Chapter 59

Application of multivariate statistical techniques in the analysis of the management of performance indicators of fixed broadband communication networks

10.56238/tfisdwv1-059

André Pedro Fernandes Neto
Federal Rural University of Semi-Árido
E-mail: andrepedro@ufersa.edu.br

Gutembergue Soares da Silva
Federal University of Rio Grande do Norte)
E-mail: guttembbergue@gmail.com.br

Fred Sizenando Rossiter Pinheiro
Federal University of Rio Grande do Norte
E-mail: fredrossiter@uol.com.br

Teófilo Câmara Mattozo
Federal University of Rio Grande do Norte
E-mail: mattozo@oi.com.br

José Alfredo Ferreira Costa
Federal University of Rio Grande do Norte
E-mail: jafcosta@gmail.com

ABSTRACT

The demand for access to telecommunications services is always growing, especially those requiring broadband infrastructure, in which there is competition between networks and services platforms, contributing to leverage transformations in the economy, society and human life, resulting from numerous technological innovations that allowed the proliferation services such as internet, mobile communications and high-speed fixed and mobile connections. As a result of the great competition in this market, companies seek to find different ways to get new customers, in addition to not losing their base to competition. In the 2000s, telecommunications companies began a process of commercializing broadband internet with a product that added value to fixed telephony, the Asymmetric Digital Subscriber Line (ADLS). This study applied methods of selection of variables, through techniques of multivariate analysis of data with the purpose of evaluating the relationship of the quality of service indicators with the dropout rate (Churn). The results obtained allow us to suggest the applicability of quality indicators in maintenance management, as well as the possibility of obtaining real economic gains for telecommunication service providers with the use of these analytical methods in their operational asset management strategies.

Keywords: Broadband, Performance Measurement, Quality of Service, Abandon Rate (Churn), Multivariate Data Analysis, Data Mining

1 INTRODUCTION

A leading and revolutionary technological resource of post-industrial society is the Internet, the worldwide computer network. Pereira and Biondi (2013) state that the Internet has definitely changed the way information circulates, economic relations, radically altering the functioning of the communication market and the way people relate and communicate. Currently, Internet access is mainly due to broadband (BL) and mobile telephony (TM) services, becoming a fundamental mechanism for the dynamics of contemporary life. The number of subscriptions of fixed and mobile GLA services in Brazil has shown constant growth, especially in worldwide harmony, for the predominance of mobile BL (ANATEL, 2020). The National Telecommunications Agency (Anatel) designates the fixed BL service as multimedia communication service (SCM) and the TM service of Brazil as Personal Mobile Service (SMP) (NANIWA R. J, 2020).

With surprising growth, broadband is no longer the predominantly open form of internet access for institutions and a niche of the privileged, and is a common good for society, with telecommunications companies starting in the 2000s a process of marketing broadband internet based on a value-added product of fixed telephony, ***Asymmetric Digital Subscriber Line*** (ADLS) and later mobile communications. In the competitive scenario of telecommunications operators, competition and increased costs, associated with attracting new customers, impused a challenge to retain increasingly demanding customers, causing companies to change their tactical and strategic behavior employed in the practice of managing the dropout rate (*Churn*) (FERNANDES, 2007).

Faced with fierce competition that organizations face in providing services, with increasingly globalized businesses, the search for maximum efficiency for their assets is a fundamental priority for their managers. The planned maintenance of technological plants, within the concept of asset management, is certainly the most relevant foundation in the search for this efficiency in all technical activities of preventive and predictive maintenance. Strategic asset management not only includes maintenance activities, but involves the organization's results, including the various business indicators, integrating the areas of marketing share, billing, costs, profit, operational and people safety, among others (KARDEC and NASCIF, 2015).

The relationship between performance indicators, quality and maintenance management is relevant for the current business environment, including in this context the companies providing services, since the formulation of strategies, operational planning, actions achievements, programs and improvement processes, in addition to changes in their operations, begin in the elaboration of metrics that can monitor their stages of development (ARNETT, 2000).

The evaluation of abandonment rates in relation to the methods associated with the management of service maintenance can show a great challenge, because if on the one hand levels of satisfaction should be guaranteed that lead to customer loyalty, on the other hand, time and money are required to be spent in the search for new customers. This research was directed to compare two methods of multivariate analysis, with the objective of providing decision-making support in the management of the operational areas of the service providers, through the analysis of performance of the quality of services versus the dropout rate. Thus, the broadband data stored by a telecommunications company were analyzed, in order to find a relationship between the data related to quality of service and the dropout rate (FERNANDES, 2007).

One of the main objectives in the application of multivariate techniques is the expansion of the researcher's explanatory ability with statistical efficiency and the possibility of addressing a range of theoretical and management issues (HAIR, 2009). In the present study, considering the characteristic nature of the existence of multiple dependent variables and interrelated questions, multiple linear regression techniques and factor analysis were used. The mutivariated statistical tools not only work with a diversity of variables as a fundamental factor in the decision-making process, but also offer a wide range of options due to the muliplicity of variables, including metric as and non-metrics (RIBAS and VIEIRA, 2011). To perform the tests and analyses, the SPSS (Statistical Package for Social Sciences) tool was widely used in the applications of analysis of research in the area of management and business (WAGNER *et al*., 2004).

The article is divided as follows. Section 2 is dedicated theoretical review addressing aspects of service delivery, maintenance management, performance indicators and the use of analysis and data techniques and multivariate modeling. In Section 3, the method of preparation of the work and the statistical resources for data modeling are presented. In section 4, the results obtained are explored and discussed. Finally, Section 5 brings the main conclusions of the work and the expectation of future work.

2 THEORETICAL REFROND

The transport infrastructure of a telecommunications network is segmented into access, aggregation, and switching tier. The function of the on-call layer is to physically connect customer equipment to carrier equipment. The physical medium (copper or fiber optic pairs) of access can be shared by multiple clients or reserved for a single customer. The aggregation layer logically connects customer equipment to network equipment, called service equipment, (which has logical interfaces dedicated to each client that route flows originating or ending in customer accesses. Switching is the layer that responds to the routings of flows originating or ending in aware accesses (EUSÉBIO, 2010).

In public switched telecommunications networks (PSTN) each customer uses a copper pair and port of the remote unit of their carrier's access equipment. For xDSL-based services, customers use the same pair of wires, which holds multiple FDM streams for voice and other services. Each of these client streams can include one or more packet streams (usually one per service, for example, internet access and one for IPTV (EUSÉBIO, 2010).

The approach used here for management indicators considers that they are defined as an integral part of the business management model, thus being structured as a relationship between two variables, in the form of numerator and denominator, in which their attributes and values are feasible measurement. According to Tachizawa *et al.* (2003), conceptually it can be said that a management model depends on measurements, information and analyses. The measurements need to be due to the organization's strategy, covering the main processes of the provision of services, as well as their results. The information needed to evaluate and improve performance includes, but is not, customer-related, product performance, operations, market, competitive and benchmarking, suppliers, employees and financial and cost aspects.

Quality management indicators are used to select objectives and metrics related to measuring the effectiveness of an organization, so that important decisions for the organization are made in a scientific way. Frequently represent a mathematical relationship in which it measures attributes of a process or business results, with the objective of comparing the values arising from real events with the pre-established goals. The descriptive characteristics of management indicators can be defined as a mathematical relationship that results in quantitative measure, identifying a state of the process or result of this and associating with pre-established numerical goals (FERNANDES, 2007).

In the telecommunications sector where competition and customer search refers to a survival issue, maintenance activity is considered strategic for these organizations. The absence of this view results in losses and reduction of profit, affecting customers, employees, investors and society. The c*hurn rate*, also known as the abandonment rate, is the metric that indicates how much a company lost customers in a given period usually in favor of a competitor. This is a very important rate for companies and should be monitored regularly to check the health of a business (IKEDA, 2006). The term in a free translation for Portuguese, when related to the environment of telecommunications companies, causes exactly what the verb means: a great "agitation" of customers in the market, causing business exchange at all times, which, in turn, lead companies to "mobilize " in search of new ways to keep their customers in their business, while seeking to seduce customers of the competition (FERREIRA, 2005).

The following factors of influence are positively correlated with the maintenance of customers, including Venetis and Ghauri (2004) and Ganesh *et al*. (2000) highlight: the quality of services provided to customers; their satisfaction and loyalty, in addition to the treatment given by employees to customers. To Neslin *et al*. (2006), one way to *manage churn is* to predict which customers are most likely to abandon the relationship with the company and work with them in order to try to avoid the occurrence of this disruption. For this, it is necessary that the company is able to perform this type of forecasting and identification, seeking to have customers who are likely to turn off, considering which of these are the ones that, in fact, generate sufficient value for the company and justify the investment in retention actions.

In recent years the activities of maintenance of services, has undergone many changes that included management as an indispensable factor to achieve the best results for maintenance and for the company as a whole. The current concept of the maintenance mission involves ensuring the availability of equipment and facilities in order to meet a production process or service with reliability, safety, environmental preservation and adequate cost (KARDEC and NASCIF, 2015).

There is a great diversity of names of the forms of maintenance, being the main types explained in NRB 5462 (ABNT, 1994), as corrective, preventive or predictive maintenance. The various types of maintenance can also be considered as maintenance policies or strategies, provided that their application is the result of a management performance of the organization based on technical-economic data (KARDEC and NASCIF, 2015). In Fernandes (2007) a classification is presented according to the type of intervention that is made in the service or installation. Thus, corrective, preventive maintenance based on time, preventive condition-based or predictive and improvement maintenance are identified, with its definitions described below:

- Corrective maintenance - intervention resulting from a failure, breakage or malfunction. A service requires corrective maintenance when it is necessary to intervene in it because it has interrupted or degraded its function;
- Time-based preventive maintenance - intervention at regular intervals of running time (e.g. weeks) or operating time (e.g. hours worked). In English known as TBM (time based maintenance);

- Preventive maintenance based on condition or predictive - intervention made according to the monitoring of certain service parameters (e.g., wear measurement or high degree of degradation in the service). In English known as CBM (condition based maintenance);
- Maintenance improvement - intervention made to change the conditions of a service in order to increase its performance, the quality of the processed products or improve some operational parameter.

In order to intervene less and less in the plant, predictive maintenance practices and monitoring of the condition of process equipment are increasingly used. The same is true of unplanned corrective maintenance, which becomes an indicator of maintenance ineffectiveness. The state of the art in maintenance by world-class companies privileges the interaction between the areas of engineering, maintenance and operation as a guarantee factor of performance goals for management indicators. In this context, the focus on business results, the main reason for achieving the necessary competitiveness of the company's survival, is obtained through joint efforts integrated by the asset management system (KARDEC and NASCIF, 2015).

The quality indicators and the measures that are part of the universe of quality management, enable the incessant improvement of service providers, who are at all times urged to change their systematics and procedures in an attempt to meet the desired customer satisfaction (FITZSIMMONS, 2004). Service management requires telecommunications operators to measure end-to-end service performance and assess network occurrences, including the overall quality of each service and business impact on a particular customer pool.

For Nasser (2022), there are basically two basic types of solutions aimed at Service Management: Failure Management with Vision of Services, which utilize the artifice of service trees to support the analysis of impact on the service, as well as the analysis of the root cause of failures and the management of Quality of Services, that uses performance and quality indicators, collected directly from the network or via interfaces with other operating systems. The two approaches are complementary, the second being more comprehensive, contemplating network failures that affect services and that can be translated into availability indicators. Table 2 describes below the broadband internet indicators used in telecommunications companies, which in this work are associated with the characterization of dependent and independent variables.

Table 2: Characterization of dependent and independent variables

Indicador	Formula dos Indicadores	Valores	Tipo de Variável
TAXA PREVENTIVA	∑ dos Reparos Preventivos / ∑ dos Reparos Corretivos do Mês Anterior	%	Independente
TAXA DE ABANDONO	Quantidade de Retiradas X 100 / Base em Serviço do Mês Anterior	%	Dependente
IGT	Reparos Aberto nos últimos 30 dias X 100 / Total de Instalação Realizadas nos Últimos 30 dias	%	Independente
TEMPO MEDIO REPARO	∑ Tempos de Reparo no mês / Base de Terminais Reclamados no mês	%	Independente
IPGC	Instal. Executados em até 3 dias corridas x 100 / Total de Instalação no Período	%	Independente
RPAV	Reparos Executados no Prazo (4 h) X 100 / Total de Reparos no Mês	%	Independente
RPDC	Reparos Executados no Prazo (8 h) X 100 / Total de Reparos no Mês	%	Independente
RPT	Reparos Executados no Prazo X 100 / Total de Reparos no Mês	%	Independente
RRAV	Reparos Reincidente < 90 dias Alto V. X 100. / Total de Reparos no Mês	%	Independente
RRDC	Reparos Reincidente < 90 dias D. C. X 100. / Total de Reparos no Mês	%	Independente
RRT	Reparos Reincidente < 90 dias X 100 / Total de Reparos no Mês	%	Independente
TAXA DE REPARO	Quantidade Reparos em 30 dias X 100 / Base em Serviço Mês	%	Independente
TEMPO DE INSTALAÇAO	∑ Tempos de Instalação no mês / Base de Terminais instalados no mês	%	Independente

Source: Prepared by the authors

Translation:
Indicator
indicator formula
Values
variable type
Independent

The advancement of information technologies (IT) as well as their use in organizations provided storage and access to large volumes of data, creating an excellent opportunity to obtain useful knowledge. These procedures highlight the need for a systematic approach focused on the data preparation process, which allows to increase confidence in the results of data analysis (COSTA *et al.*, 2014). The process of knowledge discovery in the database can be divided into three stages: Preprocessing, Data Mining and Post-Processing (GOLDSCHMIDT *et al.*, 2005). The pre-processing stage comprises the functions related to data capture, organization and processing. In this step, the data are prepared for the subsequent mining stage where the effective search for useful knowledge is carried out in the context of the research. The post-processing stage covers the treatment of knowledge obtained in data mining (COSTA *et al.*, 2014).

There is a growing need for the use of analysis and data techniques and multivariate modeling due to aspects such as the improvement of research and data collection techniques, which enable the generation of databases with sample relevance and the development of computational packages that allow the inclusion of a large amount of data (observations and variables) that lead to the development of models quickly and accurately (FÁVERO *et al.*, 2009)

The use of appropriate modeling for decision analysis has provided organizations that employ them as management tools to obtain fast and accurate results, which provide valuable information for determining new investments, for trend-making and for investigating phenomena or little-known aspects of their business. In this context, multivariate analysis has increasingly been important for decision-making, in the most varied fields of knowledge and with the increasing development of computational resources, has become more frequently used to evaluate various behaviors and trends in definitions of management strategies and marketing models in companies (FÁVERO *et al*, 2009).

The multivariate analysis techniques used here, multiple regression and factor analysis were applied to analyze and identify variables or factors related to maintenance management that had the most relevant characteristics. Multiple regression is the appropriate method when the problem involves a single dependent metric variable related to two or more variables is independent(HAIR, 2009). Factor analysis is a statistical approach used to allow the analysis of possible intre-ralations between a large number of variables in order to condense the information contained in each group of original variables into a new set of variables (factors) without loss of relevant information (VICINI and SOUZA, 2005).

3 METHOD AND MODELING

As a method applied in this work, the steps guided by Hair *et al were followed.* (2009), according to which six stages are necessary for multiple regression analysis, as well as for factor analysis. The Tabela 1 presents below, displays a summarized analysis between the two techniques, contemplating the stages mentioned, with the objective of showing the specificities of each structure. After the initial step follows the stage of data collection, which is carried out through data analysis secondary to the collection of them in a customer treatment system of a telecommunications company, sendo carefully observed the reliability of them. The statistical software chosen was *SPSS* 19 (Wagner *et al.* , 2004), due to its flexibility in data analysis and knowledge of the authors in previous research.

The sample size influences the appropriate choice of model and statistical power in the varied analysis. Small samples are not indicated for multiple analysis, as well as very large samples (more than 1000 observations), make statistical significance tests sensitive, often indicating that any relationship is statistically significant. With these samples, it should be ensured that the criterion of practical significance is met along with statistical significance. Power is directly influenced and in magnitude by sample size (MALHOTRA, 2011).

Table 1 Stages in the multiple regression decision diagram.

Multiple Regression	**Factor Analysis**
Define the problem of research and the objectives of analysis in conceptual terms.	Define a way to condense the information contained in several original variables into a smaller set of new composite dimensions or statistical variables.
Develop an analysis plan that addresses the particular issues for your purpose and project, as well as the appropriate sample size.	Develop an analysis plan that addresses the particular issues for your purpose and project, as well as the appropriate sample size.
Develop assumptions about the relationships between dependent and independent variables that affect the statistical procedure (quadratic minimums).	Develop assumptions about the interdependent relationships of variables;
Building the model now proceeds to the estimation of the model to be researched and the evaluation of the overall fit of the model.	Extract the selected factors to represent the latent structure of data.
Examine the predictive equation, and thereby evaluate the relative importance of individual variables in the overall product forecast.	Interpret the factorial solution;
Generalize the application of the model.	Validate the applicability of the model.

Source: Adapted by authors, (2009).

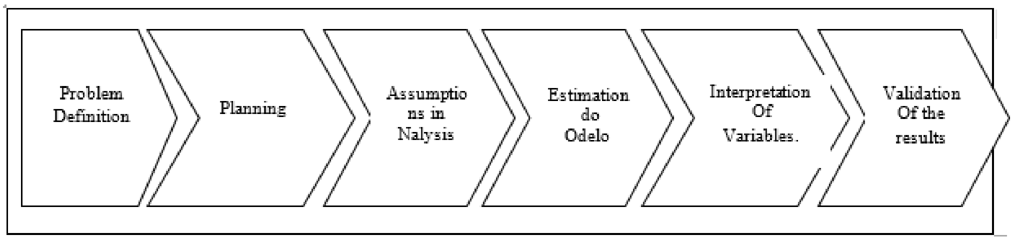
The sample size influences the appropriate choice of model and statistical power in the varied analysis. Small samples are not indicated for multiple analysis, as well as very large samples (more than 1000 observations), make statistical significance tests sensitive, often indicating that any relationship is statistically significant. With these samples, it should be ensured that the criterion of practical significance is met along with statistical significance. Power is directly influenced and in magnitude by sample size (MALHOTRA, 2011).

In this research, the measurement of the direct impact and magnitude of the power of the influence that the sample size may exert on the research was measured, i.e., the probability of being evaluated as statistically significant a deterinated level of R^2. For this was collected a total of 336 observations. In the stage of assumptios about relationships, a study is initially carried out with the multiple regression model through a sequoia search approach, known as *Stepwise* (MINGOTI, 2007). Next, the factor analysis method was used to confirm the regression equation with a set of factors and then performed the extraction of the factors in the order of their greatest importance. In the *Stepwise* method (Ribas and Vieira, 2011), in each stage, with the independent variable still absent from the equation, these variables are incorporated into the equation if they have low *values of the F* statistic (LATTIN *et al.*, 2011). The independent variables present in the regression equation are removed if their *probabilities of F become sufficiently* large, the procedure terminates when there is no longer variable eligible for inclusion or removal. The tolerance level used was 0.05.

In factor analysis, the significance test used is *bartlett's* measure lessness test (Manly, 2008) which is based on the statistical distribution of "*Chi-Square*" (Marôco, 2020) in the hypothesis (null H0), that the correlation matrix is an identity matrix (whose diagonal is 1.0 and all others equal to zero), meaning that there is no correlation between the variables (*FÁVERO et al.* , 2001). Significance values greater than 0.100 indicate that the data are not adequate for treatment with the method in question; thus the null hypothesis cannot be rejected. On the other hand, values lower than indicated allow them to reject the null hypothesis (SPSS, 2010;   HAIR *et al.*, 2009)

Next, Figure 1 explains the sequence of stages that were performed for the two multivariate analyses.

Figure 1 - The number of factors in the extraction.



Source: Adapted by Goldsimith's author es, (2005).

3.1 MODELING DATA WITH MULTIVARIATE REGRESSION ANALYSIS

Regression is the term used to designate a mathematical equation that describes the relationships between two or more variables (LARSON, 2010). Linear regression is a method used to estimate the expected value of a *Variable Y (dependent* variable), given the values of some other X variables (independent variables). Thus, given two data matrices, *X and Y*, the purpose of regression is to construct a model *Y* = *f(X)*. This model tries to explain, or predict, *the y variations* given the multivariate regression variations takes into account the various predictive variables simultaneously, modeling the dependent variable more accurately. In this work, the dependent variable is the effective sales and the group of independent variables are the indicators of sales performance. The regression model is represented by em *X*. Aequation 1 (LISBOA *et al*., 2012).

$$Y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_p x_{ip} + \varepsilon_i \,. \tag{1}$$

Where Y_i - represents the dependent variable, x_{ik} ($i = 1,\ldots, n$) are the independent variables ($k = 1, 2,\ldots, p$); β i's are the regression coefficients (unknown parameters in the model – to be estimated); ε_i is the residual, random variable that captures the behavior portion of *variable* Y_i not explained by the regression equation. The parameters of a regression model can be estimated in several ways (MANLY, 2008):

1) Least squares, minimizing the average quadratic error of waste;
2) Maximum likelihood;
3) Bayesian methods;
4) Minimizing absolute deviation.

The methods of equation 2-a and 2-b coincide for a model with the normally distributed errors. Estimates of the least squares used in this work are given by (LAROSE, 2006).

$$\hat{\beta} = \frac{\sum (x_i - \bar{x})(y_i - \bar{y})}{\sum (x_i - \bar{x})^2} . \qquad \text{(2-a)}$$

$$\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} . \qquad \text{(2-b)}$$

The least squares estimator, in matrix form, is given *by* β= (*X'X).*$^{-1}$ (*X'Y.*), where the apostrophe means transposed. According to Hair (2009), each observation has its own residue, which together produce sum of quadratic errors, a total measure of estimation errors. Three quadratic sums (SSE, quadratic sum of errors; SSR, the sum of the regression squares; SST, the total sum of squares) can be calculated as follows:

$$SSE = \sum (y - \hat{y})^2 . \qquad \text{(3-a)}$$

$$SSR = \sum (\hat{y} - \bar{y})^2 . \qquad \text{(3-b)}$$

$$SST = \sum (y - \bar{y})^2 . \qquad \text{(3-c)}$$

Regression statistics can be briefly presented using variance analysis tables - ANOVA (GOLDSCHMIDT, 2005). Average errors (e.g. MSE and MSR) are derived from equation 4. An important parameter is the coefficient of multiple determination, defined in Lisbon *et al.*, (2012), as:

$$R^2 = \frac{SSR}{SST} . \qquad \text{(4)}$$

For multiple regression, R^2 *is interpreted* as the proportion of variability in the target variable that is clarified in the linear relationship with the set of variables.

3.2 MODELING DATA WITH FACTOR ANALYSIS

Factor analysis is a generic name given to a class of multivariate methods whose main purpose is to define the adjacent structure in a data matrix, being addressed by the problem of analyzing the structure of interrelationships between a diverse cast of variables, thus defining a set of common latent dimensions and called factors (VIEIRA and RIBAS, 2011).

a) Typology

- **Exploratory Analysis:** Useful in searching for the structure in a set of variables or as a reduction method. In this perspective, we consider the information that the data offer and do not establish a priori restrictions on the estimation of the components or on the number of components to be extracted, promoting the search for structured sets of variables and the reduction of data.

- **Confirmatory Analysis:** Tests the hypotheses involving questions about, for example, which variables should be grouped into one factor or the exact number of factors, that is, it evaluates the degree to which the data meet the expected structure.

a) Objectives of Factor Analysis

Mingoti (2007), states that the main objective of a factor analysis is to describe the original variability of the dependent variable, condensing the information contained in a number of original variables, in a smaller set of factors with a minimum of loss of this information, aiming at:

- Observe a set of latent dimensions in a large set of variables - Factor analysis of type R.
- Combine or condense, a list of observations in groups - Factor analysis of type Q.
- Identify appropriate variables for further regression, correlation, or discriminant analysis.
- Generate a new set of new variables in smaller numbers, to replace another set.

The variables used in Factor Analysis are usually metric. Observations should be at least 50 cases and preferably 1 00 or more. However, dichotomous variables, although considered non-metric, can be used, as if all variables are dichotomous, more specific forms of factor analysis are indicated, such as Boolean factor analysis (VIEIRA and RIBAS, 2011).

b) Determination of Factors and Evaluation of The General Adjustment

These factors can be extracted as orthogonal or oblique. As orthogonals, they will be independent of each other and how oblique they will be correlated or dependent - which is controversial and debatable. Orthogonal factors represent a reduction of information, being good for regressions or analysis of discriminant, but may not have real meaning, after being examined the matrix of factors without rotation, and can explore the possibilities of data reduction and obtain a preliminary estimate of the number of factors to be extracted (MALHOTRA, 2011).

If the purpose of the analysis is to identify important variables for later use, the expert should examine the data matrix and select the variable with the highest factorial weight as representative of a particular dimension. However, if the objective is to create an entirely new set, with a smaller number of variables, then factor scores should be calculated and used as raw data in later analyses. In the factorial analysis, the common variance is included in the diagonal of the same matrix, before extracting the factors

(FERNANDES, 2007). At this stage, two different analyses were performed, principal component analysis and that of common factors.

To improve the analysis , one can use the device of the rotation of factors to rotate the reference axes of the factors, around the origin, until reaching an ideal position. It can be orthogonal or oblique if the axes remain 90 degrees between each other during turning or not. The objective is to facilitate the reading of the factors, because rotation leaves high factor weights in one factor and low in others, defining more clearly the groups of variables that are part of a factor studied. The oblique rotation is more realistic, but more controversial. The most used rotation method is *VARIMAX*, which simplifies the columns of the factor matrix (HAIR, 2009).

4 RESULTS AND DISCUSSIONS

Considering the confirmation requirement of the regression equation proposed through factor analysis, there is a need to specify and select the variables to be analyzed. Multiple regression and orthogonal factor analysis were selected as the multivariate techniques to be used, because they provide a means of objectively evaluating the degree and character of the relationship between dependent and independent variables. In this article, the SPSS tool was used for its ease of use, its vast literature and the current in multiple regression analysis, as well as factor analysis.

4.1 ASSUMPTIONS IN MULTIPLE LINEAR REGRESSION ANALYSIS

By analyzing the correlation matrix between the variables of the model, it can be verified that, among the independent variables, there are no values higher than 0.5, according to data obtained from *the SPSS* (NORUSIS, 2004). That is, there is no value, it can be completed by the absence of multilinearity. $r_{X_i X_j} \geq 0{,}5$

The following is an analysis of the dispersion plots for multiple regression and *THE SCREE TEST* for factor analysis. This test consists of separating trivial factors from the onset of nontrivial factors through a visual inspection of the graph. Subjective decision is based on the use of a straight line placed along the bottom part of the chart where the points form an approximately straight line. The points above the straight line are associated with nontrivial factors, while the remaining points represent the trivial factors (FÁVERO *et al.*, 2009). In *THE SCREE TEST* presented in Figure 2, it is identified that the ei values versus the number of factors in the order of extraction can determine the cutoff point, by the existence of a consistent distribution of the residues in relation to the theoretical distribution, that is, it follows approximately the shape of the normal curve. Figure 3 shows a uniform distribution trend, i.e. the points are close to a line.

Figure 2 - The number of factors in the extraction.
Source: Factor Analysis SPSS Report.

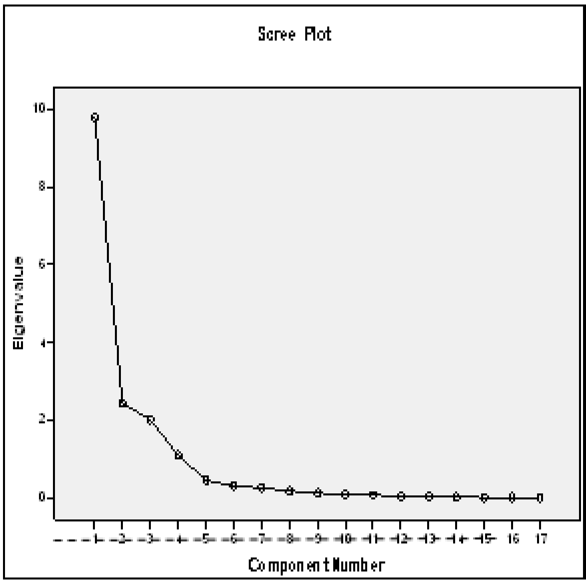
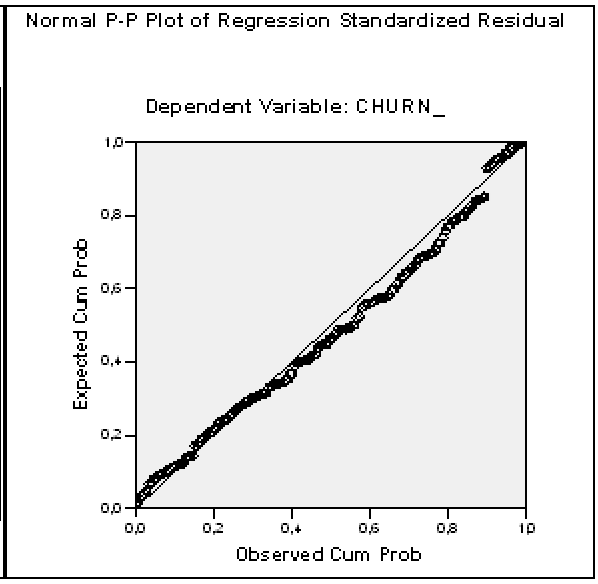


Figure 3- Normal probability PP-plot.
Source: SPSS Report.



4.2 THE EQUATION ESTIMATED BY THE MULTIPLE LINEAR REGRESSION METHOD

The main objective of regression models is to estimation the unknown parameters β. This process is common to call "model tuning to data". The coefficients of *the stepwise* model contain the estimates of the parameters and their estimates of the standard error, the estimates of the standardized coefficients and *the value of the t-Student statistic* (WAGNER *et al.* , 2004). Using the SPSS data (model coefficients), the regression equation presented below can be characterized:

$$\text{Dropout Rate} = 0.53(\text{TMI}) + 0.41\,(\text{TRP}) + 0.60(\text{TMR}) - 0.39(\text{QRPR}) \quad (7)$$

Where:
TMI – Represents the average time spent in the installation of the circuit, that is, it is the time from opening the service order to its effective closing.
TRP – Represents the repair rate performed within the agreed time with the customer, ranging in an interval of 8 hours for High Value customers and 24 hours for other customers (usually retail customers).
QRPR - This value represents the amount of preventive repairs performed in companies.
TMR – Represents the average time spent on circuit repair, that is, it is the time involved from opening the Defect Ticket to its effective closing.

4.3 INTERPRETATION THROUGH FACTOR ANALYSIS

In the interpretation of the values and in the selection of the final factor, the unrotated matrix should initially be analyzed in order to obtain a preliminary identification of the number of values to be extracted. By computing the non-rotational factor matrix, the analysis is performed only in the best particular combination of the original variables, the first factor being the best summary of linear relationships (FERNANDES, 2007). The following is a rotational method to simplify factorial and more significant solutions, in these cases orthogonal rotation using VARIMAX of factors improves interpretation, reducing ambiguities. Finally, the need to respecify the factorial model should be evaluated.

In the analysis of the correlation actress shown in table 3, it can be concluded that the factor 1 tin high coefficients s for the variables V1 (Time Of Repair), V3 (Average installation time) and a negative coefficient for V5 (Amount of Repairs Preventive). Thus, it is possible to identify with prominence, greater relevance for these three variables, and the first two account for more than 80% of the variance, and the gain obtained to the addition of the third corresponds to approximately only 10%..

Table 3: Rotated Component Matrix

	Variáveis															
	V1	V3	V5	V2	V6	V7	V8	V10	V12	V9	V11	V13	V14	V16	V17	V15
Fator 1	.803	.835	-.903	-.067	.075	-0,027	0,061	-0,053	-0,06	-0,044	0,095	0,096	0,093	0,095	0,096	0,01
Fator 2	-.014	-.164	-.119	.873	.827	0,358	0,726	-0,011	0,238	0,395	-0,191	-0,129	-0,158	-0,153	-0,122	-0,151

Extraction Method: Main Component Analysis. - Rotation Method: Varimax with Kaiser Normalization. a Rotation converged in 3 iterations. Source: Factor Analysis SPSS Report.

4.4 THE COMPARISON BETWEEN SIGNIFICANCE OF MODELS

Through factor analysis, three variables identified as the most relevant variables were selected: V1 (Mean Repair Time), V3 (Average Installation Time) and V5 (Number of Preventive Repairs).

On the other hand, in the multiple regression analysis, four variables were taught: Mean installation time (IMT), Repair rate (TRP), Mean repair time (MrT), Preventive Repair (QRPR). The following table 4 presents a comparison between the two methods:

Table 4: Comparative table factor analysis x multiple regression

Multiple Regression	**Factor Analysis**
4 Independent variables	1 Factor with 3 variables
In the statistical significant test, in the observed sample we *have:* F_{obs} = 43,367 which is higher *than* F_c=$F_{(95\% ; 4 ; 163)}$ = 2.40 (percentage value of an *F distribution with* 4 degrees of freedom in the numerator and 163 in the denominator) with this validates the alternative hypothesis that the regression is statistically significant.	In the Bartlett test, in the observed sample we have: A value lower than 0.0001, which allows confirming the possibility and adequacy of the factor analysis method for data processing.
Multivariate method of analysis that uses dependency and is measured as metric variables.	Multivariate method of analysis that uses independence and has in its structure the list of variables.
Final result: an equation with four independent variables and one dependent	Final result: three most relevant variables.
Average installation time (TMI) Average repair time (TMR) Preventive Repair (QRPR) Repair Rate (TRP)	V3 (Average Installation Time) V1 (Average Repair Time) V5 (Number of Preventive Repairs) -

Source: Prepared by the author

It is observed that the sequence of relevance of the variables is similar in the two techniques used, even considering that in factor analysis, the variables have relations of independence, differently from multiple regression, where the relationship is from one dependent variable to two or more independent. It is verified that both techniques can offer similar interpretations, since three of the four variables highlighted as the most relevant are identified and observed with this characterization in both techniques.

5 CONCLUSIONS

In this research, it was identified that control, analysis and maintenance management are critical factors for companies in the telecommunications sector in the challenge for the reduction of *churn*. In order to compete in this market, service operators have thoroughly taken action to retain valuable customers.

The techniques of multivariate analysis, multiple regression and factor analysis, which were applied to analyze and identify variables or factors related to maintenance management, which present characteristics of greater influence on business results, were relevant and effective. After completing the phases of inadequate data processing and cleaning in the creation of the models, it was demonstrated that the use of data analysis through multivariate statistical techniques can help telecommunications service operators to make efficient *churn management and control*, through some maintenance indicators. Therefore, it is evident the usefulness of the application of data mining techniques jointly or individually, in the management of maintenance management problems, and regression can be used in the development of correlations between variables, while factorial analysis can be used to condense the information.

Regarding the indicators, the quality of the services provided was identified as one of the highlight factors for *churn analysis*, both in the method using multiple regression and in the procedure using factor analysis. Thus, it was possible to observe in the multiple regression that 51.4% of abandonment can be correlated with the behavior of four quality indicators, namely: the Average Repair Time, Average Installation Time, Preventive Repairs, Repair Rate. Regarding factor analysis, the index obtained with this solution shows that 90.3% of the total variance is represented by information contained in the factorial matrix, belonging to three variables V1 (Mean Repair Time) and V3 (Average Installation Time) and a negative coefficient V5 (Amount of Preventive Repairs).

From these analyses, the importance of the indicators Mean Repair Time, Average Installation Time and Preventive Repairs are confirmed, as variables were identified as relevant in the two analysis techniques. Finally, the results obtained demonstrate the applicability of the indicators, enabling gains for service providers who decide to use these methods in their maintenance strategies. For this, it is essential that a task of provision of installation or maintenance of repair is carried out in a shorter time and that the actions of preventive provision is successful, as they positively affect customer satisfaction contributing to remain in the company.

In order to obtain a more comprehensive perspective of the performance of telecommunications companies, we hope to conduct future research by applying the same techniques of multivariate analysis to the services of the mobile broadband (TM) segment, involving indicators related to the performance of services.

REFERENCES

ANATEL **Agência Nacional de Telecomunicações**. Relatóriode Banda Larga Fixa. 2020.

ARNETT, D. B; MENON, A., WILCOX, J. B. “Using Competitive Intelligence: Antecedents and Consequences”, Competitive Intelligence Review, Vol. 11(3), 2000.

COSTA J.A.F., MATTOZO T. C., SILVA G.S., FERNANDES N., A. P., **Uma Introdução à Mineração de Dados: Conceitos e Aplicações** Ed. EdUFERSA.2014

EUSÉBIO F., Redes de Telecomunicações - Uma visão funcional e o motor do seu desenvolvimento. Ed. SÍLABO. Lisboa 2010

FÁVERO, L.P.L., BELFIORE, P.P., SILVA, F.L., CHAN, B. .L., **Análise de Dados: modelagem multivariada para tomada de decisões.** Ed. Elsevier. Rio de Janeiro.2009

FERNANDES, A. P. N., Análise dos Indicadores de Qualidade versus Taxa de Abandono Utilizando Método de Regressão Múltipla para Serviço de Banda Larga (Dissertação de Mestrado). PPgEP da UFRN. 2007.

FERREIRA, J. B., **Mineração de Dados na Retenção de Clientes em Telefonia Celular**. Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro – PUC. 2005.

FITZSIMMONS, J. A.; MONA, J., **Administração de Serviços** – Operações, Estratégias, e Tecnologia da Informação. 4°edição, São Paulo: Bookman, 2004.

GANESH, J. ARNOLD, M. J.; REYNOLDS, K. E. Understanding the Customer Base of Service Providers: Na Examination of the Differences Between Switchers and Stayers. **Journal of Marketing**, Vol. 64, 2000.

GOLDSCHMIDT, R.; PASSOS, E. **Data Mining** – Um Guia Prático, Conceitos, Técnicas, Ferramentas, Orientações e Aplicações. São Paulo: Campus, 2005.

HAIR, J.; ANDERSIN, R. E.; TATHAM, R. L. **Análise Multivariada de Dados.** 6º ed. São Paulo: Bookman,Vol.1, 2009.

IKEDA, A. A. Segure o Churn! In: LOVELOCK, Cristopher; WIRTZ, Jochen. **Marketing de Serviços: pessoas, tecnologia e resultados**. 5º ed. São Paulo: Editora Prentice Hall, 2006

LAROSE, Daniel T. **Data Mining Methods and Models**. a John Wiley & Sons, inc; 2006.

LARSON, Ron. **Estatística Aplicada.** 4º ed. São Paulo: Pearson Perntice Hall, 2010.

(LATTIN J., CARROLL, J.D., GREEN, P. E, **Análise de Dados Multivariados.** Ed. Cengage Learnig. 2011.

KARDEC, A., NASCIF, J., **Manutenção-Função Estratégica.** Ed.**Qualitymark**.2015

LISBOA, J.V., AUGUSTO, M. G. FERREIRA, P.L., **Estatística Aplicada à Gestão**. Ed. Vida Econômica. 2012

MALHOTRA, N. K. **Pesquisa de Marketing: Foco na Decisão**. 3º Ed. São Paulo: Pearson Perntice Hall, 2011.

MANLY, J.F.,**Métodos Estatísticos Multivariados - Uma Introdução.** Ed. Artmed-Bookman, 2008

MARÔCO, J., **Análise Estatística com o SPSS Statistics**. Ed. Pêro Pinheiro. 2020

MINGOTI, S. A., **Análise de Dados Através de Métodos de Estatística Multivariada**. Uma Abordagem Aplicada. Ed. UFMG 2007

NASSER R., O Desafio da Gestão dos Serviços Convergentes.TELECO www.teleco.com.br. acesso em 24.06.2022

NANIWA R. J., O Cenário do Serviço de Banda Larga e de Telefonia Móvel e a Legislação de Telecomunicações do Brasil. TCC. Engenha de Telecomunicações. UFU. 2020

NBR 5462- Confiabilidade e Mantenabilidade. ABNT.1994

NESLIN, S. A.; GUPTA, S.; KAMAKURA, W., LU, J.; MASON, C. H. Defection detection: measuring and understanding the predictive accuracy of customer churn models. **Journal of Marketing Research, Chicago**, Vol. 43, n.2, 2006.

NORUSIS, M. **SPSS 13.0 Statistical Procedures Companion.** Upper Saddle-River, N.J.: Prentice Hall, Inc. 2004.

PEREIRA S, BIONDI A, **Caminhos para a universalização da internet banda larga: experiências internacionais e desafios brasileiros.** São Paulo: Intervozes; 2013.

TACHIZAWA, T. CRUZ Jr J. B., . ROCHA J. A. **Gestão de Negócios – Visões e Dimensões da Organização.** 2° edição, São Paulo: ATLAS, 2003.

VENETIS, K. A.; GHAURI, P. N. Service quality and customer retention: building long-term relationships. **European Journal of Marketing**. Vol. 38, n. 11/12, 2004.

RIBAS, J. R. e VIEIRA, P. R., **Análise Multivariada com o uso de SPSS**. Rio de Janeiro: Editora Moderna, 2011.

VISCINI L. e SOUZA A. M., **Analise Multivariada na Prática**, UFSM, 2005

WAGNER M.B., MOTA V.T. DORNELES, C.C**., SPSS Passo a Passo - Statistical Package For The Social Sciences**. Ed. EDUCS. 2004