Measurement 86 (2016) 125–132

Contents lists available at ScienceDirect

Measurement

journal homepage: www.elsevier.com/locate/measurement

Repeatability and Reproducibility techniques for the analysis of measurement systems

Andrea Zanobini, Bianca Sereni, Marcantonio Catelani, Lorenzo Ciani *

Department of Information Engineering, University of Florence, Via Santa Marta, 3, 50139 Florence, Italy

ARTICLE INFO

Article history:
Received 12 March 2015
Received in revised form 4 February 2016
Accepted 19 February 2016
Available online 4 March 2016

Keywords:
ANOVA
Gage R&R
Repeatability
Reproducibility
Measurement uncertainty

ABSTRACT

This research was conducted with the aim of analyzing two of the main metrological characteristics of any measurement system: Repeatability and Reproducibility. Both of these features play an important role in the analysis of the measurements and they can give us a lot of information about who and what influences any measuring system. The analysis of Repeatability and Reproducibility is generally carried out through the use of the study *Gage R&R*. This study permits to understand which are the decisive factors in a measurement system, and, definitively, if the process is stable, that is under statistical control or out of statistical control.

1. Introduction

The statistical control of processes consists in a set of techniques of analysis concerning the quality of products and services, and in this particular case, of the measures. To define the concept of quality is not simple, in one of its definitions it is inversely proportional to the variability: in fact a decrease of quality corresponds to an increase in variability [1].

The SPC or *Statistical Process Control* is a process of analysis of variability, or rather to its reduction; and it uses some methods or techniques such as the *Gage R&R* to achieve that. The *Gage R&R* is a study on the variability observed in a measurement and due to the measurement system itself. *R&R* denotes Repeatability and Reproducibility, that are two characteristics of each measurement system.

In particular, Repeatability is the variation caused by the instrumentation or the variation observed when the same operator measures the same part more times with the same instrumentation.

Reproducibility is the variation caused by the measurement system or the variation observed when different operators measure the same part with the same instrumentation.

A small variability of a series of measurements is a good indicator of repeatability, meantime the reproducibility is colligated to the stability of a measurement process. The ANOVA or ANalysis Of VAriance and the DOE or Design Of Experiment are two very powerful methods to conduct a Gage R&R study [2]. The *Gage R&R* studies determine how much of variability of processes is due to the variation of the measurement system and they uses inferences technique to estimate Repeatability and Reproducibility [3,4]. In this paper the authors analyze two particular cases.

In the first, the standard "CEI ENV 13005: *Guide to the expression of uncertainty of measurement*", [5] is analyzed and in particular the Appendix *H5* referring to the "Analysis of Variance" and we propose an example concerning the calibration of a Zener sample of voltage diode showing the values of the average voltage and the standard deviation in 10 days of observation. A one-way analysis of variance

* Corresponding author.
E-mail address: lorenzo.ciani@unifi.it (L. Ciani).

http://dx.doi.org/10.1016/j.measurement.2016.02.041
0263-2241/

tests the hypothesis that the means of several populations are equal. We recall that this method is an extension of the two-sample t-test, specifically for the case where the population variances are assumed to be equal. One-way ANOVA will tell us if there are statistically significant differences among the level means. The null hypothesis for the test is that all population means (level means) are the same. The alternative hypothesis is that one or more population means differ from the others. In addition with the help of Minitab software we will determine which level means are different when (and if) differences exist.

In the second case study the data [6–8] obtained by the rise time and fall time measurement of a waveform generated by two different waveform generators have been analyzed. The Gage R&R Study is performed as a way to identify any deficiencies in the measurement system. In a Type 1 Gage Study, measurements of one part measured by one operator are analyzed to estimate the level of variation in the gage itself, the gage repeatability, and the accuracy of the measurements.

By analyzing the measurements of one reference part by one operator, we will determine whether a measuring device is capable of measuring a particular characteristic under conditions with relatively small variation.

2. ANOVA technique: some theoretical recalls

As known, ANOVA consists in a series of techniques originating by the theory of inferential statistics, that can be applied in order to evaluate and compare the variability of data. Starting from two or more populations of data, the ANOVA methods are used to estimate the differences between the sample means of these populations by analyzing the respective variances. By the evaluation of two or more different distributions, ANOVA allows to determine if such differences are random or not. ANOVA, in particular, is a technique of parametric statistical inference based on an hypothesis test [9–12]. If we consider an experiment where effects of factors are taken into account with different levels (or Treatments) for each factor, the result of such experiment is a variable response. The influence of each factor as source of variability can be analyzed by considering "a" levels and collecting "n" random observations (replicated responses for each level). The observed data can be represented as shown in Table 1.

Consequently a set of data, that can be represented by a $a \times N$ matrix, is obtained for each factor. The variability observed among these data is the result of the effort of two sources of variability: variability in the replicate responses of each level, denoted as ***within*** variability and variability among levels, called ***between*** variability. On the basis of this structure different ANOVA technique in function of the number of factors involved in the experiment can be taken into account. In particular, ***One-Way*** ANOVA is denoted if only one factor is considered in the experiment. Similarly, ***Two-Way*** ANOVA if two factors are involved and so on. As said above only one factor is considered with different $i = 1 \ldots a$ treatments. For each level n random observations are collected. The analytical model that describes the sources of variability in i-th random observation from j-th treatment in (Table 1), is given by:

$$y_{ij} = \mu + \tau_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{1}$$

In Eq. (1), μ is the general mean representing the common effect, τ_i denotes the effect due to the treatment i (treatment effect) and ε_{ij} the random error in the j-th observation due to the treatment i. Assuming $\mu_i = \mu + \tau_i$, the linear model in (1) becomes:

$$y_{ij} = \mu_i + \varepsilon_{ij} \begin{cases} i = 1, 2, \cdots, a \\ j = 1, 2, \cdots, n \end{cases} \tag{2}$$

where μ_i is the mean of the i-th treatment.

As we have just said, the experiment must to be completely randomized, or rather the observations are extracted in a completely random way. So the analysis of variance consists to perform the following hypothesis test:

$$\begin{aligned} &H_0: \ \mu_1 = \mu_2 = \cdots = \mu_a = 0 \\ &H_1: \ \mu_i \neq 0 \end{aligned}$$

where H_0 is the null hypothesis and H_1 is the alternative hypothesis.

The analysis of variance checks if the means of the a populations are equal.

It is possible to consider τ_i as the first order deviation by the general mean μ, so:

$$\sum_{i=1}^{a} \tau_i = 0 = \sum_{i=1}^{a} (\mu_i - \mu) \tag{3}$$

Consequently trying the equality between the means is trying the equality of levels effects. So the hypothesis test can be written as:

$$\begin{aligned} &H_0: \ \tau_1 = \tau_2 = \cdots = \tau_a = 0 \\ &H_1: \ \tau_i \neq 0 \end{aligned}$$

Moreover the following relationship are valid:

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \tag{4}$$

$$\overline{y}_i = \frac{y_{i.}}{n} \tag{5}$$

$$y_. = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \tag{6}$$

$$\overline{y} = \frac{y_.}{N} \tag{7}$$

Table 1
Table of data detection (one factor, a levels, n observations).

Level	Observations				Total	Expected values
1	y_{11}	y_{12}	...	y_{1n}	$y_{1.}$	$\overline{y}_{1.}$
2	y_{21}	y_{22}	...	y_{2n}	$y_{2.}$	$\overline{y}_{2.}$
...	...	...	...	...	...	...
a	y_{a1}	y_{a2}	...	y_{an}	$y_{a.}$	$\overline{y}_{a.}$