Study the parameters of a petrol engine using analysis of variance

C O Ilie¹, M Marinescu², L Barothi³, R Vilau⁴, I Lespezeanu⁵
¹,²,³,⁴,⁵ Military Automotive Engineering and Transportation Department, Military Technical Academy, 39-49 George Coșbuc Blvd, Bucharest, Romania
E-mail: ovidiuilie66@yahoo.com

Abstract. This paper presents a method to analyse the influence of various parameters on spark ignition engine operation. The mathematical tool it was used is analysis of variance. Based on statistical data, analysis of variance is a method of study the influence of factors acting simultaneously on any process. From a statistical point of view, it consists of comparing the average values of different collectives and checking of the statistical hypothesis on their homogeneity. Based on its results, the structure of the mathematical models obtained by regressions and predictions is established. Data were obtained by measuring the dynamic parameters of vehicles and engines. We have used 14th automobiles of the same type with different wages and mileages. Testing program aimed to capture a large range of operating regimes. Vehicles moved on different surfaces in different weather and traffic conditions. Some of the engine parameters like torque and power were measured on the test bench.

1. Introduction

This paper presents the analysis carried out in a phase of a vehicle dynamics research program based on statistical models. The first step was to design and apply a testing program to obtain a variety of data defining dynamic behavior of a gasoline engine vehicle. There were used 15 cars having 1.6 l engines and mileages ranging between 13500 to 115000 km. Most of the experimental data were acquired and stored in the tester intended to diagnose the cars used. There were capture with a frequency of 10 Hz. Same parameters like fuel consumption, power and engine’s torque could not be measured using the onboard system. These were calculated using static engine characteristics obtained by measuring on the test bench. Eventually, we selected 64 tests, each of them having 15 strings with 256 values. Each string represents a measured or calculated size lasts approximately 26 s. The data were saved in a tensor. The tensor was used to issue mathematical models that describe vehicle’s dynamics.[1]

2. Analysis of variance

Making statistical models, in this case regressions and predictions, requires a prior analysis of the data. [2] In other words, should be compared the samples of each parameters to see which one has a stronger influence to the studied process. The models will only be set based on the insightful series. An instrument for this is alongside the correlation and frequency analysis, ANalysis Of VAriance (ANOVA). The method consists in comparing the samples based on their means. The means of two or more groups are checked if they are significantly different from each other. Depending on the number of factors whose influence is studied there are one way, two-way ore multi-variate analyses of
variance. All of them can be performed either on samples of the same volume or using different volume samples.

3. **One-Way ANOVA (Analysis of Variance)**

The One-Way ANOVA compares the means of two or more series of the same parameter in order to determine whether there is statistical evidence that the associated population means are significantly different. Also, it is known as One-Factor ANOVA.

Samples of the same $n$ volume as are the experimental data were used to perform One-Factor analysis of variance. In this respect, it is considered a number of $k$ sub-collectivities that hypothetically have the same mean $m$ of the general characteristic $X$. One random sample of the same volume $n$ is drawn from each sub-collectivity, in order to find out if their averages differ significantly from $m$.

The initial data forms a matrix $x_{ij}$ with $n$ lines and $k$ columns, so $i=1, n$ and $j=1, m$. Random $k$ sub-collectivities reflect the influence of a certain factor acting on the studied process considered by the $X$ characteristic.

The following are defined:

- means of sub-collectivities:
  \[
  \bar{x}_j = \frac{\sum_{i=1}^{n} x_{ij}}{n}
  \]  
  \(1\)

- great mean (the mean of the means):
  \[
  \bar{X} = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij}}{nk}
  \]  
  \(2\)

The distance of each value of the characteristics from the great mean is divided in two components:

\[
 x_{ij} - \bar{X} = (x_{ij} - \bar{x}_j) + (\bar{x}_j - \bar{X})
\]  
\(3\)

Summating the squares of expression (3) results:

\[
\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{X})^2 = \\
= \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2 + 2 \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{X}) + \sum_{i=1}^{n} \sum_{j=1}^{k} (\bar{x}_j - \bar{X})^2
\]  
\(4\)

Expression (4) becomes:

\[
\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{X})^2 = \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2 + n \sum_{j=1}^{k} (\bar{x}_j - \bar{X})^2
\]  
\(5\)
because
\[ 2 \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x})(\bar{x}_j - \bar{x}) = \]
\[ = 2 \sum_{j=1}^{k} \left[ \sum_{i=1}^{n} (x_{ij} - \bar{x}_j)(\bar{x}_j - \bar{x}) \right] = 2 \sum_{j=1}^{k} \left( \bar{x}_j - \bar{x} \right) \sum_{i=1}^{n} (x_{ij} - \bar{x}_j) = 0 \]  \hspace{1cm} (6)

Expression (5) shows that the total sum of squares noted $SPA_t$ of the $X$ characteristic from the $k$ samples (sub-collectivities) of $n$ volume was decomposed in two main components:

- sum of squares within noted $SPA_r$:

\[ SPA_r = \sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2 \]  \hspace{1cm} (7)

- sum of squares between noted $SPA_f$:

\[ SPA_f = n \sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2 \]  \hspace{1cm} (8)

Thus expresion (5) becomes:

\[ SPA_t = SPA_f + SPA_r \]  \hspace{1cm} (9)

Mathematical expresions of variances of the three sums, $SPA_r$, $SPA_f$, $SPA_f$, are obtained dividing them by number of degrees of freedom involved.[4] Result:

- factorial variance $D_f$

\[ D_f = \frac{n \sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2}{k - 1} \]  \hspace{1cm} (10)

- residual variance $D_r$

\[ D_r = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2}{k(n-1)} \]  \hspace{1cm} (11)
- total variance $D_t$

$$D_t = \frac{\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x})^2}{nk - 1} \quad \text{(12)}$$

Analysis of variance consists of comparing the two variances, factorial and residual variances, based on the null hypothesis $H_0$. The null hypothesis in ANOVA is valid when all the sample means are equal, or they don’t have any significant difference. The alternate hypothesis $H_1$ is valid when at least one of the sample means is different from the rest of the sample means.[3]

If the two variances do not differ significantly (with a level of significance $\alpha$), than the Fisher's computed coefficient noted $F_c$, which is the ratio between them is close to the unit value.

If the $k$ samples are extracted from collectivities with different means than the factorial variance $D_f$ will be considerrabily higher than ten residual one $D_r$ and the calculated Fisher's coefficient $F_c$ will be greater than the table Fisher's coefficient $F_t$.

Table 1 presents all the calculus to be performed for One-Way Anova.

| Type of variance | Sum of squares SPA | Number of degrees of freedom GL | Variance $D_f$ | Fisher's coefficient (computed) $F_c$ | Fisher's coefficient (table) $F_t$ |
|------------------|--------------------|-------------------------------|---------------|-------------------------------------|----------------------------------|
| Factorial        | $n \sum_{j=1}^{k} (\bar{x}_j - \bar{x})^2$ | $v_1 = k - 1$ | $D_f$ | $F_c = \frac{D_f}{D_r}$ | $F_{\alpha,v_1,v_2}$ |
| Residual         | $\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x}_j)^2$ | $v_2 = k(n - 1)$ | $D_r$ | | |
| Total            | $\sum_{i=1}^{n} \sum_{j=1}^{k} (x_{ij} - \bar{x})^2$ | $\nu = nk - 1$ | $D_t$ | | |

The One-Way Anova has the following steps:
1. $F_c$ is computed according with the formula given in Table 1;
2. It is adopted the level of significance $\alpha$ which usually is 0,05;
3. $F_t$ is established on the basis of the tables with critical values of the distribution $F$ depending on the level of significance $\alpha$ adopted and the values of the related degrees of freedom;
4. By comparing the two values of Fisher's coefficient, it follows:
   a. If $F_c < F_t$, the null hypothesis $H_0$ with the significance level $\alpha$ is accepted, which means that the grouping factor considered did not have a large influence on the means of the samples;
   b. If $F_c >= F_t$, the null hypothesis $H_0$ with the significance level $\alpha$ is rejected, the alternative hypothesis $H_1$ is accepted, and thus the grouping factor considered has had a significant influence on the means of the samples.[1]
4. Study the parameters of petrol engine using One-Way ANOVA

ANOVA has been used in the presented research program to established the data strings that will be used to calculate mathematical models. Input (factorial) and output (end result) sizes for each model were determined. The series corresponding to the input for the 64 tests were compared using One-Way ANOVA. Based on the analysis, it was established for each model the series to be used.

An example is shown below. It was set the throttle $\xi$ as an input and series corresponding test $I_{1n}$ and $I_{2n}$ to be compared. The results are presented in table 2 and figure 1.

| Type of variance | Sum of squares $SPA$ | Number of degrees of freedom $GL$ | Variance $D$ | Fisher's coefficient (computed) $F_c$ | Fisher's coefficient (table) $F_t$ |
|------------------|----------------------|-----------------------------------|--------------|--------------------------------------|-------------------------------|
| Factorial        | 50474.6              | 1                                 | 50474.6      | 148.446                              | 0                             |
| Residual         | 173410.6             | 510                               | 340          |                                      |                               |
| Total            | 223885.1             | 511                               |              |                                      |                               |

Table 2. One-Way ANOVA for throttle $\xi$ and tests $I_{1n}$ and $I_{2n}$

![Figure 1 - Quartiles of throttles $\xi$ for tests $I_{1n}$ and $I_{2n}$](image)

Figure 1 - Quartiles of throttles $\xi$ for tests $I_{1n}$ and $I_{2n}$

Table 1 shows that $F_c > F_t$ and therefore the two means differ significantly as also seen in the figure 1. It results that the throttle $\xi$ for tests $I_{1n}$ and $I_{2n}$ considerably influences the process so both should be used to develop simple regressions models.

In the case of the same factor, throttle $\xi$ but for tests $I_{1n}$ and $I_{3n}$ is obtained values presented in table 3 and figure 2.

| Type of variance | Sum of squares $SPA$ | Number of degrees of freedom $GL$ | Variance $D$ | Fisher's coefficient (computed) $F_c$ | Fisher's coefficient (table) $F_t$ |
|------------------|----------------------|-----------------------------------|--------------|--------------------------------------|-------------------------------|
| Factorial        | 23.8                 | 1                                 | 23.8         | 0.085                                | 0.771                         |
| Residual         | 143281.2             | 510                               | 280.9        |                                      |                               |
| Total            | 143305.1             | 511                               |              |                                      |                               |
This time, $F_c < F_t$, that means the null hypothesis $H_0$ with the significance level $\alpha$ is accepted. The two samples have close means as can be seen on the figure 2. Taking into account the above mentioned exemple (where samples of tests I1n and I2n were considered), the final conclusion is that the sample of throttle $\xi$ for test I3n does not significantly influence the modeling process of the vehicle dynamics, so it can be eliminated when determining the mathematical model (by simple regression).

Similar analysis can be conducted for all tests of a specific factor and for all parameters.

5. Conclusions

The process of modeling vehicle dynamics on a statistical basis requires prior data preparation.[5] A useful tool with good effectiveness is One-Way ANOVA. It is a way to find out if a string of values or experiment results are significant for the process of modeling. In the case of vehicle dynamics it helps us to make the best data base for each type of model.

Also, it has some limitations. It requires a large amount of calculations. One-Way ANOVA tells us that at least two groups are different from each other. But it won’t tell us which groups are different.[3] Both, One-Way ANOVA and analysis of correlation used together provides useful information and leads to the construction of data bases necessary for the construction of high precision models.

References

[1] C.O. Ilie, Statistical Modeling of Vehicle Dynamics, Military Technical Academy, Bucharest, 2008

[2] B Laszlo, D Voicu 2016 “Developing a measuring system able to determine the mechanical stress on a rectangular cross – section bar”, MTA Review, Vol. XXVI, No. 1, pp. 47-52

[3] https://www.analyticsvidhya.com/blog/2018/01/anova-analysis-of-variance/

[4] L Z Rumsiski, Prelucrarea matematica a datelor experimentale, Editura Tehnică, București, 1974

[5] F Popescu, F Enache 2013 Training of RBF Neural Networks: a comparative overview, “Mircea cel Batran” Naval Academy Publishing House, Scientific Bulletin, Constanța, Romania