Number of Distinct Data Categories and Gage Repeatability and Reproducibility.

A double (but single) requirement.

Andrés Carrión García. (Corresponding author)

Centre for Quality and Change Management. Universidad Politécnica de Valencia, Spain.

Cno. de Vera s/n. 46022, Valencia. Spain

Email: acarrion@eio.upv.es , Phone num. +34 963877490, Fax. : +34 963877499

Angela Grisales del Río.

Centre for Quality and Change Management. Universidad Politécnica de Valencia, Spain.

Cno. de Vera s/n. 46022, Valencia. Spain

Abstract

Measurement Systems Analysis, as part of the Statistical Process Control practices, has been widely used in many industries, especially those related with the automotive sector. To the initial requirements for system acceptance, centered in the values of Repeatability and Reproducibility (integrated in indexes as GRR and GRR%), new elements have been introduced over the years. One of them, the Number of Distinct Data Categories (or ndc) has been treated as an independent requirement. In this paper we show that ndc and GRR aren't independent and that in fact both are linked by an exact relationship. That makes the double requirement redundant, confusing and unnecessary, except maybe to enhance some conceptual questions.

Keywords

Gage Repeatability and Reproducibility, Measurement Systems Analysis, Number of Distinct Categories.

1. Introduction

In the present industrial and manufacturing world, having measurement systems capable of providing reliable and accurate data is essential for satisfying each time more strict regulations and customer demands.

Measurement systems must be not only carefully designed, considering the elements conforming the system (equipment, people, procedures, environment, standards, …) but need also initial validation and follow up control to ensure their adequateness over time to the measuring purposes.

One of the most used methodologies for initial validation of measurement systems and continuous follow up is Measurement System Analysis (MSA) defined for example in AIAG (2010) and widely used in the automobile sector.

This methodology establishes analysis and validation criteria for variables and attributes measurement systems analysis, basing decisions on data analysis from designed experiments that consider the main factors affecting measurement systems.

While initially requirements for system acceptance were centered in the values of Repeatability and Reproducibility (integrated in indexes as GRR and GRR%), new elements have been introduced over the years. One of them, the Number of Distinct Data Categories (or ndc) has been treated as an independent requirement.

In this paper we show that ndc and GRR aren't independent and that in fact both are linked by an exact relationship. That makes the double requirement redundant, confusing and unnecessary, except maybe to enhance some conceptual questions.

2. Measurement Systems Analysis.

For variable measurement systems, two types of variation sources are considered: those affecting the mean of the readings and those affecting their variability. For the first one, characteristics as bias, linearity and stability are defined, and the criteria for acceptance are fixed. In brief, bias must be not significantly different to zero and any lack of stability and linearity must be analyzed and controlled (and if it is the case, corrected).

In what refers to the variability of measurements, the analysis is done by decomposing total variability in its components. Decomposition of data variability usually considers two sources: the parts and the measurement system. Thus:

$$\sigma_T^2 = \sigma_x^2 + \sigma_m^2 \quad (1)$$

Where σ_T^2 is the total variance of measured data; σ_x^2 is the parts variance; and σ_m^2 is the measurement system variance. The standard deviation associated to parts variability, σ_x, is also known as Part Variation (PV). Similarly, σ_T is the Total Variation (TV).

The variability related to a measurement system is usually decomposed in two components, known as Repeatability and Reproducibility, associated generally with the variation due to the equipment and the variations caused by the different operators (appraisers) using this equipment.

The first component, Repeatability, really includes some other variation sources, as the measuring procedure impact in the measurement variance and the appraiser skill. Nevertheless it is identified as Equipment Variation (EV) and this term is used as synonymous of Repeatability. A frequent assumption, that may need to be confirmed in some cases, is that EV does not change from one operator to other. Defined as the variability of data produced by a single operator using a unique gage and measuring repeatedly a single part, its exact computation depends on the method used to analyze the measurement system.

The second component of measurement system variability, Reproducibility, is caused by the presence in the system of different appraisers. A system with only one appraiser will not have Reproducibility issues. Again, not only one factor is included in Reproducibility. Apart of the differences in skill and performance of appraisers, the method used and its clear definition has an important impact on this variability component. In any case, Reproducibility is identified as Appraiser Variation (AV). Computation of AV requires considering the effect of EV over measurements obtained by the different appraisers, and its exact calculation depends again on the method used to analyze the measurement system.

Measurement System variance is the sum of Repeatability and Reproducibility variances:

$$\sigma_m^2 = EV^2 + AV^2$$

Measurement system variability is known as Gage Repeatability and Reproducibility (GRR), and is the standard deviation associated to the measurement system:

$$\sigma_m = \sqrt{EV^2 + AV^2} = GRR$$

Using the MSA terminology, expression (1) can be written as:

$$TV^2 = PV^2 + GRR^2 \qquad (2)$$

According to AIAG (AIAG, 2010), two main methods are available to compute GRR: the Average and Range Method and the ANOVA Method. Each of them has some advantages and disadvantages, but the tendency is to use with preference ANOVA, as it gives more accurate estimates of the parameters involved in measurement system analysis.

Absolute value of GRR, EV and AV are not relevant in most cases, and is the impact of this measurement system variability over the total variability what may make us to select, discard or improve a measurement system. Thus, GRR% is computed as the percentage of the total variability due to the measurement system:

$$GRR\% = 100\ \frac{\sigma_m}{\sigma_T}$$

Also EV% and AV% can be calculated (referred also to the total variability) and their relative values are essential information to address measurement system improvement efforts.

Rules of thumb to accept measurement systems are based in GRR%: Values of GRR% over 30% mean that the system cannot be accepted; Values over 10% and under 30% are associated with measurement systems that, under some circumstances and maybe with additional requirements, can be accepted; Finally, systems are clearly acceptable when GRR% is equal or lower to 10%. (AIAG, 2010).

But MSA is not only GRR, and other requirements must be fulfilled by measurement systems to be acceptable. Apart of requirements related with bias and those related with stability over time and linearity, there is a requirement related with sensitivity and practical resolution of the system.

This last one pays attention to the need of having measurement systems with the ability to distinguish differences in the parts to be measured. The known Rule of Tens, or 10-to-1 Rule, states that instrument discrimination should be at least one tenth of the tolerance (or the process variation). This rule of thumb must be considered as a practical minimum starting point for gage selection. Sensitivity is determined by inherent gage design and quality, in-service maintenance, and operating condition (AIAG, 2010).

A different requirement is related with the effective resolution, and with the number of distinct data categories that can be reliably distinguished as result of the effective resolution of the measurement system. This parameter, the Number of Distinct Categories (ndc), based on Wheeler's discrimination ratio (Wheeler and Lyday, 1989), is defined as:

$$ndc = 1.41 \, \frac{\sigma_x}{\sigma_m} = 1.41 \, \frac{PV}{GRR} \qquad (3)$$

This parameter is also related with signal to noise ratio (SNR) (Burdick, Borror and Montgomery, 2005), (Majeske and Gearhart, 2006).

As rule of thumb for accepting a measurement system, the ndc must be at least 5. For values between 2 and 5 acceptance can be conditioned to the intended use and to additional requirements.

In the automobile industry, GRR% and ndc are recommended to evaluate measurement systems. (Barrentine, 2003; Burdick, Borror and Montgomery, 2005; AIAG, 2010).

Acceptance criteria for ndc and GRR% are summarized in Table 1 (Pedott and Fogliatto, 2012) (AIAG, 2010).

Table 1. Acceptance criteria for GRR% and ndc.

	GRR%	ndc
Good	≤ 10%	≥5
Acceptable	Between 10% and 30%	Between 2 and 5
Not acceptable	> 30%	< 2

3. Relationship between ndc and GRR.

When using MSA and teaching it to quality and metrology people, a frequent question is if ndc can be acceptable while GRR% is not. While this situation has not been seen, as MSA AIAG Manuals and general MSA literature deal with these two parameters as if they were independent, the answer to this question must be yes… but then, why we never find a good ndc with a bad GRR%?

The answer is that ndc and GRR% are not independent. Moreover, there is an exact relation among them.

Using equation (2), it is not difficult to see that:

$$PV^2 = TV^2 - GRR^2 \qquad (4)$$

And substituting (4) in (3)

$$ndc^2 = 2\,\frac{TV^2 - GRR^2}{GRR^2} = 2\left(\frac{10\,000}{GRR\%^2} - 1\right)$$

$$ncd = \sqrt{2}\,\sqrt{\frac{10\,000}{GRR\%^2} - 1}$$

According to this, a table can be prepared to have an idea of how GRR% and ndc are related. Table 2 shows these values.

Table 2. GRR% and ndc values

GRR%	ndc
5	28,2
10	14,0
15	9,3
20	6,9
25	5,5
27	**5,0**
30	4,5
40	3,2
50	2,4

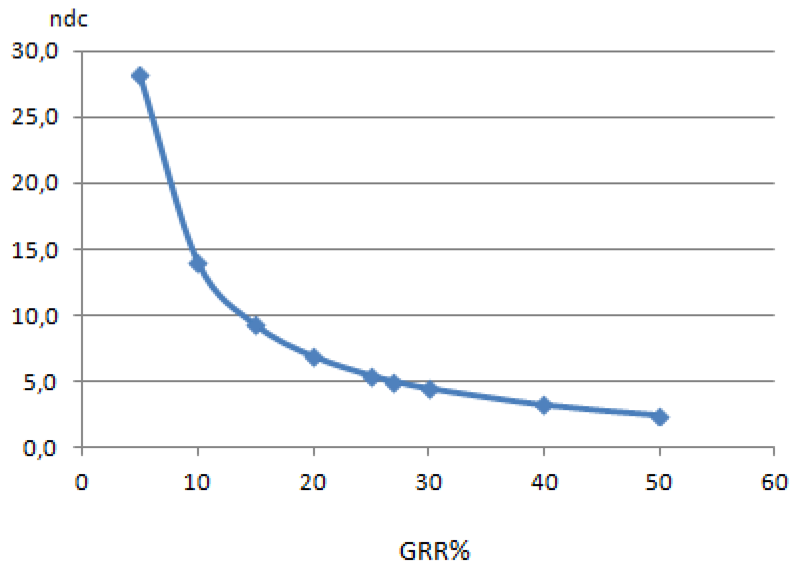

Figure 1. Relationship between ndc and GRR%.

The application of the general acceptance rules to GRR% and ndc is incoherent, as show in Table 3. To simplify comments we can use a Green, Yellow, Red terminology, where Green means clearly acceptable, Yellow means acceptable under conditions and Red means non acceptable. Some cases can illustrate this situation:

- For ndc values between 5 and 14, clearly acceptable (green) from the ndc point of view, values of GRR% are in yellow area.
- When ndc values are in the yellow area, between 2 and 5, values of GRR% are in the red area (over 30, except for a small area with GRR% between 27 and 30).
- Values of ndc must be over 14 to have a good GRR%.
- Only for GRR% values over 58, ndc will enter in the red area.

Table 3. Acceptance criteria for GRR% and ndc

GRR%	**Acceptance criteria (GRR%)**	**Acceptance criteria (ndc)**	**ndc**
5	G	G	28.2
10	G	G	14.0
15	Y	G	9.3
20	Y	G	6.9
25	Y	G	5.5
27	**Y**	**G**	**5.0**
30	Y	Y	4.5
40	R	Y	3.2
50	R	Y	2.4
58	R	R	1.98

The above presented equations are independent of the method used to compute EV, AV and PV. Both for Average and Range method and for ANOVA method, the equations are the same, so the relationship between ndc and GRR% is always valid.

4. Geometrical interpretation.

A simple graphical interpretation of the relationship between the different variances (error components) present in the model enhances the notion that all parameters are related. From expressions (1) and (2), Figure 2 can be drawn.

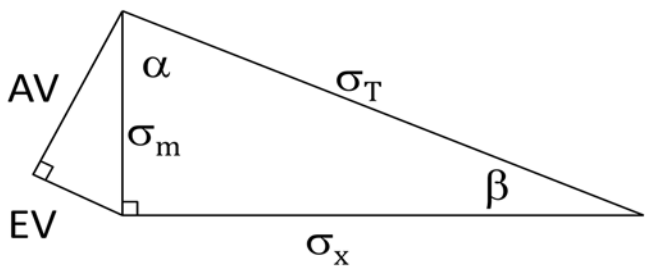

Figure 2. Variability components in MSA.

Expressions (3) and (4) have a geometrical interpretation

$$GRR\% = 100\ \frac{\sigma_m}{\sigma_T} = 100\sqrt{\cos\alpha}$$

$$ndc = 1.41\ \frac{\sigma_x}{\sigma_m} = 1.41\ \frac{PV}{GRR} = 1.41\sqrt{tg\ \propto}$$

Basic trigonometric knowledge proves that between the sinus and the tangent of an angle, exact relation exists:

$$tg\ \propto = \frac{1}{cos^2 \propto} - 1$$

And using the previous expressions to substitute sinus and tangent by their value in function of GRR% and ndc, respectively:

$$\frac{ndc^2}{2} = \frac{1}{\left(\frac{GRR\%^2}{10\ 000}\right)} - 1$$

And then:

$$ndc\ = \sqrt{2}\sqrt{\frac{10\ 000}{GRR\%^2} - 1}$$

5. Some numerical examples.

The relationship is followed in all examples revised in bibliography, as it is theoretically based and deducted. Only for illustration, some of these cases are summarized in Table 4.

Table 4. Correspondence between GRR% and ndc in the bibliography.

Reference*	Method used	GRR% value	ndc value (in the paper)	ndc computed
Plura and Klaput (2012)	A&R	17.85	7.77	7.77
Plura and Klaput (2012)	A&R	28.77	4.69	4.69
Plura and Klaput (2012)	ANOVA	17.09	8.13	8.13
Plura and Klaput (2012)	ANOVA	42.25	3.02	3.02
Hoffa and Laux (2007)**	ANOVA	97.89	0	0.29
Hoffa and Laux (2007)	ANOVA	88.75	1	0.73
Hoffa and Laux (2007)	ANOVA	1.91	74	73.81

* Both papers present more numerical cases. Only a few have been selected to illustrate correspondence between ndc values.

**The authors specifically state that ndc values are rounded to the nearest integer.

6. Discussion.

As we have shown, ndc and GRR% are linked by a mathematical expression. This makes the existence of the two acceptance criteria, as has been defined until now, incoherent, as limit values where defined, apparently, considering that both indicators where independent, producing rules of thumb with strong discrepancies.

The sensitivity and practical resolution of the measurement systems are important aspects of their adequacy for each specific use, but this question is not independent of the GRR. Having two independent rules of acceptance is not a good option. Instead, both aspects must be considered together.

In any case, probably it can be good to re-define acceptance criteria. These new criteria must take into account the relationship presented to keep integrity of the decisions.

It has been considered until now that values of ndc over 5 indicate that data can be adequate to use SPC control charts (AIAG, 2010). But, what to say about these charts, based in measurements systems that may have GRR% values just over 27%? Are the information provided by this charts really useful for control? And, what about the decisions based in such poor measurement system? Or maybe the measurement system is not so bad?

Then, does it have sense to continue paying attention to the value of ndc? Maybe yes, because, even considering that ndc says nothing that GRR% does not say, it gives information related with the usefulness of measurements for control purposes. But a rethinking may be needed in the acceptance criteria for measurement systems.

Is it possible to redefine now the acceptance criteria? According to the previous analysis, some questions are clear:

- The present limits to consider a measurement system as not acceptable are more or less coherent, as 30% of GRR% corresponds to ndc 4.48, near to the 5 of the present rule of thumb. These critical values of 30% and 5 can be maintained with low inconsistency. Maybe is enough to fix the 30% requirement as strictly lower to keep coherency ensured (ndc for GRR%=29 is 4.65≅5). This means to redefine in deep values for ndc, as until now ndc between 2 and 5 were acceptable, and if we keep the limit for GRR% acceptability in 30%, ndc will be non acceptable under 5.

- Limits for considering a measurement system as a good one need redefinition. Probably the frontier between good and acceptable systems should be near GRR% = 15%, where ndc=9.3. Requiring GRR% < 15 and ndc ≥10 can be a good option (for GRR%=14, ndc=9.97).

Table 5 presents a possible set of coherent acceptance criteria.

Table 5. Coherent acceptance criteria for GRR% and ndc.

	GRR%	ndc
Good	< 15%	≥10
Acceptable	Between 15% and 30%	Between 5 and 10
Not acceptable	≥ 30%	< 5

In any case, defining these criteria is out of the purposes of this paper, and Table 5 should be considered only as a tentative alternative, to serve for discussion.

7. Conclusions.

There exists and exact relation between ndc and GRR% measurement system, and this makes necessary to review the criteria used in MSA to accept a measurement system.

It is not correct to work with the present rules of thumb as if these parameters were independent.

This relationship opens a series of questions related with the adequacy of measurement systems, the usability of data coming from these systems and the confidence we can have in the decision making process that is based in this information.

References.

AIAG Reference manual. 2010. Measurement Systems Analysis. MSA. 4th Edition., AIAG, Southfield.

Barrentine, L. B. Concepts for R&R studies. 2nd ed. Milwaukee: ASQ, 2003.

Burdick, R. K., Borror, C. M. and Montgomery, D. C. 2003. A Review of Methods for Measurement Systems Capability Analysis. Journal of Quality Technology, v. 35, p. 342- 354.

Hoffa, D.W. and Laux, C.. 2007. Gauge R&R: An Effective Methodology for Determining the Adequacy of a New Measurement System for Micron-level Metrology. Journal of Industrial Technology. Vol. 23, N. 4.

Majeske, K. D. and Gearhart, C. 2006. Approving Measurement Systems when Using Derived Values. Quality Engineering, 18, p. 523-532.

Plura, J. and Klaput. P. 2012. Influence of the interaction between parts and appraisers on the results of Repeatability and Reproducibility analysis. Quality Innovation Prosperity XVI - 1, p. 25-36.

Pedott, Alexandre Homsi and Fogliatto, Flávio Sanson. 2012. Estudos de repetitividade e reprodutividade para dados funcionais. Produçao. ahead of print Epub 09-Nov-2012.

Wheeler, D. J., & Lyday, R. W. (1989). Evaluating the measurement process. Knoxville, TN. SPC Press.