Using Excel for Dynamic Analysis of Variance and Unplanned Multiple Comparisons Procedures

K. Thomas Klasson

United States
Department of
Agriculture

Agricultural
Research Service

Southern Regional
Research Center

Technical Report
May 2020

The Agricultural Research Service (ARS) is the U.S. Department of Agriculture's chief scientific in-house research agency. Our job is finding solutions to agricultural problems that affect Americans every day from field to table. ARS conducts research to develop and transfer solutions to agricultural problems of high national priority and provide information access and dissemination of its research results.

K. Thomas Klasson is a Supervisory Chemical Engineer at USDA-ARS, Southern Regional Research Center, 1100 Robert E. Lee Boulevard, New Orleans, LA 70124; email: thomas.klasson@usda.gov

Using Excel for Dynamic Analysis of Variance and Unplanned Multiple Comparisons Procedures

K. Thomas Klasson

U.S. Department of Agriculture
Agricultural Research Service
Southern Regional Research Center
New Orleans, Louisiana, USA

Technical Report
May 2020

Abstract

Microsoft Excel is often used for data storage, calculations, data charting, etc. but not for complicated statistics. While some statistical functions and data analysis tools exist within the software, it is often not sufficient for most statistical evaluations. Therefore, a dynamic spreadsheet was developed that carried out standard analysis of variances evaluation and post hoc tests using unplanned multiple comparisons procedures (UMCPs). It was done by minimal programming, utilizing the existing functions of Excel, so that it can also be used as a teaching tool for UMCPs. The spreadsheet is dynamic and is updated automatically as existing raw data are revised or new data are entered. The spreadsheet is useful for all agricultural experimentalists who evaluate the impact on subjects receiving different treatments.

K. Thomas Klasson, Using Excel for Dynamic Analysis of Variance and Unplanned Multiple Comparisons Procedures

1. INTRODUCTION

Unplanned multiple comparisons procedures (UMCP); i.e., pairwise comparisons of mean results from individuals receiving different types of treatment are often the cornerstone of applied research. Determining that one treatment is better than another is often the desired outcome and a starting point for additional experimental studies. Consider the example in Figure 1, where hypothetical data for tomato plants receiving different types of fertilizers have been plotted. The results indicate that fertilizers 3 and 6 were significantly better than fertilizer 2 and 5 but were not statistically different than fertilizers 1 and 4.

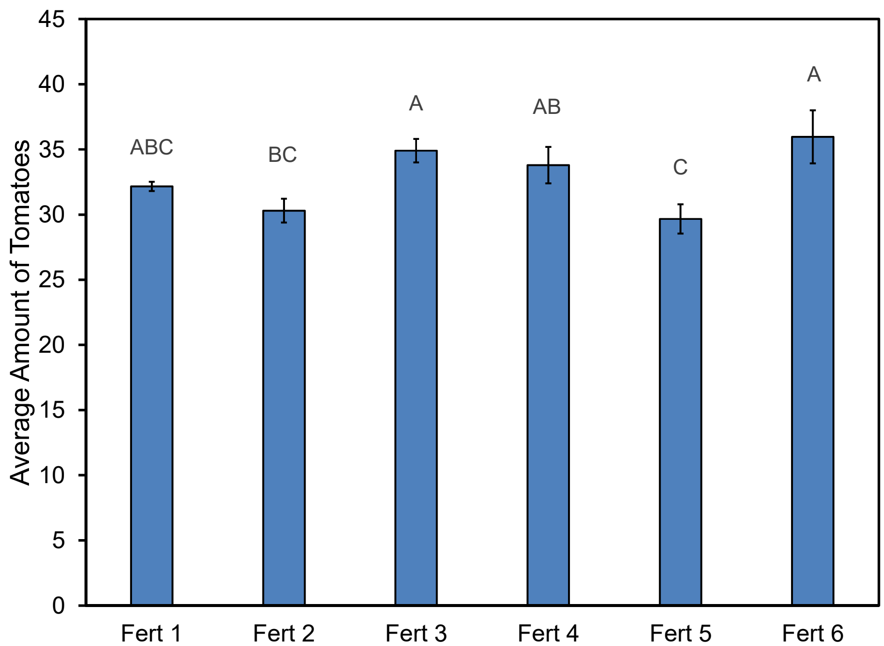


Figure 1. Hypothetical results showing tomato yield from plants receiving different fertilizers. Each fertilizer was applied to three plants and the average yield is shown. Letters above the bars indicate the result of the Fisher Least Significant Difference (LSD) post hoc statistical test with α equal to 0.05 and the error bars correspond to the standard error of the mean.

While the results displayed in Figure 1 represent the conclusion of the study, the data evaluation to get to this point usually begins with a standard one-way analysis of variance (ANOVA) approach and if the ANOVA indicates that the means are not the same, further testing is done by post hoc analysis by UMCPs which leads to a conclusion as the one shown in Figure 1.

Microsoft Excel (e.g., version 2007, 2013) offers some statistical functions. In addition, a few statistical tests (e.g., standard one-way ANOVA) are available in its Analysis ToolPak add-in. However, the ToolPak tests yield static results; i.e., if the data serving as input to the test are changed, the result of the test does not change unless we rerun the test. It would be far more useful if the results were dynamic and updated automatically when changes to the input data occur. Furthermore, UMCP tests are lacking from Excel. Within this manuscript, we will develop a dynamic spreadsheet for one-way analysis of variance and the post hoc analysis of means. Part of the reason this topic was chosen is that it is one of the most common tests performed by experimentalists at all levels. We see them reported at high school science fairs and in scientific articles. While there are dedicated statistical software that carry out UMCPs, they are often expensive or less user friendly than Excel.

2. SIMULATED EXPERIMENT

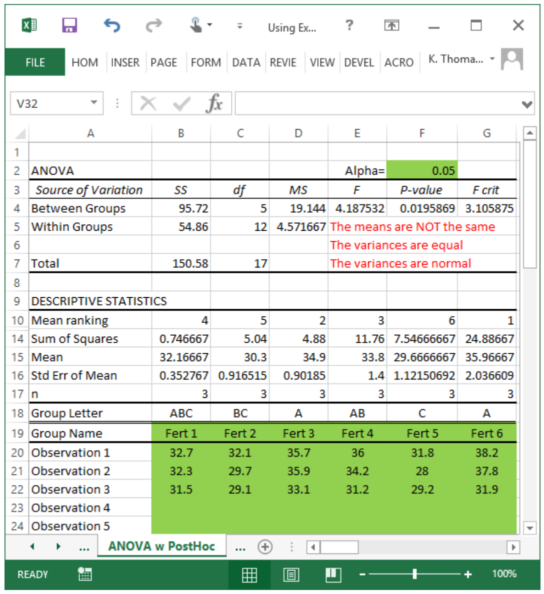
In order to have a more practical discussion, consider the hypothetical example in Figure 1. A tomato grower tested the impact of six different fertilizers on plant productivity. The grower seeded 18 different pots and randomly assigned three pots to each fertilizer. Let us assume that all the plants took root and received the same amount of sunlight and water (and other growth factors). The only difference was which type of fertilizer the plant received. At the end of the season, the grower tabulated the total quantity of tomatoes harvested from each plant. The outcome of the hypothetical experiment is listed in Table 1. While the number of plants (3) in each group (each group received a different fertilizer) was the same, this is not true for all experiments so and the methods discussed herein can also be used with different group sizes as long as only one factor (here fertilizer) is changed.

Table 1. Hypothetical data of the harvested amounts from tomato plants receiving different fertilizer treatments. The average and standard deviation are also shown.

	Fert 1	**Fert 2**	**Fert 3**	**Fert 4**	**Fert 5**	**Fert 6**
Rep 1	32.7	32.1	35.7	36.0	31.8	38.2
Rep 2	32.3	29.7	35.9	34.2	28.0	37.8
Rep 3	31.5	29.1	33.1	31.2	29.2	31.9
Avg.	32.17	30.30	34.90	33.80	29.67	35.97
S.D.	0.61	1.59	1.56	2.42	1.94	3.53

3. DEVELOPMENT OF A DYNAMIC SPREADSHEET FOR UMCP

Our challenge is to create a dynamic spreadsheet that is capable of performing all the necessary statistics needed to evaluate the results and provide the necessary information to construct a figure as the one shown in Figure 1. In Figure 2, a portion of such a spreadsheet is shown and the group names and data from the hypothetical tomato experiment have been added in range B19:G22. The spreadsheet is described below and the workbook is available via this link.



	A	B	C	D	E	F	G
1							
2	ANOVA				Alpha=	0.05	
3	*Source of Variation*	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
4	Between Groups	95.72	5	19.144	4.187532	0.0195869	3.105875
5	Within Groups	54.86	12	4.571667	The means are NOT the same		
6					The variances are equal		
7	Total	150.58	17		The variances are normal		
8							
9	DESCRIPTIVE STATISTICS						
10	Mean ranking	4	5	2	3	6	1
14	Sum of Squares	0.746667	5.04	4.88	11.76	7.54666667	24.88667
15	Mean	32.16667	30.3	34.9	33.8	29.6666667	35.96667
16	Std Err of Mean	0.352767	0.916515	0.90185	1.4	1.12150692	2.036609
17	n	3	3	3	3	3	3
18	Group Letter	ABC	BC	A	AB	C	A
19	Group Name	Fert 1	Fert 2	Fert 3	Fert 4	Fert 5	Fert 6
20	Observation 1	32.7	32.1	35.7	36	31.8	38.2
21	Observation 2	32.3	29.7	35.9	34.2	28	37.8
22	Observation 3	31.5	29.1	33.1	31.2	29.2	31.9
23	Observation 4						
24	Observation 5						

Figure 2. A section of our desired spreadsheet. The green sections are for user input.

3.1. Calculating Basic Descriptive Statistical Information from the Data

In the Excel spreadsheet, basic statistical information in the range B10:G17 is extracted or calculated from the data. The three hidden rows, rows 11-13, will be explained later.

Range B17:G17: The number of data values (n_j) for each group is determined using the Excel COUNT function.

Range B15:G16: The group means ($y_{\cdot j}$) and standard error of the means are calculated for each group using the Excel AVERAGE and STDEV functions but only if the group contains values (i.e., $n_j \neq 0$).

Range B14:G14: The sum of squares within each group is calculated using the Excel DEVSQ function but only if the group contains values (i.e., $n_j \neq 0$). The purpose of this calculation will be evident later.

Range B10:G10: The means are ranked in order of mean value, with a rank of 1 corresponding to the highest mean. The Excel function RANK is used. The hidden rows (11-13) can be reviewed by the reader. There it is shown how, when two means have the same rank, Excel functions are used to slightly and randomly adjust those means and perform the final rank. A random small fraction of the mean sum square error within groups is used as the adjustment. This adjustment of the means is not used beyond the ranking mechanisms. The purpose of the ranking will be evident later.

3.2. The ANOVA Table

The standard ANOVA test is often the starting point of any ANOVA procedure and is a requirement for some post hoc tests. The ANOVA table for the data in our example can be seen in Figure 2 in range A4:G7. The significance level (α) is entered into cell F2. The following procedures were used to calculate each cell value in the ANOVA table.

Cell B4: The sum of squares between the groups. From each group mean ($y_{\cdot j}$), the overall mean ($y_{\cdot\cdot}$) is subtracted, the difference is squared, and multiplied with the number of data point in that group (n_j). This procedure is done for each group and the resulting values are added together. The mathematical formula is

$$\sum_j n_j \left(y_{\cdot j} - y_{\cdot\cdot}\right)^2 \tag{1}$$

But in the spreadsheet we take advantage of the fact that this value can be obtained by subtracting cell B5 from cell B7.

Cell B5: The sum of squares within the groups is also known as sum of the square errors. For each data value (y_{ij}), its group mean ($y_{\cdot j}$) is subtracted and the difference is squared. This is done for all the data values and these are added together. The mathematical formula is

$$\sum_j \sum_i \left(y_{ij} - y_{\cdot j}\right)^2 \tag{2}$$

In the spreadsheet, the sum of squares for each group was calculated in range B14:G14 and the sum of this range is entered into cell B5.

Cell B7: The total sum of squares. From each data value (y_{ij}), the overall mean ($y_{\cdot\cdot}$) is subtracted and the difference is squared. This is done for all the data values and these squares are added together. The mathematical formula is

$$\sum_j \sum_i \left(y_{ij} - y_{\cdot\cdot}\right)^2 \tag{3}$$

Excel has a built-in function, DEVSQ, for exactly this task and we use that function on the entire range of possible data cells.

Cell C4: This is the degrees of freedom between the groups (df_b). This value is equal to the number of groups in the experiment (g) minus one. The number of means are counted using the COUNT function.

Cell C5: This is the degrees of freedom within the groups (df_w). This value is equal to the total number of data points (N) minus the number of groups (g). But in the spreadsheet we take advantage of the fact that this number can be calculated as the difference between cells C7 and C4.

Cell C7: This is the total degrees of freedom for the experiment. This value is equal to the total number of data points (N) minus one. In Excel, the number of data points in each group (n_j) are listed in range B17:G17 which are summed and one is subtracted in cell C7.

Cell D4: This is the mean squares between the groups. It is equal to cell B4 divided by cell C4.

Cell D5: This is the mean squares within the groups (MS_w). It is equal to cell B5 divided by cell C5.

Cell E4: This is the F-value. This value is equal to the cell D4 divided by cell D5.

Cell G4: This is the critical F-value for this experiment. The Excel function FINV is used to calculate the critical F-value based on α , d_{fb}, and df_w. If the F-value is less than the critical F-value, then there is no statistical difference between any of the groups and there is no need to do post hoc comparisons procedures. The result of the F test is shown in cell E5.

Cell F4: This is the *p*-value for the experiment and the Excel function FDIST is used to find the *p*-value. While this value is not used in this spreadsheet, it is often part of ANOVA tables.

3.3. Unplanned Multiple Comparisons Procedures

If the evaluation from the ANOVA table indicates that the means of the groups are statistically not the same ($F > F_{crit}$), unplanned comparisons procedures are done to see which of the means are statistically different and which means are statistically the same. In these procedures, every mean is systematically compared to every other mean and differences are noted. The method by which to prove differences between means is to compare the absolute value of the difference between two means (e.g., mean *a* and mean *b*) to a critical value (*CV*), which is calculated as follows:

$$CV = SF \times \sqrt{MS_w \times \left({}^{1}/_{n_a} + {}^{1}/_{n_b}\right)} \tag{4}$$

Where *SF* is a type of safety factor, which depends on the specific procedure that serves as the test. If the absolute difference between two means exceeds *CV*, the means are statistically different. Below are examples of safety factors by the Fisher Least Significant Difference (LSD), Fisher-Hayter, Tukey Honest Significant Difference (HSD), and Scheffe tests [1, 2].

Fisher LSD $$SF = \sqrt{F(\alpha, 1, df_w)} \tag{5}$$

Fisher-Hayter $$SF = Q(\alpha, g-1, df_w) \times \sqrt{{}^{1}/_{2}} \tag{6}$$

Tukey HSD $$SF = Q(\alpha, g, df_w) \times \sqrt{{}^{1}/_{2}} \tag{7}$$

Scheffe $$SF = \sqrt{(g-1)F(\alpha, g-1, df_w)} \tag{8}$$

Each of the statistical tests above uses either the Fisher–Snedecor Distribution (F) or Studentized Range Distribution (Q) tables that are available in most statistics test books. In the Excel spreadsheet, the F-distribution is directly available via the FINV function, and the Q-distribution is approximated with an Excel user-defined function, QTAB. Details of the user-defined functions QXLA and QTAB have recently been published [3, 4]. As another option, Q-values can be calculated in Excel using an Excel add-in available on the internet [5].

The selection and the calculation of the appropriate safety factor is performed in range MJ:M10 of the Excel spreadsheet (Figure 3) and can be explained as follows.

Range M4:M10: The *SF*-value is calculated for each of the tests described in Equations 4-8 using data from other parts of the spreadsheet and the FINV or QTAB functions.

Cell J10: The user choses the desired statistical test in cell J10 using the ID's listed in range J4:J8.

Cell M10: Based on the choice in cell J10, the *SF*-value is automatically chosen in cell M10 from the values in range M4:M8 using the INDEX function. There is also an additional test (ID=5), the Ryan-Einot-Gabriel-Welsch Q (REGWQ) test, that the user can select. We will discuss this test shortly.

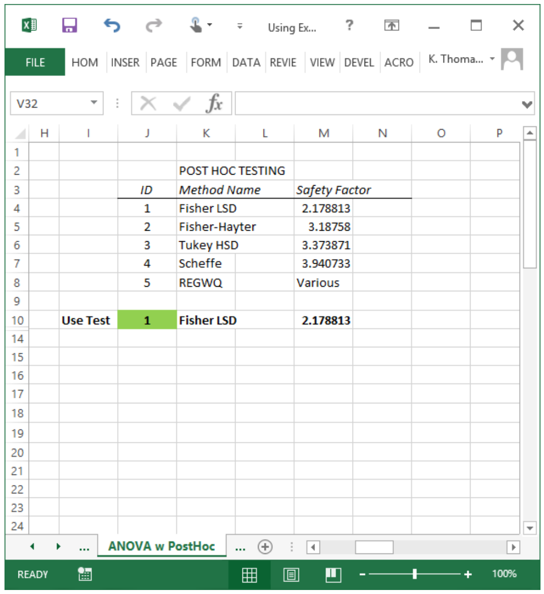


Figure 3. The section of the spreadsheet where the statistical multiple comparisons procedure is selected and most *SF*-values are calculated.

3.4. Comparing Means by the Step-down Procedure in Excel

For efficiency and other considerations, the systematic pairwise comparisons of means are done in a step-down procedure. In summary, this method begins by arranging the means in order of rank before comparisons are made on subsets of means in decreased order of subset size. The method is illustrated in Figure 4, top left. Here, the mean with largest value (rank 1) is compared with the mean with the lowest value (ranked last; e.g., rank 6 in the above example). If the means are found to be different, the next comparisons are made between the means ranked 2 and 6, and between the means ranked 1 and 5. The comparisons continue with smaller and smaller subsets of groups until all possible comparisons have been made. If, during this process, a comparison indicates that the means are statistically the same, then all the means within that subset are also statistically the same and no other comparisons are made within this subset. For example, if the means ranked 1 and 4 were found to be the same, then the means ranked 1, 2, 3, and 4 are the same and no other comparisons are necessary within this subset of groups. This corresponds to the dotted triangle in Figure 4, top left, with the 1 and 4 comparison at the triangle's apex. Similarly, if the means ranked 3 and 5 were found to be the same, then the means ranked 3, 4, and 5 are the same and no other comparisons are necessary within this subset. This corresponds to the dashed triangle in Figure 4, top left. And, if the means ranked 4 and 6 were found to be the same, then the means ranked 4, 5, and 6 are the same and no other comparisons are necessary within this subset. This corresponds to the solid triangle in Figure 4, top left. The pyramid-type arrangement of the comparisons in Figure 4, top left, can easily be translated into Excel by rotating and realigning the information. This procedure is demonstrated in the remainder of Figure 4. The three triangles (with dotted, dashed, and solid lines) indicate that there are three groupings of means. Each of them would correspond to a letter grouping as was shown in Figure 1. Where groups ranked 1 through 4 were assigned the letter A, the groups ranked 3 through 5 were assigned the letter B, and the groups ranked 4 through 6 were assigned the letter C.

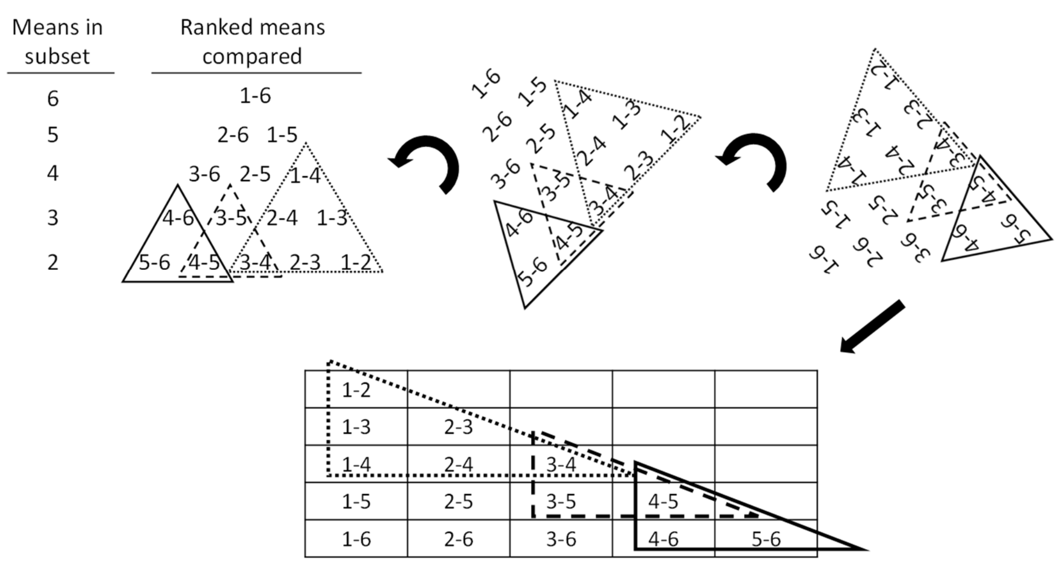


Figure 4. Demonstration of the step-down procedure and how it is translated into Excel.

The step-down procedure can be used with any of the comparisons tests listed above in Equations 4-8. It is also necessary for some of the more advanced procedures such as Student-Newman-Keuls and Ryan-Einot-Gabriel-Welsch F or Q tests [1].

The pairwise comparisons of means are performed in range T5:Y10 (Figure 5) using the step-down procedure explained above.

Ranges T3:Y4 and R7:S10: The means ($y_{\cdot j}$) and the number of values in each mean (n_j) and transferred according to the rank from range B10:G17 to ranges T3:Y4 and R7:S10 using the HLOOKUP function.

Range T5:Y10: The comparisons that take place in this range are carried out by an Excel equation that is rather complex with nested IF functions and takes several things into account, namely:

1. If a mean does not exist, then the comparison is not made. Instead the cell is left empty.
2. If the content of the cell immediately to the left or below contains "NSD" (which indicates that we are within a subgroup where no comparison needs to be made according to the step-down procedure) or if the difference between the means is not greater than *CV* (Equation 4) then "NSD" is displayed in the cell. If however the above criteria are not true, it indicates that there is a difference between the means and the symbol "**" is displayed in the cell.
3. All necessary variables (SF, MS_W, $y_{\cdot a}$, $y_{\cdot b}$, n_a, n_b) needed for the comparison are obtained from the different part of the spreadsheet. If the desired test is any other test than the REGWQ test, the *SF* value is obtained from cell M10. For the REGWQ test, the *SF* value is obtained elsewhere (see below).

The Ryan-Einot-Gabriel-Welsch Q (REGWQ) test is one of the recommended tests [1] but it is more complex to implement. In the REGWQ test, the safety factor (*SF*) is not constant as with the other tests (Equation 5-8). Instead, the *SF* depends on how many means are in the subgroup being compared. For example, if we compare the means ranked 1 and 4, there are 4 means in that subgroup. And, if we compare means ranked 3 and 5, there are 3 means in that subgroup. The safety factor for REGWQ is calculated as

REGWQ $$SF = Q(\beta, m, df_w) \times \sqrt{1/2} \quad (7)$$

Where m is the number of means in the subgroup and $\beta = \alpha$ if $m = g$ or $m = g-1$, otherwise $\beta = 1 - (1-\alpha)^{m/g}$. For each possible comparison, a separate *SF* is needed.

Range AB:AH10: This range is used to calculate *SF*-values for the REGWQ test, according to Equation 7, and contains *m*-values in range AB6:AB10, *β*-values in range AC6:AC10, and *SF*-values in range AD6:AH10. These REGWQ *SF*-values are accessed when needed in range T5:Y10 if the test selected in cell J10 is REGWQ.

3.5. Grouping Means Together and Assigning Letters

Finding the triangles and assigning the letters (Figure 4) is done in range T17:Y19. One can visually identify the triangles of "NSD's" in range T5:Y10 (Figure 5, emphasized). We can refer to the beginning of a triangle as the left-hand side of the triangle. For example, triangle 1 (Figure 5) begins in column T and stretches to column W. Triangles 2 and 3 begin in columns V and W, respectively, and stretch to columns X and Y, respectively. The process of identifying the triangles in the spreadsheet is based on counting the number of "NSD" in columns T to Y and determining if an existing triangle is continuing or a new one is starting.

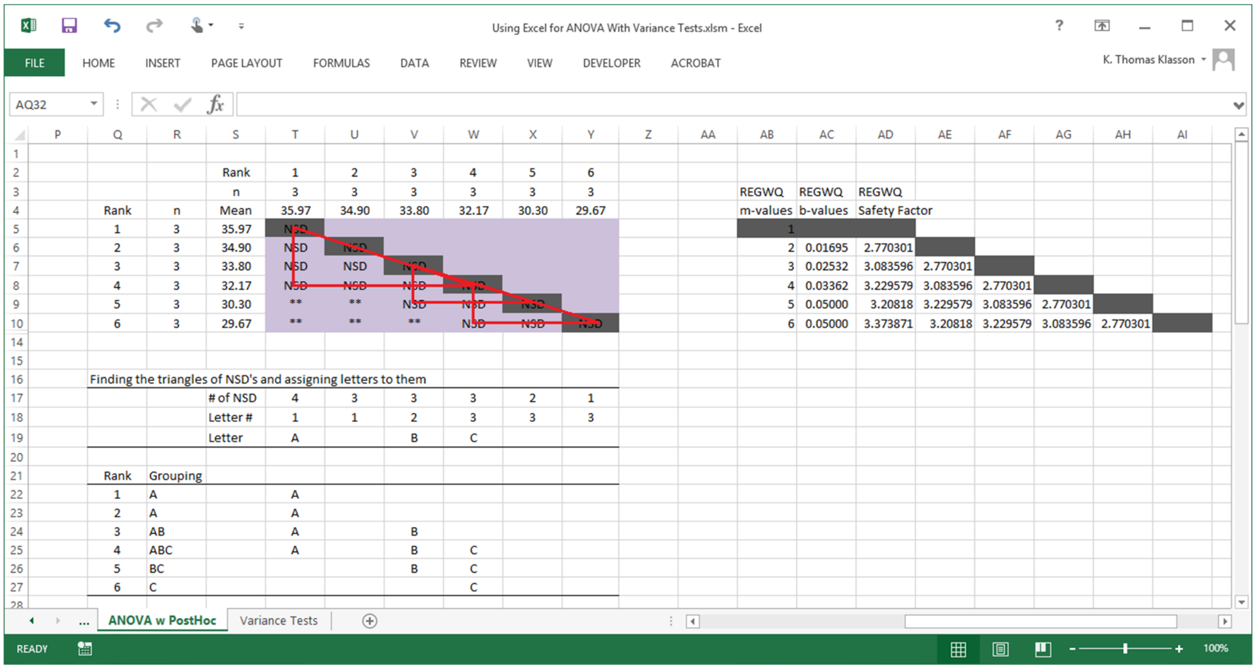
Range T17:Y17: The number of "NSD's" are counted in each column for rows 5 to 10 above using the COUNTIF function.

Range T18:Y18: The number of triangles and where they begin are determined based on the pattern of numbers in the row above (range T17:Y17). The reader is encouraged to consult the Excel equations to see how this is done.

Range T19:Y19: Once the triangles have been found, it is rather easy to assign letters (A, B, C, etc.) to the triangles and also to the highest ranked mean that is part of the triangle, which is also the mean in the column where the triangle begins. In our example, A is assigned to mean ranked 1, B is assigned to mean ranked 3, and C is assigned to mean ranked 4. We use the CHOOSE function to assign letters based on the numerical values in range T18:Y18.

Range T22:Y27: In this section, we assign letters to the individual means belonging to a triangle.

Range R22:R27: The individual letters for a group are then combined using the CONCATENATE function. The result of the grouping can then finally be transferred by the Excel INDEX function to the section of the spreadsheet (range B18:G18) close to the group names.



Using Excel for ANOVA With Variance Tests.xlsm - Excel

	Q	R	S	T	U	V	W	X	Y	AB	AC	AD	AE	AF	AG	AH
2			Rank	1	2	3	4	5	6							
3			n	3	3	3	3	3	3	REGWQ	REGWQ	REGWQ				
4	Rank	n	Mean	35.97	34.90	33.80	32.17	30.30	29.67	m-values	b-values	Safety Factor				
5	1	3	35.97	NSD						1						
6	2	3	34.90	NSD	NSD					2	0.01695	2.770301				
7	3	3	33.80	NSD	NSD	NSD				3	0.02532	3.083596	2.770301			
8	4	3	32.17	NSD	NSD	NSD	NSD			4	0.03362	3.229579	3.083596	2.770301		
9	5	3	30.30	**	**	NSD	NSD	NSD		5	0.05000	3.20818	3.229579	3.083596	2.770301	
10	6	3	29.67	**	**	**	NSD	NSD	NSD	6	0.05000	3.373871	3.20818	3.229579	3.083596	2.770301
16	Finding the triangles of NSD's and assigning letters to them															
17			# of NSD	4	3	3	3	2	1							
18			Letter #	1	1	2	3	3	3							
19			Letter	A		B	C									
21	Rank	Grouping														
22	1	A		A												
23	2	A		A												
24	3	AB		A		B										
25	4	ABC		A		B	C									
26	5	BC				B	C									
27	6	C					C									

Figure 5. Section of spreadsheet which performs the step-down multiple comparisons procedure, calculates safety factors for the REGWQ procedure, and assigns group letters to means that are statistically the same.

4. DISCUSSION

The spreadsheet that was developed is dynamic. If the data in range B20:G39 is replaced with other data, the rest of the spreadsheet updates automatically with the selected comparison test. It is easy to choose a different UMCP test and see the outcome without re-running the analysis. The automatic assignment of letters to grouping is particularly helpful as this is often desired in order to report and discuss results. The spreadsheet was constructed to handle six groups with up to 20 observations in each group but the concepts presented within this article, together with the downloadable Excel workbook, would make it easy to expand the spreadsheet to handle more groups or observations.

5. SUPPLEMENTARY MATERIALS

The spreadsheet discussed in this manuscript is available with this paper. The file is downloadable from this link and was created using Microsoft Excel 2013. Also included in the workbook is a spreadsheet that evaluates the equality and normality of the in-group errors of the data using Levene's test [6] for equality and the expanded Shapiro-Wilk test [5, 7] for normality. In-group error equality and normality are pre-requisites for ANOVA and the UMCP evaluations discussed here.

NOMENCLATURE

i = counter within a group

j = group counter

y_{ij} = data value of point i in group j

$y_{\cdot j}$ = mean value of data in group j

$y_{\cdot\cdot}$ = overall mean of all the data values (not necessarily the mean of means)

g = number of groups

n_j = number of data points in group j

N = total number of data points in the entire data set

MS_w = mean sum of squares within the groups (from an ANOVA table)

df_w = degrees of freedom within the groups (from an ANOVA table)

ACKNOWLEDGEMENTS

Mention of trade names or commercial products in this publication is solely for the purpose of providing specific information and does not imply recommendation or endorsement by the U.S. Department of Agriculture. USDA is an equal opportunity provider and employer.

REFERENCES

[1] Day RW, Quinn GP. Comparisons of treatments after an analysis of variance in ecology. Ecological Monographs. 1989;59:433-63.
[2] Richter SJ, McCann MH. Using the Tukey–Kramer omnibus test in the Hayter–Fisher procedure. British Journal of Mathematical and Statistical Psychology. 2012;65:499-510.
[3] Klasson KT. QXLA: Adding upper quantiles for the Studentized range to Excel for multiple comparison procedures. Journal of Statistical Software. 2018;85:1-9.
[4] Klasson KT. A simple but accurate Excel user-defined function to calculate studentized range quantiles for Tukey tests. New Orleans, LA: USDA, Agricultural Research Service, Southern Regional Research Center; 2019. p. 11.
[5] Zaiontz C. Real statistics using Excel, http://real-statistics.com; 2007 [accessed June 24 2018].
[6] Draper NR, Hunter WG. Transformations: Some Examples Revisited. Technometrics. 1969;11:23-40.
[7] Royston P. Approximating the Shapiro-Wilk W-test for non-normality. Statistics and Computing. 1992;2:117-9.