ARM, ANOVA and all the Rest ...

More and more participants in our seminars have been worried lately about the procedures and interpretation of results relating to the well-known “type-2 study”, “R&R” or “GRR method”.

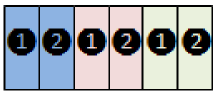
The days when calculation methods were explained with the help of a thumbnail sketch and some ranges are long gone. Back then, the magic of the truth was hidden in two k factors. Ever since the release of the 4th edition of the AIAG MSA manual, the method of ANOVA has always been recommended whereas ARM is only allowed in case of an “emergency”.

This fact seems to be justified by the fact that ARM only determines EV and AV components (repeatability and reproducibility) whereas the method of ANOVA also evaluates interactions IA. So this is enough information to enter into the discussion, maybe even enough to end it, but what is really behind it?

First of all, here is a brief diagram describing type-2 study. Typically, 2-3 operators take 2-3 repeated measurements on ten different workpieces. This is an ordinary full factorial design of experiments; however, we do not like to underline this information in our MSA seminars.

Example: 3 operators take 2 repeated measurements on 10 parts.

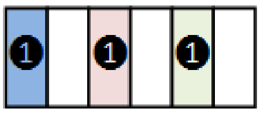
Operator A		Operator B		Operator C	
Trial 1	Trial 2	Trial 1	Trial 2	Trial 1	Trial 2
1-A1	1-A2	1-B1	1-B2	1-C1	1-C2
2-C1	2-C2	2-B1	2-B2	2-C1	2-C2
3-C1	3-C2	3-B1	3-B2	3-C1	3-C2
4-C1	4-C2	4-B1	4-B2	4-C1	4-C2
5-C1	5-C2	5-B1	5-B2	5-C1	5-C2
6-C1	6-C2	6-B1	6-B2	6-C1	6-C2
7-C1	7-C2	7-B1	7-B2	7-C1	7-C2
8-C1	8-C2	8-B1	8-B2	8-C1	8-C2
9-C1	9-C2	9-B1	9-B2	9-C1	9-C2
10-C1	10-C2	10-B1	10-B2	10-C1	10-C2

This is a simplified illustration of the diagram: 

What is the scope of ARM?

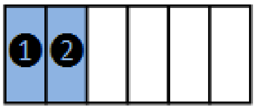
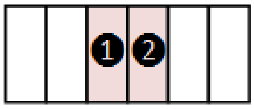
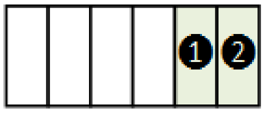
This method tries to answer the following questions:

- How precisely can the operator repeat his measurement?
  In order to answer this question, the average ranges between



and

are compared. The result is referred to as “repeatability“.

- To what extent are the operators‘ measurements comparable to one another?
  In order to answer this question, the range between
   ,  and 
  is assessed. The result is referred to as “reproducibility“.

Any other influences are not considered - just because nobody asks for them.

The next step is to convert the ranges to standard deviations. Everyone knows that even in case of identical standard deviations, various subgroups might have different ranges. For this reason, we base our decision on the w distribution and choose the standard deviation that is most likely to lead to the range we gained. This method is generally accepted, however, it takes some courage. Do you generally assume that you will always draw the likeliest number in the lottery? That would be a pity because you would never win anything.

What is the scope of the method of ANOVA?

The method of ANOVA follows a different approach. First, “all variation” of the 60 measured values is lumped together. Then you start fishing in troubled waters.

- You may calculate the part variation from the means of the 10 parts and deduct it from the total variation.
- You may calculate reproducibility from the means of the operators and deduct it from the total variation.
- At this point, we still leave out interactions...
- What remains of the total variation has to refer to repeatability.

Thus the variation of repeatability is “the rest“ also including all those influences the ARM does not consider. This is the main difference between the method of ANOVA and ARM.

What do EV and AV mean?

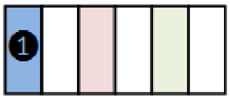
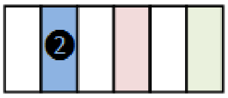
It was a very unfortunate choice to name these factors like this.

- EV = equipment variation = repeatability
- AV = appraiser variation = reproducibility

At least the meaning of these terms makes us stumble. EV means “equipment variation“ relating to the variation of the measuring equipment. Unfortunately, we therefore relate this variation component to a cause without any reason. And it is almost the same with the term AV, appraiser variation. It need not be the operator’s fault if the measured values of the operators are not reproducible.

Example: In one of our seminars, participants had the task to measure the height of a centering bush made of aluminum. They were allowed to make mistakes because you normally learn from them. So they decided that the operators take both their repeated measurements one after another. The first operator touched the aluminum bush with the outer edges of the outside measuring jaws. For this reason, he could not repeat his measurement.

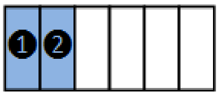
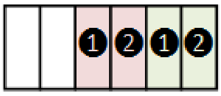
This was definitely an error caused by the operator, but

 did not comply with 

which led to an equipment variation that was too high in terms of measurement repeatability. It is definitely wrong to call this variation caused by the operator “equipment variation”.

After the first operator had watched both other operators using the micrometer correctly and achieving far better results, we discussed the results. The first operator wanted to repeat his measurements. He took the micrometer and started measuring – without resetting the micrometer that had not been adjusted to zero in the meantime.

Now he was able to repeat his measurements accurately. He has learned from his earlier attempt, but his values were not comparable to the other ones because of instrumental drift. This means that

 did not comply with 

which led to an AV that was too high in terms of measurement reproducibility. It is just as wrong to call this variation caused by instrumental drift “appraiser variation”.

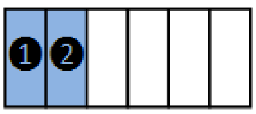
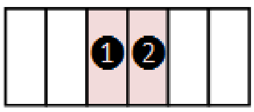
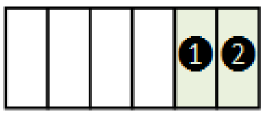
However, we have to accept these terms because even VDA Volume 5 uses them.

What is an interaction IA?

Even in this case, you have to be careful with the term. The term “interaction“ is understood in a very general sense and must not be mistaken for an interaction used in the classical statistical design of experiments. The statistical design of experiments distinguishes between two-factor, three-factor or any other multiple-factor interactions. A two-factor interaction exists, for example, when we add a specified amount of air at different temperatures. The change in the volume depends on the temperature and the whole issue can be described by means of a formula. This interaction is a systematic interaction that may be applied in production to find the optimum of a process.

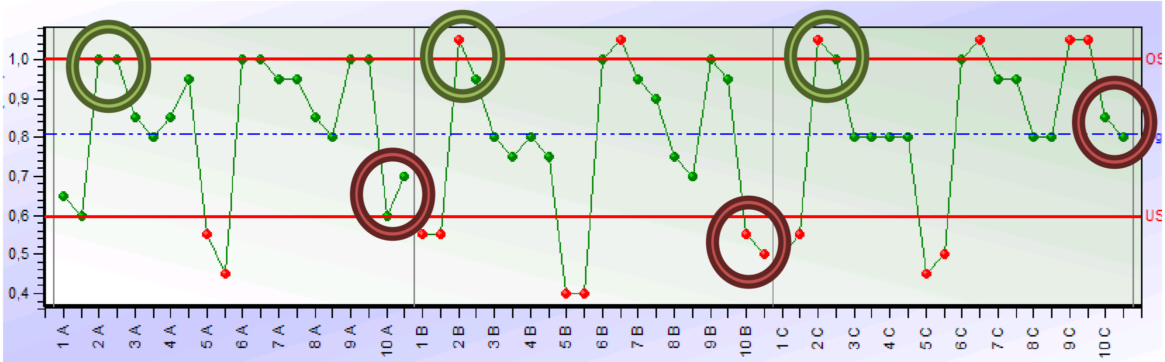
The measurement system analysis uses the term “interaction” to signify that the variation of one or several of the ten parts differs significantly from expected reproducibility. There are various reasons for this problem, some of them are random, but almost all of them are annoying. Consequently, the interaction is not a property or an “accepted spread“ of the measurement system, such as repeatability, but more of an evidence indicating a still undetected error in the system. The full extent of the actual impacts of this error might not be known yet.

Example: In our experiment, three operators took two repeated measurements on each one of the ten parts. This led to six measured values per part. Based on these values, we now calculate the reproducibility from the mean deviations of

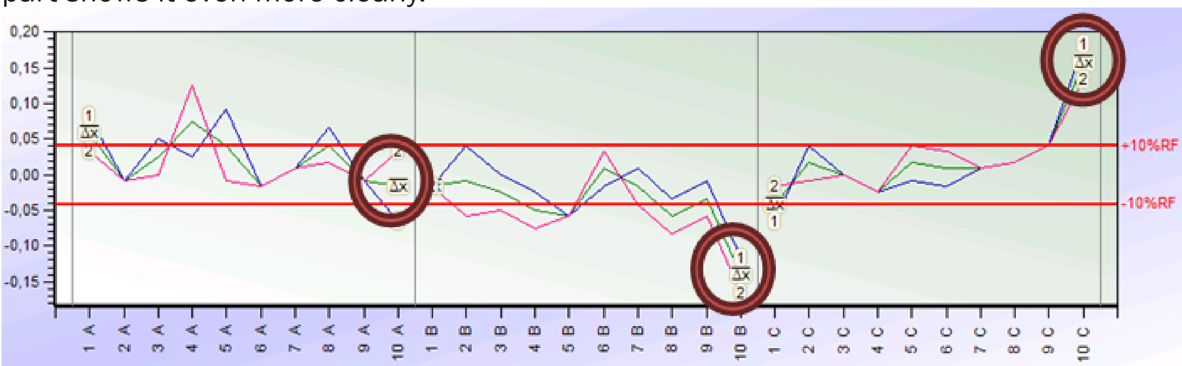
 ,  and 

We compare the variation of each one of these ten parts to the expected mean deviation. If it is obvious that the variation of some single parts does not comply with the expected mean deviation at all, the “appraiser variation“ of these single parts is significantly different from AV. In this case we have to assume that another influence affects these appraiser-part combinations, which is simply called interaction.

Example: Three operators A, B and C measure each one of 10 parts twice.



This diagram makes it clear that the tenth part shows a huge difference that is more considerable compared to all the other parts (as an example, we also highlighted the second part). The diagram below displaying the deviations of each individual measurement from the mean of the respective part shows it even more clearly.

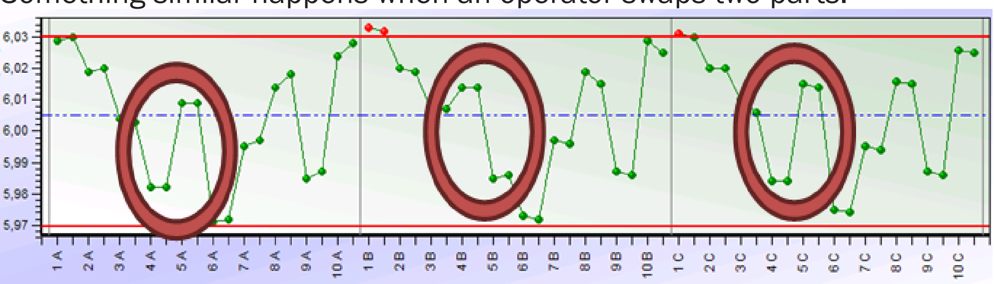


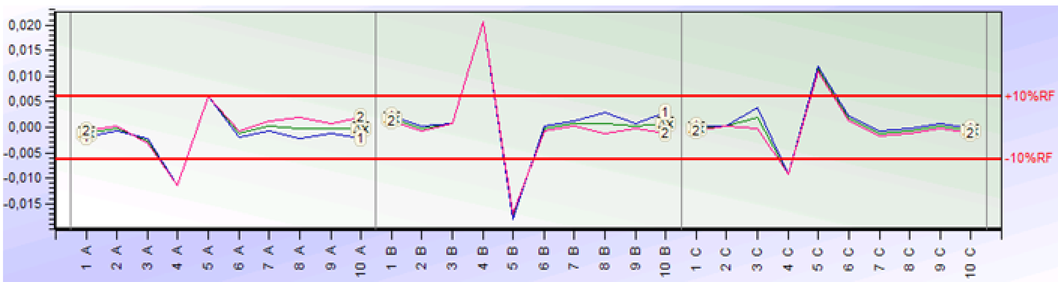
The result indicates a high interaction IA.

0,00129	0,0359	EV	=	0,148 ≤ **0,185** ≤ 0,247	%EV	= 46,29%
0,000912	0,0302	AV	=	0,000 ≤ **0,156** ≤ 1,117	%AV	= 38,89%
0,00223	0,0473	IA	=	0,131 ≤ **0,243** ≤ 0,412	%IA	= 60,87%
0,00444	0,0666	R&R	=	0,307 ≤ **0,343** ≤ 1,159	%R&R	= 85,79%

For instance mishandling or problems with the clamping device might cause this interaction.

Something similar happens when an operator swaps two parts.

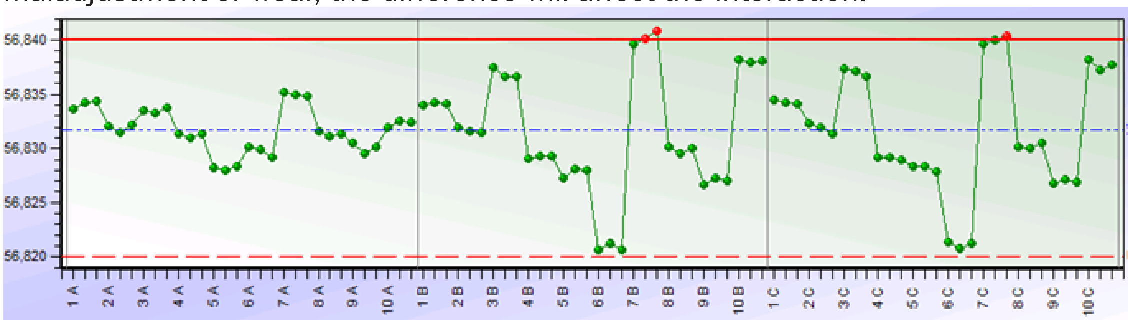


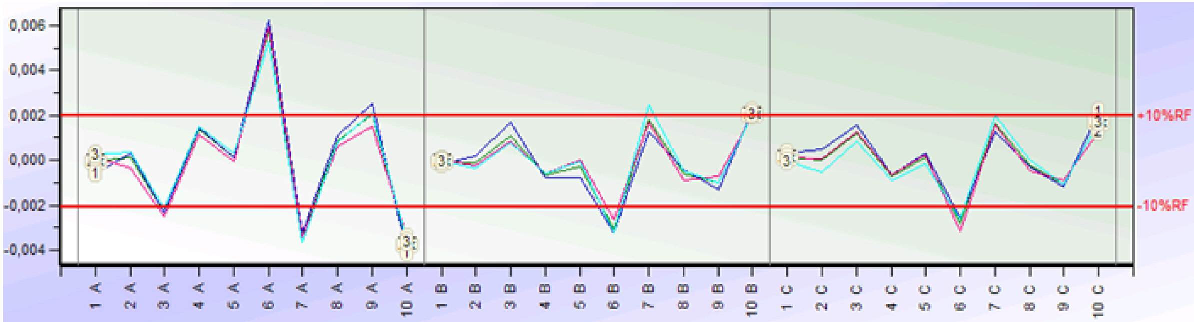


0,00000175	0,00132	EV = 0,00545 ≤ **0,00682** ≤ 0,00911	%EV = 11,36%	
0,000	0,000	AV = 0,000 ≤ **0,000** ≤ 0,0294	%AV = 0,00%	
0,0000615	0,00784	IA = 0,0268 ≤ **0,0404** ≤ 0,0634	%IA = 67,33%	
0,0000633	0,00795	R&R = 0,0390 ≤ **0,0410** ≤ 0,0505	%R&R = 68,29%	

This is only a trivial error within the operation and does not indicate a problem of the measurement system; nevertheless the error should be detected.

Even if an operator (here: A) measures a considerably different part variation because of mishandling, maladjustment or wear, the difference will affect the interaction.





0,000000114	0,000337	EV = 0,00147 ≤ **0,00174** ≤ 0,00212	%EV = 8,69%	
0,000	0,000	AV = 0,000 ≤ **0,000** ≤ 0,000	%AV = 0,00%	
0,00000596	0,00244	IA = 0,00890 ≤ **0,0126** ≤ 0,0192	%IA = 62,89%	
0,00000608	0,00246	R&R = 0,0121 ≤ **0,0127** ≤ 0,0123	%R&R = 63,49%	

It becomes clear that the interaction is not a phenomenon that is easy to describe. The interaction rather indicates that the data shows perturbations not corresponding to repeatability or reproducibility. We require further inspections to find the source of error. It would be wrong to accept this interaction as an "ordinary variation of the measurement system". The interaction rather points out that further perturbations exist and these perturbations have to be inspected prior to the release of the measurement process.

However, we would like to point out again that even the EV and AV must not be attributed to the equipment or the operator without any reason, just because they are referred to as equipment and appraiser variation!

So the method of ANOVA is preferable?

Yes, for sure. ARM may be allowed if someone is in the jungle far away from any civilization and has to make a quick measurement system analysis with the help of a pencil, some paper and a solar calculator. (The author asks pardon for these polemizing statements, but it is just so amusing to him.) However, even in this case, there are ANOVA tables available that quality engineers had used long since the computer was invented. The only acceptable excuse for using ARM is that it is a hard and slow process for a company to set a course for something new, similar to an ocean steamer. In return, it is important to change the course now in order that the company will not suffer the same fate as the Titanic in the foreseeable future.

In order to sum up the approach of the method of ANOVA explained before:

- You calculate the part variation from the means of 10 parts and deduct it from the total variation.
- You calculate reproducibility from the means of the operators and deduct it from the total variation.
- If the variation of at least one part deviates significantly from the mean reproducibility, you should calculate the corresponding interaction and deduct it from the total variation.
- What remains of the total variation refers to repeatability.

The “if ... significantly“ indicates that the calculated interaction is evaluated by means of an F-test. In case the difference is so slight that it might be a random difference, you do not have to report this value separately, but you may combine the random residuals and add it to repeatability.

Another advantage of the method of ANOVA is that it can be expanded in order to consider more components. According to VDA Volume 5, the experiment may include additional influence components, such as

- different measuring equipment
- different places of measurement on the object
- different measurement methods.

You may determine the influence of each component with the help of the method of ANOVA and report them separately. However, each influence component to be inspected increases the number of required measurements. A higher number of measurements affect the method to find a reasonable design of experiments, but that is another story we will tell on another day...

TEQ

TEQ Training & Consulting GmbH
Eisleber Str. 2
69469 Weinheim / Germany
+ 49 6201 3941-15

Interested in this topic?
www.teq.de
Contact the author
morteza.farmani@teq.de