Choice of parameters for an R&R study and its impact on the measurement capability

Christophe Dubois[1], *Peggy* Courtois[1]

[1]Deltamu, 48 rue de Sarliève, 63800 Cournon d'Auvergne, France

Abstract. When you perform a traditional R&R (Repeatability and Reproducibility), the first practical question that comes to mind is "How many times "n" do we repeat the measurements, and how many operators "p" should we consider?". As the quality of the standard deviations obtained during this study depends on the parameters n and p, a large number of measurements is required. This is obviously not possible due to environment, time and cost aspects. The question is then, how to optimise the parameters "n" and "p", for an R&R study? To answer this question, statistics are useful tools, which, combined with your practical skills, will help you to understand the outcome of the study and rationalise your choice.

1   What is an R&R study?

An R&R study is a simple experimental design allowing an experimental study with two sources of variations of a measurement system: the Repeatability and the Reproducibility, hence its name.
The definition of Repeatability is usually well understood. It is the observed variation over several successive measurements, with fixed parameters. The VIM [1] gives the following definition:

> *Condition of measurement, out of a set of conditions that includes the same measurement procedure, same operators, same measuring system, same operating conditions and same location, and replicate measurements on the same or similar objects over a short period of time*

This definition provides information on the condition of measurement, but misses its purpose. The repeatability conveys two aspects. First, it gives the minimum dispersion of the studied process. Independently of the chosen parameters (operators, etc.), the measurement process will always have at least this dispersion. It is similar to the minimum background noise of the measurement. Second, it shows the dispersion of parameters, which impacts the measurement and what we do not control.

The definition of Reproducibility is a little bit more problematic. For the VIM, all parameters must vary:

> *Condition of measurement, out of a set of conditions that includes different locations, operators, measuring systems, and replicate measurements on the same or similar objects*

This definition is used in ISO 5725 [5] to describe the dispersion of the global result. However, it contrasts with other documents such as the MSA [2], where the

reproducibility consists in varying one parameter (e.g. the operator).

In the following sections, for the sake of clarity, I only use repeatability, intermediate measurement precision, and total dispersion.

2   Why do we do an R&R study?

An R&R study helps to understand our measurement process. In a dispersion study, as the repeatability is our minimal dispersion, we can identify a source of variation (or noise) by comparing our results to our repeatability. Without our repeatability, we cannot conclude if this dispersion is pure noise or a proper source of variation.
The intermediate measurement precision helps to characterise the impact of one specific parameter. It usually represents the inter-operator dispersion in the mechanical industry, but it can also be an inter-equipment, or an inter-time dispersion. Therefore, the choice of the parameter for the intermediate measurement precision depends on what we want to observe. In general, we choose the most impacting parameter on the dispersion of the measurement process.

With both repeatability and intermediate measurement precision of the most impacting parameter, we obtain an experimental estimation of the measurement uncertainty. This estimation may be coarse, but remains easy to determine; we consider the other sources of variation negligible. By comparing this "uncertainty" U to the tolerance we want to measure, we determine the process capability C. We have just described the conventional approach to estimate the capability of a measurement process:

---

*   Corresponding author: cdubois@deltamu.com ; pcourtois@deltamu.com

$$C = \frac{TolanceInterval}{2U} \qquad (1)$$

The deficiency of an R&R is that it is restraint to study one and only one parameter (addition to the repeatability). The fact that it is relatively simple to implement is, however, a real benefit of the method. If we want to study several parameters, we can use more advanced experimental designs, such as the ANOVA (ANalysis Of Variance). The last but not the least, an R&R study gives information on the dispersions only, not the trueness of the measurement.

3   Balanced R&R study

An R&R study is a balanced experimental design where the measurements are structured in a table (see Table 1).

Table 1. Balanced R&R

Operators Trials	A	B	C	D
1	11,04	10,98	11,20	10,23
2	10,84	11,14	11,45	9,95
3	11,06	11,16	11,21	10,56

Note : In the following sections, « n » is the number of trials, while « p » is the number of operators.

We can ask ourselves one question. If we keep the same amount of measurements (n x p = 12), why not asking the operator A to do 9 measurements (to estimate the repeatability), and ask the three other operators to do only one measure each (to estimate the intermediate measurement precision)? The answer is simple. This approach leads to a loss of information. The obtained repeatability would be the operator A's repeatability, which cannot be compared to the others' repeatability (as we have only one measure for the operators B, C, and D). We also have little knowledge regarding the dispersion of the other operators B, C, and D (again, no repeatability to set the basis).
By balancing the experimental design, we check the homogeneity of the dispersion for each operator. We also have a better knowledge of their measurement without losing too much information on the final repeatability.

4   Quality of the estimators

Experimental design only provides information on the repeatability and the intermediate measurement precision. The quality of these estimators depends on the large number of measurements. How many measurements are required and how to divide them up in an experimental design?

4.1   Degree of freedom

The degree of freedom is a statistical tool used to tackle this problematic. It can define the quality of an estimator based on the number of independent samples (i.e. data). The degree of freedom for a mean of n independent samples is n.
The degree of freedom for a standard deviation of n independent samples is $n-1$. Indeed, we look for the behaviour of the differences from the mean $\left(x_i - \bar{x} \right)$:

$$s = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}} \qquad (2)$$

Each sample is compared to the mean, which it is part of. The n differences are thus not completely independent; if one x_i changes, the mean is impacted and thus all differences as well. We say that we lose a degree of freedom.
The determination of the degree of liberty for the different parameters in an R&R study is described in the following section.

4.1.1   Degree of freedom for the mean of standard deviations

To determine the repeatability of an R&R study, we average the mean of the variances of each operator:

$$s_r^2 = \frac{\sum_{i=1}^{p} s_i^2}{p} \qquad (3)$$

with s_i, being the standard deviation of n trials, and p, the number of operators.

The degree of freedom is obtained by multiplying the degree of freedom of the standard deviation with that one of the mean, such as we have *(n-1)*p*.

4.1.2   Degree of freedom for the standard deviation of means

In an R&R study, we calculate the dispersion of the operators with the standard deviation of the means:

$$s_{\bar{x}} = \sqrt{\frac{\sum_{i=1}^{p} (\bar{x}_i - \bar{\bar{x}})^2}{p-1}} \qquad (4)$$

with $\bar{x}_i$, the mean of the operator i, $\bar{\bar{x}}$, the mean of $\bar{x}_i$, and p, the number of operators.
The degree of freedom is *p-1*. The information coming from the means is not taken into account. In fact, we only determine the quality of the standard deviation of 3 values. Intuitively, however, we would like to know the standard deviation for the inter-operator, which is not the same. The latter is obtained such as:

$$s_{op}^2 = s_x^2 - \frac{s_r^2}{n} \qquad (5)$$

with s_{op}, the standard deviation of the inter-operator, and n, the number of trials by operator.

In that case, the impact of the mean occurs with the repeatability, and the number of trials for the mean.

The estimation of the degree of freedom is no longer straightforward (see following section). Nevertheless, as we have only p operators, the degree of freedom will not exceed $p-1$.

4.1.3　Degree of freedom for a sum of variances

In an R&R study, we carry out sums of variances, such as:

$$s_R^2 = s_{op}^2 + s_r^2 \qquad (6)$$

The estimation of the degree of freedom for s_R is not simple. It is not only dependent on the degree of freedom of s_{op} and s_r, but also on their values. For example, let us have a measurement process for which the dispersion results from the inter-operator impact. An R&R study is realized with $p = 3$ operators and $n = 100$ measurements by operator. In that case, the repeatability is well defined, but the knowledge of s_R is limited by the number of operators, its degree of freedom will be close to 2 ($p-1$).

An approximation of the degree of liberty can be obtained with Welch-Satterthwaite's equation [3], with independent standard deviations:

$$\frac{s_{total}^4}{df_{total}} = \sum_i \frac{s_i^4}{df_i} \qquad (7)$$

with dl_i, the degree of freedom of the standard deviation.

4.1.4　Degree of freedom in an R&R study

In a classical R&R study, we can easily determine the degrees of freedom for the repeatability and standard deviations of means (close to the inter-operator standard deviation, but more complex to determine).

For a total number of measurements, we can estimate the impact of the choice of n and p.

Table 2 : Degree of freedom for 30 measurements

Number of measurement	p	n	Degree of freedom Repeatability	Degree of freedom Standard deviation $\overline{x_i}$
30	15	2	15	14
30	10	3	20	9
28	7	4	21	6
30	6	5	24	5
30	5	6	25	4
28	4	7	24	3
32	4	8	28	3
30	3	10	27	2

We can see that the repeatability is far less impacted than the standard deviation of the means. Thus, in an R&R study, for a total number of measurements (often restraint for a matter of cost), it is better to increase p than n.

Of course, we need to take into account business and technical constraints. For example, it is not always convenient to monopolize 9 operators for an R&R study. It is also very important to evaluate the ratio of each standard deviations. If we know that the operator has no impact, we would rather focus on the knowledge of the repeatability and thus n.

4.2　Evaluation of the estimator

The degree of freedom provides a quality scale for the estimator, but how to obtain a proper range, which is easier to understand physically?

4.2.1　Distribution of the estimator

By definition, an estimator gives a representative value of the true value we are looking for. By repeating an experience, we obtain several estimators, probably slightly different from one another. The goal is to determine the variations of the estimators around the true value. This information defines the quality of the estimator. In fact, we look for the distribution of the estimators.

4.2.2　Khi-square ($\check{\chi}$) distribution

Statistics provide interesting information. The experimental variance coming from n samples, has a distribution proportional to a χ^2 distribution with $n-1$ degree of freedom. The distribution depends of course on the quality of the estimator, and thus on the degree of freedom studied in the previous section.
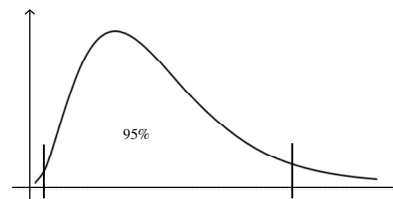
Figure 1 : Distribution and confidence interval

As we know the properties of the χ^2 distribution, we can calculate the confidence interval of the variance.

$$\left[\sqrt{\frac{df * s^2}{F_{KhiSquare(\text{dl})}^{-1}(1 - \frac{\alpha}{2})}} ; \sqrt{\frac{df * s^2}{F_{KhiSquare(\text{dl})}^{-1}(\frac{\alpha}{2})}} \right] \qquad (8)$$

with dl, the degree of freedom of the standard deviation, α, the risk, and $F_{KhiDeux(\text{dl})}^{-1}(1 - \frac{\alpha}{2})$, being the inverse χ^2 distribution.

This equation can be used for the repeatability in the case of a balanced R&R study.

4.2.3   Complex cases

When the standard deviation comes from several variances, the study is more complex. The literature [4] gives approximate values. ISO 5725-1 [5] provides a table giving the relative error of dispersions as a function of n and p.

5   MSA R&R study

The MSA recommends to take for an R&R study 10 manufactured parts (q), each measured 3 times (n) by 3 different operators (p), to finally have 90 measurements.

Table 3 : MSA R&R study

Appraiser /Trial #		PART									
		1	2	3	4	5	6	7	8	9	10
A	1	0.29	-0.56	1.34	0.47	-0.80	0.02	0.59	-0.31	2.26	-1.36
	2	0.41	-0.68	1.17	0.50	-0.92	-0.11	0.75	-0.20	1.99	-1.25
	3	0.64	-0.58	1.27	0.64	-0.84	-0.21	0.66	-0.17	2.01	-1.31
	Average										
	Range										
B	1	0.08	-0.47	1.19	0.01	-0.56	-0.20	0.47	-0.63	1.80	-1.68
	2	0.25	-1.22	0.94	1.03	-1.20	0.22	0.55	0.08	2.12	-1.62
	3	0.07	-0.68	1.34	0.20	-1.28	0.06	0.83	-0.34	2.19	-1.50
	Average										
	Range										
C	1	0.04	-1.38	0.88	0.14	-1.46	-0.29	0.02	-0.46	1.77	-1.49
	2	-0.11	-1.13	1.09	0.20	-1.07	-0.67	0.01	-0.56	1.45	-1.77
	3	-0.15	-0.96	0.67	0.11	-1.45	-0.49	0.21	-0.49	1.87	-2.16
	Average										
	Range										

The MSA R&R study is a special case. It introduces an additional parameter of intermediate fidelity, the dispersion of production of the parts. However, only the repeatability and the inter-operator dispersions are studied, as in a classic R&R study.

$$GRR^2 = EV^2 + AV^2 \qquad (9)$$

with

   EV :   standard deviation of repeatability
   AV :   standard deviation of intermediate fidelity
   GRR : intermediate fidelity of measurement process

The additional factor introduces the measurand effect (shape defect for example) within the repeatability. This is why the MSA is suggesting taking 10 parts in order to have a representative sample of possible imperfection.

If we know that there is no impact of the parts on the measurement, it becomes possible to adjust the number of parts required.

Degrees of freedom of these elements:

$$df(EV) = (n-1)*p*q = 60 \qquad (10)$$
$$df(s_{\bar{x}}) = p-1 = 2 \qquad (11)$$

with

$s_{\bar{x}}$: standard deviation of the operator's average, close to the value of AV

The obtained EV is relatively reliable but that the quality of the AV is relatively poor.

It is a pragmatic choice made for the world of automotive production measurement, mainly dimensional. The measurement uncertainty in this case is mainly due to the defect of shape of the produced parts, the repeatability. The inter-operator effect is lower, but above all, it can be difficult to monopolize several production operators in order to perform a capability measurement test.

A compromise must be made in the R & R parameter choice with the following elements:

- Industrial constraints (cost, time of measurements)
- Main dispersion source (based on expertised)
- Degree of freedom (equation (10) and (11))

Example 1:
There is no form defect known on the parts, so the number of part can be reduce to 5.
The dispersion of the measurement is mainly due to the inter-operator effect and 5 operators can be monopolized.
The optimize experiment study :

   q = 5 number of parts
   p = 5 number of operators
   n= 3 number of trial
   Total = 75 measures
   df (repeatability)= 50
   df($s_{\bar{x}}$)=4

In this case, the repeatability remains of good quality, the inter-operator effect is better defined and the total number of measurements is reduced.

Example 2 :
The process is mainly impacted by repeatability, little operator effect. The parts have no form defect and it is a small series production where parts are expensive.
The optimize experiment study:

   q = 3 number of parts
   p = 3 number of operators
   n= 10 number of trial
   Total = 90 measures
   df (repeatability)= 81
   df($s_{\bar{x}}$)= 2

In this case, with the same number of measures as a classical MSA, the quality of repeatability is improved.

6   Conclusion

To choose the parameters of an experimental study and to improve the relevance of the results, we must not only rely on the statistical properties such as the degree of freedom, but we must also understand the physical impact on the results and have expertise in the field of the expected result to guide the decision.

References

1. JCGM 100 International vocabulary of metrology – Basic and general concepts and associated terms (VIM) 3rd edition

2. AIAG (2010). Measurement Systems Analysis, MSA (4th ed.). Automotive Industry Action Group

3. WELCH, B. L. (1936), *J. R. Stat. Soc. Suppl.* **3**, 29-48; (1938), *Biometrika* **29**, 350-362; (1947), *ibid.* **34**, 28-35
   - FAIRFIELD-SMITH, H. (1936), *J. Counc. Sci. Indust. Res. (Australia)* **9**(3), 211
   - SATTERTHWAITE, F. E. (1941), *Psychometrika* **6**, 309-316; (1946) *Biometrics Bull.* **2**(6), 110-114

4. Modelisation et estimation des erreurs de mesure, Michèle Neuilly, CETAMA ed. Technique & documention.

5. ISO 5725 Accuracy (trueness and precision) of measurement methods and results