

## WORKING GROUP – LAUNDRY DETERGENT TEST PROTOCOLS

### User manual for the Evaluation Macro for Comparative Performance Tests

***Last revised: April 2016***

This Excel macro is a simple tool to evaluate the performance data resulting from using the A.I.S.E. minimum test protocol for comparative laundry detergent performance. It allows looking for statistical significant differences based on Tukey's HSD (Honest Significant Difference). It is flexible to group stains/items together and to look for different responses like Stain removal, whiteness maintenance parameters, etc. It contains also a tool to identify single outliers within a group replicates based on the Dixon test.

#### Creating a new set of data for evaluation:

**Step 1) Click on the Sheet: Data:**

**Step 2) Press button: Design new test'**

Fill data table (only the yellow fields)					
	Source 1	Products 8	Fabric/stain 14	Rep. external 6	Rep. internal 1
counts	1				
# of data rows:	Lab	Product A	tea	1	1
672		Product B	coffee	2	
# of data sets:		Product C	make-up	3	
2		Product D	dirty motor oil	4	
		Product E	fruit juice	5	
		Product F	chocolate	6	
		Product G	grass		
Data table is:		Product H	grass/mud		
<b>Complete</b>			tomato puree		
# of data points:			blood		
1344			red wine		
Selected data:			used cooking oil		
1			vegable oil		
			carrot baby food		

Design new test

Fill table with data

Dixon test for outliers


Do you want replace outliers by the average?

☒ No
☐ Yes

Screen and select item

If you are sure, press **Yes** - all the data in the data table from a previous test will be deleted. The labels like Fabric/stain will be kept but can be modified in the next step

Design/Input of a new test



Designing and starting a new test  
will automatically delete the data table!

Do you want to continue?

### Step 3) Fill in Parameter table

There are several columns to fill in with your testing parameter: Only the **yellow cells** should be changed in this step.

**Source:** The current version only works with one data source. So add only one cell with a text.

**Products:** You can add different product as a text. It is recommended to use short alpha-numerical description like A, B, C, A1, A2, B1, B2, or a short name like: Product A, Product B, etc. The total number of products can be between 2 to 26 for one test. However, high numbers may slow down the computer.

**Fabric/Stain:** You can specify the stains by any text as shown in the example. The total number of stains can vary between 1 and 26 depending on the design. Again, high numbers may slow down the computer.

Fill parameter table - only the yellow cells					
counts	Source 1	Products 8	Fabric/stain 14	Rep. external 6	Rep. internal 1
# of data rows:	Lab	Product A	tea	1	1
672		Product B	coffee	2	
# of data sets:		Product C	make-up	3	
2		Product D	dirty motor oil	4	
Data table is:		Product E	fruit juice	5	
Empty		Product F	chocolate	6	
# of data points:		Product G	grass		
0		Product H	grass/mud		
Selected data:			tomato puree		
1			blood		
			red wine		
			used cooking oil		
			vegable oil		
			carrot baby food		

**Rep. external:** Every stain needs to be measured at replicates either external or internal. The number needs to start with 1 and ending with the total number of external replicates.

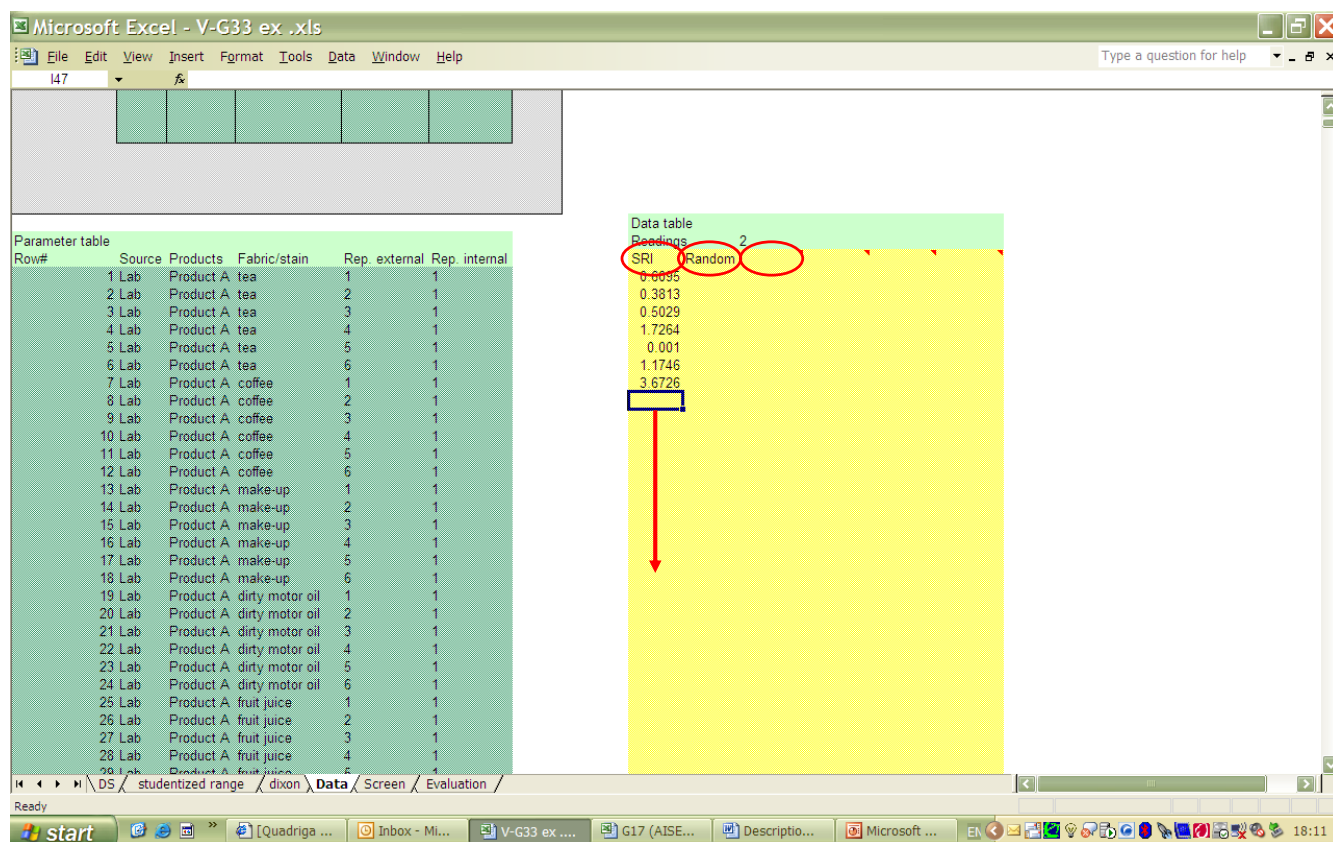
**Rep. internal:** The same as before. The total number of replicates for the test is defined as "Rep. external x Rep. internal" – in this example it is 6.

The number of expected data values will be calculated from the set of parameters entered, and should not be changed manually (see red circles). In this example there are 672 rows of data expected and for each row we have added 2 responses (e.g. stain removal % at 30C and at 40C).

#### Step 4) Press button: Fill table with data

The parameter table will be created, after which you can add the response data sets. For every response like SRI (Stain Removal Index) you can add the data in the below column. Write also the name of the response in the first row, like “SRI” in the example. In total 6 different responses can be added.

Important: The data in the columns below need to correspond with the parameter table. You can use the Sort tool (Data -> Sort) in a separate Excel file to get them in the right sequence/order. The set needs to be complete. If you are missing one data point, you can use an average value of the replicates to estimate one missing data point.



The screenshot shows a Microsoft Excel spreadsheet with two tables. The 'Parameter table' is on the left, and the 'Data table' is on the right. The 'Data table' has a yellow background and a red arrow pointing to a cell in the 'Random' column.

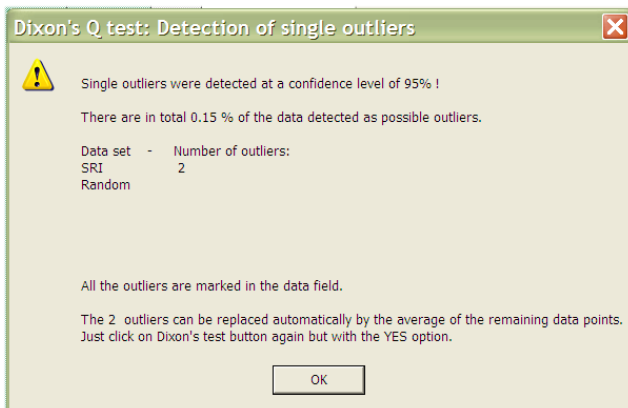
Row#	Source	Products	Fabric/stain	Rep. external	Rep. internal
1	Lab	Product A	tea	1	1
2	Lab	Product A	tea	2	1
3	Lab	Product A	tea	3	1
4	Lab	Product A	tea	4	1
5	Lab	Product A	tea	5	1
6	Lab	Product A	tea	6	1
7	Lab	Product A	coffee	1	1
8	Lab	Product A	coffee	2	1
9	Lab	Product A	coffee	3	1
10	Lab	Product A	coffee	4	1
11	Lab	Product A	coffee	5	1
12	Lab	Product A	coffee	6	1
13	Lab	Product A	make-up	1	1
14	Lab	Product A	make-up	2	1
15	Lab	Product A	make-up	3	1
16	Lab	Product A	make-up	4	1
17	Lab	Product A	make-up	5	1
18	Lab	Product A	make-up	6	1
19	Lab	Product A	dirty motor oil	1	1
20	Lab	Product A	dirty motor oil	2	1
21	Lab	Product A	dirty motor oil	3	1
22	Lab	Product A	dirty motor oil	4	1
23	Lab	Product A	dirty motor oil	5	1
24	Lab	Product A	dirty motor oil	6	1
25	Lab	Product A	fruit juice	1	1
26	Lab	Product A	fruit juice	2	1
27	Lab	Product A	fruit juice	3	1
28	Lab	Product A	fruit juice	4	1
29	Lab	Product A	fruit juice	5	1
30	Lab	Product A	fruit juice	6	1

Readings	2
SRI	Random
0.6095	
0.3813	
0.5029	
1.7264	
0.001	
1.1746	
3.6726	

### Step 5) Scroll back to the top part of the sheet and press button: Dixon test for outliers

Only if the data set is complete (see red circle on the left) the test for outliers can be done. Run this test for every new set of data.

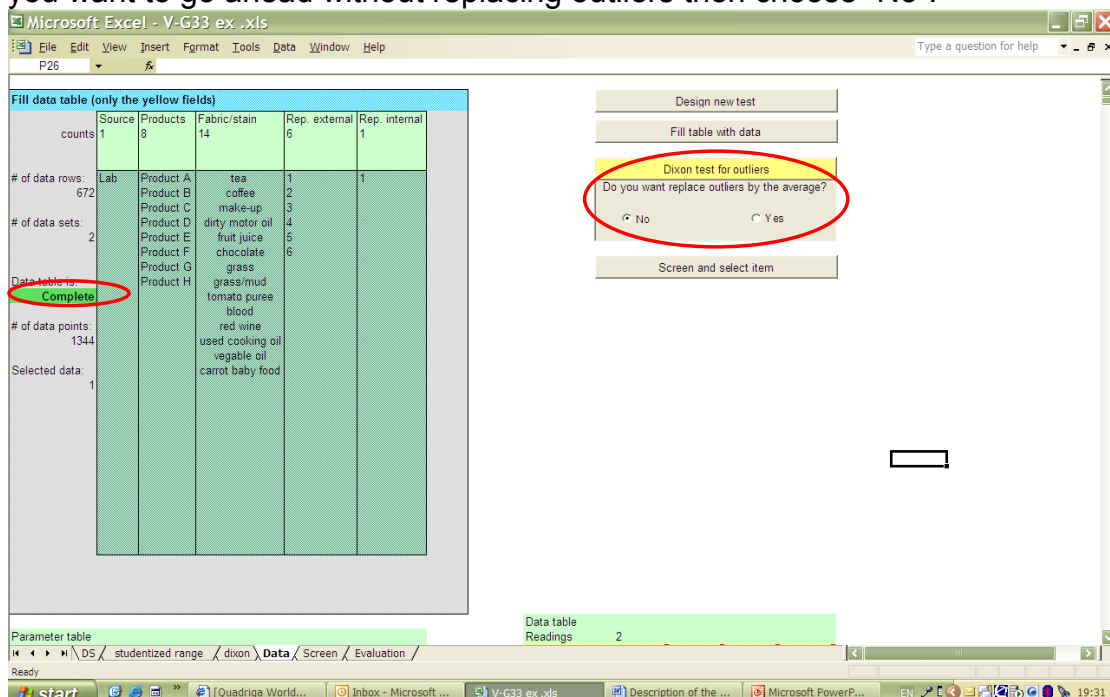
The Dixon's outlier test is a simple test to assess if one (and only one) observation from a small set of replicate observations can be rejected or not. Usually, an outlier is defined as an observation that is quite different from the main "body" of data. See more at link: [Dixon test](#).



This test is intended to reduce the risk of overlooking obvious outliers due to mistakes in the measurement (wrong sample), typing error, or lost data (wrong value) during the process of data gathering. It is not intended to clean up data of a poorly executed test. Therefore every outlier should be checked as a possible mistake during the measurement or data handling. At this point in time you still have the option to manually replace the outlier value in the dataset. Do this if you have checked values that were identified as outliers, and that you have

produced new values for (e.g. by re-measuring, etc).

When you are satisfied with the integrity of every data point, chose whether to replace the remaining outliers with the average of the other replicate results. Although the default is on "Yes", if you want to go ahead without replacing outliers then choose "No".

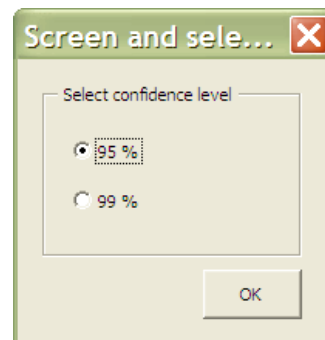


A message box appears where you can select the level of confidence for the Dixon test, either 95% or 99%. The recommendation is to select 95%. After pressing OK another message box opens with the details of the test. In the example there are 2 outliers in the first data set (SRI) and none in the second set (Random).

## Step 6) Press button: Screen and select item

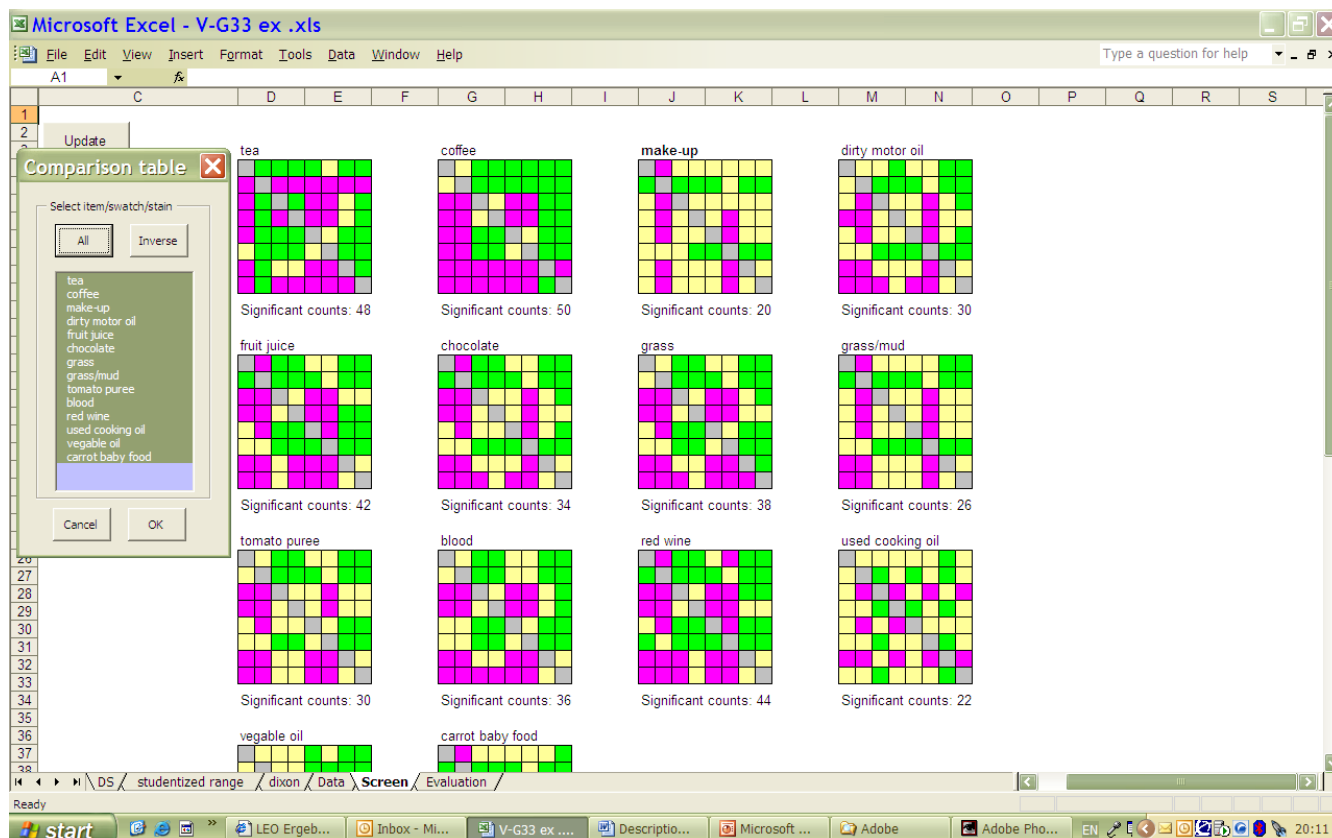
A message box will appear where you can select the confidence level, either 95% or 99%. Again 95% is recommended. You can change this confidence level later again if you want to see the effect of the confidence level in your evaluation.

After pressing OK, a new sheet automatically appears:



## 1) Selecting items in the screen window

Every stain is presented as a square. In every square, smaller squares with different colors are plotted. They are used as a simple indication for significant differences based on HSD (Honest Significant Difference). Green means that there is a significant positive difference. Purple means that there is a significant negative difference. Yellow means that there is no significant difference between the products. The orders of the boxes are according the data table, from top to the bottom and from left to right, in this example from product A to product H.





## Choosing the stains

The AISE minimum protocol for comparative laundry detergent performance testing recommends the following selection of individual stains, and possible ways of clustering them:

Stains	Standard Stains			Hand-made Stains* (ex Warwick-Equest)	Stain classes Consumer denomination/Chemical nature
Tea		WFK 10J		WE5LTWKC	Drink / Bleachable
Coffee			CFT KC-H109	WE5ECWKC	Drink / Bleachable
Red wine			CFT KC-H026	WE5RWWKC	Drink / Bleachable
Fruit juice			CFT CS15		Drink / Bleachable
Tomato puree				WE5TPWKC	Food / Bleachable
Carrot baby food				WE5IACBFWKC	Food / Bleachable Enzymatic
French Squeezy Mustard				WE5FSMWKC	Food / Bleachable Enzymatic
Chocolate		WFK 10Z	CFT CS44		Food / Enzymatic
Grass	EMPA 164		CFT CS08		General soil / Bleachable Enzymatic
Grass/Mud				WE5GMWKC	General soil / Bleachable Enzymatic Particulate
Blood				WE5DASBWKC	General soil / Enzymatic
Unused motor oil	EMPA 106	WFK 10 RM	CFT C01		Grease, Oil / Greasy Particulate
Frying fat (Hamburger grease)				Burnt Beef WE5BBWKC (on WHITE cotton)	Grease, Oil / Greasy Enzymatic
Make up	EMPA 143/2	WFK 10MU	CFT CS17		Cosmetics / Greasy Particulate

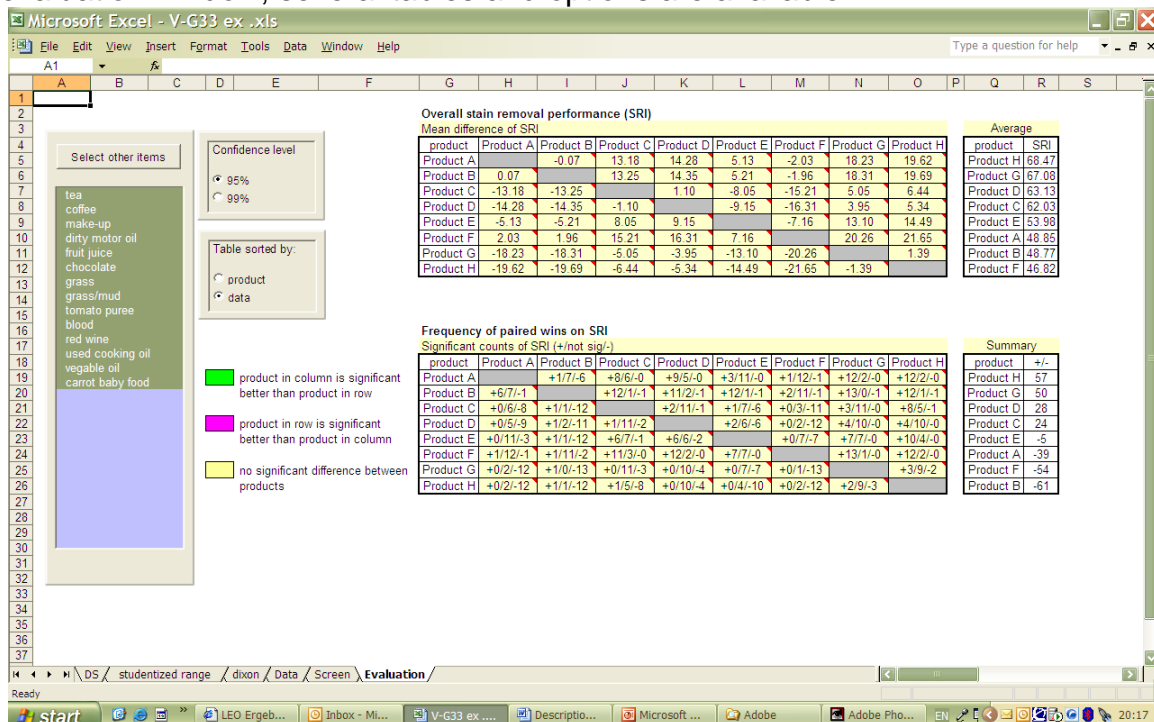
### Example: Red Wine Stain

The red wine stain belongs to the cluster of the bleachable stains. To understand the influence of the performance on wine on the performance on “bleachables” or “overall”, one does the following:

- Select Red Wine and press “OK”. Refer to the section “Evaluating the data” to interpret.
- Select the stains in the bleachables cluster (tea, coffee, red wine, fruit juice, tomato puree, French squeezy mustard, grass, grass/mud) and run the analysis. Compare the results and aim to describe and rationalize the differences.
- Select all stains and run the analysis one more time. Again compare the results and aim to describe and rationalize the differences.

## 2) Evaluating the data

In the evaluation window, several tables and options are available.



**Overall stain removal performance (SRI)**

Mean difference of SRI

product	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Product A		-0.07	13.18	14.28	5.13	-2.03	18.23	19.62
Product B	0.07		13.25	14.35	5.21	-1.96	18.31	19.69
Product C	-13.18	-13.25		1.10	-8.05	-15.21	5.05	6.44
Product D	-14.28	-14.35	-1.10		-9.15	-16.31	3.95	5.34
Product E	-5.13	-5.21	8.05	9.15		-7.16	13.10	14.49
Product F	2.03	1.96	15.21	16.31	7.16		20.26	21.65
Product G	-18.23	-18.31	-5.05	-3.95	-13.10	-20.26		1.39
Product H	-19.62	-19.69	-6.44	-5.34	-14.49	-21.65	-1.39	

**Frequency of paired wins on SRI**

Significant counts of SRI (+/not sig/-)

product	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Product A		+1/7/-6	+8/6/-0	+9/5/-0	+3/11/-0	+1/12/-1	+12/2/-0	+12/2/-0
Product B	+6/7/-1		+12/1/-1	+11/2/-1	+12/1/-1	+2/11/-1	+13/0/-1	+12/1/-1
Product C	+0/6/-8	+1/1/-12		+2/11/-1	+1/7/-6	+0/3/-11	+3/11/-0	+8/5/-1
Product D	+0/5/-9	+1/2/-11	+1/11/-2		+2/6/-6	+0/2/-12	+4/10/-0	+4/10/-0
Product E	+0/11/-3	+1/1/-12	+6/7/-1	+6/6/-2		+0/7/-7	+7/7/-0	+10/4/-0
Product F	+1/12/-1	+1/11/-2	+11/3/-0	+12/2/-0	+7/7/-0		+13/1/-0	+12/2/-0
Product G	+0/2/-12	+1/0/-13	+0/11/-3	+0/10/-4	+0/7/-7	+0/1/-13		+3/9/-2
Product H	+0/2/-12	+1/1/-12	+1/5/-8	+0/10/-4	+0/4/-10	+0/2/-12	+2/9/-3	

**Summary**

product	+/
Product H	57
Product G	50
Product D	28
Product C	24
Product E	-5
Product A	-39
Product F	-54
Product B	-61

**Confidence level**

☒ 95%  
☐ 99%

**Table sorted by:**

☐ product  
☒ data

**Legend:**

- product in column is significant better than product in row
- product in row is significant better than product in column
- no significant difference between products

### Confidence level and sorting data

Chose the Confidence level (default at 95% - recommended to keep it at this level), and how the data must be sorted (default at “data”).

Changing either option causes an immediate re-calculation of evaluation tables. The recommended level is 95% confidence.

### Interpretation of results

Two tables summarize the results of the statistical analysis. It is important to consider the information in both tables, as they give an equally pertinent but different view on the results of the test. Because of this, it is to be expected that the relative rankings between products can be different between the two tables. If this is the case, it is important to identify the root cause by looking at the results for a specific

cluster of stains or for one, or more, stains individually. Working with this macro in this way will enable a balanced assessment of the test.

#### First data table: Overall stain removal performance

This table compares the products. For example, product C is 13.18 SRI units better than Product A, while Product B provides nearly the same level of removal as product A (-0.07 SRI units). To the right, the absolute SRI values of every product on the stains are selected. In this example, product H shows the highest overall performance while Product F the lowest. Statistically significant differences can only be shown when selecting a single stain. When selecting more than one stain no indication of statistical difference is given (as in the example below)

#### **Overall stain removal performance (SRI)**

##### Mean difference of SRI

product	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Product A		-0.07	13.18	14.28	5.13	-2.03	18.23	19.62
Product B	0.07		13.25	14.35	5.21	-1.96	18.31	19.69
Product C	-13.18	-13.25		1.10	-8.05	-15.21	5.05	6.44
Product D	-14.28	-14.35	-1.10		-9.15	-16.31	3.95	5.34
Product E	-5.13	-5.21	8.05	9.15		-7.16	13.10	14.49
Product F	2.03	1.96	15.21	16.31	7.16		20.26	21.65
Product G	-18.23	-18.31	-5.05	-3.95	-13.10	-20.26		1.39
Product H	-19.62	-19.69	-6.44	-5.34	-14.49	-21.65	-1.39	

##### Average

product	SRI
Product H	68.47
Product G	67.08
Product D	63.13
Product C	62.03
Product E	53.98
Product A	48.85
Product B	48.77
Product F	46.82

#### Second data table: Frequency of paired wins

The second table shows the results for every paired product comparison and for every selected stain. As for the table above, all (14) stains were selected in this example. Product C is significantly better than product A on 8 stains and no significant difference can be detected on 6 stains, therefore it is quoted “+8/6/-0”. To the right, is a sum total of all these significant wins and losses. Product H is best with an overall +57 counts of significant wins versus all other products, while product B loses overall 61 times. For example C = +24 and is calculated as following (from A to H): (8-0) + (12-1) + (1-2) + (6 -1) + (11-0) + (0-3) + (1-8) = +24.

#### **Frequency of paired wins on SRI**

##### Significant counts of SRI (+/not sig/-)

product	Product A	Product B	Product C	Product D	Product E	Product F	Product G	Product H
Product A		+1/7/-6	+8/6/-0	+9/5/-0	+3/11/-0	+1/12/-1	+12/2/-0	+12/2/-0
Product B	+6/7/-1		+12/1/-1	+11/2/-1	+12/1/-1	+2/11/-1	+13/0/-1	+12/1/-1
Product C	+0/6/-8	+1/1/-12		+2/11/-1	+1/7/-6	+0/3/-11	+3/11/-0	+8/5/-1
Product D	+0/5/-9	+1/2/-11	+1/11/-2		+2/6/-6	+0/2/-12	+4/10/-0	+4/10/-0
Product E	+0/11/-3	+1/1/-12	+6/7/-1	+6/6/-2		+0/7/-7	+7/7/-0	+10/4/-0
Product F	+1/12/-1	+1/11/-2	+11/3/-0	+12/2/-0	+7/7/-0		+13/1/-0	+12/2/-0
Product G	+0/2/-12	+1/0/-13	+0/11/-3	+0/10/-4	+0/7/-7	+0/1/-13		+3/9/-2
Product H	+0/2/-12	+1/1/-12	+1/5/-8	+0/10/-4	+0/4/-10	+0/2/-12	+2/9/-3	

##### Summary

product	+/-
Product H	57
Product G	50
Product D	28
Product C	24
Product E	-5
Product A	-39
Product F	-54
Product B	-61

These win/equal/losses counts are based on statistically significant differences versus all the stains and products and depend on the confidence level. Changing this level will automatically update the tables.





International Association for Soaps,  
Detergents and Maintenance Products

As the protocol itself emphasizes, it is critical to assess the test on all the data, i.e. by looking at the statistical analysis of a combination of all the stains (for the stain removal performance), and all the whiteness tracers (for the whiteness maintenance performance). Assessing performance on stain clusters and individual stain has second, resp. third priority. In case of ambiguity, we recommend to use expert judgment (e.g. watch out in the case of near complete removal of the stain where instrumental assessment may exaggerate differences...). If still in doubt, do more testing in order to come to a robust final conclusion (e.g. visual inspection of stains,...).

The above needs to be taken in the context of the “A.I.S.E. Laundry Detergent Testing Guidelines – Minimum requirements for comparative detergents testing”.