Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

"Human inside"

From the potential ad campaign

Dossier

Aleksey Popelyukhin is a Senior Vice-President of Technology with the Sam Sebe LLC and a Vice-President of Information Systems with the Commercial Risk Re in Stamford, Connecticut. He holds a Ph.D. in Mathematics and Mathematical Physics from Moscow University (1988).

His actuarially related achievements include:

- *Prize for the best 1997 article in the "Data Management discussion paper" program entitled "The Big Picture: Actuarial Process from the Data Management point of view" (1996)*

- *Creation and distribution of the popular actuarial utilities like Triangle Maker™ (1994) and Triangle Maker™ Pro (1997), Actuarial Toolchest™ (1998) and Enabler™ (1999)*

- *Design, development and coding of the 2nd and 3rd (current) generation of the very powerful and flexible actuarial software package called Affinity (1996)*

- *Promotion (through his papers and presentations) of his notions like Ideal Actuarial System and Data Quality Shield, and paradigms like object-oriented actuarial software and data-driven visualization.*

Aleksey is presently developing an integrated pricing/reserving/DFA computer system for reinsurance and also an action/adventure computer game tentatively called "Actuarial Judgement". Dr. Popelyukhin is an active member of several scientific societies and an author of almost 20 scientific publications.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Abstract

No one would argue that there are limits on how much information a human being can perceive, process and comprehend. Even as advances in computer technology throw more and more data at actuaries, these limits stay the same. It is time to delegate to computers the very important task of *presentation of information*.

The article will try to demonstrate how existing *data-driven* technologies can help to evolve an *Ideal Actuarial System* from an actuarial tool into a company's *Alarm System.* Utilizing tools readily available to everyone who owns a contemporary Office Suite package, actuaries can present information in such a way that the effectiveness of *Corporate Decision Making, Data Error Detection* and suitability of *Actuarial Algorithms* will increase dramatically.

Actuarial results, properly combined, summarized and filtered **by importance,** may be arranged into a so-called *Digital Dashboard* that serves as a *portal* into the wealth of detailed actuarial information and the calculations behind it. This article itself can be considered as a portal that refers actuaries to the wealth of information on visualization techniques and data-driven technologies.

Let Me See:
Visualizing Actuarial Information

Aleksey S. Popelyukhin, Ph.D.

Introduction

"Let's see…"

Graeae to Polyphemus

The Actuarial Process, like every analytical process, consists of three major stages:

- *Data Collection, Cleanup and Transformation*

- *Application of Algorithms and Methods*

- *Representation of Results*
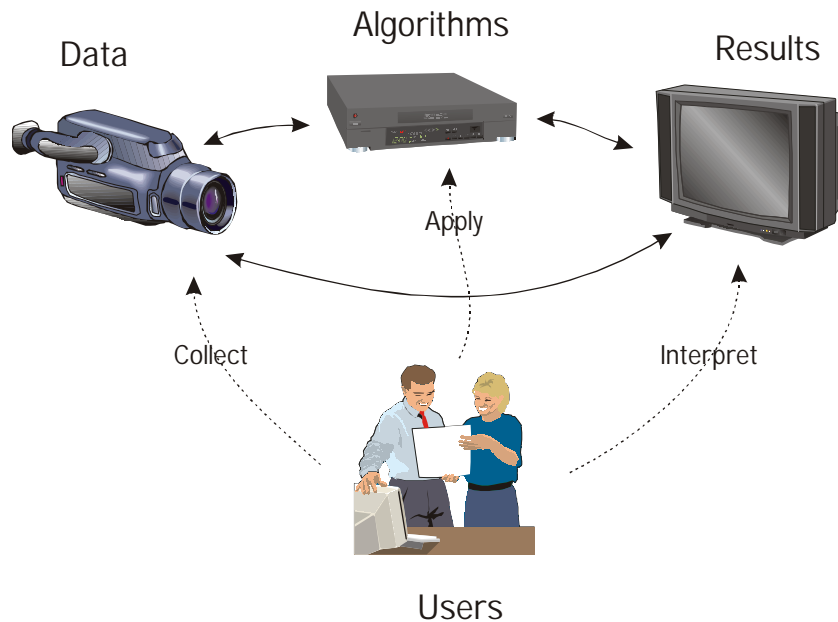
All these stages require human interaction:

Figure 1

The amount of information to be processed is so vast, though, that there is no way to do it without computers.

But, computers are used strikingly differently throughout the process: in the first two stages we *serve* computers (feeding them with data and supplying them with actuarial parameters selection), and only in the last stage computers *serve* us (providing us with answers for making decisions). The irony here is that first two stages (*data transformation* and *application of algorithms*) have been automated to a fair degree, while *presentation of results* stage still remains the least computerized of all, and tools available for reporting and visualization are severely underused by actuaries.

This article is the first attempt to correct this situation.

- ***Approaches***

The amount of numeric information at all stages of the Actuarial Process is either exceedingly massive (raw data), overwhelming the recipient, or too small (summary) to adequately represent all the nuances and data patterns.

Human perception relies heavily on Short-Term Memory (STM) – a small "buffer" where external information is recognized and perceived (see [1]). Unfortunately, STM is limited in capacity – it can hold only 5-7 similar items at once. Another problem with STM is that new information replaces old (it's just a "buffer"), and new information is coming in continuously. Every change (in color, size or position) attracts a person's attention and changes the focus of his perception. Even in the best-case scenario, without any distractions, STM can hold information only 30 seconds or so♣.

A few approaches seem to alleviate these limitations of human perception (and even exploit its weaknesses to attract a person's attention toward the important information): visualization, adaptive reporting and alarm systems. Indeed, "traditional" ways of displaying just myriad "boring" numbers in a spreadsheet are not adequate anymore. Without proper assistance, it is practically impossible to notice imperfections in the data; the inapplicability of a particular actuarial technique; or to pinpoint a claim, line or location that demonstrate unusual development.

The solution lies in augmenting standard report techniques with the information ***filtered by importance***. It means that only a few outstanding items with ***alarming*** behavior show up (or somehow get highlighted) in the report. The task of defining alarms and assigning levels of importance to actuarial results lies squarely on actuarial shoulders.

It is important to realize that nowadays, with the proliferation of Office Suites software, a wide variety of visualization tools is within the reach of every actuary. Almost every chapter of this article is illustrated by an example from an Office program. Equally important, one can safely assume that everybody understands the text of the BASIC[i] program. Coding in VBA has become a skill nearly as essential and vital as reading and writing.

♣ Conduct the following experiment: read a new telephone number digit-by-digit and then, (without attempting to repeat it, combine digits into groups or make associations and other mnemonic rules) after 30 seconds pass, try to dial it. Even without distraction it is practically impossible - that is probably why 411-type services repeat telephone numbers at least two times.

Adaptive Reporting

Every company almost certainly has established a fixed way of presenting the results of actuarial analysis. The overwhelming majority of these presentations are *static* reports with predefined content and layout: think of it as a list of reserves for 100 lines of business or a list of net present value of premium for 1000 treaties. There is nothing wrong with that way of presenting information, except that human perception cannot effectively span beyond 5-7 similar items. Nobody can guarantee that equal attention will be paid to every item in the long, monotonous report. To alleviate this problem, sometimes information is presented in a summarized form without important details. Either way, important information about the 68th LOB or the 729th treaty may escape the reader's attention.

The solution lies in the use of data-driven technologies to create *dynamic* or *adaptive* reports. Reports whose size, shape and format adapts to the data. Placing these reports in an interactive environment such as a spreadsheet allows the user to interact dynamically with the report (effectively creating a whole family of reports rather than a single one), shaping it to the level of detail that suits the user.

A partial list of data-driven implementations found in spreadsheets includes:

- *Filtering,*

- *Outlining,*

- *Sorting,*

- *Conditional Formatting and*

- *OLAP-enabled tools.*

- **Filtering**

The simplest and most straightforward way to reduce the amount of information displayed in the report is Filtering. If information is organized as a list or a table in a spreadsheet and there is an easy way to define relevant subset, then Filtering fits the bill.

The AutoFilter feature of Microsoft Excel is a powerful and elegant implementation of Filtering. Accessible and customizable through either an interface or the VBA "macro" language, AutoFilter serves as an ideal Filtering tool[*].

---

[*] Reporting and Visualization tools, including Filtering, are available in many products, not only in Excel. In fact, database products with built-in SQL language provide much more powerful and robust Filtering tool. However, these products may be outside the reach of many actuaries.

One can use this tool, for example, to limit the visible data table to a particular LOB[ii], location and Open/Closed status. Another interesting use of Filtering tools, and Excel's AutoFilter in particular, is checking for all distinct values in a list: AutoFilter's drop-down boxes display a sorted list of all distinct values present in a column. The fastest way to check if a huge spreadsheet populated with the data contains all requested LOB's, States or Policy Numbers is to initiate AutoFilter and click on the down-arrow in the corresponding column.

	A	B	C	D	E	F	G
1	Company ▼	State ▼	LOB ▼	Paid/Inc ▼	AccYear ▼	EvalYear ▼	NetLoss ▼
112	ABC	CT	(All)	Paid	1988	1988	666,434
113	ABC	CT	(Top 10...)	Paid	1988	1989	1,143,933
114	ABC	CT	(Custom...)	Paid	1988	1990	1,342,429
115	ABC	CT	AL	Paid	1988	1991	1,432,131
116	ABC	CT	GL / WC	Paid	1988	1992	1,525,594
117	ABC	CT	(Blanks)	Paid	1988	1993	1,625,430
118	ABC	CT	(NonBlanks)	Paid	1988	1994	1,682,298
119	ABC	CT	GL	Paid	1988	1995	1,702,690
120	ABC	CT	GL	Paid	1988	1996	1,766,216
121	ABC	CT	GL	Paid	1988	1997	1,730,972
122	ABC	CT	GL	Paid	1989	1989	704,188
123	ABC	CT	GL	Paid	1989	1990	1,272,152
124	ABC	CT	GL	Paid	1989	1991	1,464,204

Figure 2

Example 1. For an illustration of using AutoFilter for something less straightforward than plain-vanilla filtering (i.e., by LOB or Location), observe how to use it to filter losses in the 90th percentile of their Incurred Value:

```
Const colLOSS As Integer = 6        'Loss value is located in the column number colLOSS
Sub CreativeUseOfAutoFilter()
    Dim rRange As Range,nRows As Long
    ActiveSheet.Cells(1, 1).Select
    With Selection
      nRows = .CurrentRegion.Rows.Count
      Set rRange = .Offset(1, colLOSS - 1).Resize(nRows - 1, 1)
      .AutoFilter
      .AutoFilter Field:=colLOSS, Criteria1:=">" & Application.Percentile(rRange, 0.9)
    End With
End Sub
```

Filtering is a fast and effective way to cut down the amount of data displayed. However, if a filtered subset is still too large, or there is a need to see different levels of detail for different groups of data, or the user has to see different aggregate values (subtotals, averages), then Outlining or Pivot Tables would be a better choice of tools.

- *Outlining*

Outlining is a hierarchical representation of data organized into a list or table, with the ability to hide or display details of all or selected groups on any level of the hierarchy. Every user of Windows Explorer or any other File Directory tool is familiar with the notion of Outlining. Excel's implementation of Outline allows both horizontal (rows/records) and vertical

(columns/fields) outlining. Along with the ability to display detailed records, Excel's Outline supports aggregate functions such as sum, count, average, standard deviation, etc. This capability may become handy in situations when different members of a hierarchy are treated at different levels of detail.

Example 2. A screenshot below illustrates Outline's advantages over AutoFilter: it allows the display of different levels of detail for different LOB's on the same screen. While WC information is shown up to the State and Accident Year levels, GL and AL are shown only as aggregates without unnecessary details.

Figure 3

Example 3. Microsoft Access 2000's tables with properly defined "master-detail" relationships can display records in an Outline fashion too, making table navigation as simple and intuitive as browsing a directory tree in Windows Explorer.
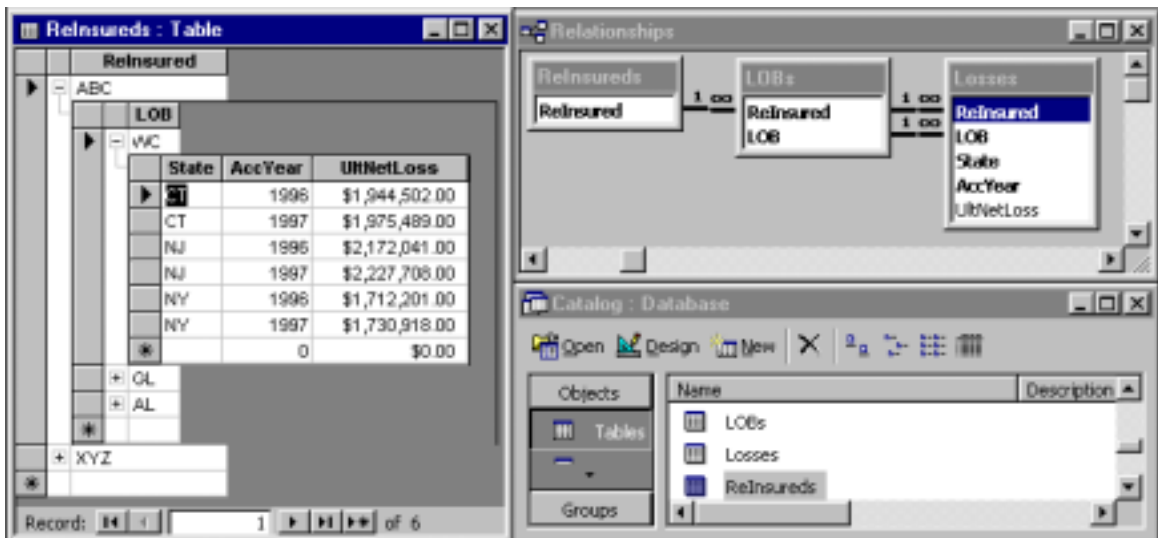

Figure 4

5

- ***Sorting***

Sorting is a powerful technique that brings the most important records to the top of the display or, in the case of printed report, to the first page. Sorting does not reduce the amount of data displayed, but it assures that the first several records get more attention before the report reader gets tired. Consequently, the main skill of using Sorting is in defining what constitutes *importance*. An ability to rank information in accordance to its significance, and to identify which actuarial measurements affect decision-making *the most* is a yardstick that separates great report designers from mere mortals.

"Combined Loss Ratios," "Net Profitability" and "Reserve Adequacy" – all these measurements may serve as actuarial significance indicators. Both more sophisticated ones like "Percent Change of the Current Estimate of the Net Ultimate Loss Hindsight Estimated" or simpler ones like "Time since the Latest Loss Run" help to sort data and generate useful actuarial information.

UW Year	Policy	Premium	Aggr Limit	Aggr Ret	Prob to lose 20% Prem
1998	NWA-00019	$ 2,944,691	$ 5,700,000	$ -	0.22
2000	NWA-00003	$ 913,502	$ 2,500,000	$ -	0.21
2000	NWA-00007	$ 6,736,411	$ 8,100,000	$ -	0.21
1998	NWA-00011	$ 4,226,719	$ 9,900,000	$ -	0.21
2000	NWA-00030	$ 1,224,858	$ 2,100,000	$ -	0.19
2000	NWA-00006	$ 7,994,304	$ 24,600,000	$ -	0.18
1998	NWA-00012	$ 1,594,191	$ 8,500,000	$ 2,800,000	0.18
1998	NWA-00002	$ 5,818,746	$ 29,100,000	$ 9,700,000	0.17
2000	NWA-00015	$ 9,135,813	$ 31,600,000	$ -	0.16
1998	NWA-00010	$ 4,637,832	$ 5,200,000	$ -	0.16
2000	NWA-00017	$ 8,455,004	$ 45,200,000	$15,000,000	0.16
2000	NWA-00001	$ 5,817,872	$ 33,200,000	$11,000,000	0.16
1998	NWA-00023	$ 8,313,738	$ 13,800,000	$ -	0.15
1998	NWA-00032	$ 254,889	$ 900,000	$ -	0.15
1998	NWA-00024	$ 2,723,106	$ 10,500,000	$ -	0.14
1999	NWA-00028	$ 6,413,104	$ 23,100,000	$ -	0.13

Sort   20%

Aleksey Popelyukhin: Enter desired % and press "Sort" button to resort

Figure 5

Example 4. In a situation where a new significance indicator has to be added or modified, there is the need to copy an analytical expression into all the cells adjacent to the existing data table. The code below helps in these situations:

```
Sub FillAdjacentColumn()
    With ActiveCell
        .Resize(.CurrentRegion.Rows.Count + .CurrentRegion.Row - .Row, 1).Formula _
        = .Formula
    End With
End Sub
```

- ***Conditional Formatting***

Conditional Formatting is a feature of Microsoft Excel that allows users to define the font, color, border and background pattern of a cell as a function of the values in other cells. When values in the referenced cells change so does the conditional format. Despite some limitations (currently,

Excel currently supports up to 3 variable formats per cell), this feature opens unprecedented creative possibilities for report designers. Combined with other reporting techniques like Filtering, Sorting and Pivot Tables, Conditional Formatting is indispensable for attracting the report reader's attention to the most crucial information. Due to its dynamic nature, Conditional Formatting can serve as a building block for an actuarial Alarm System (a cell automatically becomes red when an actuarially significant value becomes too high or too low).

Given that the format condition's formula can be any expression that uses user-defined functions along with built-in ones, the use of Conditional Formatting is limited only by one's imagination: it's use can range from data-error detection to Thomas Mack-like assumption testing to statistical outlier warnings (see [2]).

> Example 5. A powerful yet simple application of Conditional Formatting would be highlighting statistical outliers. For example, it would be convenient to see values in a triangle of age-to-age factors that are too far (more than 2 or 3 standard deviations) from the column's average. Outliers like these frequently trigger an additional investigation of data that usually produces interesting results:

> Example 6. As an example of the user-defined format condition formula, one can use an algorithm that assigns different types to the treaties based on the relationships among parameters like Premium, Ultimate Loss, Aggregate Limit and others. Every treaty's record in this kind of report can be formatted in accordance with assigned type. And as the estimate of the ultimate net loss changes, the type of the treaty (potentially) changes, and so does the formatting.

Conditional Formatting used in combination with other reporting techniques is especially powerful. Sorting a list of records by one criterion (e.g., Net Profitability) and Conditionally Formatting them according to another criterion (e.g., treaties with a particular insurance company can create quite an impressive display and produce an easy-to-comprehend overall picture.

- *OLAP*

OLAP stands for Online Analytical Processing and there are many tools from numerous vendors that provide OLAP functionality. However, in order to demonstrate the accessibility of these tools, only one particular implementation of OLAP will be considered – Microsoft Excel's Pivot Tables.

A Pivot Table is essentially a mechanism for displaying multidimensional information in 2-D. Before data can be used in a Pivot Table, it is converted into multi-D array called an OLAP cube, where the table's fields become dimensions. That is, a $1,000,000 1999 WC California paid loss as of 6/30/2000, becomes an element (value, measure) of the OLAP cube with dimensions such as State, LOB, Accident Year, Evaluation Date stored at the intersection of the "CA," "WC," "1999," and "6/30/2000" members of these respective dimensions). There are only two meaningful ways to display multi-D info in 2-D: cross-section and projection with aggregation – and both ways are implemented in Excel's Pivot Tables. When used to display a cross-section of the data, a Pivot Table serves as a Filtering tool; when used to display aggregations (subtotals, averages, etc.), Pivot Table serves as an Outline[∗]. Pivot Tables also support Sorting and, to some degree, Conditional Formatting. So, in effect, this tool incorporates all the adaptive reporting techniques described above, and as such it should serve as a **preferred choice** for any report designer. In addition to that, Pivot Tables are capable of using external data sources and, consequently, are subject to fewer limitations than Excel's Sort, Filter and Outline.

[∗] Bear in mind, that for true Outlining, it is necessary to support hierarchies. Starting with Excel 2000, Pivot Tables' dimensions do support hierarchies. See [3] for examples of actuarial hierarchies.

The total flexibility of Pivot Tables may cause some problems for actuaries.

First, it is unclear which dimensions to choose for display and which ones for aggregation (or cross-section) to get actuarially meaningful results. Also, actuaries should define additional (calculated) fields with some kind of "actuarial significance" indicator, which can be later used in Sorting, Filtering or Conditional Formatting.

Second, unlike other professions where creating a Pivot Table is a destination – a final act of the analytical process – actuaries frequently use aggregated data as a starting point of their analysis (see [4]). If created properly, a Pivot Table can serve as a convenient storage for actuarial triangles with selective drill-down capabilities. One can create an OLAP Cube hierarchy in such a way that any suspicious element of the triangle can be drilled-down for details up to individual Claims and even individual payments level:
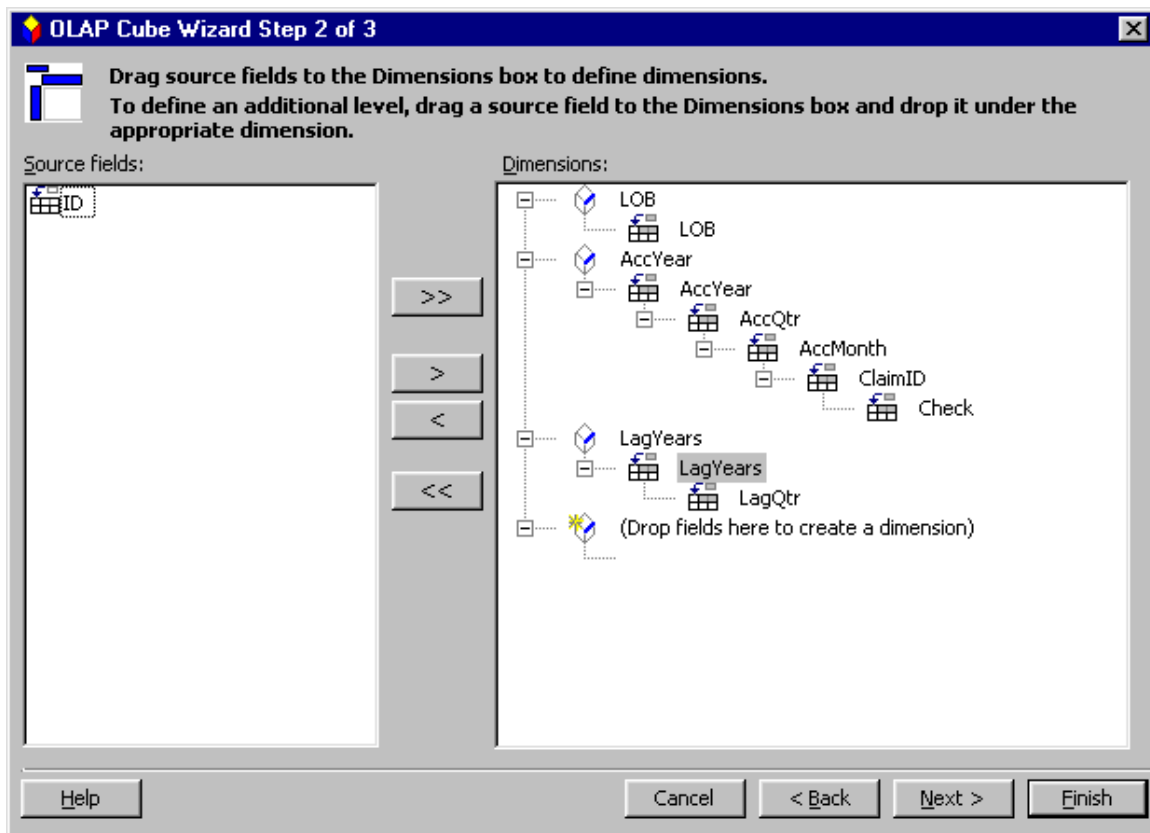
	A	B	C	D	E	F	G	H	I
1	LOB	WC ▼							
2									
3	Sum Of Loss					LagYears ▼			
4	AccYear ▼	AccQtr	AccMonth	ClaimID	Check	1	2	3	4
5	1996 *					$2,166,733	$411,335	$408,024	$366,587
6	1997 *					$1,094,331	$381,124	$117,174	
7	1998 *					$1,225,399	$67,132		
8		2 *				$39,240	$67,132		
9		3 *				$19,429			
10		4 *				$1,166,730			
11			11 *			$70,027			
12			12 *			$1,096,702			
13				C-00248 *		$1,070,676			
14					1023	$54,176			
15					1036	$12,514			
16					1694	$14,577			
17					1784	$989,409			
18				C-00581 *		$26,027			
19	1999 *					$2,047,854			

Figure 6

The problem is that with an unpredictable shape/size of the Pivot Table it is hard to incorporate its content into subsequent calculations. One workaround is to use the GETPIVOTDATA spreadsheet function, while another is to use the Pivot Tables' Calculated Fields – an ability to add fields/dimensions to the Pivot Table that are calculated on the fly.

> Example 7. Sometimes, a Calculated Field is the only mechanism to add new dimensions correctly. As an example, consider the Loss Ratio field. A Loss Ratio like any ratio is a nonlinear operation and, consequently cannot be summarized properly: a sum of ratios *is practically never*♣ a ratio of the sums. That is where Excel's Pivot Tables make a clear distinction between input fields (for subtotals, the *sums of ratios* are calculated) and Calculated Fields (for subtotals, the *ratios of sums* are calculated).

♣ Unless the ratios have the same absolute values and different signs.

PolicyID	UWyear	QShare	Claims	NetPremium	NetUltLoss	NetUltALAE
NWA-0960503	1997	50%	163	$ 685,337	$ 1,392,731	$ 840,791
	1998	50%	166	$ 1,111,135	$ 1,479,987	$ 765,345
	1999	50%	172	$ 1,149,034	$ 1,496,641	$ 765,345
NWA-0960503 Total			501	$ 2,945,506	$ 4,369,359	$ 2,371,481
NWA-0960511	1997	80%	622	$ 9,675,608	$ 11,289,812	$ 578,824
	1998					
	1999					
NWA-0960511 Total						
NWA-0960520	1996					
	1997					
	1998					
	1999					
NWA-0960520 Total						
NWA-0960542	1997					
	1998					
NWA-0960542 Total						
NWA-0960545	1997					
	1998					
	1999					
NWA-0960545 Total						
NWA-0960585	1997					
	1998	75%	318	$ 2,448,384	$ 2,842,144	$ 154,825
	1999	75%	331	$ 2,877,218	$ 3,115,155	$ 331,944
NWA-0960585 Total			942	$ 7,014,261	$ 8,980,813	$ 582,093

Insert Calculated Field

Name: NetLossRatio

Formula: = (NetLoss+ NetALAE) / NetPrem

Fields:
- PTDDeductible
- INCDeductible
- PTDExcess
- NetALAE
- NetPrem
- NetLoss
- PTDDedLayer2
- INCDedLayer2

[Insert Field]   [Add]   [Delete]   [OK]   [Close]

Example 8. Once a (calculated on the fly) Loss Ratio field is added to the Pivot Table, one can use it for Sorting and Filtering. A screenshot below illustrates 3 important features of Pivot Tables simultaneously. First, by the simple dragging of the field label, one can rearrange the Pivot Table from a "Policies by Underwriting Years" to an "Underwriting Years by Policies" view. Second, one can Sort a field by the results of on-the-fly calculations: in our case we sorted Policies by Loss Ratios in descending order. And, third, one can Filter by the results of on-the-fly calculations; in our case we chose to display just the 5 worst policies per underwriting year based on the Loss Ratio indicator. Note that we could choose any indicator (like Net Profitability or Discounted Loss Ratio) that is available in the Pivot Table as an input field or a dynamically calculated field. As one can see, the impressive reporting tools are all there; the quest is on for actuarial indicators.

	UWyear	PolicyID	QShare	Claims	NetPremium	NetUltLoss	NetUltALAE	Net LossRatio
3								
4	1996	NWA-0960520	80%	4,027	$ 7,375,016	$ 9,354,430	$ -	126.84%
5		NWA-0960622	75%	17,568	$ 24,643,120	$ 27,504,869	$ 3,110,613	124.23%
6		NWA-0960587	80%	13,123	$ 28,932,277	$ 34,141,643	$ 1,674,944	123.79%
7		NWA-0950781	90%	230	$ 1,653,244	$ 1,820,253	$ -	110.10%
8		NWA-0940604	10%	2,912	$ 15,714,567	$ 16,917,095	$ -	107.65%
9	1996 Total			37,860	$ 78,318,224	$ 89,738,090	$ 4,785,557	120.69%
10								
11	1997	NWA-0971113	75%	1,447	$ 1,396,321	$ 1,948,560	$ -	139.55%
12		NWA-0950696	10%	1,067	$ 10,600,577	$ 11,858,992	$ 1,980,625	130.56%
13								130.37%
14								127.73%
15								127.51%
16	1997 T							129.18%
17								
18								139.91%
19								137.97%
20								136.76%
21								135.30%
22								134.09%
23	1998 T							136.13%
24								
25								138.87%
26								138.11%
27								137.14%
28								136.91%
29								136.72%
30	1999 T							138.08%

PivotTable Field Advanced Options ? X

Page field options

- (•) Retrieve external data for all page field items (faster performance)
- ( ) Query external data source as you select each page field item (requires less memory)
  - [ ] Disable pivoting of this field (recommended)

AutoSort options
- ( ) Manual
- ( ) Ascending
- (•) Descending

Using field:
Net LossRatio ▼

AutoShow options
- ( ) Manual
- (•) Automatic

Show: Top ▼ 5 ▲▼

Using field:
Net LossRatio ▼

NetUltLoss
NetUltALAE
Net LossRatio
NetProfitability
Discounted LossRatio

Visualization

Seeing is believing

Popular belief

Visualization (see [5]) is the process of exploring, transforming and viewing data as images to gain understanding and insight into the data. Studies in human perception, computer graphics, imaging, numerical analysis, statistical methods and data analysis have helped to bring visualization to the forefront. Images have unparalleled power to convey information and ideas. Informally, visualization is the transformation of data or information into pictures… it engages the primary human sensory apparatus, *vision*, as well as the processing power of the human mind. The result is an effective medium for communicating complex and/or voluminous information.

As the amount of data overwhelms the ability of the human to assimilate and understand it, there is no escape from visualization. So actuaries have to develop conventions for representation of their data and results.

There exist a multitude of visualization approaches: mapping scalars to colors, contouring (isosurfaces), glyphs (arrows of different color, length, direction), warping (display of different stages in the motion), displacement plots, time animations, streamlines (particle traces) and tensor algorithms.

There are unavoidable problems with multidimensional visualization: projection (only two or three dimensions are available) and understanding (we as humans do not easily comprehend more than three dimensions or three dimensions plus time animation). Projections can be implemented if the three most important dimensions are identified in such a way that the remaining dimensions can be ignored. Once again, it is an actuarial task to choose these dimensions.

- ***Charts***

This paper is too short to discuss all the possible uses of charts and diagrams in the actuarial process. A great number of wonderful books (see [6]-[8]) explain which type of chart to use in which situation: a line chart for displaying increases, bar for shares, pedestal for ranks, Gantt for schedules, etc… However, it is up to actuaries to decide how to display triangular data.

> Example 9. By examining the 3-D chart representing logarithms of age-to-age factors, one can formulate a hypothesis about changes in calendar year trends (arguably not immediately evident from looking at the raw data): the last 4 years have different trend than the rest of the triangle. Rotating the graph for the look from another angle allows one to confirm or discard that hypothesis. The final check comes from the algorithm described, for example, in [9].
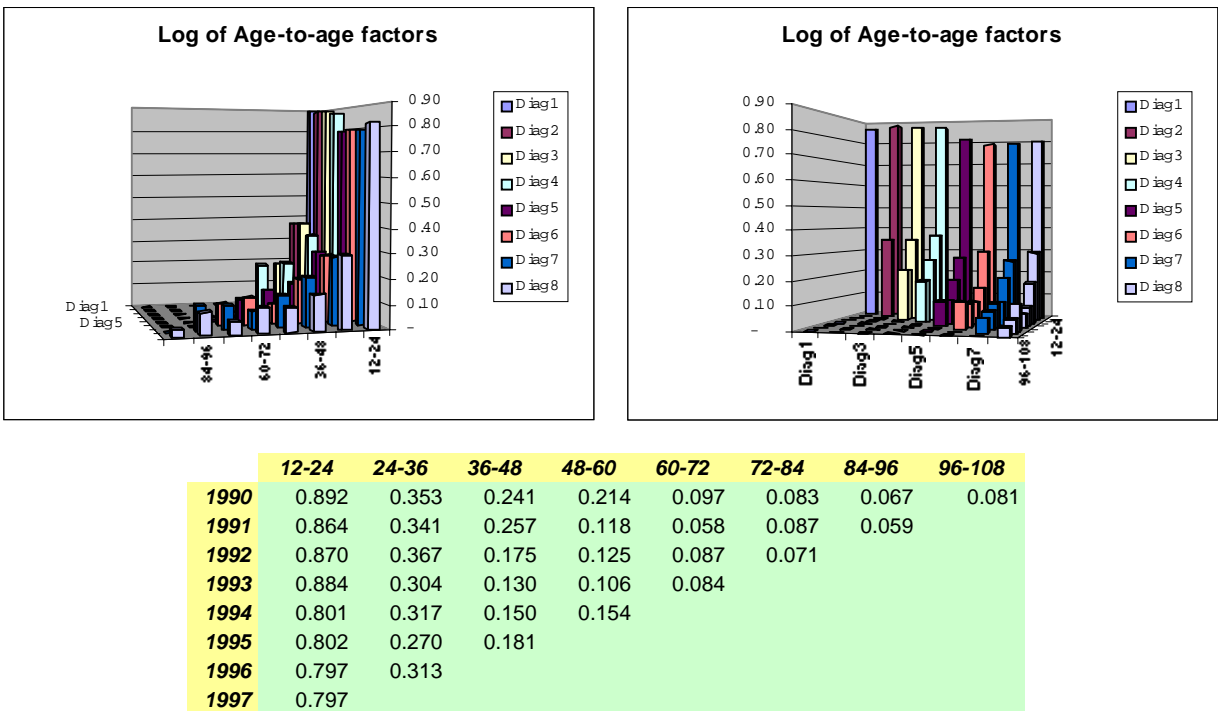

	12-24	24-36	36-48	48-60	60-72	72-84	84-96	96-108
1990	0.892	0.353	0.241	0.214	0.097	0.083	0.067	0.081
1991	0.864	0.341	0.257	0.118	0.058	0.087	0.059	
1992	0.870	0.367	0.175	0.125	0.087	0.071		
1993	0.884	0.304	0.130	0.106	0.084			
1994	0.801	0.317	0.150	0.154				
1995	0.802	0.270	0.181					
1996	0.797	0.313						
1997	0.797							

Figure 8

For almost 86 years since 1914, CAS members have not yet agreed on the standard graphical representation of a triangle (one of the most basic actuarial notions). The author firmly believes that properly displayed triangles may reveal some important trends that are not evident in numerical representations.

Not only does visualization inspire useful hypotheses for actuarial analysis: sometimes it's the only way to deal with the data*. Developments in new areas of actuarial science present new challenges in demonstrating findings. DFA[iii] that deals with many less traditional notions, such as scenarios and strategies, is a good example of such a challenge.

> Example 10. The majority of actuarial information contains a location code associated with the values. Legislation requirements, types of coverage, rates, exposures and loss performance differ from region to region. Geo-coding swiftly emerges as one of the hottest actuarial applications. Yet it is hard to imagine how one can notice trends and dependencies of geographically related data without visualization. Microsoft MapPoint – ideologically, an integral part of Microsoft Office – provides the means for precisely that type of visualization. For example, a map of WC Ultimate development factors by State based on NCCI data (see [10]) is presented below.

Figure 9

- **Animation**

If a picture is worth a thousand words, then an animation is worth a thousand pictures.

Animation is the best way to exploit an aspect of the human psyche called "selective attention" – people readily react (by shifting the focus of their attention) to any movement, including change in color, size or position. Animation is suitable for visualization of the range of uncertainty and/or development – two of the most important actuarial phenomena. While not a standard feature of a plain-vanilla spreadsheet, animation is nevertheless quite within reach of every Microsoft Excel user.

---

* So-called "quarterly" triangles, sometimes studied by actuaries, can easily reach size of 60x60 or more, which makes them impractical for examining by traditional means (in a spreadsheet), yet visualization techniques would shine in this circumstance.

Example 11. Below is the code for conceptual animation:

```
Sub AnimationInExcel()
Dim i As Long
    Call SetupAnimation      'see Appendix 2
    Application.Iteration = True
    Application.MaxIterations = 1000
    Application.MaxChange = 0.1
    For i = 1 To 1000
        Calculate
    Next i
End Sub                      'do not forget to restore original Calculation Mode!!!
```

Taking a tip from the computer games industry, animation can be **effectively** used for visualization of the simulation process (see [11]). Indeed, with the growing importance of DFA and other non-analytical modeling techniques, simulations are steadily becoming a technique of choice for the majority of actuaries. Currently, however, all intermediate steps used in simulation are hidden from the user – such a wealth of information is, essentially, discarded. Use of animation may prevent this "waste" of intermediate calculations. Animated display of simulation's steps may help the user to visualize the dynamics of the simulated process or appreciate the range of uncertainty in simulated scenarios.

- *Interactive selection*

Another application of visual technologies is the interactive selection of actuarial parameters. For example, selecting parameters by moving points directly on the graph of a development pattern appears to be much more intuitive and convenient than typing numbers into spreadsheet cells.
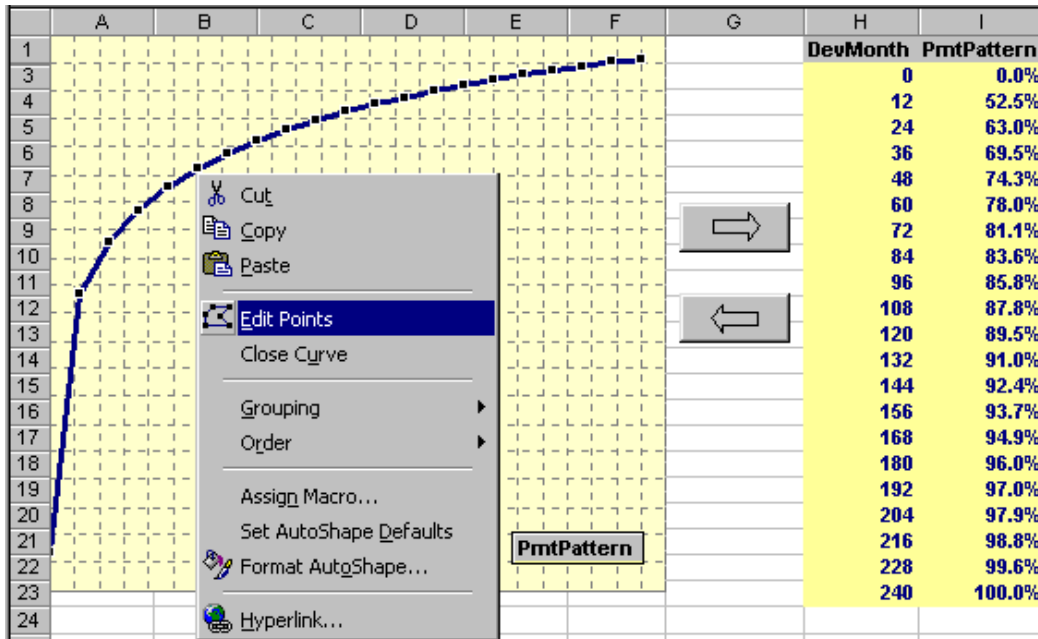

	DevMonth	PmtPattern
	0	0.0%
	12	52.5%
	24	63.0%
	36	69.5%
	48	74.3%
	60	78.0%
	72	81.1%
	84	83.6%
	96	85.8%
	108	87.8%
	120	89.5%
	132	91.0%
	144	92.4%
	156	93.7%
	168	94.9%
	180	96.0%
	192	97.0%
	204	97.9%
	216	98.8%
	228	99.6%
	240	100.0%

Figure 10

15

This functionality – a two-way link between the graphical display of numbers and their values in a spreadsheet – is not yet as user-friendly as other Office tools. But given that this feature is available in numerous other applications, it is only a matter of time until Visual Editing becomes an equal member of the Office tools roster and actuaries are able to incorporate interactive graphical manipulation of numbers into their spreadsheets.

Example 12. Even though "Visual Chart Editing" is not a standard Office feature, with some amount of VBA programming, it is possible to establish a two-way link between numbers and shapes in Excel (below is the code behind the buttons from Figure 10):

```vba
Sub VisualizationFromShapeToSpreadsheet()
    Dim cell As Range, node As ShapeNode, n As Integer
    Set cell = ActiveCell
    ActiveSheet.Range("Coordinates").CurrentRegion.Clear
    ActiveSheet.Shapes("InteractiveSelect").Select
    n = 0
    For Each node In ActiveSheet.Shapes("InteractiveSelect").Nodes
        n = n + 1
        ActiveSheet.Range("Coordinates").Cells(n, 1) = node.Points(1, 1)
        ActiveSheet.Range("Coordinates").Cells(n, 2) = (210 - node.Points(1, 2)) / 200
    Next node
    cell.Select
End Sub

Sub VisualizationFromSpreadsheetToShape()
    Dim c As Range, cell As Range
    Set cell = ActiveCell
    On Error Resume Next
    ActiveSheet.Shapes("InteractiveSelect").Delete
    With ActiveSheet.Shapes.BuildFreeform(msoEditingAuto, 0, 0)
        For Each c In ActiveSheet.Range("Coordinates").CurrentRegion.Resize(, 1)
            .AddNodes msoSegmentLine, msoEditingAuto, c.Value, 210 - 200 * c.Cells(1, 2).Value
        Next c
        .ConvertToShape.Select
    End With
    Selection.Name = "InteractiveSelect"
    ActiveSheet.Shapes("InteractiveSelect").Nodes.Delete 1
    Selection.Placement = xlFreeFloating
    cell.Select
End Sub
```

Alarm Systems

- ### *Data Quality*

Companies that build data warehouses and clean up their data soon realize that the majority of their data comes from external sources (TPA's[iv], industry bodies, self-insureds), which are neither clean nor in a single format. It is time to combine efforts and make sure that every source can supply high quality data in a timely manner.

There are some recommendations on data quality procedures by IDMA[v] (see [12], [13]) and data elements' definitions by ISO[vi] and NCCI[vii] (see [10], [14]), but they are not part of everyday life in every data collection entity. In fact, a study of more than 40 TPA's (see [15]) showed that practically every one of them has failed even the most primitive data quality checks.

> Example 13. An Alarm System that is worth its while should trigger some action when a problem is found. Painting some cells in a spreadsheet is a good example of such an action, but automatically sending an e-mail with the description of the problem would be much more effective. The code below continues Example 5: first, it checks a triangle of age-to-age factors for outliers and then it sends an e-mail to the System Administrator with the addresses of all problematic cells:

Data Quality can be tested both on the detail level and on pre-aggregated levels (see [15]). In both cases, reporting techniques like Filtering, Sorting and Conditional Formatting may help attract attention to the problem (the same applies to visualization techniques, which can help to pinpoint a problem). One can calculate changes in case reserves and sort claims in descending order by that field to bring the largest outstanding claims on top. Or, using conditional formatting, one can highlight outliers among age-to-age factors (see Example 5).

- ***Algorithms Applicability***

Closely related to Data Quality tests on the pre-aggregated (as opposed to detailed) level is actuarial assumption testing (see [15]). Indeed, a monotonically increasing number of claims can be both a data quality test and a requirement for the applicability of the Berquist-Sherman algorithm. The same for the assumption of lognormality in ICRFS[viii], which coincides with the check that requires incremental gross payments to be positive. An Alarm System may warn users about *Thomas Mack*-style test (see [2], [16]) failures.

```vba
Sub AlarmEvent()
    Dim c As Range, sAlarm As String
    Dim otlApp As Outlook.Application, eMail As Object
    For Each c In Range("Tri").SpecialCells(xlCellTypeSameFormatConditions)
        c.Select
        Range("Temp").Formula = c.FormatConditions(1).Formula1
        If Not Application.IsError(Range("Temp").Value) Then
            If Range("Temp").Value Then
                sAlarm = sAlarm & "Problem at :" & c.Address & Chr(13)
            End If
        End If
    Next c
    If sAlarm & "" <> "" Then                      'if there is any problem
        Set otlApp = New Outlook.Application        'launch Outlook
        Set eMail = otlApp.CreateItem(olMailItem)   'and create e-Mail message
        With eMail
            .To = "samsebe@consultant.com"
            .Subject = "ALARM from " & ActiveWorkbook.Name & "!" & ActiveSheet.Name
            .Body = sAlarm
            .Send                                    'send e-Mail
        End With
        Set eMail = Nothing
        otlApp.Quit
        Set otlApp = Nothing
    End If
End Sub
```
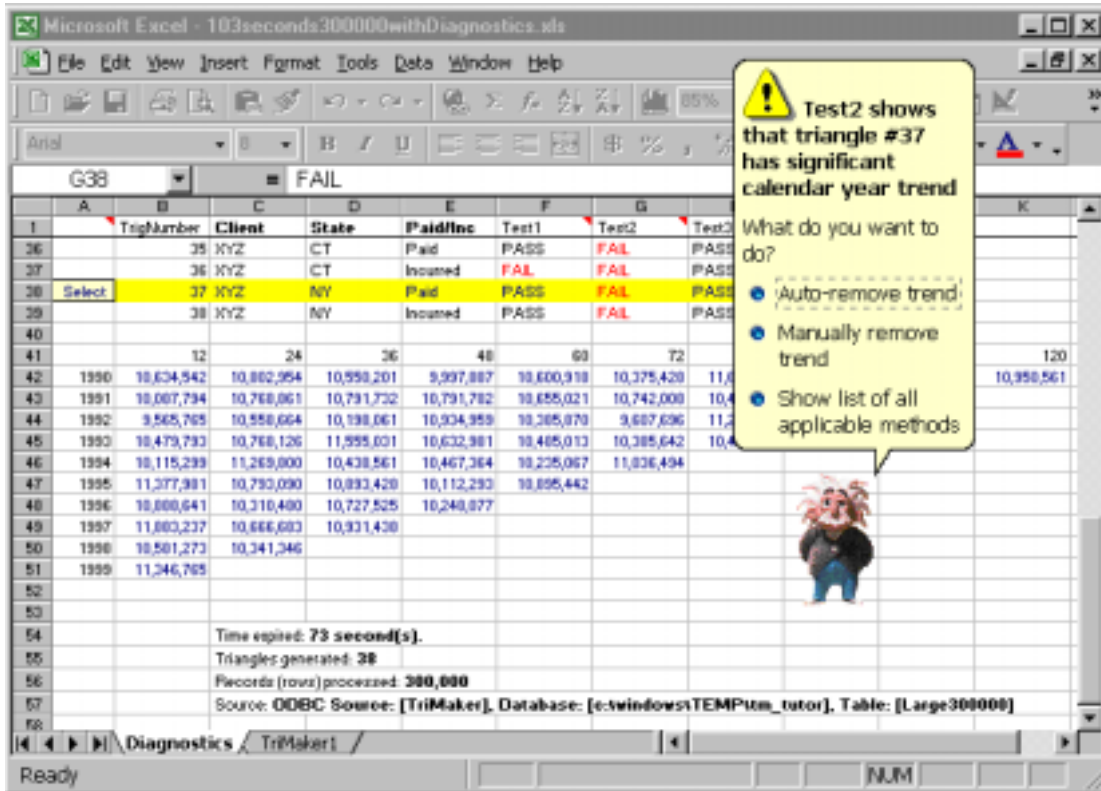
Figure 11

- ***Digital Dashboard***

Digital Dashboard is Microsoft's name for a portal that consolidates the most important personal, professional, corporate and external information with an immediate access to analytical and collaborative tools. In a single view, the user can see charts, Alarm messages, Pivot Tables, calendars, etc. Thus, Digital Dashboard looks like an obvious place for all important reports and alarms. Dashboard's space limitations re-emphasize the necessity of smart and space-conscious reporting techniques: Dashboard's start screen is the place for the most important information presented in the most concise way.

While every reporting and visualization technique described in this paper is powerful and effective, it is their combinations (Filtering + Sorting, Pivot Tables + Conditional Formatting, etc + etc…) that convert a flood of data into truly useful and indispensable information. Digital Dashboard – a "combination of combinations" of reporting tools - is just a very logical extension of the mechanisms that make this information immediately available and accessible. By the same token, Digital Dashboard is a most natural interface for an Alarm System. Not only can it display all types of alarms in a single location, but also – thanks to its portal capabilities – it can provide links to detailed information that triggered an alarm.

Figure 12

With the proliferation of the Internet, portal interfaces have become very popular: an ability to organize a wealth of information into a concise and focused display is very appealing. In fact, this article itself is organized as a portal into a wealth of information on reporting and visualization techniques: it is just a concentrated extract of the most important facts about tools available to actuaries, and as such it serves as a starting point for further research.

Actuarial Significance

"The whole system makes me feel… insignificant."

Ant Z.

Actuaries have a lot of work to do before any of the aforementioned techniques can be used to generate useful reports and visualizations.

For adaptive reporting, actuaries have to decide which measurements are of actuarial significance, so reports can be filtered or sorted accordingly. Every step of the actuarial process requires different significance indicators. For the *data preparation* step, actuaries should find measurements that will catch severe data errors that may considerably affect the consequent application of actuarial methods. For the *results presentation* step, actuaries should define (and calculate during the *application of algorithms* step) indicators that will aid decision-making. That will help to concentrate the attention of the report readers on important issues and will energize and strengthen the decision-making process.

Another important area that needs assistance from actuaries is in Alarm Systems. Nobody likes false alarms. It is actuary's job to come up with and fine-tune alarm definitions, to determine which combination of circumstances should trigger an actuarial alarm and attract immediate attention.

Selection of these *most important* variables depends on the available data and the goal of the display, and is clearly an actuarial task. It would make sense for actuaries to develop conventions that cover most situations. Unfortunately, to the best of the author's knowledge, this work has not even started.

Conclusion

"The gods help them that help themselves."

Aesop

A list of easily accessible presentation techniques with examples of their uses should help actuaries to realize what tools are available for Actuarial Reporting. But it is up to actuaries to express themselves using these tools. Indeed, reporting techniques described in this article are so flexible it does not make sense to use a limited number of pre-designed "canned" reports anymore. In addition to that, reporting tools are incredibly interactive – they were designed in order to give the end user (i.e., actuary) report-creation power. And they are so easy to use – it is a sin not to use them.

The importance of presentation skills is severely underestimated by actuaries: it is conceivable that quite a few companies would still remain solvent if actuaries-in-charge could **convincingly present** results of their analysis. If actuaries expect computers to be effective helpers in reporting and visualization tasks, they have to define "actuarially significant" information and learn how to present it in the most "attention-grabbing" way. That would insure that actuarial analysis is indeed used as a solid foundation for the company's decision-making process.

Acknowledgements

Appendix 0

Readers of the Acrobat version of this paper (downloaded from the www.casact.org website) can copy code snippets from the text and paste them into Excel's VBA Editor as described below.

Appendix 1

To run subroutines described throughout the article:

- *launch Excel, preferably version 97 or better*

- *press Alt-F11 to start VBA[ix] Editor*

- *find a project with the name of your current spreadsheet or, if you prefer, Personal.xls*

- *right-click on the project and choose Insert/Module*

- *type or copy/paste the desired fragment of the code into the Module window*

- *launch code from the Excel: press Alt-F8, select macro name and press Run*

Appendix 2

```vba
Sub SetupAnimation()
Dim sName As String
    With ActiveSheet
        .Cells(1, 2).Formula = "T"
        .Cells(1, 3).Formula = "X"
        .Cells(2, 2).Formula = "12"
        .Cells(3, 2).Formula = "24"
        .Cells(10, 3).Formula = "1"
        .Cells(1, 1).FormulaR1C1 = "=RC+0.01"
        Range("B2:B3").Select
        Selection.AutoFill Destination:=Range("B2:B10"), Type:=xlFillDefault
        Range("C2:C9").Select
        Selection.FormulaR1C1 = "=1-EXP(-1/R1C1/(1-RC[-1]/R10C2)/(1-RC[-1]/R10C2))"
    End With
    sName = ActiveSheet.Name
    Charts.Add
    ActiveChart.ChartType = xlXYScatterSmooth
    ActiveChart.SetSourceData Source:=Sheets(sName).Range("B1:C10")
    ActiveChart.Location Where:=xlLocationAsObject, Name:=sName
    With ActiveChart.Axes(xlValue)
        .MinimumScale = 0
        .MaximumScale = 1.2
        .MinorUnit = 0.04
        .MajorUnit = 0.2
        .CrossesAt = 0
    End With
End Sub
```

Bibliography

(in the order of reference)

[1] Theo Mandel. *The elements of User interface Design*. Wiley Computer Publishing, 1997

[2] Thomas Mack, *Measuring the Variability of Chain Ladder Reserve Estimates*. CAS, 1993

[3] Aleksey S. Popelyukhin. *On Hierarchy of Actuarial Objects: Data Processing from the Actuarial Point of View.* CAS Forum, Spring 1999

[4] Aleksey S. Popelyukhin. *The Big Picture: Actuarial Process from the Data Processing Point of View.* Library of Congress, 1996

[5] Will Schroeder, Ken Martin, Bill Lorensen. *Visualiazion Toolkit*. Prentice Hall, 1998

[6] Gene Zelazny. *Say it with Charts: The Executive's Guide to Visual Communication*. McGraw-Hill, 1996

[7] Robert L. Harris. *Information Graphics: A Comprehensive Illustrated Reference*. Management Graphics, 1997

[8] Tufte Edward R. *Visual Exlplanations.* Graphics Press, 1997

[9] Ben Zehnwirth, *Probabilistic Development Factor Models with Applications to Loss Reserve Variability, Prediction Intervals, and Risk Based Capital*

[10] *NCCI Annual Statistical Bulletin, 1998 Edition.* NCCI, 1999

[11] Aleksey S. Popelyukhin. *SimActuary: The Game We Can Play.* Submitted to CAS Forum, Spring 2003

[12] *Insurance Data Quality.* Marr, Richard, ed. IDMA, 1995

[13] *Data Quality Certification Model (Framework and Guidelines) for Insurance Data Management.* IDMA, ISO, 1995

[14] *Standards and Guidelines for Data Elements.* ISO, 1981

[15] Aleksey S. Popelyukhin. *Watch your TPA: A Practical Introduction to Actuarial Data Quality Management.* CAS Forum, Spring 1999

[16] Gary G. Venter, *Checking Assumptions of Age-to-Age Factors.* CLRS, 1994

Acronyms

[i] **BASIC - Beginner's All-purpose Symbolic Instruction Code**: one of the earliest and simplest high-level programming languages – still a very popular choice among educators.

[ii] **LOB – Line of Business**: here, a type of insurance coverage like Workers' Compensation or Professional Liability.

[iii] **DFA – Dynamic Financial Analysis**: a process for analyzing the financial condition of an insurance entity.

[iv] **TPA – Third Party Administrator**: a company in the business of handling day-to-day activities and/or providing services on insurance claims. Consequently, TPA is a primary source of actuarial data. See [14].

[v] **IDMA - Insurance Data Management Association**: an independent nonprofit association dedicated to increasing the level of professionalism in insurance data management. http://www.ins-data-mgmt.org .

[vi] **ISO – Insurance Services Office**, Inc.: leading supplier of statistical, actuarial, underwriting, and claims information. http://www.iso.com .

[vii] **NCCI – National Council on Compensation Insurance**, Inc.: a value-added collector, manager, and distributor of information related to workers' compensation insurance. http://www2.ncci.com/ncciweb .

[viii] **ICRFS – Interactive Claims Reserving Forecasting System**: commercially available statistical modeling framework from Insureware. http://www.insureware.com .

[ix] **VBA – Visual BASIC for Applications**: version of BASIC language embedded into host application (i.e. Excel) with the access to host's objects - a better "macro language". http://www.microsoft.com .