STAT COE-Report-13-2017

Calculating a Consensus of Opinion

Authored by: Aaron Ramert

30 August 2017

Revised 17 October 2018

The goal of the STAT COE is to assist in developing rigorous, defensible test strategies to more effectively quantify and characterize system performance and provide information that reduces risk. This and other COE products are available at www.afit.edu/STAT.

STAT Center of Excellence
2950 Hobson Way – Wright-Patterson AFB, OH 45433

Table of Contents

Revision 1, 17 Oct 2018: Formatting and minor typographical/grammatical edits.

Executive Summary

There is a continuing push for rigor in acquisition system testing. Gains have been made toward this end but there still exists situations where expert opinion is the best or only option to gather data. This best practice introduces a method for mathematically calculating a consensus of opinion using a weighted least-squares function.

Keywords: Subject matter expert, consensus, ratio-scale matrix, multi-criteria decision, factor reduction

Introduction

The Department of Defense (DoD) testing community strives to make tests more rigorous and objective in order to comply with regulations and provide leaders with accurate, relevant, and defensible information to aid in decision making. Rocket engineer Wernher Von Braun is reported to have said, “One test is worth a thousand expert opinions."[1] The motivation behind this quote is intuitively obvious to members of the testing community. An opinion, even an extremely well-informed opinion, lacks explicit analytical calculations and objectivity and includes some degree of bias, whether intentional or not. Conversely, a well-designed and properly executed test can calculate a performance metric with great precision and certainty and without subjectivity.

The trend within the testing community has favored hard data over subject matter experts (SMEs), but there still exist situations where the experience and educated opinion of experts is necessary. Designing rigorous and informative tests is generally easier for systems with a small number of factors, especially when ranges of all factors can be determined by an engineering process. However, there are other systems which possess more factors than can be effectively tested or factors with ranges that are too broad and must be narrowed in scope. It is in these situations that testers must rely on the experience of a SME.

A common practice to minimize the bias introduced by a SME is to solicit the opinions of several SMEs to obtain an average opinion. A complication that can emerge using this approach is trying to find the consensus opinion from a diverse group of SMEs, each with unique backgrounds and goals. The test designers then need to evaluate which SME’s opinion to use or how best to combine them. This best practice outlines a multiple-criteria decision model and provides a mathematical solution to differences of opinion through the use of a weighted least-squares function and Ratio-Scale (RS) matrices to calculate a relationship between factors. The method still relies on human judgment and does not

[1] Von Braun, Wernher. "Famous Quotes." Goodreads. n.p., n.d. Web. 21 August 2017. <https://www.goodreads.com/quotes/search?utf8=%E2%9C%93&q=von+braun&commit=Search>.

eliminate bias, but it adds rigor to the valuation of multiple expert opinions with the goal of finding the best solution from the combined wisdom of SMEs.

Background

The Purpose of Calculated Consensus

When developing the factors and levels for a test plan, it may not be obvious which factors should be included in a test. In the field of acquisition program testing, the Capabilities Development Document (CDD), key performance parameters (KPPs), key system attributes (KSAs), and other references may provide a thorough list of factors to test. This list of potential factors is normally too large for every factor and every interaction to be tested. In some cases, a screening design will eliminate the less influential factors, so a thorough test can be designed for the remaining ones. Often however, it is necessary to rely on personal, professional, and industry experience to determine the test factors, factor interactions, or the important ranges of specific factors that are to be tested. The sources of these opinions are generically referred to as SMEs. Using SME input introduces bias and is counter to the DoD guidance for rigor in the acquisition process, but is sometimes unavoidable.

When a group of people come together to express their opinion, we can be sure that there will be some level of disagreement. I am sure that many readers have memories of some very dynamic disagreements among experts. In some cases, the person with the loudest voice wins, in some cases it is the person with the greatest seniority, and in other cases the argument is never resolved. The technique described here seeks to assign a value to the total disagreement between participants' priority of factors with the other SMEs, and then adjust the aggregate priority to minimize that disagreement and reach the best consensus. Note that a consensus is a general agreement or group solidarity in sentiment and belief. Consensus does not mean complete agreement. The method contained in this article was derived primarily from the book *Ordinal Information and Preference Structures* by Cook and Kress (1992). It has a mathematically intensive explanation, but the practical application is straightforward and clear. The calculations, done on a spreadsheet, are clearly presented in the example at the end of this best practice. The method has been implemented in Excel and is available for download at the STAT COE website (https://www.afit.edu/STAT/stattools.cfm).

Least-Squares Distance

The most frequently used method to measure closeness and accuracy in n-dimensional space is least-squares distance (Kennedy, 2011). This method is easiest to imagine in two dimensional space as shown in Figure 1. The figure illustrates simple regression analysis where data are plotted within an x-y coordinate system, and a straight line captures the trend between the two variables. The trend line is placed so it has minimum squared error; in other words, each point is, on average, closest to the least squares line compared to other potential lines.

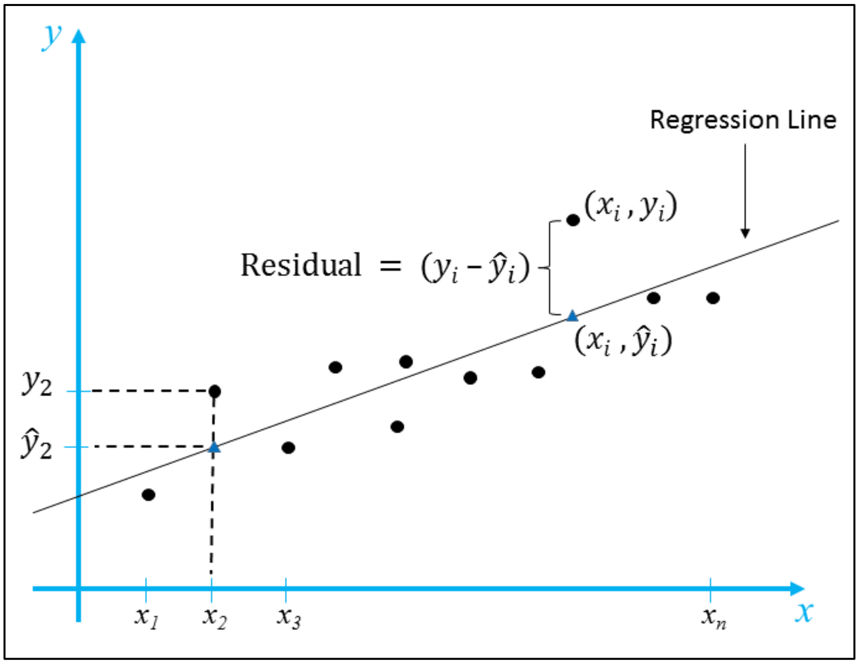

Figure 1: Linear Regression

Equation (1) shows a simplified version of the calculation used to place the line. The value E is the total squared error and is calculated by measuring each residual (the distance along the y-axis from each of the n points (y) to its respective predicted value on the line ($\hat{y}$), squaring it, and summing all the squared values. The linear regression line seeks to minimize this value E. Because the distance is squared, a point with twice the distance will have four times the error value. This is the reason outliers, points abnormally far from the norm, often have disproportionately large effects in regression analysis.

$$E = \sum_{i=1}^{n} (y_i - \hat{y}_i)^2 \tag{1}$$

The next step in linear regression is to calculate the best placement of the line where E is minimized.

Weighted Least-Squares Distance

A weighted least-squares distance expands the least-squares concept by placing more weight, i.e. value, on some data points compared to others. Not only do we account for the residual error, but we also account for the importance of the observation. In Equation (2), this is accounted for by multiplying each squared distance by the importance, or weight of each point (w_i). With this method, if two points had an equal but opposite distance from the norm, the trend line would not be in the middle, but would be biased closer to the point with the larger weight.

$$W = \sum_{i=1}^{n} w_i (y_i - \hat{y}_i)^2 \tag{2}$$

Ratio-Scale Matrix

The mathematical tool used to compare the relative values of the potential factors is a Ratio-Scale (RS) matrix. The numbers within the matrix represent the value of one preference over another. They are calculated as $d_{ij} = i/j$, the preference of the factor in column i over the factor in row j. The RS matrix is always positive because the information value of all factors is greater than zero. A factor can have no impact, but it cannot have negative impact. There can never be a situation where testing an additional factor will give you less information about a system. The values across the diagonal of the RS are the inverse of each other. This is because if factor i is assessed to be twice as important as factor j, then factor j must be half as important as factor i. We express this mathematically by stating that $d_{ij} = 1/d_{ji}$ for all i, j. Because of this property, only the upper triangle of the matrix is used in the subsequent calculations, and therefore, as Equation (3) shows, only the upper triangle of the RS matrix is calculated. Equation (3) also shows that the diagonal of the RS matrix is 1. This is because the value of any factor divided by itself is one.

$$D = \begin{pmatrix} d_{11} & d_{12} & d_{13} & d_{14} \\ d_{21} & d_{22} & d_{23} & d_{24} \\ d_{31} & d_{32} & d_{33} & d_{34} \\ d_{41} & d_{42} & d_{43} & d_{44} \end{pmatrix} \xrightarrow{\text{simplified to}} D = \begin{pmatrix} 1 & d_{12} & d_{13} & d_{14} \\ & 1 & d_{23} & d_{24} \\ & & 1 & d_{34} \\ & & & 1 \end{pmatrix} \tag{3}$$

It is necessary for each SME to create their own RS matrix (creation methods are discussed later). Equation (4) is an example of one for SME u with six factors. Only the upper triangle is calculated.

$$D_u = \begin{pmatrix} 1 & d_{12}^u & d_{13}^u & d_{14}^u & d_{15}^u & d_{16}^u \\ & 1 & d_{23}^u & d_{24}^u & d_{25}^u & d_{26}^u \\ & & 1 & d_{34}^u & d_{35}^u & d_{36}^u \\ & & & 1 & d_{45}^u & d_{46}^u \\ & & & & 1 & d_{56}^u \\ & & & & & 1 \end{pmatrix} \tag{4}$$

Minimizing Equation

The final step is to plug the data into Equation (5) and calculate the ranking of the factors that is least objectionable: the consensus value of the factor. The variables in Equation (5) are as follows. There are n columns in each RS matrix, indexed with i. There are n rows in each RS matrix, indexed with j. By restricting the value of i to start at two, and for j to be less than or equal to i in Equation (5), we restrict our inputs to the upper triangle of each RS matrix. The preference of factor i over factor j for SME u is d_{ij}^u and is taken from the SME's RS matrix, like the one in Equation (4). The weight of SME u is w_u and is defined for each SME. There are m SMEs in the group, indexed with u. The final consensus value of factor i is k_i. The final consensus value of factor j is k_j. The value of all factors must add up to one. This is done for ease of analysis. We find it easy to think of values in terms of percentages. When all factors add to one then the decimal value assigned to each one is easily converted to a relative percentage. The values of k will be adjusted by the software to an optimum value. The calculations seek to minimize the total consensus value (Equation (5)) by systematically changing the value of each factor, k.

$$\min_{k} \sum_{u=1}^{m} \sum_{j<i} \sum_{i=2}^{n} w_u \left(d_{ij}^{u} - \frac{k_i}{k_j} \right)^2 \qquad (5)$$

$$\text{subject to} \sum_{i=1}^{n} k_i = 1$$

Method

Overall Process

A specific use for this method is a situation where there are more potential factors than can be tested. The best solution is to find a way to test them all, but if that cannot be done this is a rigorous method for a group of SMEs to come together in a brainstorming session and develop a list of factors. Each SME should then produce a hierarchical list of those factors and convert it into an RS matrix. Someone in a position of leadership within the program can weight the SMEs (though this method will also work if all the SMEs are weighted equally). The data can then be entered into a spreadsheet similar to the ones shown and also offered on the STAT COE website. When all of the data are entered, the consensus values for the factors are almost instantly calculated. Each step is described in more detail in the following sections.

Develop a List of Factors

The first step is for a group of SMEs to brainstorm and create a list of factors. In this phase, it is only necessary to show there is reasonable cause for a factor to be tested. Keep in mind that these are only factors which need to be ranked. Consider a system which had three factors with a proven need to be tested and five more which many thought should be tested. Only the additional five factors would be included in the list.

Have the SMEs Rank the Factors and Create a Ratio-Scale Matrix

A SME can simply create a RS matrix. They ask themselves "How much more important is factor i when compared to factor j?" and write down the number. This is the quickest and simplest method, but can lead to mathematical inconsistencies. For example, a SME may decide factor A is twice as important as factor B (A=2B), and factor B is twice as important as factor C (B=2C). The mathematical conclusion is that factor A is four times as important as factor C (A=2(2C)). But a SME looking at each factor comparison individually may not go through these calculations and whimsically decide that factor A is only three times as important as factor C. Although this is an inconsistency, Equation (5) will still calculate consensus values for the factors. The resulting consensus values, however, may not be as accurate as possible.

Two methods to avoid inconsistencies are 1) first rank the list of factors or 2) weight the list of factors. Consider a list of four factors. They can be ranked from one to four in order of importance. This is the easiest method but offers no insight to their relative importance. The factors can also be weighted. This method may take longer, but yields all the information a ranked list does plus the insight into how much more important each factor is. Compare the two methods in Figure 2. In the ranked list, we know that Factor B is more important than Factor D, but we have no information on the magnitude of that difference. The weighted list has the same order of preference, but we also learn that Factor B is slightly more than twice as important as Factor D. This added information will make the subsequent RS matrices more precise. It is important for all of the SMEs to use the same scale. We recommend that weights range from one to ten, but any scale will work.

Rank	Factor
1	Factor C
2	Factor B
3	Factor D
4	Factor A

Weight	Factor
7	Factor C
5	Factor B
2	Factor D
1	Factor A

Figure 2: Ordered List of Factors

To convert the lists into RS matrices requires only simple math. For the ranked factors, you need to first convert the ranks into a weight. This is done by reversing the numbers. This means that for the ranked list in Figure 2, Factor C has a weight of 4 and Factor B has a weight of 3 and so on. A spreadsheet can automate much of the work.

Weight the SMEs

As shown in Equation (5), there are provisions to weight the SMEs in this consensus calculation. One option is to give them all a weight of one so each opinion is equal. Another option is to make most of them one but add (or subtract) some value to the few SMEs whose opinions are clearly more (or less) informed or valued than the others. The final option is to create a unique weighted value to each one. Warning, this could lead to a recursive situation where you need another set of RS matrices to rank the SMEs.

Calculate the Consensus

Equation (5) is solved in the Microsoft Excel spreadsheet using the Solver add-in. It iterates through all possible values of k so that the objective value (the number calculated by Equation (5)) is the smallest one possible. The objective value is not important, only the consensus values of each of the factors (which create the minimum value) are.

Conclusion

Calculating a consensus is offered as a technique to combine several opinions through a mathematical method that minimizes the total disagreement of the group. It gives deference to SMEs perceived to have different levels of insight. It can add rigor to a part of the process normally lacking it, but still remains a largely subjective process. Although it can rigorously calculate a consensus, there is no guarantee that the perfect consensus represents the truth. A fully worked example is shown below, which demonstrates how the excel consensus tool functions.

Tutorial

Situation

The following example is provided for a better understanding of consensus calculations. All calculations are included in the Microsoft Excel file. Suppose a new bicycle is developed, and the manufacturer can only afford to test four factors, but they aren't sure what the four most important factors are. A group of SMEs with diverse backgrounds and experience come together and develop a list of six factors. In alphabetical order, the factors are:

1. Aerodynamic Drag
2. Ease of Shifting
3. Grip Comfort
4. Noise from Chain
5. Seat Comfort
6. Handlebar Vibrations

From this list of six potential factors, they must then decide what the most important four are. Figure 3 shows the excel spreadsheet with all required data entered.

Enter the Data into the Spreadsheet

We assume that the SMEs each weight the factors independently. That is what is depicted in Figure 3. The solutions for SMEs who rank the factors and who create inconsistent RS matrices are included in the excel file, but are not shown in this document.

The setup displayed in Figure 3 is explained in order of the process. Down the left side of the figure, we can see five colored boxes. Each box is an RS matrix and is labeled for a specific SME. Additionally, the rows and columns of each RS matrix are labeled with the appropriate factor. On the left of each RS matrix, we see the weight of each factor. The various d_{ij} values are automatically calculated from these weights. To the right of each box, we see the weight of each SME.

The boxes on the right of Figure 3 are slightly more complicated. The box on the top is labeled $j = 1$, and is multicolored. The top line has calculations from SME 1, the next line from SME 2, and so on for all

SMEs. The left column is labeled k_2/k_j. Since this is box $j = 1$, the left column is k_2/k_1; the weighted least-squares distance value of Factor 2 over Factor 1 for each SME. The value below the label, in this example 0.01, is the value of Equation (5) when $u = 1$, $i = 2$, and $j = 1$. The number below this (0.81) is the same calculation for SME 2, where $u = 2$, $i = 2$, and $j = 1$. The number to the right (0.05) is for SME 1, but compares Factor 1 over Factor 3, where $u = 1$, $i = 3$, and $j = 1$. The multicolored boxes underneath are the same, except for a different value of j. There are fewer values calculated in each subsequent multicolored box because we are only calculating the upper triangle of the RS matrices.

At the top of the boxes on the right, we have the consensus values of the factors. We begin by assigning them arbitrary values between zero and one. Values outside this range won't allow the Solver add-in to work properly. To the right of these values is the sum of all consensus factor values. For the final solution, this total should be one. Finally, at the upper left in a yellow box is the objective function, the value of Equation (5) for this example. The number by itself is meaningless, but when it is reduced to the lowest possible value, we know we have consensus values for the factors which present the least total amount of disagreement from the group of SMEs. A value of zero indicates unanimous agreement between all SMEs on all factors. Note that in the spreadsheet, even with complete agreement, you will still have a small objective value because of rounding errors.

Sample problem with 5 SMEs and 6 Weighted Factors

Objective =	222.51

Factor:	AD	ES	GC	N	SC	V	
	k_1	k_2	k_3	k_4	k_5	k_6	$\sum k$
Consensus Value:	0.160	0.160	0.160	0.160	0.160	0.160	0.960
Rank:	1	1	1	1	1	1	

	SME 1	AD	ES	GC	N	SC	V
9	Aerodynamic Drag	1	0.9	0.78	0.67	0.56	0.22
8	Ease of shift		1	0.88	0.75	0.63	0.25
7	Grip Comfort			1	0.86	0.71	0.29
6	Noise				1	0.83	0.33
5	Seat Comfort					1	0.4
2	Vibrations						1

W1: 1.1

	k_2/k_j	k_3/k_j	k_4/k_j	k_5/k_j	k_6/k_j
SME1	0.01	0.05	0.12	0.22	0.67
SME2	0.81	0.16	0.64	0.01	0.01
SME3	0.25	2.25	0.56	0.25	0.06
SME4	0.06	0.02	0.06	0.02	0.02
SME5	0.03	0.11	0.25	0.44	0.69

$j=1$

	SME 2	AD	ES	GC	N	SC	V
10	Aerodynamic Drag	1	0.1	0.6	0.2	0.9	0.9
1	Ease of shift		1	6	2	9	9
6	Grip Comfort			1	0.33	1.5	1.5
2	Noise				1	4.5	4.5
9	Seat Comfort					1	1
9	Vibrations						1

W2: 1.0

	k_3/k_j	k_4/k_j	k_5/k_j	k_6/k_j
SME1	0.05	0.12	0.22	0.67
SME2	25.00	1.00	64.00	64.00
SME3	16.00	6.25	4.00	2.25
SME4	0.09	0.00	0.09	0.09
SME5	0.04	0.16	0.36	0.64

$j=2$

	SME 3	AD	ES	GC	N	SC	V
4	Aerodynamic Drag	1	0.5	2.5	1.75	1.5	1.25
2	Ease of shift		1	5	3.5	3	2.5
10	Grip Comfort			1	0.7	0.6	0.5
7	Noise				1	0.86	0.71
6	Seat Comfort					1	0.83
5	Vibrations						1

W3: 1.0

	k_4/k_j	k_5/k_j	k_6/k_j
SME1	0.02	0.09	0.56
SME2	0.44	0.25	0.25
SME3	0.09	0.16	0.25
SME4	0.18	0.00	0.00
SME5	0.06	0.25	0.56

$j=3$

	SME 4	AD	ES	GC	N	SC	V
8	Aerodynamic Drag	1	1.3	0.88	1.25	0.88	0.88
10	Ease of shift		1	0.7	1	0.7	0.7
7	Grip Comfort			1	1.43	1	1
10	Noise				1	0.7	0.7
7	Seat Comfort					1	1
7	Vibrations						1

W4: 1.0

	k_5/k_j	k_6/k_j
SME1	0.03	0.49
SME2	12.25	12.25
SME3	0.02	0.08
SME4	0.09	0.09
SME5	0.11	0.44

$j=4$

	SME 5	AD	ES	GC	N	SC	V
6	Aerodynamic Drag	1	0.8	0.67	0.5	0.33	0.17
5	Ease of shift		1	0.8	0.6	0.4	0.2
4	Grip Comfort			1	0.75	0.5	0.25
3	Noise				1	0.67	0.33
2	Seat Comfort					1	0.5
1	Vibrations						1

W5: 1.0

	k_6/k_j
SME1	0.40
SME2	0.00
SME3	0.03
SME4	0.00
SME5	0.25

$j=5$

Figure 3: Excel Setup for Bicycle Example (Weighted Factors)

The Excel worksheet is interactive. By changing the weights of the factors to the left of each SME box or by changing the weight of any SME, the solution will change. The consensus values, however, will not automatically change.

Use the Solver Add-in to Calculate an Answer

Solver is a built-in tool in Microsoft Excel. It has several methods of changing the variables in an equation to find the best answer. The Solver tool is listed under the 'Analysis' tab within the 'DATA' ribbon. If Solver is not accessible at this location, it may need to be added. This can be done in the few steps listed in the following link : https://support.office.com/en-us/article/load-the-solver-add-in .

To use Solver, you need to define the value to be optimized, what the goal is, the values to be manipulated in search of the goal, any constraints, and the calculation method for Solver to use.

After all of the data have been entered into the spreadsheet as shown in Figure 3, we can use Solver to find a solution. Start by double clicking on Solver, and you will see the parameter box shown in Figure 4. Set the objective value to the location of the objective value in the spreadsheet; in this case, cell E4. Next, in the 'To' line, click the radio button next to 'Min' because our goal is to minimize the objective value. We want this to happen by changing the Consensus Value of each factor which are cells M5, N5, O5, P5, Q5, and R5 in this example. This is abbreviated as 'M5:R5' in the 'By Changing Variable Cells' entry blank. There are two constraints for this problem. The first is that the total of all the consensus values must equal 1. For this, we click on the 'Add' button on the right and type 'T5=1'. The second constraint is that all of the Consensus factor values must be non-negative. This is done by simply checking the box below the constraints. Finally, we select the "GRG Nonlinear" solving method and click the 'Solve' button.

The GRG Nonlinear function looks for a locally optimal solution. The Solver web page explains, "Figuratively, this means that the Solver has found a "peak" (if maximizing) or "valley" (if minimizing) -- but there may be other taller peaks or deeper valleys far away from the current solution." The web page further explains that, "The best that current nonlinear optimization methods can guarantee is to find a locally optimal solution. We recommend that you run the GRG Solver starting from several different sets of initial values for the decision variables -- ideally chosen based on your own knowledge of the problem. In this way you can increase the chances that you have found the best possible `optimal solution'" (Frontline Solvers, 2017).

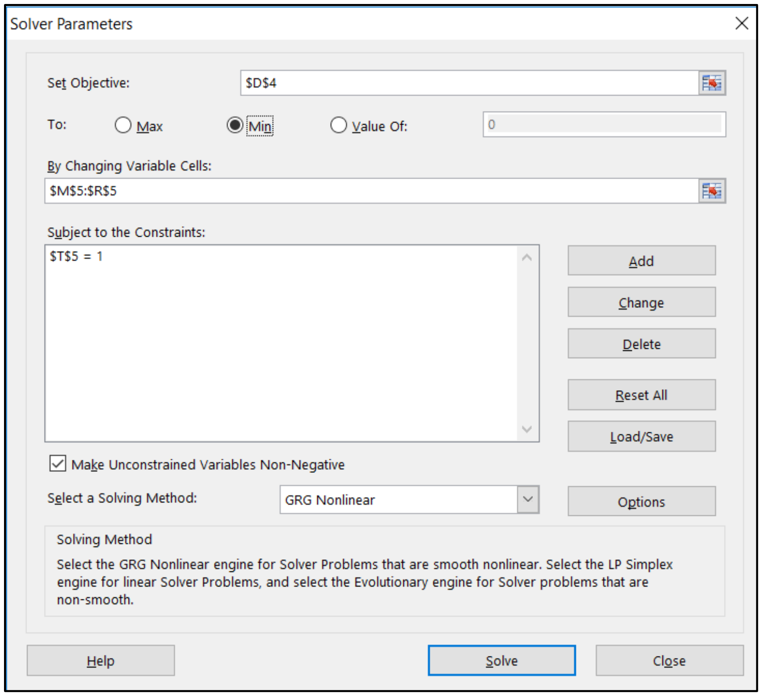

Figure 4: Solver Add-in setup to solve sample problem

Solver will succeed or fail to converge to a solution in a few seconds. If Solver converges, you will get a results message like the one on the left of Figure 5. If it fails to converge, you will get a message like the one on the right. If Solver fails, first try some different consensus values and try again. If that does not work after a couple of attempts, you must check all of the data entries in the spreadsheet and the Solver parameters until you find the error.

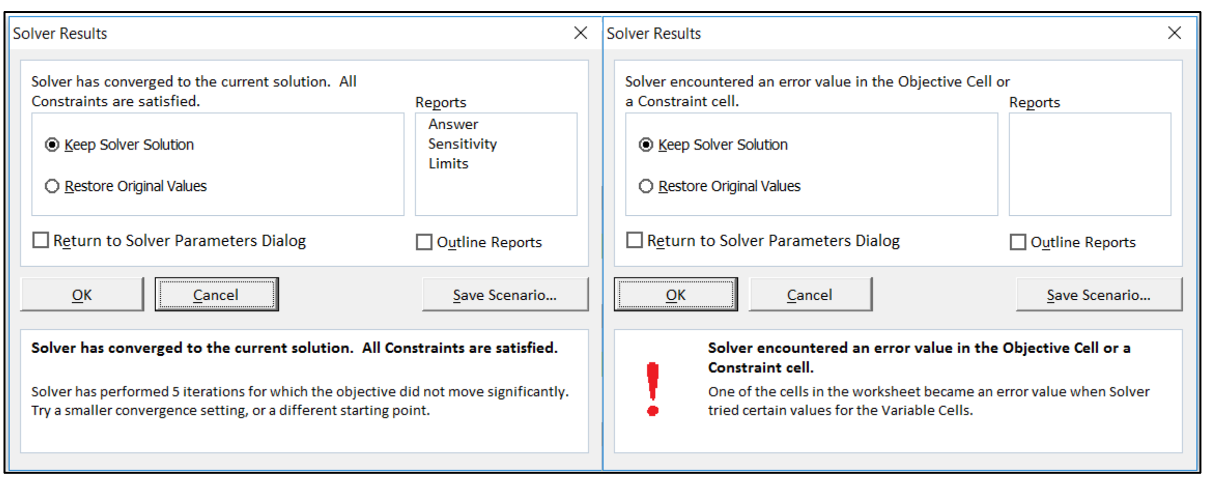

Figure 5: Examples of good and bad Solver results

For this example, Solver produces the results shown in Figure 6. The Rank line ranks the factors from most important to least. For this example, ease of shifting and noise from gears were determined to be the least important factors. The bar chart is included at the bottom of the sample spreadsheet. Note that the Objective value, which is not shown, is less than the original value and has been minimized.

There are also examples with generic factors which are ranked and with inconsistent RS matrices in the STAT COE consensus Excel tool. In addition, there are three more examples with fewer factors and SMEs. Remember that if you change any values in the spreadsheet after a solution is calculated, you need to rerun Solver and calculate an updated minimum.

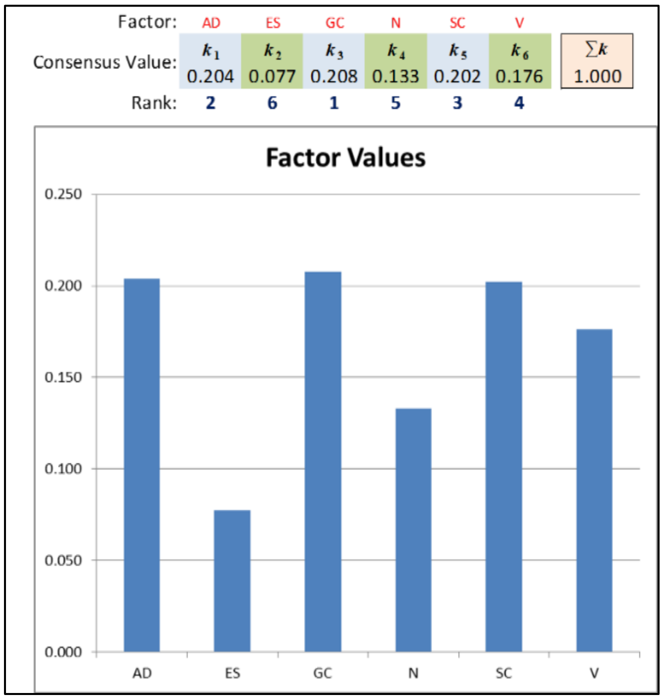

Figure 6: Solver Solution to Example Problem

References

Cook, Wade D., and Moshe Kress. Ordinal information and preference structures: decision models and applications. Prentice-Hall, Inc., 1992.

Frontline Solvers. "Standard Excel Solver - GRG Nonlinear Solver Stopping Conditions". 30 August 2017. https://www.solver.com/standard-excel-solver-grg-nonlinear-solver-stopping-conditions

Kennedy, Peter. *A guide to econometrics.* 6th ed. Malden, MA: Blackwell Publishing, 2011. pp. 12-17.