Mess-System-Analysis discrete

Requirement and related topics

The basics of statistics and especially statistical distributions are advantageous for these descriptions. Further topics are:

www.weibull.de/COM/Measurement_System_Analysis.pdf

Introduction

Measuring system analysis, or MSA for short, is the analysis of the capability of a measuring device and the measuring process. A discrete measurement is understood to be a result based on good/bad, 0/1, etc. If possible, measurement using quantitative measured values is preferable. Only if there is no other option should the discrete measurement be used. This is e.g. in most cases a subjective assessment. The next level is an ordinally scaled measurement from 3 levels, e.g. the evaluation according to school grades..

Purpose and usefulness

In a discrete MSA for examinations and processes it should show that the measuring system or the tester is better than a random result.

Basics

These procedures are described below:

- **Gage R&R for discrete attributes**
- **Kappa method (Fleiss & Cohens)**
- **Bowker method**
- **Kendalls concordance (ordinal data)**

Gage R&R for discrete attributes

In the Gage R&R procedure for discrete features, several appraisers each assess different parts twice. This could e.g. checks on parts that are either intact or faulty. The appraisers must not know which part they are facing and the order must be random. Please note the information on the test objects in a later chapter.
If there are deviations within an appraisers or between different appraisers, these are counted.

In the procedure with a reference value, the appraisers are presented with random samples of which they also do not know the classification. At least 30 random samples are recommended, which each appraisers has to evaluate twice. The results are shown in a table:

Mess-System-Analysis discrete

	Reference	Appraiser A		Appraiser B		Appraiser C	
1	passed	passed	passed	passed	passed	passed	passed
2	failed	failed	failed	failed	failed	failed	passed
3	passed	passed	passed	passed	passed	passed	passed
4	passed	passed	passed	passed	passed	passed	passed
5	passed	failed	failed	failed	failed	failed	failed
6	passed	passed	passed	failed	failed	passed	passed
7	passed	passed	passed	passed	passed	passed	passed
8	failed	failed	failed	failed	failed	failed	failed

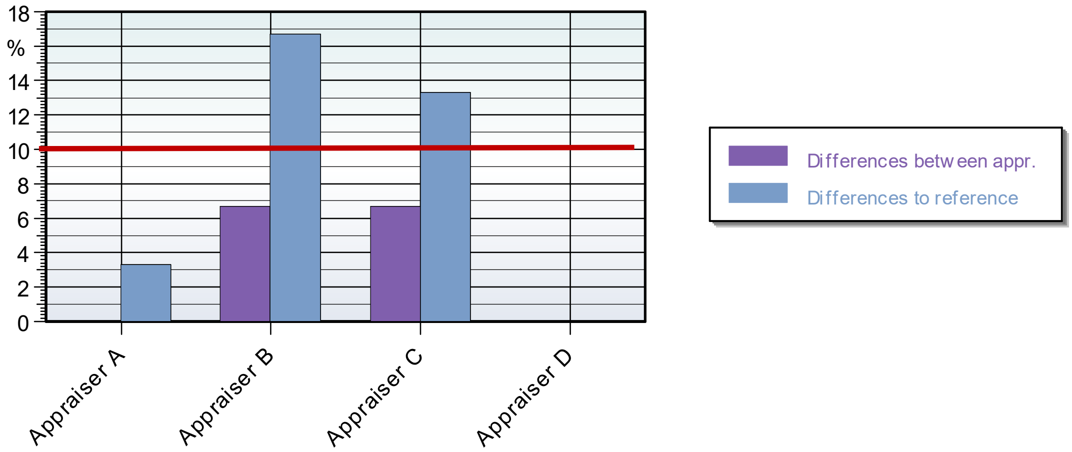
deviations:

within appraisers

to reference

between appraisers

A requirement can be that the ratio of different results to the number of parts should not be greater than 5% or 10%.
In this example, appraiser B and C have more than 10% deviations from the reference



Kappa method

In the so-called Kappa method, not only the deviations are counted, as in the previous illustration. The question arises which proportion would have been possible only by chance. The following parameter, Kappa, takes this into account:

$$\kappa = \frac{p_o - p_c}{1 - p_c}$$

p_o : observed concordance
pc : random possible concordance

Kappa represents the correspondence between different appraisers, or to a reference, minus a random correspondence. The requirements for Kappa are usually graded as follows:

$\kappa < 0{,}7$	⇨	not capable
$0{,}7 \leq \kappa < 0{,}9$	⇨	conditionally capable
$0{,}9 \leq \kappa < 1{,}0$	⇨	capable

You can determine the correspondence of each appraiser with himself, i.e. whether each object / part is always rated the same (repeatability). The appraiser must not know which part is currently being assessed. The next step is to check how the appraisers behave towards each other (reproducibility). The so-called **Fleiss-Kappa** is used for these two questions. The **Cohen's Kappa** is relevant for questions about the compliance of each tester with a reference.
The table to be evaluated is basically structured in the same way as with the Gage R&R discrete method. Here, however, 0 and 1 are used instead of good / bad. You can also do 3 repetitions here.

In the following example, 30 parts were rated by 3 appraisers:

lower confidence level
standard deviation
analog normalv. $z = \kappa/SE$
error probability

	correct	%	Fleiss k	95% CI	SE	z	p-val
Appraiser A	29	96,7	0,955	0,643	0,10541	9,06	0,000
Appraiser B	28	93,3	0,910	0,597	0,10541	8,63	0,000
Appraiser C	29	96,7	0,954	0,642	0,10541	9,05	0,000

At the end of the table, the null hypothesis is tested that the matches differ significantly from a random result (p-value <0.05).
It should be noted here that these statistics cannot be compared with the limits shown by Kappa! Only the actual kappa value provides the decision as to whether the measurement is capable.

A similar calculation is described in the AIAG guideline Measurement System Analysis (4th Edition). In contrast to the variant shown here, only the relationships between 2 appraisers can be used in pairs.

Bowker method

In the so-called Bowker procedure, there are also good / bad assessments. The appraisers can have the same results, different or inconsistent results. At least 40 different test objects are tested 3 times by 2 appraisers. Each of the 40 results is divided into three classes:

class 1 : all 3 repetitions results have passed
class 2 : within all 3 repetitions different results
class 3 : all 3 repetitions results have failed

The result is shown in the form of a crosstab:

		Appraiser B		
		failed	mixed	passed
Appraiser A	failed	5	5	2
	mixed	7	1	10
	passed	0	3	7

For the test, the symmetry is tested using the χ^2 distribution. The null hypothesis is that the frequencies nj, i to be expected in the table are symmetrical to the main diagonal. This null hypothesis is rejected if the following test variable with 3 degrees of freedom is greater than the value from the χ^2 distribution

$$\chi^2 = \sum_{i>j} \frac{(n_{i,j} - n_{j,i})^2}{n_{i,j} - n_{j,i}}$$

Kendalls concordance (ordinal scaling)

With the help of the following Kendall test, an MSA for ordinally scaled data is possible. This could e.g. be a subjective test with a defined ascending evaluation index, such as school grades.

At least 30 parts or test objects are required for this. An ideal distribution of the properties is important (realistic spread). As far as possible, all grades should be given with the same frequency. Here, too, of course, the testers receive random parts in a random order whose properties they do not know.
The evaluation is based on the ranking of the evaluation. The non-parametric test with the help of the so-called Kendall's concordance coefficient W is defined as follows:

$$W = \frac{12 \sum_{i=1}^{n} R_i^2 \; - \; 3\, m^2 n\, (n+1)^2}{m^2 n\, (n^2+1) - m \sum_{j=1}^{m} T_j}$$

mit

$$T_j = \sum_{k=1}^{g_j} (t_{jk}^3 - t_{jk})$$

R_i : rank for the part

t_{jk} : number of part with the same ranking (length of rank ties / bindings)

g_j : number of rank ties of appraiser j

n : number of parts

m : number of appraisers

k : number of rankings

The concordance coefficient W is χ²-distributed with the degrees of freedom $f = n - 1$.

$$\chi^2 = \mathrm{m\,(n-1)\,W}$$

how the p-value can be determined. Here the null hypothesis is defined that there is no match between the appraisers. If the p-value < α (0.05), the concordance is sufficient.

The ranking is formed according to an ascending numerical evaluation. The following example shows a sorted evaluation. Since the rating 3 occurs 3 times, the middle ranking 4 is to be assigned 3 times. If the 3 only existed twice, the ranking numbers would be… 2; 3.5; 3.5; 5; ...
The number of rank ties is here $g_j = 1$ and the rank tie length $t_{jk} = 3$. This results in

No.	assess-ment	ranking
1	1	1
2	2	2
3	3	4
4	3	4
5	3	4
6	4	6

$$T_j = \sum_{k=1}^{g_j} (t_{jk}^3 - t_{jk}) \;=\; \sum_{k=1}^{1} (3_1^3 - 3) \;=\; 24$$

This test allows relatively large deviations from the testers before the p-value reaches 0.05. It makes sense to define an additional requirement for a specific number of matches, e.g. All appraisers must have rated at least 10% of the parts equally.

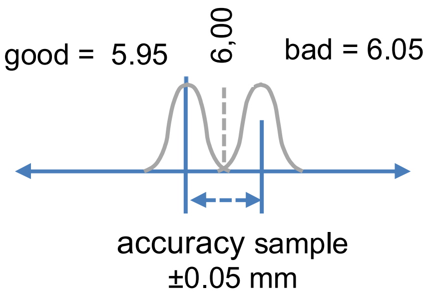
The same calculation can be used between the appraisers, but also against repetitions of each appraiser. A test against a reference assessment is also possible. However, reference values are rarely available for subjective evaluations..

Notes on the test objects

For the mentioned procedures, test objects, samples or parts which are required the question is how their properties should be.

It makes no sence to make sampes or pattern so bad that it will always be recognized as bad under all circumstances. It is the same with good samples. All patterns should be close to the border between good and bad so that the possibility of a wrong assessment can also be "observed".

The physical property is to be selected in such a way that a clear assignment is still possible. Example: The task is to use a self-made gauge to test whether a hole in a sheet metal (sample) has a maximum diameter of 6.0 mm. If the gauge goes into the hole, it is considered too big and therefore not OK. (bad). The question is with what accuracy the drilling of the samples can be produced. Is used as a measuring device e.g. a caliper, the accuracy here is 0.05mm. For a "safe" differentiation, an good bore must have 5.95mm and a bad one must have 6.05m.



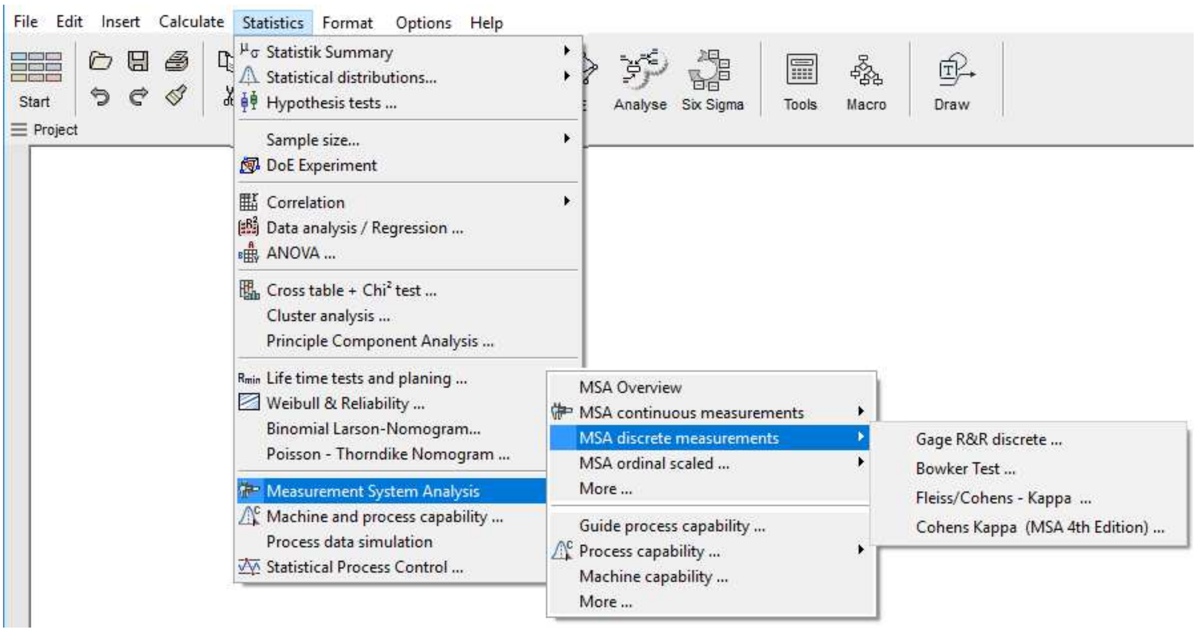
If you now put the bad samples to 6.05, the probability of getting 6.00 or less is 2.5% (95% corresponds to ± 0.05 mm). But there is also a probability of 2.5% to get incorrectly a bad good unintentionally.

Furthermore, the accuracy of the gauge is included. While the errors of the different patterns are distributed over their number, there is only one gauge that always has the same error (systematically error). If now the distribution of the samples is 50% good and 50% bad, the error in the gauge shifts the proportion of observed good and bad patterns one-sided. The accuracy of the gauge should therefore be significantly better.

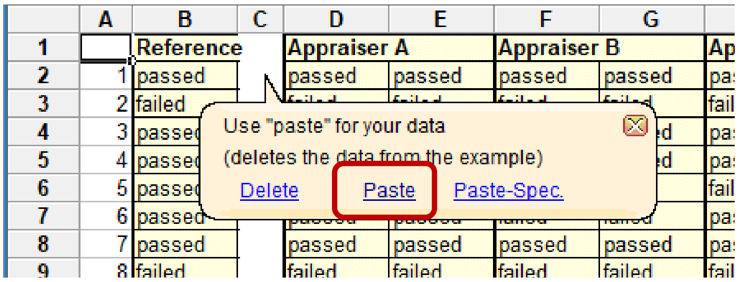
Mess-System-Analysis discrete

Using Visual-XSel 15.0/16.0

www.crgraph.com



After selecting the template you want, use the Paste link to insert data from the clipboard or enter it manually



Possibly further fields (highlighted in yellow) have to be filled in. Then start the macro with F9, or with the icon macro.