

Article

An Approach to the Analysis of Causes of Delays in Industrial Construction Projects through Planning and Statistical Computing

Carlos Araújo-Rey *  and Miguel A. Sebastián 

Department of Construction and Manufacturing Engineering, Universidad Nacional de Educación a Distancia (UNED), C/ Juan del Rosal, 12, 28040 Madrid, Spain; msebastian@ind.uned.es

* Correspondence: caraju6@alumno.uned.es; Tel.: +34-91-398-6458

Abstract: The analysis of the planning activities of industrial construction projects can help to evaluate some of the causes that have an impact on the variation of execution times and can also contribute to identifying those activities and components that are most likely to experience or cause delays. Data analysis is facilitated by the use of techniques based on statistical programs, allowing delays to be unequivocally linked to the different elements that make up these projects. In a theoretical study, a simulation is carried out with data that are hypothetical but consistent with real projects, which are transformed and standardized before being uploaded to the statistical software. Using the statistical software's graphical interface, the data set is analyzed from a descriptive point of view, unraveling the relationships between variables and factors by means of contingency tables and scatter plots. Using other techniques such as the comparison of variables and correlation studies, as well as linear regression and variance analysis, the characteristics are evaluated and the differences in project delays are investigated in order to determine, after the fact, which components have the highest rates of delay in execution times.

Keywords: construction delays; industrial projects; planning; statistical computing



Citation: Araújo-Rey, C.; Sebastián, M.A. An Approach to the Analysis of Causes of Delays in Industrial Construction Projects through Planning and Statistical Computing. *Sustainability* **2021**, *13*, 3975. <https://doi.org/10.3390/su13073975>

Academic Editor: Maria Socorro Garcia Cascales

Received: 1 March 2021

Accepted: 31 March 2021

Published: 2 April 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

A project can be defined as a “unique, temporary, multidisciplinary and organized endeavor to realize agreed deliverables within pre-defined requirements and constraints” [1]. Industrial construction projects, which fall under this definition, have their own peculiarities, one of the most relevant being their complexity due to the multiple variables, components, and activities that must be managed, coordinated, and controlled in each phase. They are part of various construction markets, but different from civil construction, infrastructure, or residential buildings, being their purpose to deliver a functioning facility or asset to End-Users. This type of projects encompasses a wide variety of fields of application and sectors, including but not limited to power generation, nuclear plants, industrial plants, renewables, power transmission and distribution, water treatment, Oil and Gas fields development and processing plants or treatment units [2–6].

In this study, there is considered hypothetical planning data from a set of power generation projects (PXE) and one non-power generation project (INE), all of them carried out under the EPC (Engineering, Procurement, and Construction) form of contracting arrangement. The EPC contractor is responsible for all the activities during the execution phase, including the design engineering, the procurement, and supply of the necessary equipment and materials, the construction, along with the installation, commissioning and start-up of the facility or asset in order to complete the scope of works necessary for the handover of the project to the Client.

The prominence of EPC projects is clear from the figures for the international procurement market, which shows that EPC contracts were worth an estimated USD 7.60 trillion

at the end of 2019 [7], although this project model is coming under increasing pressure due to low productivity, tight profit margins or lack of digitization [8]. Improving the efficiency of projects to minimize problems between the parties involved, optimize execution, and control operating and maintenance costs become an indispensable task given the multitude of simultaneous activities that are extremely complicated to manage [9–11], especially since they are heavily conditioned by cost margins and deadlines.

Numerous methodologies and tools have been developed and implemented to manage projects, despite which the number of failed projects remains high [1]. Industrial construction projects are no exception, where a clear example of this can be found in EPC contracts, which very often experience high-cost overruns and significant delays [12–15]. Over the last years, diverse studies have tried to identify the variety of reasons for delays in projects and proposed methods to mitigate them by analyzing different aspects and factors [16–21]. According to the literature review, there are major factors causing a delay in the completion of construction projects [22,23] such as construction mistakes and defective works, delays in approving design documents and in payments, change orders, difficulties in financing project, ineffective project planning, and scheduling, late procurement, and delivery of materials, low productivity, material shortages, mistakes and deficiencies in design documents, poor communication and coordination with other parties, poor site management and supervision, price escalation, or unreliable subcontractors.

While in the case of industrial construction projects, additional factors can be included [20,24–28], such as a change in laws and regulations, confiscation of the bid guarantee, contractor's incompetent technology, delay of design approval from consultant, incomplete onshore fabrication, inaccurate contractor cost estimates, inadequate baseline schedule development and updating by contractors, inadequate contractor experience, or insufficient and inexperienced owner's technical personnel.

It seems clear then, as stated by many authors, that one of the main causes of delays is the project planning, but few of them have verified the effect of the activities of the schedules in the delays of the industrial construction projects [18–21,23,25,27–30].

The importance of planning and scheduling understood as the tool that aims to identify dependencies between activities, allocate resources, determine the start and finish dates, and thus calculate the duration of the entire project and each one of its phases and parts [31–33], is evident throughout the different stages of an industrial project. Commercial software available on the market, such as Primavera P6 or MS Project [34,35], is widely used to develop these schedules, from which it is possible to extract information and learn valuable lessons once project execution is complete.

The monitoring and control processes of these complex construction projects must consider variables related to the fundamental constraints, such as the execution times of activities, delays, or activities that pose a higher risk of generating cost overruns and/or delays [4,15,23,36]. These delays are also a major source of complaints and disputes [37] concerning both costs and deadlines and can involve not only the principal parties to the contract but subcontractors as well [38], which highlights the need to analyze and identify the causes of the delays [39–41].

It is believed that by analyzing the activities included in the planning of industrial projects it will be possible to evaluate some of the reasons that can influence variations in execution times and help to identify those activities and components that are likely to experience or cause execution delays, making an important contribution to overcoming these project limitations. In an effort to examine the impact of these delays, it is also considered necessary to develop a planning analysis methodology to assess the causes of delays occurring during the execution phase.

The emergence of technologies aimed at the construction sector seems to offer significant advantages for implementation in industrial projects, such as the use of 4D software and Building Information Modelling (BIM) methodology as a support for planning, coordination, and project management [42]. It also seems inevitable to start exploring the feasibility of using new techniques for data analysis, based on innovative tools such as

data mining, Big Data [43,44], or statistical software, which provides utilities for processing the parameters of various models [45].

In this study, the R-Commander [46] graphical interface of the R programming language [47] will be used to provide an analytical view of certain elements that cause delays during the development of industrial projects, in an effort to describe and classify the types of activities that tend to be delayed during project stages, and to identify, to the extent possible, those components that are most significant in this respect using different metrics. In general, the use of statistical software can help organizations obtain valuable information which contributes to the know-how that facilitates the correct assessment of project risks from an execution perspective and adds value through lessons learned that can be applied to future projects.

While in absolute terms it seems straightforward to identify activities that are completed after their originally scheduled completion date, it is also important to discern whether it is the implementation of the activity itself that was delayed or whether, on the contrary, the delay could have been due to issues not attributable to the activity itself. Despite the heterogeneity of the scope of industrial construction projects [3,48–50] and the relationships between scheduled activities, there is no denying the importance of determining not only the start of an activity but also the loss of float or reduction of the planned duration, which can even influence the modification of the tasks associated with each activity to mitigate or avoid delays.

This study focuses on the relationships between and effects of the most important factors (System, Discipline, Specialty, Component) of these projects, and the influence which delays in predecessor activities can have on delays in successive activities. It also examines how to connect delays to the different components of a project at the planning level, starting with the most generic ones such as Systems and moving on to the disaggregated ones such as units or equipment. The aim is not to analyze project delays in general but rather to determine which activities and types of activities tend to experience execution delays during a project, considering the activities, dates, durations, relationships (unique predecessor/successor), and the different elements involved in the planning. The workflow of this approach to the analysis through planning is shown in the Figure 1.

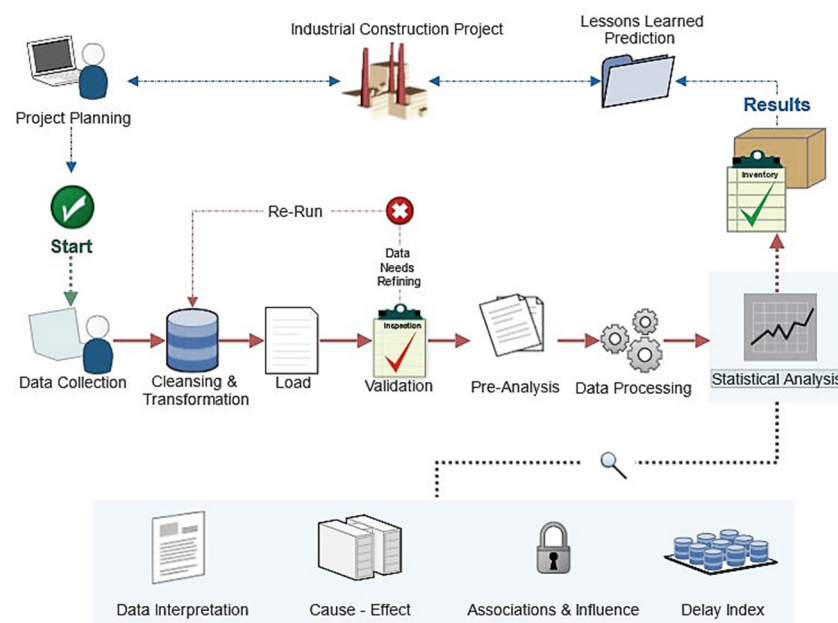


Figure 1. Methodology workflow.

It is believed that this research could lead to an approach for visualizing certain factors that cause delays in these types of projects and even serve as an analysis methodology if properly developed.

2. Materials and Methods

2.1. Data Collection

For the development of the study, hypothetical data from industrial project plannings were used as a starting point, based on and consistent with real projects, treating the information generically and focusing mainly on basic parameters such as the duration of activities and delays, the relationships between activities (predecessor/successor) and the different components and equipment included in each plan. The projects considered were quite heterogeneous in terms of their scope of application, resources, etc. Nevertheless, they were considered to be representative of the industrial construction field since they shared at least the following common characteristics:

- They would be part of an EPC-type contract.
- They would be subject to planning and a control and monitoring schedule.
- They would include activities such as:
 - Engineering: basic, detailed, Construction and Commissioning support.
 - Procurement: equipment and material purchases, manufacturing, supply.
 - Construction: civil, electrical, instrumentation and control (I&C), mechanical.
 - Commissioning and Start-up.
- There were unique relationships between activities.

2.2. Data Cleansing and Transformation

The first step was to prepare the information from the source data. Given the variety of industrial construction projects that exist, there may be different original data formats (software/PDF/spreadsheet), so they had to be converted to a common format. The medium used to contain and prepare the data before loading them into the statistical software was a spreadsheet.

Although it was possible to perform these tasks in the software itself, before the data were imported, they were cleansed of errors and inconsistent values, removing the parameters that were not going to be used in the study. Other parameters were used instead, such as:

- Difference in duration of activities (initial phase/advanced phase).
- Delay in the start and finish of an activity (difference of initial/advanced phase).
- Delay of the predecessor activity.

This last parameter was of utmost importance, as it tells how an activity was impacted by a predecessor (or predecessors). Once the format had been standardized, the next process was the identification and grouping of:

- Systems: associating similar terms to the extent possible.
- Disciplines: Procurement, Engineering, Construction, among others.
- Specialties: according to EPC scheme; Civil, Electrical, I&C, Mechanical.
- Components: basic units or elements of the Systems.

In many cases, the classifications were determined by the planning structure and type of project; in others, the most general criteria possible were assumed in order to identify under which heading various activities or components were included. Within Systems, the main ones and equipment were considered while the rest were included in sections that were as generic as possible in order to cover the diversity of components contained in the plannings.

Since the time interval between plannings is not the same for every project, weighted magnitudes were calculated from the existing data in order to obtain a more harmonized criterion for the values of the study variables. The word "Weighted" was added to these new magnitudes to differentiate them from the original ones.

2.3. Data Import into R

R-Commander, a graphical interface that covers a majority of the most common statistical analyses in drop-down menus without the need to write code, was used for statistical analysis and the creation of graphs, in conjunction with R-Studio, an integrated development environment for the free programming language R [47,51,52].

The data were loaded by choosing the option “Import data set” in the menu, Data→Import Data→From Excel. After the import, the software itself established the qualitative and quantitative variables, which were called factors and variables in this software, respectively, the most significant of these being:

- {6} “System”, {7} “Specialty”, {8} “Discipline”, {9} “Component” for factors.
- {15} “Delay Start”, {17} “Delay Finish”, {19} “Duration Difference”, {21} “Delay Predecessor” for variables.

The factors were chosen by selecting those that better define the structure of the data set since they were widely present in these types of projects, which benefits the homogenization of the original data and then obtaining results common to the majority of industrial construction projects. Although not listed above, there are other less relevant factors, such as “Project Type”, “Project”, “ID_Activity”, “ID_Predecessor” and “Activity_Name”, rarely used. The description of the variables was as follows:

- Delay Start: difference between the planned start date of the planning activity and the actual start date, measured in days. Data is obtained from each schedule.
- Delay Finish: difference between the planned finish date of the planning activity and the actual finish date, measured in days.
- Duration Difference: difference between the expected duration of an activity and the actual duration, measured in days.
- Delay Predecessor: expected end date of the activity preceding the activity and its actual end date, measured in days. The purpose of considering this variable is to try to evaluate the influence that the delay of the activity has on the subsequent one.

What was intended by focusing on the variables chosen was to reduce the number of elements to handle in the analysis to the lowest possible. In this way, instead of working with many variables, only a few were used that group most of the information, simplifying the analysis of the planning. This also allowed for generating other variables if necessary, as done with the weighted ones, which were calculated by dividing the variable by the duration in months of the project.

The first verification was a summary of the active data set from Statistics→Summaries→Active data set. The general information of the data set comprised a total of 6216 entries corresponding to five projects, including minimum and maximum values, first and third quartiles, the median, the mean, and the missing values. The same summary showed the frequency distribution of the main items in each category.

One important piece of information that this initial analysis revealed was the dispersion of data and the existence of many missing values in relation to the delays in the start and finish of the activities. It can also be seen that these were left-skewed asymmetric distributions of values where the mean was greater than the median, as shown in the Figure 2. This can also be seen in the Numerical Summaries of the variables, such as the one for “Delay Finish” and “Duration Difference”, where the dispersion measurements extracted using the menu option Statistics→Summaries→Numerical summaries can be visualized.

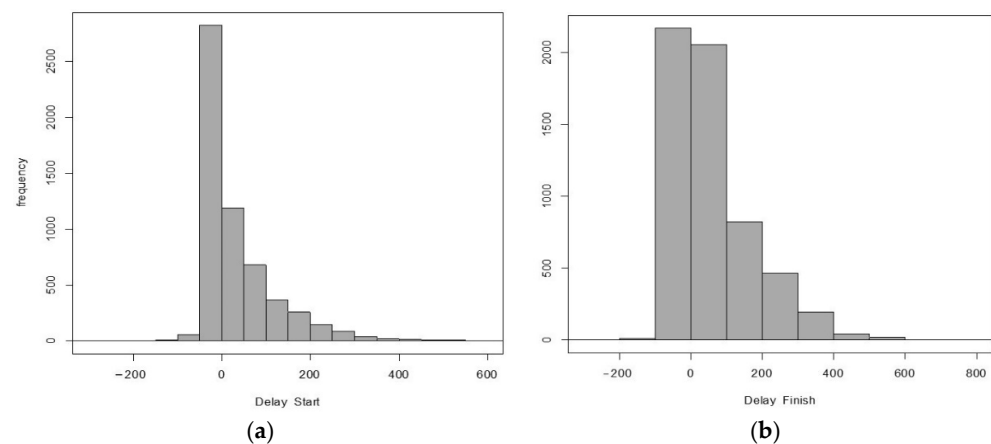


Figure 2. Histograms (own elaboration): (a) Delay Start; (b) Delay Finish.

The next step was to apply filters to reduce anomalies that could be due to issues not directly related to the execution of the project itself (force majeure, financial problems, onsite risks, etc.), and to limit existing outliers given the dispersion of data. At the same time, possible errors were minimized using statistics and inference. Following an iterative process to estimate the best fit, the filtering for “Delay Finish Weighted” was performed for values $30 > X > -10$, while for “Duration Difference Weighted”, filtering was performed for the interval $30 > X > -10$. The final number of rows compared to the spreadsheet dropped from 6216 to 5145, with the total number of activities decreasing by 17.2%. A new overview of the active dataset was then generated, summarized in the Table 1, which showed minimum delay values reduction.

Table 1. Comparison Delay Start/Delay Finish after filtering.

Delay Start	Filtered	Initial	Delay Finish	Filtered	Initial
Min.	−140.0	−290.00	Min.	−158.00	−288.00
Q ₁	0.0	0.00	Q ₁	0.00	0.00
Median	0.0	0.00	Median	28.00	28.00
Mean	42.8	43.77	Mean	72.29	70.71
Q ₃	60.0	63.00	Q ₃	112.00	111.00
Max.	535.0	555.00	Max.	538.00	735.00
NA’s	-	487	NA’s	-	415

2.4. Statistical Analysis

Once the data set was configured, different statistical studies were carried out. First, a more detailed summary of the new set was drafted as the starting point for the analysis from a descriptive point of view, with statistics for the complete set. Then, the key characteristics of the duration and delay variables were reviewed in order to describe them using a small number of descriptors. This exercise helped to visualize trends and to summarize and characterize data and interpret them. The key conclusions of this section showed:

1. A (persisting) lack of uniformity of the factors under study in the dataset and the dispersion of variable values.
2. A reduction in the average duration of activities as the project progresses.
3. Shorter delays in completing activities in the advanced stages than at the start.

For this summary, it was started with the options in *Statistics* → *Summaries* → *Numerical summaries* to evaluate the centrality and dispersion of the variables and the effectiveness of the filter applied to reduce outliers, relying on the graphs available in the *Graphs* menu:

- Statistics: frequency tables and numerical characteristics of position, centrality, and dispersion such as mean, median, maximum, and minimum, quartiles, or skewness.

- Graphs: scatterplots and plots of means which facilitate the transmission and presentation of information in a visual way.

The next step was the comparison of variables and correlation using the different options in the *Statistics*→*Summaries* menu, where associations between variables and factors are checked using:

- Contingency tables.
- Numerical summaries of variables grouped by factors.
- Variable correlation matrix.

This was used to check the relationship between different variables in order to determine the existence of a cause-effect relationship. The following conclusions were drawn from the results obtained, among others:

1. The farther along with in the phases of a project, the greater the delay by Discipline.
2. There was a direct relationship between a delay in predecessor activity and a delay in the completion date of the next activity.

As an example of possible combinations for segregating information in this and subsequent sections, the data was usually filtered for Construction Discipline, Mechanical Specialty, or a combination of both, or Engineering and Civil.

Linear regression analyses were then conducted to determine the function that interprets the relationship between the dependent and the independent variables. In addition to providing information on the residuals, the results were used to obtain the coefficient of determination, R^2 , which allowed for studying the goodness-of-fit of the model, as well as the values of the test statistics and the corresponding p -values [53] in the following cases:

- Single, between predecessor and successor.
- Two-degree, predecessor-successor-subsuccessor: the relationships between an activity, its successor, and the successor's successor must be established.
- Multiple, with grouping by factors.

The key conclusions of this section were as follows:

1. As the delay in the predecessor activity increases, the delay in the activity under review increases.
2. The duration of an activity increases in direct relation to a delay in the Finish date.
3. The more distant the degree of relationship between activities, the smaller the effect which the delay of the predecessor has on the subsequent successor.
4. No conclusive results can be drawn on the impact of second-degree successor activities or the multiple linear regression developed.

With regard to the analysis of variance to assess the differences in delay per project, it was considered that there were significant statistical differences between them, either in general or those of the PXE type.

Finally, it was concluded by determining the Components with the greatest completion delays and their relationships with the following stages. To that end, a new parameter called Delay Index was defined, which was calculated using median values and the interquartile range (IQR), finding that the main activities related to the elements considered in this index were the ones concerning Mechanical Insulation and one type of Turbine. When the predecessor activities of this Turbine were analyzed, it was observed that Mechanical Supply activities as those with the highest incidence.

3. Results

3.1. Descriptive Analysis

Given the high number of factors and variables, only some of the cases were shown in this section. For the descriptive analysis, a summary of the active data set was created to calculate basic statistics for these factors and variables. The information of some of the main factors is shown in the Table 2 below:

Table 2. Data summary-System/Specialty/Discipline.

System ¹	Frequency	Specialty	Frequency	Discipline	Frequency
General	755	Civil	960	Commissioning	714
Electrical system	546	Electrical	1126	Construction	1312
Mech. assembly	421	General	157	Engineering	1617
Civil works	324	I&C	574	General	5
Supply	287	Mechanical	2328	Milestones	2
Mech. system	263	Milestones	0	Procurement	1495

¹ Only the Systems with the highest number of observations.

What was observed was a lack of uniformity of the data in terms of the Specialty and Discipline factors, as was to be expected due to the different types of input data.

A summary was obtained for each of the variables in order to begin assessing the centrality and dispersion of these variables and to draw initial conclusions. In the case of the evolution of the durations and delays of the Table 3, it seemed clear that as the stages of the project advance (Design Engineering, Procurement, Construction, Commissioning), the average duration of the activities was reduced, either due to adjustments as the project reaches the final stages or due to the needs of the project to make up for cumulative delays.

Table 3. Data summary. Durations and Delays.

Summary	Duration Advance Stage	Duration Early Stage	Delay Start Weighted	Delay Finish Weighted
Min.	−11	0	−23.40	−9.40
Q ₁	13	18	0	0
Median	39	49	0	2.60
Mean	71.28	100.80	3.17	5.44
Q ₃	97	156	4.70	9.40
Max.	678	921	30.80	29.90

On the other hand, the delay in starting an activity at an advanced stage was less than the delay in finishing. In other words, although there was relatively little delay in starting activities compared to what was planned at the beginning of the project, the delay in completion increased significantly for the cases under study. Add to this the information provided by the percentiles as well as positive skewness with the most extreme values above the mean, and it was confirmed that there was a shift in the delay of activities which was more pronounced for the variable Finish. This is presented in Figure 3 and Table 4:

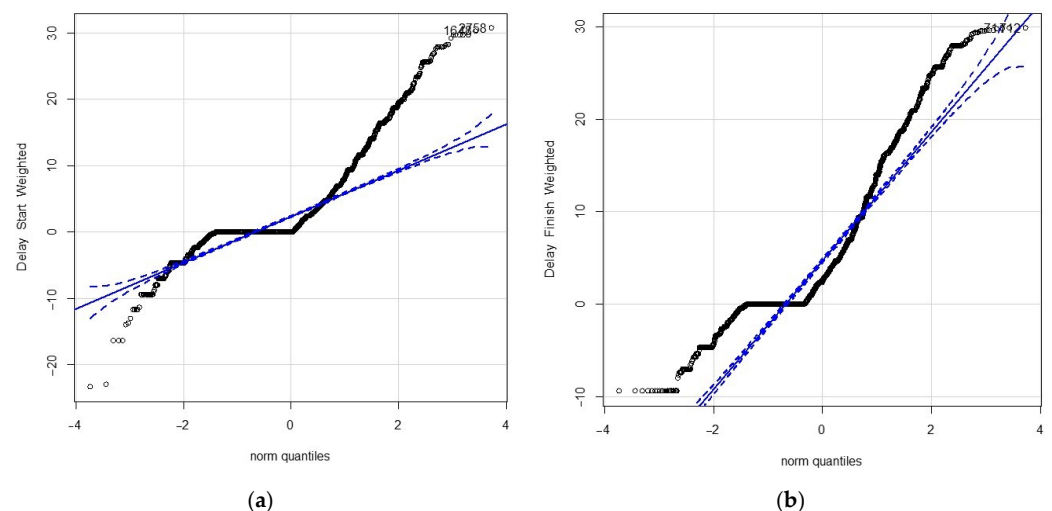
**Figure 3.** Quantile-comparison plot Delay Weighted (own elaboration): (a) Start; (b) Finish.

Table 4. Data summary. Skewness.

	Skewness	0%	25%	50%	75%	100%
Delay Finish Weighted	1.15	−9.4	0	2.6	9.4	29.9
Delay Start Weighted	1.58	−23.4	0	0.0	4.7	30.8
Duration Difference Weighted	2.17	−9.4	0	0.0	2.4	29.7

It was also clear from these values that the dispersion of the variables, despite data cleaning and filtering, was still significant and that, in some cases, there were quite a few outliers, which was also visible in the different descriptive analyses.

3.2. Comparison of Variables and Correlation

Contingency tables can be used to infer information on the activities with the greatest relative weight in relation to the factors. Specifically, in the projects analyzed and for the Discipline and Specialty factors, these activities were Procurement and Commissioning within the Mechanical Specialty, with Civil Engineering activities related to the Construction phase also having particular relevance in the projects, as shown in Table 5. This was consistent with the sector in which industrial construction projects were carried out, where it was common to start from scratch with earth movement for the Civil Engineering and Construction portions and the need for extensive Mechanical equipment to do the work. It was also noteworthy that the importance of Civil Engineering activities disappeared during commissioning, as is logical. As for the general part of the project, the Electrical Specialty was the one that has the greatest influence on the activities.

Table 5. Contingency table Discipline/Specialty (%).

Discipline/Specialty	Civil	Electrical	General	I&C	Mechanical
Commissioning	0.30	27.90	2.20	15.80	53.80
Construction	38.40	16.60	0.90	6.90	37.20
Engineering	20.90	21.60	5.40	14.10	38.00
General	0	40.00	40.00	0	20.00
Procurement	7.80	23.90	2.50	9.60	56.30

By way of example, in the case of Systems filtered by the Engineering Discipline, the activities related to I&C were seen to have greater relative weight in Engineering activities than in the rest of the Specialties. This was largely due to the importance of the Distributed Control System (DCS), which was critical to the operation of an industrial plant.

Moving on to other numerical summaries, in the comparison of delays by Discipline, it seemed clear according to Figure 4, that the delays increase as the project advances through the different phases.

Using the value of the weighted variable in this case, what it showed was that the delay in Engineering < Procurement < Construction < Commissioning, which is in line with the different stages of project development according to the theoretical phases of a construction project of this kind. The same occurs when predecessor activities are analyzed: the further along the project is, the longer the delay.

It was also necessary to verify the relationship between some variables and others, i.e., to discern whether there was indeed a cause-effect relationship that was appreciable from an analytical or statistical point of view. A correlation matrix was used for this purpose.

According to Table 6, the linear correlation coefficient of Predecessor Delay and Duration Difference was 0.10, a very weak association, but for Delay Finish it was intermediate at 0.55. Another intermediate result, with a value of 0.59, was the correlation between Delay Finish and Duration Difference. In other words, there was a direct relationship between a delay in predecessor activity and the completion of the next one, and thus an increase in the duration of the successor activity.

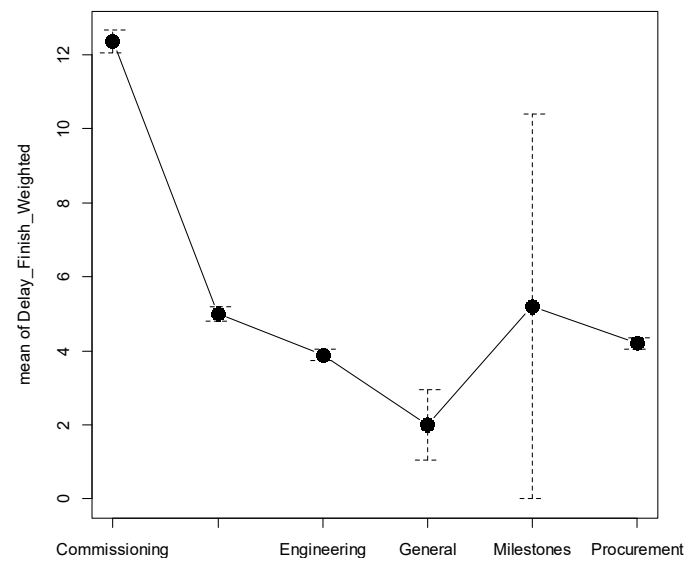


Figure 4. Plot of means Delay Finish Weighted by Discipline (own elaboration).

Table 6. Correlation matrix. Delay Finish/Delay Predecessor/Duration Difference.

Overall	Delay Finish W.	Delay Predecessor W.	Duration Differ. W.
Delay Finish W.	1	0.5525808	0.5972869
Delay Predecessor W.	0.5525808	1	0.1006063
Duration Difference W.	0.5972869	0.1006063	1
Mechanical			
Delay Finish W.	1	0.616303	0.5834116
Delay Predecessor W.	0.616303	1	0.113232
Duration Difference W.	0.5834116	0.113232	1
Construction & Mech.			
Delay Finish W.	1	0.5332073	0.4861093
Delay Predecessor W.	0.5332073	1	0.1948184
Duration Difference W.	0.4861093	0.1948184	1

It was possible to filter using different factors, such as Specialty or Discipline, to see how the different variables behave as the level of detail increased. The delay in the predecessor activity, in the case of Specialty = Mechanical, had a greater impact on the delay of Mechanical activities than on the overall activities. However, when the Mechanical Specialty was observed for Construction activities only, the relationship between the predecessor and the successor activity was somewhat weaker.

All of these statistical analyses provide useful data for similar construction projects. By knowing which activities are critical, it is possible to identify which predecessor activities to focus on in order to improve efficiency and take preventive actions to limit cumulative delays.

3.3. Linear Regression

3.3.1. Delay Finish

After obtaining the correlation results, the linear regression model was reviewed with respect to the totality of the activities, using Delay Finish in relation to Delay Predecessor as the dependent variable. The following results were obtained with weighted variables as shown in Table 7. In this model, the p -values that help to resolve these contrasts were in both cases, 2×10^{-16} , a value less than 0.05 [54,55].

Table 7. Linear regression–Delay Finish~Delay Predecessor.

Residuals	Min	Q ₁	Median	Q ₃	Max
	−40.21	−2.58	−1.78	1.59	27.12
Coefficients	Estimate	Std. Error	t value	Pr(> t)	
(Intercept)	2.58	0.12	22.18	<2 × 10 ^{−16}	
Delay Predecessor W.	0.56	0.01	39.29	<2 × 10 ^{−16}	
Multiple R-squared	0.3053				

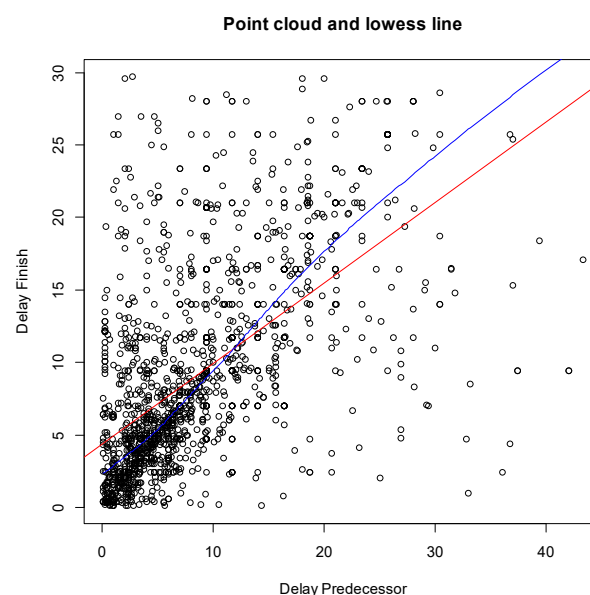
Thus, considering a significance level of 5%, the null hypothesis would be rejected in both cases, concluding that there was a linear relationship between the variables. Therefore, as the delay in the predecessor activity increased, the delay also increased and the linear model can be written as follows, considering the model function $y = \alpha + \beta x$ which describes a line with slope β (i.e., regression coefficient = Delay Predecessor Weighted coefficient) and y-intercept α [56]:

$$\text{Delay Finish Weighted} = 2.58300 + 0.55581 \cdot \text{Delay Predecessor Weighted}$$

The smaller the residuals, the better the fit of the model to the data and the more accurate the predictions made using the model, as the residuals are the differences between the observed responses of the explanatory variables and the prediction calculated using the regression function [57]. The standard error of the residuals indicates the dispersion of the residual values so that the better the fit, the smaller the standard error. In this case, the standard error of the residuals was 5.966.

The value of R^2 (Multiple R-squared) ranged between 0 and 1 [36,54], so values close to 1 indicated a good fit of the linear model to the data. In this case, 30.53% of all variability related to the delay in the completion of Mechanical Construction activities can be explained by the delay of the predecessor.

The model, fitted to the point cloud with the addition of a smoothed lowess line [58], is plotted in Figure 5. All data must be positive in order to obtain such a lowess line, so the data table was transformed to positive by filtering for Delay Predecessor Weighted and Delay Finish Weighted >0. In addition, it was filtered by Specialty for this example.

**Figure 5.** Point cloud and lowess line (blue) Delay Finish~Delay Predecessor (own elaboration).

In the case of the Duration Difference weighted variables explained by Delay Finish, the p -values were also less than 0.05, so there was a linear relationship between the variables, as stated in Table 8.

Table 8. Linear regression. Duration Difference~Delay Finish.

Residuals	Min	Q ₁	Median	Q ₃	Max
	−18.18	−1.72	0.43	1.23	23.83
Coefficients	Estimate	Std, Error	t value	Pr(> t)	
(Intercept)	−0.43	0.08	−5.63	1.89×10^{-8}	
Delay Finish Weighted	0.50	0.01	60.16	$<2 \times 10^{-16}$	
Multiple R-squared	0.4130				

Although the descriptive section concluded that, in general, activities tended to get shorter as the project progresses, the regression also showed that the duration of activities whose completion date had been delayed increases, despite the fact that corrective measures should be taken to compensate for the delay caused by reducing the duration of the activity:

$$\text{Duration Difference Weighted} = -0.42594 + 0.49593 \cdot \text{Delay Finish Weighted}$$

3.3.2. Delay with Two-Variable Filtering

To evaluate the option of obtaining a higher level of detail in the results, different estimations of linear regression models were performed by changing the variables. In another representation, shown in Table 9, the regression was again filtered, as in the previous sections, by Construction and Mechanical to check the behavior of the model.

Table 9. Linear regression. Delay Finish~Delay Predecessor (Construction & Mechanical).

Residuals	Min	Q ₁	Median	Q ₃	Max
	−26.25	−3.42	−1.35	2.54	25.53
Coefficients	Estimate	Std, Error	t value	Pr(> t)	
(Intercept)	2.99	0.38	7.87	3.13×10^{-14}	
Delay Predecessor W.	0.54	0.04	12.76	$<2 \times 10^{-16}$	
Multiple R-squared	0.2843				

The p -values were 3.13×10^{-14} and 2×10^{-16} , meaning that in this case, the impact of predecessor activities was lower than in the general case, 28.43%, obtaining the following linear model:

$$\text{Delay Finish Weighted} = 2.99338 + 0.53913 \cdot \text{Delay Predecessor Weighted}.$$

There were numerous possibilities and combinations. Depending on the variables, different filters can be applied to obtain the results that allowed them to be properly analyzed.

3.3.3. Impact on Successive Activities

Linear regressions between the predecessor (O) of an activity (I) and the delay in the completion of its successor (II) were also evaluated. For activities where the predecessor was an activity in the Engineering Discipline and the Mechanical Specialty, it was found that the variability of the delay in the completion of the successor II activity was only due to the initial predecessor 11.98% of the time, as can be seen in Table 10. It followed that the more distant the relationship between activities, the smaller the effect which the delay of the predecessor had on the subsequent successor (II). However, as more filters were applied and the level of detail increased, the null hypothesis cannot always be discarded.

Table 10. Linear regression. Delay Finish Successor II~Delay Predecessor.

Coefficients	Estimate	Std, Error	t Value	Pr(> t)
Mech. Engineering				
(Intercept)	3.82	0.54	7.09	1.76×10^{-11}
Delay Pred Weigh Successor	0.42	0.08	5.54	8.65×10^{-8}
Multiple R-squared	0.1198			
Mech. Engineering & Cons				
(Intercept)	7.99	2.68	2975.00	0.0177
Delay Pred Weigh Successor	0.08	0.14	0.56	0.591
Multiple R-squared	0.03768			
Civil Engineering & Cons				
(Intercept)	3.93	1.24	3.16	0.00258
Delay Pred Weigh Successor	0.20	0.10	2.00	0.05083
Multiple R-squared	0.07006			

In the case of Mechanical Engineering on the table above (Mech. Engineering & Cons), it does not have a significant effect on second-degree Construction activities, while in the case of the impact of Civil Engineering on the Construction subsequent successors (Civil Engineering & Cons), this conclusion is not as obvious since the p -values are bordering on the acceptable limit.

3.3.4. Multiple Linear Regression

The next observation refers to the development of a multiple regression model. It contains more than one independent or explanatory variable, which could even be qualitative (factor), and which determines the value of the variable to be analyzed. The effect of a predecessor activity, segmented by Discipline, on the completion of the second-degree successor activity (subsuccessor) is studied and shown in Table 11.

Table 11. Multiple Regression. Delay Finish Successor II~Delay Predecessor.

Residuals		Min	Q ₁	Median	Q ₃	Max
		−30.09	−4.00	−1.53	2.61	41.01
Coefficients		Estimate	Std, Error	t value	Pr(> t)	
Discipline Predecessor	(Intercept)	3.93	0.25	15.74	$<2 \times 10^{-16}$	
	[T. Construction]	0.07	0.41	0.16	0.87	
	[T. General]	−3.93	2.02	−1.95	0.052	
	[T. Milestones]	−3.68	0.92	−4.01	0.000062	
	[T. Engineering]	0.46	0.37	1.22	0.22	
	[T. Commissioning]	−13.56	1.41	−9.62	$<2 \times 10^{-16}$	
Multiple R-squared		0.05746				

Looking at the p -values, only delays in the Commissioning predecessor and, to a lesser extent, Milestones, seemed to have an influence on the delay of the subsuccessor, but the number of observations within the dataset was very low (around 2.1%), so this kind of analysis did not appear to yield conclusive results.

3.4. One-Way Analysis of Variance (ANOVA)

This analysis of variance made it possible to compare different groups in relation to a variable. In the case shown in the Table 12, the different projects in the study (factor) were compared with respect to the delay in the completion of activities (variable).

Table 12. ANOVA of delay by project.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Project	4	49,467	12,367	274.3	$<2 \times 10^{-16}$
Residuals	5140	231,774	45		
Project ID	mean	sd	data:n		
1P	2.64	3.28	2222		
2P	12.37	8.43	618		
3P	6.30	9.01	697		
4P	5.85	8.52	1228		
5P	7.54	6.99	380.00		

The groups to be compared should be normally distributed and homogeneous, but because they were large in size it was less important to ensure these two assumptions since ANOVA is usually a fairly “robust” technique, behaving well with respect to transgressions of normality [59,60]. The corresponding hypotheses are:

Hypothesis 1 (H1). *The delay in the completion of activities is the same in all projects.*

Hypothesis 2 (H2). *Some are different (there are differences between at least some of the five projects).*

According to the data obtained in the table above, the mean delays in the completion of activities differed. It can therefore be concluded that there were statistically significant differences between the projects with respect to the delay variable, since $F(4,5140) = 274.3$ (not equal to 1), with $p < 0.05$ (2×10^{-16}) [54,61]. This was the expected result, given the non-uniform nature of the projects and their source data.

The same was true for PXE-type projects where, as shown in Table 13 and despite similarities, there were significant differences in the magnitude of the delays in general and for the Disciplines in particular. In the latter case, the F-value was lower than before, but far from a value of 1.0.

Table 13. ANOVA of power generation projects (PXE) project delays by Specialty.

	Df	Sum Sq	Mean Sq	F Value	Pr(>F)
Specialty	4	4975	1243.8	26.02	$<2 \times 10^{-16}$
Residuals	3912	186,966	47.8		
Specialty	mean	sd	data:n		
Civil	3.462878	5.582319	695		
Electrical	5.148743	6.638735	915		
General	3.585315	5.397978	143		
I&C	7.322566	7.392519	452		
Mechanical	5.745035	7.501107	1712		

This type of analysis was useful for evaluating the projects for which it was appropriate to work with statistically more homogeneous parameters when the information was to be included in a dataset to be studied.

3.5. Component Lag Analysis

To conclude the statistical analyses, a new numerical summary was carried out to calculate the mean and standard deviation values of the delays in the completion of the activities for equipment, materials, etc., under the Components heading of the data table. The median was also extracted for each one. Since they were widely dispersed values, it was quite indicative of the ones that need to be reviewed.

Given that the average variation between the mean and the median of the weighted lag variable per element yielded a value of 69.07%, the latter was used along with the interquartile range (IQR) to determine the Components with the highest lag index.

$$\text{Index} = \text{Median} + \text{IQR}$$

The IQR interquartile range was the difference between the third and first quartiles to estimate the dispersion of data distribution, highly recommended when the measure of central tendency used was the median. Considering that the total number of rows was 5145 with 96 types of Components, those with a frequency of occurrence of at least half the average, i.e., 26.80, were evaluated:

$$\text{Frequency} = \text{Rows}/(\text{No. Components} \cdot 2) = 5145/(96 \cdot 2) = 26.80$$

obtaining the ranking by index on Table 14, where the activities related to Insulation and Turbine A would have the highest rate of delay:

Table 14. Components with the highest delay index.

Component	n	Median	IQR	Index
Insulation	31	16.60	11.55	28.15
Turbine Type A	87	11.70	15.95	27.65
Valves	104	9.40	16.10	25.50
Cleaning and treatments	118	7.00	16.40	23.40
DCS	55	6.60	16.10	22.70
Regulation	27	9.10	12.15	21.25
Chemical dosing and sampling	131	7.00	12.30	19.30
Filtration and ventilation	106	7.00	11.70	18.70
HVAC	38	8.35	8.83	17.18
Pressure devices and vessels	29	7.00	9.40	16.40

However, not every activity has a predecessor and they are not always homogeneous even if they do exist. By way of example, also presented in Table 15, using the source data on the spreadsheet, for Turbine A, there were various types of predecessor activities. These ranged from Procurement activities for the Turbine itself (6 Procurement and 2 Supply-related) to lifting elements in the case of the Crane System in the Construction Discipline. The Specialties were Civil, Electrical, Mechanical and one General which corresponds to Basic Engineering.

From there, it became possible to establish criteria for relationships and the possibility of examining the impact of each predecessor on successive Turbine activities. Successive analyses can be carried out, e.g., how each predecessor activity by Discipline (or System or Specialty) influences the cumulative delay of other activities, either in absolute or relative terms, weighted values, as a percentage or other metrics suitable for quantifying and evaluating such influence. Table 16 shows the average impact of the delay in a predecessor activity on the completion of DCS activities by Discipline.

Table 15. Component/predecessor relationships.

Discipline	Procurement			Construction		
	System	Mech. System	Supply	Cranes	Civil Works	Mech. System
Predecessor	<i>Turbine A</i>	<i>Turbine A</i>	<i>Turbine A</i>	<i>Buildings</i>	<i>Turbine A</i>	
Civil					2	1
Electrical						
General						
Mechanical	6	2	1			1
Engineering			Commissioning			
System	General	Civil Works	Mech. System	Turbine Type A		
Predecessor	<i>Basic Engineering</i>	<i>Buildings</i>	<i>Turbine A</i>	<i>Gas/fluids</i>	<i>Turbine A</i>	
Civil		2	1			
Electrical				1		
General	1					
Mechanical			1		1	

Table 16. Average delay predecessor.

Discipline	Procurement		Construction			Engineering	Comm.
	DCS	Others	DCS	Civil Works	Trans.	DCS	Buildings
Predecessor	DCS	Others	DCS	Civil Works	Trans.	DCS	Buildings
Average Delay Predecessor	40	0	63	128	178	11	−13
Average Delay Finish	54	123	56	148	175	34	−9

4. Discussion

The analysis of planning is a tool that can be used to identify key points affecting the development and efficiency of industrial construction projects, as the amount of information and lessons learned that can be obtained are significant. However, the level of digitization of these types of projects remains low, which is a major handicap for management and control, as well as for subsequent diagnosis. It is also one of the reasons for the productivity gap compared to other industrial sectors [62].

In this study, the use of advanced statistical software for data analysis was examined as a way of contributing to closing this gap and to assess the suitability of its use as part of a methodology that can lead to more effective identification of the elements and causes of delays during the execution phase, based on the study of scheduled activities.

It was possible to extract relevant information from these plannings in a fast and effective way, demonstrating the ease of use of the chosen interface and showing the multiple possibilities it can offer, although statistical knowledge is necessary to take actions or interpret results. In addition, programming is required for options not currently included in the menus, but the capacity and, above all, the speed of calculation allowed complex and repetitive operations to be carried out in a relatively short period of time.

As a starting point for the work, the task of extracting and standardizing project data from different areas was an arduous one, so the possibility of automating this process would be a substantial improvement. This is where the development of new technologies can play an important role, both in the initial phases of extracting common information (relating concepts) and later in the transformation of data before loading it for subsequent analysis using data mining, Big Data, or machine learning tools. The use of automation routines would simplify work time, providing continuous, reproducible, and repeatable analysis, while reducing errors.

Regarding the actual development of the study, descriptive analysis was used to define the characteristics of the data set for the study and to assess the dispersion of its

components using a small number of descriptive statistics. The results showed there was a need to homogenize the data and eliminate anomalies in order to extract more conclusive results. This task was performed using filters and establishing new weighted variables. The initial analysis showed that there was still a lack of uniformity of the factors in the data set and dispersion of the values of the variables. This first analysis also served as the basis for representing trends and for summarizing and characterizing data on the variables relating to the duration and delays of planning activities as the different phases of the projects progressed.

It was subsequently possible to deepen the analysis through the use of variable comparison and correlation tools. It was observed that as the project advanced through the different stages there was a greater incidence in the delays of activities with respect to the original project planning. To a large extent, this was caused by delays in predecessor activities, as a consequence of which the completion date of the next planned activity was shifted.

Linear regression showed that those activities whose completion dates were delayed also experienced an increase in duration so that the delay has a dual effect. It was also found that the more distant the degree of relationship between activities, the smaller the effect which the delay of the predecessor has on the subsequent successor. However, it was not possible to draw conclusive results regarding the impact of second-degree successor activities in all the combinations of Specialties/Disciplines or from the multiple linear regression carried out, either because of the low number of observations available in the dataset or due to the statistical parameters which were bordering on the acceptable limit.

Regarding the analysis of variance, the key conclusion is that this type of analysis offers the possibility of determining the projects whose parameters are statistically more homogeneous, in order to choose the most similar projects and extrapolate the results to others with comparable characteristics.

When evaluating the causes of delay by Component, the establishment of the “Index” parameter made it possible to identify those components with the highest rate of delay. From there, it was possible to determine which activities influenced their behavior and the impact on the activities with which each Component was directly related.

Regarding the aspects not addressed in the study, it would be feasible, although more complex, to expand it so that each activity would have more than one dependency (predecessor activity). This would also be subject to availability in the source data but would increase the options and the capacity to generate usable results. As the quality and quantity of the source data increases, so does the refinement of the model in relation to sub-successor activities, which in the study were found to be insignificant (Section 3.3.3).

As far as planning is concerned, although no tool of this kind was reviewed in this study, the use of new tools in the construction sector such as 4D software, which combines three-dimensional systems with time as a fourth dimension, is becoming widespread. Among the advantages of 4D software are the increased efficiency in the planning process of construction projects and more efficient monitoring and control of progress [63]. The use of 4D software adapted to the field of industrial projects which provides the required level of planning detail would mean that there would be much more available data and relationships between activities, leading to a much more robust model. Furthermore, as part of the process of digitizing the industrial projects sector, it is expected that the development of new technologies will provide new tools that will improve the efficiency of complex construction projects and increase the information available for analysis, such as advances in the field of Artificial Intelligence (AI) planning or the use of BIM [64–67].

5. Conclusions

This study presented an approach based on a planning analysis methodology to assess the certain elements that can influence variations in execution times during the development of construction industrial projects, and to identify, to the extent possible, those components that are most significant in this respect using different techniques, as well

as the impact in other components, considering the activities, dates, durations, relationships and the different elements part of the project schedules.

As part of the conclusions of this initial phase of the analysis, it was concluded that with this methodology, it was possible to determine the Specialties, Disciplines, and Systems whose activities had more influence on the rest of the factors. In the case of Specialties:

- Mechanical activities were almost 54% of Commissioning, 38% of Engineering, and above 56% of Procurement,
- Civil activities represent more than 38% of Construction.

This basic evidence from planning can provide the projects with operational information to define the necessary resources leading to an efficient development of the projects. Other information available from this data—which are also useful for other projects—are related to the Systems with the greatest relative weight in Disciplines. Although the Specialties of the Systems are varied, still the outstanding influence of the Mechanical part can be found, being the more relevant Systems for Commissioning, the turbines A and B and boiler; civil works, power source installation and water systems for Construction; electrical system, mechanical assembly and civil works for Engineering; electrical and mechanical systems for Procurement.

As stated before, by identifying the critical activities it is possible to focus on predecessor activities in order to improve efficiency and take preventive actions to limit cumulative delays. The lag in the predecessor activity had a greater impact on the delay of Mechanical Specialty than in I&C, with a similar effect in Electrical, while the Civil Specialty presented the lower ratio according to the results.

Linear regression was conducted to determine the function that interprets the relationship between the dependent and the independent variables in different cases, from single to multiple linear regression. The key conclusions of this section are as follows:

- The linear regression equations established that the delay of activity increases its duration and is directly proportional to its predecessor.
- The more distant the degree of relationship between activities, the smaller the effect which the delay of the predecessor has on the subsequent successor.

For instance, in activities where the predecessor was an activity in the Mechanical Specialty, the most influent one according to the contingency tables, the variability of the delay in the completion of the second-degree successor as a result of the first predecessor is reduced up to three times compared to the direct case. Based on this knowledge, it is possible to predict and control those activities with a higher probability of affecting other elements.

As a final point, the use of the component lag analysis Index was found valid to specifically determine the Components with the highest delay occurrence. In this particular case, it is reflected again the importance of Mechanical activities and components (insulation, turbines, valves).

In relation to the limitations, firstly, the data is based on and in general consistent with real projects, but some information and results could be considered as inexact due to that nature. The aim is to focus on the feasibility of using this methodology rather than considering the contributions of the numerical results. The use of statistical software facilitates the capture of valuable information for organizations. However, it is worth noting that in order to refine the model and be able to extrapolate the results to the industrial projects sector, it would be advisable to study a larger number of plans. As indicated above, the international procurement market was worth \$7.60 trillion at the end of 2019 for EPC projects [7]. Without distinguishing between fields and assuming an average price of \$400 million per contract, the number of projects would be somewhere around 19,000. With a confidence level of 95% and a margin of error of 5%, a sample of at least 379 projects would be required. For a 90% confidence level and a 10% margin of error, data from at least 68 of these projects would be required.

At the same time, the wide variety of sectors where industrial projects are carried out conditions the extrapolation of findings from one project to another, not only by segment but also by company.

Finally, it was noted that there are multiple possibilities and combinations for studying variables and that different filters can be applied to refine the study of these variables. However, as these filters are applied, the amount of data is reduced, so the results were not found to be relevant to the case study. Something similar happens with multiple linear regression, where the scarce number of observations available did not allow the quantitative results to be considered conclusive, beyond the validity of the methodology itself.

Author Contributions: Conceptualization, C.A.-R. and M.A.S.; methodology, C.A.-R. and M.A.S.; validation, C.A.-R. and M.A.S.; formal analysis, C.A.-R. and M.A.S.; investigation, C.A.-R. and M.A.S.; resources, C.A.-R. and M.A.S.; data curation, C.A.-R. and M.A.S.; writing—original draft preparation, C.A.-R. and M.A.S.; writing—review and editing, C.A.-R. and M.A.S.; visualization, C.A.-R. and M.A.S.; supervision, M.A.S.; project administration, M.A.S.; funding acquisition, C.A.-R. and M.A.S. All authors have read and agreed to the published version of the manuscript.

Funding: The research was funded by the Spanish Association of Project Management and Engineering (AEIPRO), the Department of Construction and Manufacturing Engineering of the Faculty of Industrial Engineers of the National University of Distance Education (UNED) and the UNED International School of Doctorate (EIDUNED).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work has been developed within the framework of the “Doctorate Program in Industrial Technologies” of the UNED. The authors would like to thank the UNED International School of Doctorate (EIDUNED).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. IPMA. *Project Excellence Baseline for Achieving Excellence in Projects and Programmes*; IPMA: Amsterdam, The Netherlands, 2016; ISBN 978-94-92338-17-4.
2. Mukilan, K.; Rameshbabu, C.; Velumani, P. A Modified Particle Swarm Optimization for Risk Assessment and Claim Management in Engineering Procurement Construction Projects. *Mater. Today Proc.* **2020**, *42*, 786–794. [[CrossRef](#)]
3. Berenguel-Felices, F.; Lara-Galera, A.; Guirao-Abad, B.; Galindo-Aires, R. Contracting Formulas for Large Engineering Projects. The Case of Desalination Plants. *Sustainability* **2020**, *12*, 219. [[CrossRef](#)]
4. Urgiles, P.; Sebastian, M.A.; Claver, J. Proposal and Application of a Methodology to Improve the Control and Monitoring of Complex Hydroelectric Power Station Construction Projects. *Appl. Sci.* **2020**, *10*, 7913. [[CrossRef](#)]
5. Tsolas, I.E. Benchmarking Engineering, Procurement and Construction (EPC) Power Plant Projects by Means of Series Two-Stage DEA. *Electricity* **2020**, *1*, 1. [[CrossRef](#)]
6. Eke, J.; Yusuf, A.; Giwa, A.; Sodiq, A. The Global Status of Desalination: An Assessment of Current Desalination Technologies, Plants and Capacity. *Desalination* **2020**, *495*, 114633. [[CrossRef](#)]
7. Compass International Global Construction Newsletter: July to September. 2019. Available online: <https://compassinternational.net/global-construction-newsletter-july-to-september-2019/> (accessed on 22 April 2020).
8. Wagner, R.F. EPC 4.0: The Quest for Reducing CAPEX in EPC Projects. *Organ. Technol. Manag. Constr. Int. J.* **2020**, *12*, 2245–2255. [[CrossRef](#)]
9. Losada-Maseda, J.J.; Castro-Santos, L.; Graña-López, M.Á.; García-Diez, A.I.; Filgueira-Vizoso, A. Analysis of Contracts to Build Energy Infrastructures to Optimize the OPEX. *Sustainability* **2020**, *12*, 7232. [[CrossRef](#)]
10. Konior, J.; Szóstak, M. The S-Curve as a Tool for Planning and Controlling of Construction Process-Case Study. *Appl. Sci.* **2020**, *10*, 2071. [[CrossRef](#)]
11. San Cristóbal, J.R.; Carral, L.; Diaz, E.; Fraguera, J.A.; Iglesias, G. Complexity and Project Management: A General Overview. *Complexity* **2018**, *2018*, 1–10. [[CrossRef](#)]
12. Wagner, R. EPC-Projects, a Class of Its Own. Available online: <https://www.ipma.world/epc-projects-class/> (accessed on 8 June 2019).
13. Habibi, M.; Kermanshachi, S.; Rouhanizadeh, B. Identifying and Measuring Engineering, Procurement, and Construction (EPC) Key Performance Indicators and Management Strategies. *Infrastructures* **2019**, *4*, 14. [[CrossRef](#)]

14. Moon, E. Delays in Global Engineering Procurement and Construction Projects: Main Factors in Project Management. Master's Thesis, Metropolia University of Applied Sciences, Helsinki, Finland, 2020.
15. Callegari, C.; Szklo, A.; Schaeffer, R. Cost Overruns and Delays in Energy Megaprojects: How big is big enough? *Energy Policy* **2018**, *114*, 211–220. [[CrossRef](#)]
16. Çevikbaş, M.; Işık, Z. An Overarching Review on Delay Analyses in Construction Projects. *Buildings* **2021**, *11*, 109. [[CrossRef](#)]
17. Rooholelm, V.; Sheikh Aboumasoudi, A. Designing Pattern for Project Delay Analysis with Several Periods Extension of Time, Continuous and Discrete (Pr-to-Pr). *Comput. Ind. Eng.* **2020**, *146*, 106576. [[CrossRef](#)]
18. Su, Y.; Lucko, G.; Thompson, R.C. Apportioning Contract Float with Voting Methods to Correlated Activities in Network Schedules to Protect Construction Projects from Delays. *Autom. Constr.* **2020**, *118*, 103263. [[CrossRef](#)]
19. Kaming, P.F.; Koesmargono, A.; Wahyu Aji, B. Delay Model for Engineering Procurement Construction (EPC): A Case of Liquefied Natural Gas (LNG) Projects in Indonesia. *MATEC Web Conf.* **2019**, *270*, 05010. [[CrossRef](#)]
20. Banobi, E.T.; Jung, W. Causes and Mitigation Strategies of Delay in Power Construction Projects: Gaps between Owners and Contractors in Successful and Unsuccessful Projects. *Sustainability* **2019**, *11*, 5973. [[CrossRef](#)]
21. Koulinas, G.K.; Xanthopoulos, A.S.; Tsilipiras, T.T.; Koulouriotis, D.E. Schedule Delay Risk Analysis in Construction Projects with a Simulation-Based Expert System. *Buildings* **2020**, *10*, 134. [[CrossRef](#)]
22. Abdellatif, H.; Alshibani, A. Major Factors Causing Delay in the Delivery of Manufacturing and Building Projects in Saudi Arabia. *Buildings* **2019**, *9*, 93. [[CrossRef](#)]
23. Yap, J.B.H.; Goay, P.L.; Woon, Y.B.; Skitmore, M. Revisiting Critical Delay Factors for Construction: Analysing Projects in Malaysia. *Alexandria Eng. J.* **2021**, *60*, 1717–1729. [[CrossRef](#)]
24. Pall, G.K.; Bridge, A.J.; Gray, J.; Skitmore, M. Causes of Delay in Power Transmission Projects: An Empirical Study. *Energies* **2019**, *13*, 17. [[CrossRef](#)]
25. Long, R.J. *Typical Problems Leading to Delays, Costs Overruns, and Claims on Process Plant and Offshore Oil & Gas Projects*; Long International: Littleton, MA, USA, 2018.
26. Ramli, M.Z.; Malek, M.A.; Hamid, B.; Roslin, N.T.; Roslan, M.E.M.; Norhisham, S.; Mohd, M.F. Influence of Project Type, Location and Area towards Construction Delay: A Review on Significance Level of Delay Factors. *Int. J. Eng. Technol.* **2018**, *7*, 392–399. [[CrossRef](#)]
27. Umar, A.A.; Rizeiqi, R.K.; Badr, A. Major Causes Assessment of Construction Delays. *J. Eng. Proj. Prod. Manag.* **2020**, *10*, 179–186. [[CrossRef](#)]
28. Yang, Y.; Tang, W.; Shen, W.; Wang, T. Enhancing Risk Management by Partnering in International EPC Projects: Perspective from Evolutionary Game in Chinese Construction Companies. *Sustainability* **2019**, *11*, 5332. [[CrossRef](#)]
29. Hai Nguyen, L. Empirical Analysis of a Management Function's Failures in Construction Project Delay. *J. Open Innov. Technol. Mark. Complex.* **2020**, *6*, 26. [[CrossRef](#)]
30. Issa, U.H.; Mosaad, S.A.A.; Salah Hassan, M. Evaluation and Selection of Construction Projects Based on Risk Analysis. *Structures* **2020**, *27*, 361–370. [[CrossRef](#)]
31. CoEPM2. *The PM2 Project Management Methodology Guide*; European Commission: Luxembourg, 2016; ISBN 9789279638725.
32. Kabirifar, K.; Mojtahedi, M. The Impact of Engineering, Procurement and Construction (EPC) Phases on Project Performance: A Case of Large-Scale Residential Construction Project. *Buildings* **2019**, *9*, 15. [[CrossRef](#)]
33. Jo, S.H.; Lee, E.B.; Pyo, K.Y. Integrating a Procurement Management Process into Critical Chain Project Management (CCPM): A Case-Study on Oil and Gas Projects, the Piping Process. *Sustainability* **2018**, *10*, 1817. [[CrossRef](#)]
34. Firnanto, A.; Adityatama, D.; Siregar, R.; Tanjung, E.; Purba, D. *Optimizing Project Management Software to Manage Well Planning: A Case Study of Deepwater Exploration Drilling Campaign in Indonesia*; IPA: Jakarta, Indonesia, 2020.
35. Gharaibeh, H. Evaluating Project Management Software Packages Using a Scoring Model—A Comparison between MS Project and Primavera. *J. Softw. Eng. Appl.* **2014**, *07*, 541–554. [[CrossRef](#)]
36. Kim, M.H.; Lee, E.B. A Forecast Model for the Level of Engineering Maturity Impact on Contractor's Procurement and Construction Costs for Offshore EPC Megaprojects. *Energies* **2019**, *12*, 2295. [[CrossRef](#)]
37. Muhamad, N.H.; Mohammad, M.F.; Ahmad, A.C.; Ibrahim, I.H. Delay Analysis Methodologies (DAMs) in Delivering Quality Projects: Contractors and Consultants' Perceptions. *Procedia. Soc. Behav. Sci.* **2016**, *222*, 121–131. [[CrossRef](#)]
38. Söderlund, M. Claim Management in EPC-Projects. In *A Case Study of Claim Causes and Claim Management during Plant Installation*; Uppsala Universitet: Uppsala, Sweden, 2018.
39. Arun, A. Statistical Methods for Construction Delay Analysis. *IOSR J. Mech. Civ. Eng.* **2013**, *9*, 58–62. [[CrossRef](#)]
40. Yi, D.; Lee, E.B.; Ahn, J. Onshore Oil and Gas Design Schedule Management Process through Time-Impact Simulations Analyses. *Sustainability* **2019**, *11*, 1613. [[CrossRef](#)]
41. Kim, M.H.; Lee, E.B.; Choi, H.S. A Forecast and Mitigation Model of Construction Performance by Assessing Detailed Engineering Maturity at Key Milestones for Offshore EPC Mega-Projects. *Sustainability* **2019**, *11*, 1256. [[CrossRef](#)]
42. Gong, P.; Zeng, N.; Ye, K.; König, M. An Empirical Study on the Acceptance of 4D BIM in EPC Projects in China. *Sustainability* **2019**, *11*, 1316. [[CrossRef](#)]
43. Brakstad, P.S. *Digitalization in EPC Projects, a Case Study of Aibel's Dogger Bank Project*; University of Stavanger: Stavanger, Norway, 2020.

44. Son, B.Y.; Lee, E.B. Using Text Mining to Estimate Schedule Delay Risk of 13 Offshore Oil and Gas EPC Case Studies during the Bidding Process. *Energies* **2019**, *12*, 1956. [[CrossRef](#)]
45. Lüdecke, D.; Ben-Shachar, M.; Patil, I.; Makowski, D. Extracting, Computing and Exploring the Parameters of Statistical Models Using R. *J. Open Source Softw.* **2020**, *5*, 2445. [[CrossRef](#)]
46. Fox, J. Using the R Commander: A Point-and-Click Interface for R. *J. Stat. Softw.* **2016**, *75*, 1–4. [[CrossRef](#)]
47. R-Project R: The R Project for Statistical Computing. Available online: <https://www.r-project.org/> (accessed on 21 April 2020).
48. Langhe, A.; Minde, P. Comparative Analysis of Three Major Type of Contracts with Case Study. *Int. Res. J. Eng. Technol.* **2018**, *5*, 237–240.
49. Pícha, J.; Tomek, A.; Löwitt, H. Application of EPC Contracts in International Power Projects. *Procedia Eng.* **2015**, *123*, 397–404. [[CrossRef](#)]
50. McNair, D. *EPC Contracts in the Oil and Gas Sector Sector*; PricewaterhouseCoopers: Melbourne, Australia, 2016; pp. 1–38.
51. Fox, J. The R Commander: A Basic-Statistics Graphical User Interface to R. *J. Stat. Softw.* **2005**, *14*, 1–42. [[CrossRef](#)]
52. Gandrud, C. *Reproducible Research with R and RStudio*; Chapman and Hall/CRC: Boca Raton, FL, USA, 2018. [[CrossRef](#)]
53. Agrawal, S. P Value and R Squared. Available online: <https://www.rpubs.com/shubh2565/pvalue-rsquared> (accessed on 12 February 2020).
54. Ahmed Marzouk, O. Case Studies of Statistical Analysis in Engineering. *Int. J. Stat. Distrib. Appl.* **2017**, *3*, 32. [[CrossRef](#)]
55. McLeod, S.A. What a P-Value Tells You About Statistical Significance. *Simply Psychol.* **2019**, *05*, 1–2.
56. Apeagee, B.; Haaor, S. A Logistic Regression Model of Road Traffic Fatalities in Benue State: Implication to Public Health. *Niger. Ann. PURE Appl. Sci.* **2020**, *3*, 46–52. [[CrossRef](#)]
57. Guthrie, W.F. NIST/SEMATECH E-Handbook of Statistical Methods. *Autom. Constr.* **2012**, *1*, 1243. [[CrossRef](#)]
58. Rossiter, D.G. *An Example of Statistical Data Analysis Using the R Environment for Statistical Computing*; University of Twente: Twente, The Netherlands, 2017.
59. Mair, P.; Wilcox, R. Robust Statistical Methods in R Using the WRS2 Package. *Behav. Res. Methods* **2020**, *52*, 464–488. [[CrossRef](#)]
60. Blanca, M.J.; Alarcón, R.; Arnau, J.; Bono, R.; Bendayan, R. Non-Normal Data: Is ANOVA Still a Valid Option? *Psicothema* **2017**, *29*, 552–557. [[CrossRef](#)] [[PubMed](#)]
61. Kim, H.-Y. Analysis of Variance (ANOVA) Comparing Means of More than Two Groups. *Restor. Dent. Endod.* **2014**, *39*, 74. [[CrossRef](#)]
62. ProjectTeam and Partners. *Innovation Project EPC 4.0 'Unleashing the Hidden Potential' Extract of the Final Report*; ProjectTeam and Partners: Glenwood Springs, CO, USA, 2019.
63. Candelario-Garrido, A.; García-Sanz-Calcedo, J.; Reyes Rodríguez, A.M. A Quantitative Analysis on the Feasibility of 4D Planning Graphic Systems versus Conventional Systems in Building Projects. *Sustain. Cities Soc.* **2017**, *35*, 378–384. [[CrossRef](#)]
64. Liu, N.; Kang, B.G.; Zheng, Y. Current Trend in Planning and Scheduling of Construction Project Using Artificial Intelligence. *IET Conf. Publ.* **2018**, *2018*, 1–6. [[CrossRef](#)]
65. Bilal, M.; Oyedele, L.O.; Qadir, J.; Munir, K.; Ajayi, S.O.; Akinade, O.O.; Owolabi, H.A.; Alaka, H.A.; Pasha, M. Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends. *Adv. Eng. Inform.* **2016**, *30*, 500–521. [[CrossRef](#)]
66. Gerbert, P.; Castagnino, S.; Rothballer, C.; Renz, A.; Filitz, R. *Digital in Engineering and Construction. The Transformative Power of Building Information Modeling*; The Boston Consulting Group Inc.: Boston, MA, USA, 2016; pp. 1–22.
67. Yaseen, Z.M.; Ali, Z.H.; Salih, S.Q.; Al-Ansari, N. Prediction of Risk Delay in Construction Projects Using a Hybrid Artificial Intelligence Model. *Sustainability* **2020**, *12*, 1514. [[CrossRef](#)]