Part 2

2. *Non-linear Models*

   Many other ANOVA models may be used in a statistical analysis embedded system besides linear models. Nesting becomes problematical in these cases since the nature of the factors moves from a simple combination of factors to functions of multiple factors. Furthermore, the orthogonalization of the sum of squares partition of the dataset variation requires significantly different calculations in non-linear models compared to linear models. This issue pertains to the independent and covariate variables in the model, and not to the dependent variable (which may be arbitrarily transformed without materially affecting the analytical nature of the embedded system calculations).

3. *Confounding*

   Covariates may be introduced into ANOVA models to assess the extent of spurious relationships that may be found in a given dataset. Designating a factor as a covariate to another particular factor, or groups of factors, in the configuration file, and processed accordingly within the correct analytical structures, may help explain the relationship between the primary factors and the dependent variable of interest.

4. *Consolidation Policies*

   When an F-test statistic if formed based on the EMS terms, and that ratio is deemed not significant in the dataset, then alternate methods are possible for consolidating the sum of squares of the not-significant factor into those that are deemed significant. These other procedures could be based on other statistics or on non-analytical policies. The consolidation of model terms based on approximate F-test statistics may also be defined.

Source	df	SS	MS	F	T*	(N):D**	Signf	EMS
C_i	1	1.270020	1.270020	0.173464	A	$C + CTW\ (2) : CT + CW\ (2)$	0.147822	$\sigma_\epsilon^2 + 64\sigma_{CTW}^2 + 128\sigma_{CW}^2 + 128\sigma_{CT}^2 + 256\phi_C$
H_j	1	11.556028	11.556028	1.334421	A	$H + HTW\ (2) : HT + HW\ (2)$	0.571628	$\sigma_\epsilon^2 + 64\sigma_{HTW}^2 + 128\sigma_{HW}^2 + 128\sigma_{HT}^2 + 256\phi_H$
$M_{k(j)}$	2	34.051479	17.025740	7.352680	A	$M + MTW\ (4) : MT + MW\ (4)$	0.960432	$\sigma_\epsilon^2 + 32\sigma_{MTW}^2 + 64\sigma_{MW}^2 + 64\sigma_{MT}^2 + 128\phi_M$
T_l	1	0.196095	0.196095	0.004164	E	$- : TW$	0.040964	$\sigma_\epsilon^2 + 128\sigma_{TW}^2 + 256\sigma_T^2$
$V_{m(l)}$	2	18.027029	9.013515	0.408761	E	$- : VW$	0.290156	$\sigma_\epsilon^2 + 64\sigma_{VW}^2 + 128\phi_V$
W_n	1	9.105778	9.105778	0.193355	E	$- : TW$	0.263330	$\sigma_\epsilon^2 + 128\sigma_{TW}^2 + 256\sigma_W^2$
Z_o	1	0.124376	0.124376	2.083245	A	$Z + TWZ\ (2) : TZ + WZ\ (2)$	0.675667	$\sigma_\epsilon^2 + 64\sigma_{TWZ}^2 + 128\sigma_{WZ}^2 + 128\sigma_{TZ}^2 + 256\phi_Z$
CH_{ij}	1	20.034451	20.034451	0.675036	A	$CH + CHTW\ (2) : CHT + CHW\ (2)$	0.402998	$\sigma_\epsilon^2 + 32\sigma_{CHTW}^2 + 64\sigma_{CHW}^2 + 64\sigma_{CHT}^2 + 128\phi_{CH}$
$CM_{ik(j)}$	2	11.592288	5.796144	0.719985	A	$CM + CMTW\ (4) : CMT + CMW\ (4)$	0.378978	$\sigma_\epsilon^2 + 16\sigma_{CMTW}^2 + 32\sigma_{CMW}^2 + 32\sigma_{CMT}^2 + 64\phi_{CM}$
CT_{il}	1	5.938320	5.938320	9.260131	E	$- : CTW$	0.795068	$\sigma_\epsilon^2 + 64\sigma_{CTW}^2 + 128\sigma_{CT}^2$
$CV_{im(l)}$	2	14.278725	7.139362	0.606677	E	$- : CVW$	0.377597	$\sigma_\epsilon^2 + 32\sigma_{CVW}^2 + 64\phi_{CV}$
CW_{in}	1	5.080078	5.080078	7.921802	E	$- : CTW$	0.780074	$\sigma_\epsilon^2 + 64\sigma_{CTW}^2 + 128\sigma_{CW}^2$
CZ_{io}	1	51.447830	51.447830	4.995749	A	$CZ + CTWZ\ (2) : CTZ + CWZ\ (2)$	0.833215	$\sigma_\epsilon^2 + 32\sigma_{CTWZ}^2 + 64\sigma_{CWZ}^2 + 64\sigma_{CTZ}^2 + 128\phi_{CZ}$
HT_{jl}	1	9.010013	9.010013	11.838619	E	$- : HTW$	0.816764	$\sigma_\epsilon^2 + 64\sigma_{HTW}^2 + 128\sigma_{HT}^2$
$HV_{jm(l)}$	2	2.852114	1.426057	0.170341	E	$- : HVW$	0.145548	$\sigma_\epsilon^2 + 32\sigma_{HVW}^2 + 64\phi_{HV}$
HW_{jn}	1	0.220282	0.220282	0.289437	E	$- : HTW$	0.313726	$\sigma_\epsilon^2 + 64\sigma_{HTW}^2 + 128\sigma_{HW}^2$
HZ_{jo}	1	1.500778	1.500778	0.614077	A	$HZ + HTWZ\ (2) : HTZ + HWZ\ (2)$	0.380451	$\sigma_\epsilon^2 + 32\sigma_{HTWZ}^2 + 64\sigma_{HWZ}^2 + 64\sigma_{HTZ}^2 + 128\phi_{HZ}$
$MT_{k(j)l}$	2	4.381854	2.190927	0.093627	E	$- : MTW$	0.085612	$\sigma_\epsilon^2 + 32\sigma_{MTW}^2 + 64\sigma_{MT}^2$
$MV_{k(j)m(l)}$	4	9.397420	2.349355	0.557530	E	$- : MVW$	0.292669	$\sigma_\epsilon^2 + 16\sigma_{MVW}^2 + 32\phi_{MV}$
$MW_{k(j)n}$	2	6.614473	3.307236	0.141332	E	$- : MTW$	0.123831	$\sigma_\epsilon^2 + 32\sigma_{MTW}^2 + 64\sigma_{MW}^2$
$MZ_{k(j)o}$	2	53.019146	26.509573	2.345424	A	$MZ + MTWZ\ (4) : MTZ + MWZ\ (3)$	0.745184	$\sigma_\epsilon^2 + 16\sigma_{MTWZ}^2 + 32\sigma_{MWZ}^2 + 32\sigma_{MTZ}^2 + 64\phi_{MZ}$
TW_{ln}	1	47.093513	47.093513	5.839737	E	$- : \epsilon$	0.981077	$\sigma_\epsilon^2 + 128\sigma_{TW}^2$
TZ_{lo}	1	10.620288	10.620288	0.273042	E	$- : TWZ$	0.306058	$\sigma_\epsilon^2 + 64\sigma_{TWZ}^2 + 128\sigma_{TZ}^2$
$VW_{m(l)n}$	2	44.101646	22.050823	2.734368	E	$- : \epsilon$	0.933800	$\sigma_\epsilon^2 + 64\sigma_{VW}^2$
$VZ_{m(l)o}$	2	4.363565	2.181783	2.129573	E	$- : VWZ$	0.680467	$\sigma_\epsilon^2 + 32\sigma_{VWZ}^2 + 64\phi_{VZ}$
WZ_{no}	1	8.110378	8.110378	0.208513	E	$- : TWZ$	0.272279	$\sigma_\epsilon^2 + 64\sigma_{TWZ}^2 + 128\sigma_{WZ}^2$
CHT_{ijl}	1	26.028112	26.028112	362.883667	E	$- : CHTW$	0.949044	$\sigma_\epsilon^2 + 32\sigma_{CHTW}^2 + 64\sigma_{CHT}^2$
$CHV_{ijm(l)}$	2	1.766320	0.883160	0.132420	E	$- : CHVW$	0.116936	$\sigma_\epsilon^2 + 16\sigma_{CHVW}^2 + 32\phi_{CHV}$
CHW_{ijn}	1	3.757226	3.757226	52.383198	E	$- : CHTW$	0.905922	$\sigma_\epsilon^2 + 32\sigma_{CHTW}^2 + 64\sigma_{CHW}^2$
CHZ_{ijo}	1	19.422028	19.422028	3.790019	A	$CHZ + CHTWZ\ (2) : CHTZ + CHWZ\ (2)$	0.791233	$\sigma_\epsilon^2 + 16\sigma_{CHTWZ}^2 + 32\sigma_{CHWZ}^2 + 32\sigma_{CHTZ}^2 + 64\phi_{CHZ}$
$CMT_{ik(j)l}$	2	17.825098	8.912549	1.992833	E	$- : CMTW$	0.665869	$\sigma_\epsilon^2 + 16\sigma_{CMTW}^2 + 32\sigma_{CMT}^2$
$CMV_{ik(j)m(l)}$	4	26.480011	6.620003	2.119655	E	$- : CMVW$	0.757620	$\sigma_\epsilon^2 + 8\sigma_{CMVW}^2 + 16\phi_{CMV}$
$CMW_{ik(j)n}$	2	10.698954	5.349477	1.196135	E	$- : CMTW$	0.544655	$\sigma_\epsilon^2 + 16\sigma_{CMTW}^2 + 32\sigma_{CMW}^2$
$CMZ_{ik(j)o}$	2	30.373911	15.186955	0.625301	A	$CMZ + CMTWZ\ (4) : CMTZ + CMWZ\ (3)$	0.322638	$\sigma_\epsilon^2 + 8\sigma_{CMTWZ}^2 + 16\sigma_{CMWZ}^2 + 16\sigma_{CMTZ}^2 + 32\phi_{CMZ}$
CTW_{iln}	1	0.641278	0.641278	0.079520	E	$- : \epsilon$	0.221574	$\sigma_\epsilon^2 + 64\sigma_{CTW}^2$
CTZ_{ilo}	1	5.990126	5.990126	0.178607	E	$- : CTWZ$	0.254164	$\sigma_\epsilon^2 + 32\sigma_{CTWZ}^2 + 64\sigma_{CTZ}^2$
$CVW_{im(l)n}$	2	23.535971	11.767985	1.459265	E	$- : \epsilon$	0.766307	$\sigma_\epsilon^2 + 32\sigma_{CVW}^2$
$CVZ_{im(l)o}$	2	14.288625	7.144312	1.101904	E	$- : CVWZ$	0.524241	$\sigma_\epsilon^2 + 16\sigma_{CVWZ}^2 + 32\phi_{CVZ}$
CWZ_{ino}	1	11.021512	11.021512	0.328627	E	$- : CTWZ$	0.330849	$\sigma_\epsilon^2 + 32\sigma_{CTWZ}^2 + 64\sigma_{CWZ}^2$
HTW_{jln}	1	0.761070	0.761070	0.094375	E	$- : \epsilon$	0.240791	$\sigma_\epsilon^2 + 64\sigma_{HTW}^2$
HTZ_{jlo}	1	3.251250	3.251250	0.828660	E	$- : HTWZ$	0.469292	$\sigma_\epsilon^2 + 32\sigma_{HTWZ}^2 + 64\sigma_{HTZ}^2$
$HVW_{jm(l)n}$	2	16.743515	8.371758	1.038123	E	$- : \epsilon$	0.644889	$\sigma_\epsilon^2 + 32\sigma_{HVW}^2$
$HVZ_{jm(l)o}$	2	0.479770	0.239885	0.262338	E	$- : HVWZ$	0.207819	$\sigma_\epsilon^2 + 16\sigma_{HVWZ}^2 + 32\phi_{HVZ}$
HWZ_{jno}	1	5.581976	5.581976	1.422703	E	$- : HTWZ$	0.554723	$\sigma_\epsilon^2 + 32\sigma_{HTWZ}^2 + 64\sigma_{HWZ}^2$
$MTW_{k(j)ln}$	2	46.800991	23.400496	2.901732	E	$- : \epsilon$	0.943866	$\sigma_\epsilon^2 + 32\sigma_{MTW}^2$
$MTZ_{k(j)lo}$	2	28.061979	14.030990	0.736428	E	$- : MTWZ$	0.424105	$\sigma_\epsilon^2 + 16\sigma_{MTWZ}^2 + 32\sigma_{MTZ}^2$
$MVW_{k(j)m(l)n}$	4	16.855438	4.213860	0.522531	E	$- : \epsilon$	0.280767	$\sigma_\epsilon^2 + 16\sigma_{MVW}^2$
$MVZ_{k(j)m(l)o}$	4	21.815788	5.453947	0.680911	E	$- : MVWZ$	0.359337	$\sigma_\epsilon^2 + 8\sigma_{MVWZ}^2 + 16\phi_{MVZ}$
$MWZ_{k(j)no}$	2	1.077110	0.538555	0.028267	E	$- : MTWZ$	0.027489	$\sigma_\epsilon^2 + 16\sigma_{MTWZ}^2 + 32\sigma_{MWZ}^2$
TWZ_{lno}	1	38.896198	38.896198	4.823246	E	$- : \epsilon$	0.968787	$\sigma_\epsilon^2 + 64\sigma_{TWZ}^2$
$VWZ_{m(l)no}$	2	2.049033	1.024516	0.127043	E	$- : \epsilon$	0.119267	$\sigma_\epsilon^2 + 32\sigma_{VWZ}^2$
$CHTW_{ijln}$	1	0.071726	0.071726	0.008894	E	$- : \epsilon$	0.074978	$\sigma_\epsilon^2 + 32\sigma_{CHTW}^2$
$CHTZ_{ijlo}$	1	2.650753	2.650753	0.108854	E	$- : CHTWZ$	0.202577	$\sigma_\epsilon^2 + 16\sigma_{CHTWZ}^2 + 32\sigma_{CHTZ}^2$
$CHVW_{ijm(l)n}$	2	13.338731	6.669365	0.827021	E	$- : \epsilon$	0.561872	$\sigma_\epsilon^2 + 16\sigma_{CHVW}^2$
$CHVZ_{ijm(l)o}$	2	5.163595	2.581798	0.088222	E	$- : CHVWZ$	0.081070	$\sigma_\epsilon^2 + 8\sigma_{CHVWZ}^2 + 16\phi_{CHVZ}$
$CHWZ_{ijno}$	1	8.898926	8.898926	0.365437	E	$- : CHTWZ$	0.345593	$\sigma_\epsilon^2 + 16\sigma_{CHTWZ}^2 + 32\sigma_{CHWZ}^2$
$CMTW_{ik(j)ln}$	2	8.944601	4.472301	0.554579	E	$- : \epsilon$	0.425227	$\sigma_\epsilon^2 + 16\sigma_{CMTW}^2$
$CMTZ_{ik(j)lo}$	2	61.080589	30.540295	2.934711	E	$- : CMTWZ$	0.745852	$\sigma_\epsilon^2 + 8\sigma_{CMTWZ}^2 + 16\sigma_{CMTZ}^2$
$CMVW_{ik(j)m(l)n}$	4	12.492604	3.123151	0.387280	E	$- : \epsilon$	0.182255	$\sigma_\epsilon^2 + 8\sigma_{CMVW}^2$
$CMVZ_{ik(j)m(l)o}$	4	74.679626	18.669907	1.880092	E	$- : CMVWZ$	0.722050	$\sigma_\epsilon^2 + 4\sigma_{CMVWZ}^2 + 8\phi_{CMVZ}$
$CMWZ_{ik(j)no}$	2	0.314332	0.157166	0.015103	E	$- : CMTWZ$	0.014878	$\sigma_\epsilon^2 + 8\sigma_{CMTWZ}^2 + 16\sigma_{CMWZ}^2$
$CTWZ_{ilno}$	1	33.538052	33.538052	4.158819	E	$- : \epsilon$	0.955543	$\sigma_\epsilon^2 + 32\sigma_{CTWZ}^2$
$CVWZ_{im(l)no}$	2	12.967214	6.483607	0.803987	E	$- : \epsilon$	0.551706	$\sigma_\epsilon^2 + 16\sigma_{CVWZ}^2$
$HTWZ_{jlno}$	1	3.923501	3.923501	0.486526	E	$- : \epsilon$	0.513291	$\sigma_\epsilon^2 + 32\sigma_{HTWZ}^2$
$HVWZ_{jm(l)no}$	2	1.828828	0.914414	0.113390	E	$- : \epsilon$	0.107168	$\sigma_\epsilon^2 + 16\sigma_{HVWZ}^2$
$MTWZ_{k(j)lno}$	2	38.105530	19.052765	2.362600	E	$- : \epsilon$	0.904457	$\sigma_\epsilon^2 + 16\sigma_{MTWZ}^2$
$MVWZ_{k(j)m(l)no}$	4	32.039116	8.009779	0.993237	E	$- : \epsilon$	0.588937	$\sigma_\epsilon^2 + 8\sigma_{MVWZ}^2$
$CHTWZ_{ijlno}$	1	24.351477	24.351477	3.019656	E	$- : \epsilon$	0.914931	$\sigma_\epsilon^2 + 16\sigma_{CHTWZ}^2$
$CHVWZ_{ijm(l)no}$	2	58.529305	29.264652	3.628905	E	$- : \epsilon$	0.972540	$\sigma_\epsilon^2 + 8\sigma_{CHVWZ}^2$
$CMTWZ_{ik(j)lno}$	2	20.813154	10.406577	1.290447	E	$- : \epsilon$	0.723662	$\sigma_\epsilon^2 + 8\sigma_{CMTWZ}^2$
$CMVWZ_{ik(j)m(l)no}$	4	39.721260	9.930315	1.231389	E	$- : \epsilon$	0.703022	$\sigma_\epsilon^2 + 4\sigma_{CMVWZ}^2$
$\epsilon_{p(ijklmno)}$	384	3096.699219	8.064321					σ_ϵ^2

* (E) Exact; (A) Approximate – (df); (U) Untestable
** No Numerator Given For Exact Tests

Figure 8: Example Of A Complex Summary Table

5. *Alternate Approximate F-Test Formation*

   When an exact F-test is not available, the Welch-Satterthwaite and balanced degrees of freedom approaches to form approximate F-test statistics may be replaced by other methods that may be conditional and dependent on other aspects of a statistical analysis besides the EMS terms. Each of these approaches have its own analytical characteristics that would be addressed in the theoretical formulation and planning stages.

6. *Multiple And Dependent Significance Levels*

   The significance level used with any dataset may not only depend on the model used, but also consolidation policies, interaction levels, the history of the test results, etc. Dependent significance levels and multiple levels that change as the analysis proceeds would dramatically increase the intelligence of the data analyst build into the embedded system.

7. *Anomalous Data Detection And Defense*

   "There is always something wrong with the data" is a common wisdom among data analysts. Anomalous data detection and policies for defending against their effects within statistical analyses would be a significantly powerful value-added feature of any such embedded system. The utility of including data quality filtering, repair, and compensation methods within the embedded system must be balanced against the advantages of performing such screening on the dataset prior to its use in the embedded system.

8. *Process Signaling*

   Every embedded system operates in the context of a control system that is in turn part of a larger operational system. An automated ANOVA embedded system may be defined to not only make calculations related to the variability contributions of terms in a model, but also signaling to the control system and to the operating system the presence of precursors, occurrences, and effects of analytical situations encountered during its operation that might otherwise not be sensed by those other systems.

9. *Dynamic Display Of Summaries And Reports*

   Static displays of the summary and disposition tables do not allow a data analyst the advantages of dealing with analytical situations of interest as they occur during a statistical analysis. A dynamic display of these reports, changing in real time with opportunities for interrupts from parallel processes, would provide the data analyst advanced methods for fine-tuning the analytical methods implemented in the embedded system.

10. *Use Of Non-numeric Datasets*

    An advanced use of embedded systems goes beyond the strict use of numeric datasets. Non-numeric information in the form of Unicode characters, images, reductions of datasets, state symbols, etc., would provide for arbitrarily complex embedded system functionality.

11. *Recalculation Triggers In Pipelining Processors*

    Just as databases have triggers, so may embedded systems have triggers that operate only under particular circumstances or after a specific set of actions from parallel processes. Pipelining processes would require this sort of interaction capability. In fact, defining so-called "micro-operations" at the processor level is an analogy to casting a statistical calculation as a sequence of minimal instructions. This allows the controlling parallel processor to perform at the processing rate of the most rudimentary micro-operation, thereby allowing for a net higher processing rate as compared to processing one complete instruction at a time.

12. *Updates Upon Annexing And/Or Deleting Data*

    To reduce the inefficiencies and redundancies commonly found in statistical analyses, many such calculations may be completed more effectively and efficiently by simply updating previously calculated values when new data is annexed to existing datasets and/or old data is deleted from them. This would be especially useful for embedded systems that monitor a process and react to a data stream rather than a static dataset. Embedded systems involving datasets with moving windows and snapshots of states would also benefit from this capability.

7. For More Information

The full documentation of a high-level programming implementation of the automated ANOVA embedded system described in this paper consists of four copyrighted PQI Consulting technical memoranda using the CWEB literate programming documentation system. Those memoranda and the high-level ANSI C implementation code are available under the GNU General Public License Version 3 (GNU GPLv3) by request from the author at info@pqic.com or at the PQI Consulting mailing address.

REFERENCES

Knuth, D. E., *Computers & Typesetting, Volume A: The TEXbook*, Reading, MA:Addison-Wesley, 1984 (numerous editions outstanding)

Knuth, D. E., *Literate Programming, CSLI Lecture Notes* #27, Stanford, CA: Center for the Study of Language and Information, 1992

Knuth, D. E., Levy, S., *The CWEB System of Structured Documentation*, Reading, MA: Addison-Wesley, 1993

Free Software Foundation, *The GNU General Public License*, Boston, MA: Free Software Foundation, Version 3 dated June 29, 2007, accessed through http://www.gnu.org/licenses/gpl.html

Institute of Electrical and Electronics Engineers, *IEEE Standard Verilog Hardware Description Language*, New York, NY: IEEE Computer Society, 2001 (edition of January 7, 2004)

The Unicode Consortium, *The Unicode Standard, Version 5.0*, Fifth Edition, Reading, MA: Addison-Wesley, 2006 (ASCII is a subset of the UTF-8 scheme within Unicode)