Anomaly detection in the industrial processes

Federico Trotta

Description of the project

THE PROJECT:

This project is my graduation thesis. I graduated in Mechanical Engineering with a thesis "Analysis of production data with Data Science techniques" and it is an analysis of real production data. The result was a method for anomaly detection in industrial processes.

THE IDEA:

An industrial product is typically the result of different stages of production. In assembly lines, in general, we try to make sure that the time to conclude a production phase (cycle time) is the same for all the production phases of that product. In SMEs, however, it often happens that an operator carries out different assembly phases on different days. The question I asked myself is: can we understand which operator is the most suitable to carry out a production phase, by analyzing the history of its cycle times?

Analysis description

I chose a product that is made in three production phases: two assembly phases and one functional test phase.

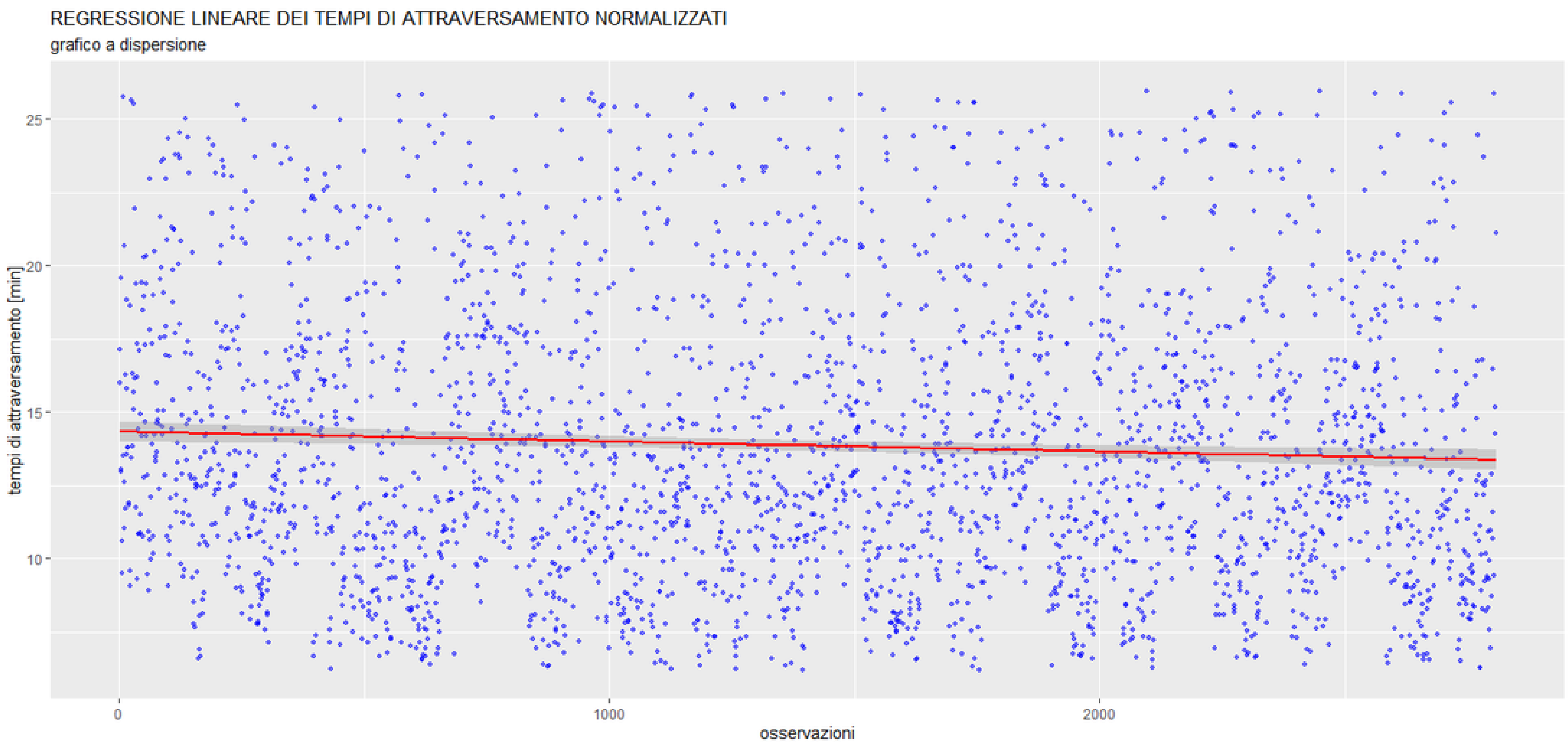
In the months before my study, some work had been done to minimize the cycle time of the testing phase; therefore, the first check to be made was that on the total time to make the product (process lead time).

After that, I took into consideration the first two production phases which can be variable according to the operator; so I studied the registered cycle times, for various operators.

For each operator, I have analyzed, in the beginning, the cycle times with linear regression; then with the polynomial one, then with the splines.

Process lead time analysis

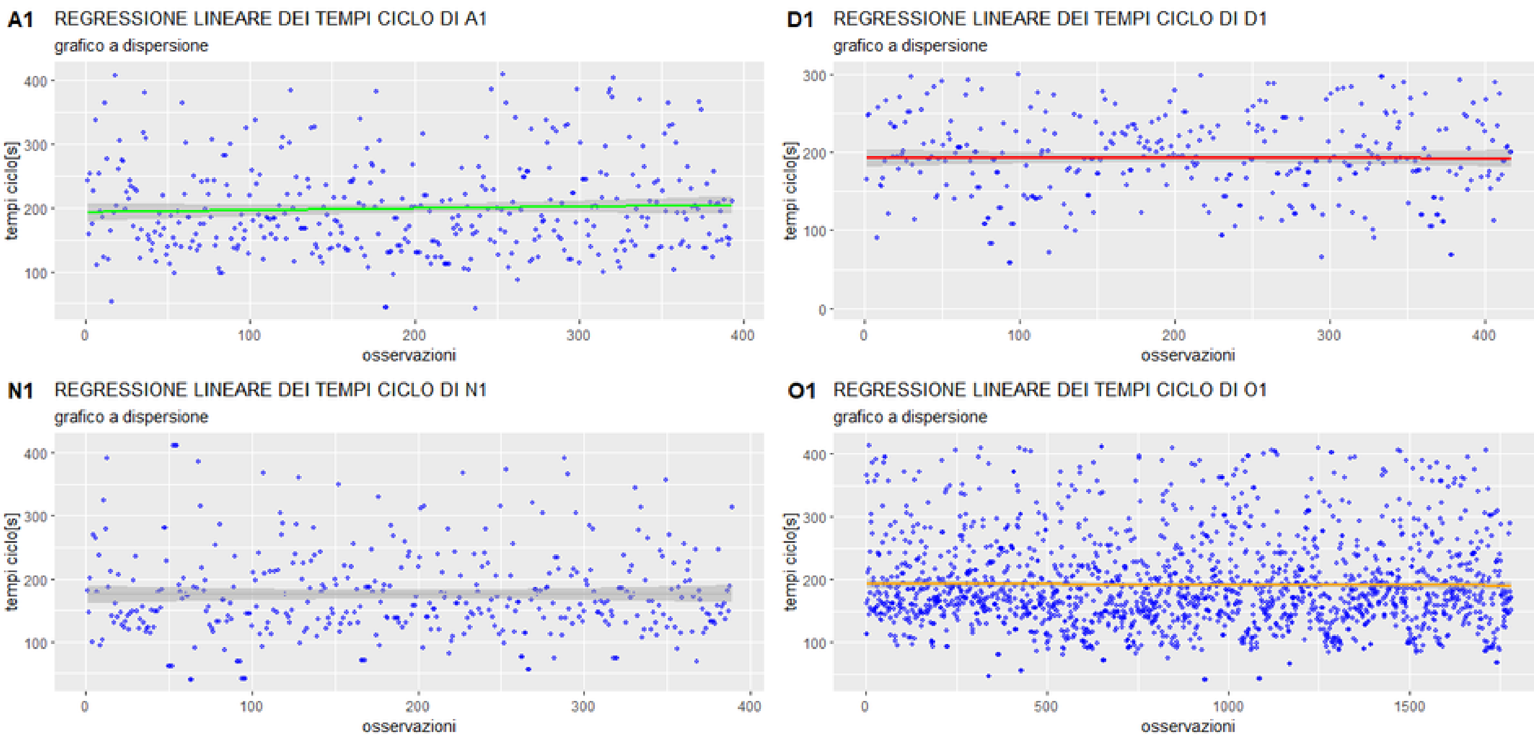
As expected, from the analysis of the linear regression of the process lead times (total times, recorded in months, to manufacture the product) I obtained a linear regression line with a negative coefficient; this means that, over the months, the process lead time has decreased (because the cycle time of the testing phase has decreased as months went by)



First production phase analysis: linear regression

I have identified the operators who, over time, produce the greatest number of observations (they are those who, more than the others, work in this production phase) and they are 4.

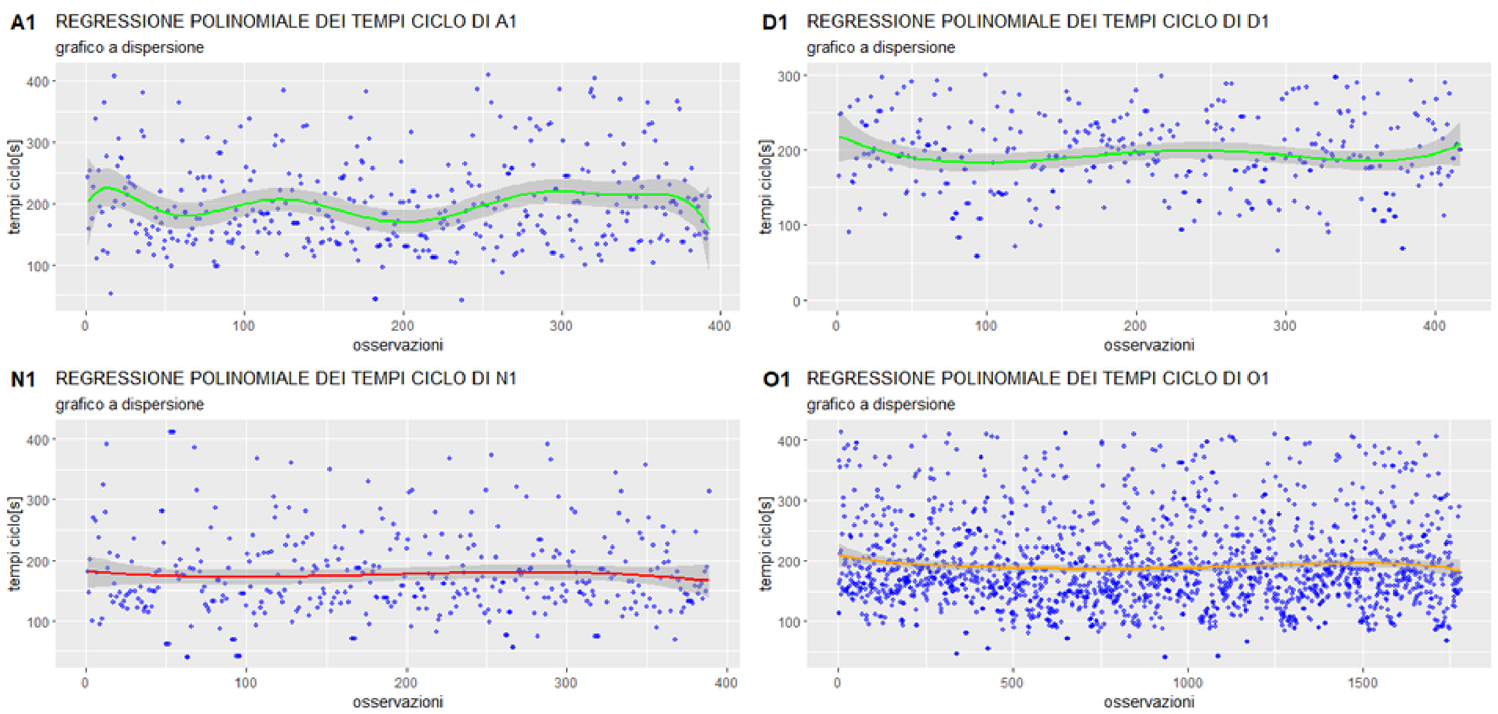
I calculated the linear regressions of their cycle times. All four have values of R^2 close to zero, and the straight lines are, in fact, all horizontal. It seems, therefore, that they all remain constant in their work (the regression is horizontal), which is what one would like in industry; but this analysis is not enough. So I tried polynymial regression.



First production phase analysis: polynomial regression

I tried different degrees of polynomial, up to the optimal one, also using R ^ 2 as a parameter.

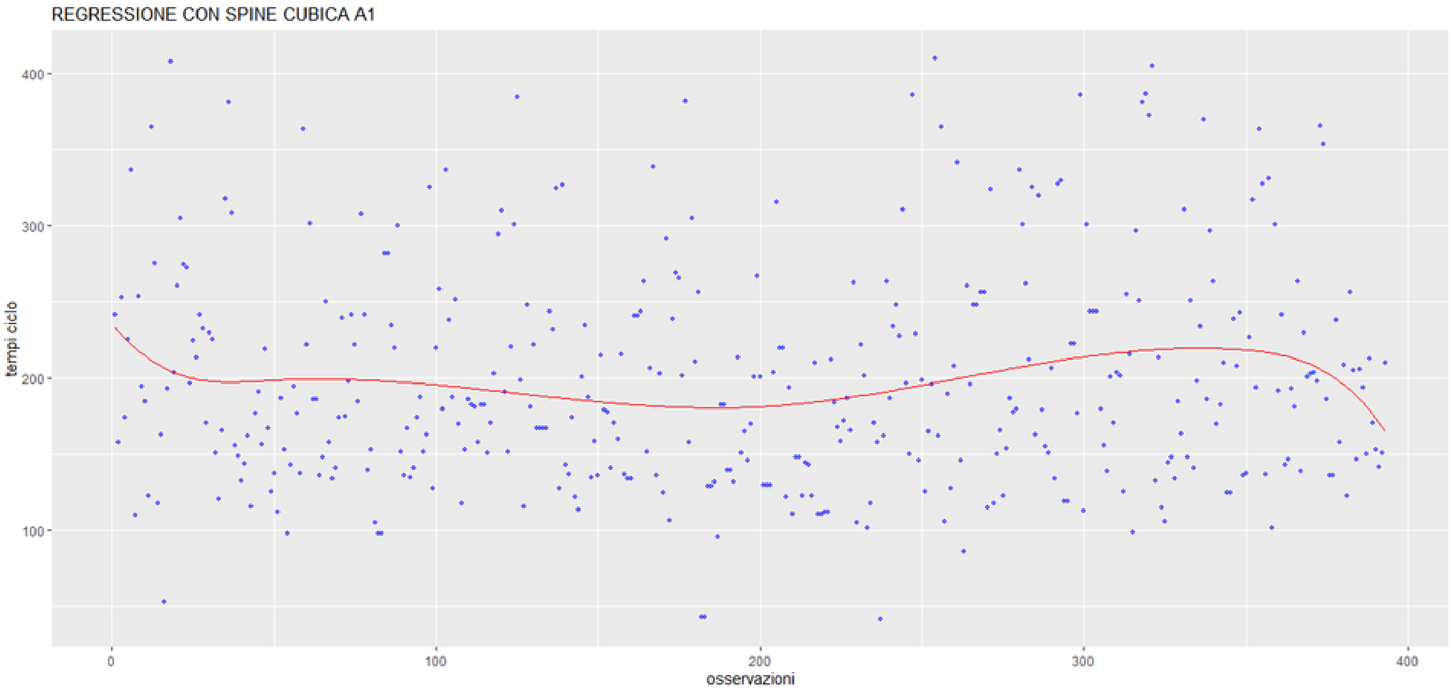
As we can see, for the operators N and O the polynomial practically becomes a straight line; D has some oscillations; A clearly has fluctuations in the cycle time records. To be on the safe side, I used the splines which are mathematically "more manageable" at will.



First production phase analysis: splicne regression

By applying a cubic spline to the times of operator A we now have greater certainty that its cycle time has some oscillations, which means that this operator has not remained constant over time in carrying out this production phase.

By studying, instead, the splines of the other operators we practically obtain straight lines, so we can consider them as constants in carrying out their processing phase.



Conclusions and possible applications

The second phase of assembly has been studied in the same way, so it has not been reported here.

This study is nothing more than Anomaly Detection in industrial processes because it allows us to identify those operators who, in carrying out a certain production phase, do not remain constant at the level of cycle time.

In fact, a horizontal regression line means that there is no correlation between the temporal trend of the observations and the cycle time recorded; and if polynomials and splines also tend to be straight, it means that that operator is really (quite, within a certain tolerance) constant in carrying out that production phase, which is exactly what one wants on an industrial level.

This type of analysis, therefore, together with a qualitative analysis of the result, allows to identify the fastest and most constant operators; being able to compare the results obtained in terms of time with the qualitative ones (the best operator is the fastest and most constant, and at the same time generates a product without quality defects), it is possible to identify the most "suitable" operators to carry out a certain production phase, adding this information, for example, in a skill matrix (the skill matrix tells me the qualities of the operator, but does not measure them)