ANOVA Techniques for Reliability Analysis

Marcantonio Catelani, Lorenzo Ciani, Bianca Sereni and Andrea Zanobini

Department of Information Engineering, University of Florence,
via S. Marta 3, 50139, Florence, Italy
marcantonio.catelani@unifi.it, lorenzo.ciani@unifi.it, andrea.zanobini@unifi.it

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 reliability of the measurement systems 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 is very useful because it permits us to understand which are the decisive factors to state the reliability of a measurement system.

I. Introduction

The statistical control of processes [1-2] consists in a set of techniques of analysis concerning the quality of products and services, and in this particular case, of the measures [3-5]. 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.

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 very powerful methods to conduct a study *Gage R&R*. The *Gage R&R* studies determine how much of variability of processes is due to the variation of the measurement system and they uses technique as ANOVA to estimate Repeatability and Reproducibility.

II. ANOVA technique: some theoretical recall

The *ANOVA* consists in a series of techniques originating by the inferential statistics theory, that are interested in the comparison of data variability. These methods are used when there are two or more populations to estimate the differences between their sample means; analyzing the respective variances to achieve its purpose. By the evaluation of two or more different distributions, *ANOVA* allows to determine if the differences are random or not.

ANOVA is based on a technique that compares the sample means to estimate the data variation. It obtains this aim decomposing the variability in *between* and *within*.

ANOVA is a process of statistical inference, in particular it is a technique of parametric statistical inference based on an hypothesis test.

If we suppose to have only a factor characterized by a levels or treatments and n observations for each level we must consider that the answer to each of the a levels is a random variable. The observed data can be represented in the following table:

Table I. Table of data detection

LEVELS	*OBSERVED VALUES*				*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}	$y_{a.}$	$\overline{y}_{a.}$

In general the observations can be described through a statistical linear model:

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

Considering that $\mu_i = \mu + \tau_i$, this model becomes:

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

Each level can be considered belonging to a population with mean μ_i and variance σ^2 .
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:

$$H_0 : \mu_1 = \mu_2 = \cdots = \mu_a = 0$$
$$H_1 : \mu_i \neq 0$$

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:

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

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

$$H_0 : \tau_1 = \tau_2 = \cdots = \tau_a = 0$$
$$H_1 : \tau_i \neq 0$$

Moreover the following relationship are valid:

$$y_{i.} = \sum_{j=1}^{n} y_{ij} \qquad \overline{y_i} = \frac{y_{i.}}{n}$$

$$y. = \sum_{i=1}^{a} \sum_{j=1}^{n} y_{ij} \qquad \overline{y} = \frac{y.}{N}$$

with: $N=an$ and $i=1,2,\ldots,n$

Regarding the total variability of the samples, we can decompose in a variability due to the treatments and in a variability relative to the errors. The variability can be described by the sums of the squares of the deviations of the average values, then we can write [1]:

$$\sum_{i=1}^{a} \sum_{j=1}^{n} \left(y_{ij} - \overline{y}\right)^2 = n \sum_{i=1}^{a} \left(\overline{y}_i - \overline{y}\right)^2 + \sum_{i=1}^{a} \sum_{j=1}^{n} \left(y_{ij} - \overline{y}_i\right)^2$$

We can therefore write:

$$SS_T = SS_{Treatments} + SS_E$$

Where:

$$SS_T = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}\right)^2$$

$$SS_{Treatments} = n\sum_{i=1}^{a}\left(\bar{y}_i - \bar{y}\right)^2$$

$$SS_E = \sum_{i=1}^{a}\sum_{j=1}^{n}\left(y_{ij} - \bar{y}_i\right)^2$$

Considering the expected values we can write the following equality:

$$E\left(SS_{Treatments}\right) = (a-1)\sigma^2 + n\sum_{i=1}^{a}\tau_i^2$$

If the null hypothesis is true, $\tau_i = 0$, so:

$$E\left(\frac{SS_{Treatments}}{a-1}\right) = \sigma^2$$

While it is true the alternative hypothesis:

$$E\left(\frac{SS_{Treatments}}{a-1}\right) = \sigma^2 + \frac{\sum_{i=1}^{a}\tau_i^2}{a-1}$$

Then we can define:

$$MS_{Treatments} = \frac{SS_{Treatments}}{a-1}$$

And:

$$MS_E = \frac{SS_E}{a(n-1)}$$

In the final analysis, if we assume, as said, that each of the *a* populations is described by a normal distribution, if the null hypothesis is true, then the ratio:

$$F_0 = \frac{SS_{Treatments}/(a-1)}{SS_E/[a(n-1)]} = \frac{MS_{Treatments}}{MS_E}$$

has a Fisher distribution with *(a-1)* and *a(n-1)* degrees of freedom.
In conclusion, the answer to the hypothesis testing will be based on the *p-value*: observing the *p-value* relative to the value of *F*, obtained from statistical tables or software, and comparing it with the level of significance α, we can decide whether to accept or reject the null hypothesis. In particular, we accept the null hypothesis if the *p-value* is greater than the significance level, while we accept the alternative hypothesis if the *p-value* is less than α. Refusing, or rather not accept the null hypothesis means that we can conclude that among the observed data there may be significant differences.
Intuitively it can be stated that, if the null hypothesis is true, so the data differ for the effect of random factors, and the changing the levels of the factor does not change the response, but if the null hypothesis is false means that the variability total phenomenon can be attributed to systematic factors. So the method allows to value if the difference between the variances of two distributions, namely the total variability, is due to chance or is significant.

III. A practical example

Trying to make a practical application of the ANOVA we firstly analyze the standard "CEI ENV 13005: *Guide to the expression of uncertainty of measurement*", and in particular the Appendix *H5* where it says about the "Analysis of Variance" and in which it is proposed an example of this technique.
The example, that is proposed here, concerns the calibration of a sample of voltage Zener diode: shows the values of the average voltage and the standard deviation in the 10 days of observation.

Table II. Expected voltage and mean standard error

Day	*Expected Voltage [V]*	*Mean standard Error [μV]*
1	10.000172	60
2	10.000116	77
3	10.000013	111
4	10.000144	101
5	10.000106	67
6	10.000031	93
7	10.000060	80
8	10.000125	73
9	10.000163	88
10	10.000041	86

Assuming, therefore, that the data were normally distributed with mean and standard deviation known and provided by the Standard, we have obtained, with Minitab, the 5 daily values of tension that correspond to 50 observed values
It was implemented with *MINITAB* a study *ANOVA One-Way* to a confidence interval of 95% and 97.5%.

```
One-way ANOVA: osservazioni versus giorno

Source  DF         SS         MS     F      P
giorno   9  0,0000003  0,0000000  2,88  0,010
Error   40  0,0000004  0,0000000
Total   49  0,0000007

S = 0,0001054   R-Sq = 39,35%   R-Sq(adj) = 25,71%


                             Individual 97,5% CIs For Mean Based on
                             Pooled StDev
Level  N     Mean   StDev    --------+---------+---------+---------+-
 1     5  10,0002  0,0001      (*)
 2     5  10,0001  0,0001      (*
 3     5  10,0000  0,0002     *)
 4     5  10,0002  0,0001      (*)
 5     5  10,0001  0,0001      (*
 6     5  10,0000  0,0001     (*
 7     5  10,0000  0,0001     (*
 8     5  10,0001  0,0001     (*)
 9     5  10,0001  0,0001      *)
10     5   9,9999  0,0001    (*)
                             --------+---------+---------+---------+-
                                 10,0008   10,0020   10,0032   10,0044

Pooled StDev = 0,0001
```

Figure 1. ANOVA 95% confidence level

```
One-way ANOVA: osservazioni versus giorno

Source  DF         SS         MS     F      P
giorno   9  0,0000003  0,0000000  2,88  0,010
Error   40  0,0000004  0,0000000
Total   49  0,0000007

S = 0,0001054   R-Sq = 39,35%   R-Sq(adj) = 25,71%


                             Individual 95% CIs For Mean Based on Pooled StDev
Level  N     Mean   StDev    --------+---------+---------+---------+-
 1     5  10,0002  0,0001      (*)
 2     5  10,0001  0,0001      (*
 3     5  10,0000  0,0002     *)
 4     5  10,0002  0,0001      (*)
 5     5  10,0001  0,0001      (*
 6     5  10,0000  0,0001     (*
 7     5  10,0000  0,0001     (*
 8     5  10,0001  0,0001     (*)
 9     5  10,0001  0,0001      *)
10     5   9,9999  0,0001    (*)
                             --------+---------+---------+---------+-
                                 10,0008   10,0020   10,0032   10,0044

Pooled StDev = 0,0001
```

Figure 2. ANOVA 97.5% confidence level

The result is *p-value* = 0.010 in both tables. This value should be compared with the levels of significance α which are 0.05 to 0.025. Since, in both cases, the value found is lower than α, we can reasonably reject the null hypothesis (ie the means are all equal) and assert that there are significant differences between the observations found in several days.

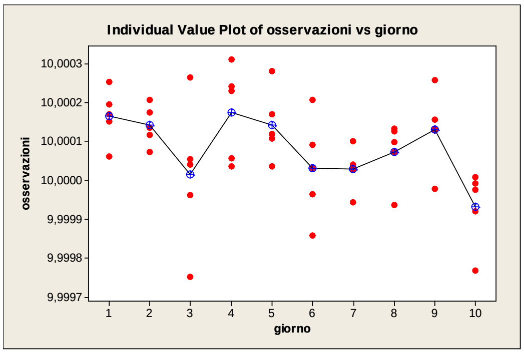

Figure 3. Individual Value Plot 95% confidence level

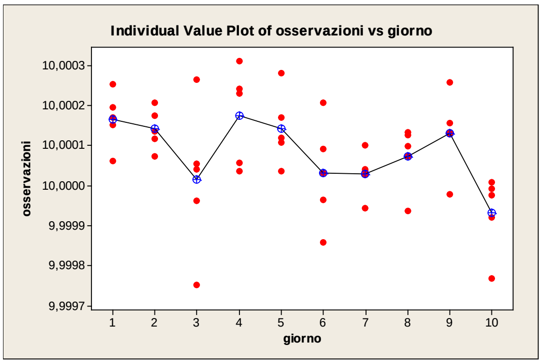

Figure 4. Individual Value Plot 97.5% confidence level

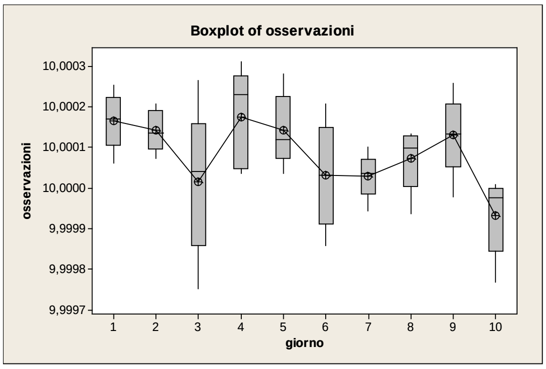

Figure 5. Boxplot 95% level

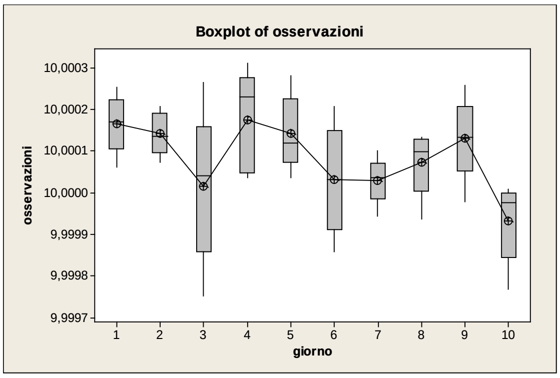

Figure 6. Boxplot 97.5% level

In the *Value Plot of observations vs day* (Fig. 3, 4) and *Boxplot of observations* (Fig. 5, 6) we can to note that the daily averages have a non-linear, but random progress, and this indicates that there are differences among the various days.
In conclusion it may be stated that, as can be observed from both the ANOVA tables and the graphs, there are no differences between the study ANOVA led to a confidence level of 95% and the other to a level of 97.5%.
The result found that the variability between different days results statistically significant compared to the daily variability and also the variability of the observations made on different days have a statistically significant effect on the process itself.

IV. Conclusions

Analysis of *Repeatability* and *Reproducibility* is currently carried out through the use of Gage *R&R* study, the main topic of this research. Its importance lies in the fact that this technique allows us to understand what are the deciding factors to assess the reliability of a measurement system, and especially if the process is stable, that is under statistical control or out of statistical control.
It was still the analysis with the Minitab software that allowed us to experience how good repeatability was an indicator of a significant variability between measurements performed by the same operator, while a poor reproducibility indicated a significant variability for the measures carried out by different operators.

References

[1] Douglas C. Montgomery, "*Statistical Quality Control*", McGrawHill, 2006.
[2] Navidi William, "*Statistics for Engineers and Scientists*", McGrawHill, 2011.
[3] M.Catelani, L.Ciani, A.Zanobini, "Uncertainty Interval Evaluation Using The Chi-Square And Fisher Distributions In The Measurement Process", Metrology and Measurement Systems, Vol. XVII (2010), No. 2, pp. 195-204.
[4] E. Balestrieri, M. Catelani, L. Ciani, S. Rapuano, A. Zanobini, "Word Error Rate measurement uncertainty estimation in Digitizing Waveform Recorders", Measurement 46 (2013), pp. 572-581, DOI 10.1016/j.measurement.2012.08.016.
[5] E. Balestrieri, M. Catelani, L. Ciani, S. Rapuano, A. Zanobini, "The Student's t distribution to measure the word error rate in analog-to-digital converters", Measurement, Volume 45, Issue 2, February 2012, Pages 148-154, ISSN 0263-2241, DOI 10.1016/j.measurement.2011.03.002.