Contemporary Engineering Sciences, Vol. 11, 2018, no. 62, 3089 - 3100
HIKARI Ltd, www.m-hikari.com
https://doi.org/10.12988/ces.2018.87312

Use of Statistical Tools for Market Research as a Differentiating Factor in Business Process Training

Osvaldo Ricardo Chamorro Altahona

Universidad del Atlántico, Facultad Ciencia Economicas, Colombia

William Niebles Nuñez

Universidad de Sucre, Programa de administración de empresa, Colombia

Hugo G. Hernández P.

Universidad del Atlántico, Facultad Ciencia Economicas, Colombia

Abstract

The dynamics of formation of the industrial engineer is projected towards scenes of major competitiveness, so in the measure that the requirements of the economies and productive areas change, also they must change the ways of approaching and giving solution to the requirements of interest. With the purpose of analyzing how they come renewing aspects related to the training to discipline of this professional, there are exposed later the results of a bibliographical review that it took as a purpose, to penetrate into the tools that from the statistics the industrial engineer must receive to give response to the market researches. The results allow to view a sinnúmero of instruments that can use as forceful form for the generation of new alternatives in the constant improvement that is looked in an inherent way from this field of the engineering.

Keywords: training, statistical tools, market researches, industrial engineer

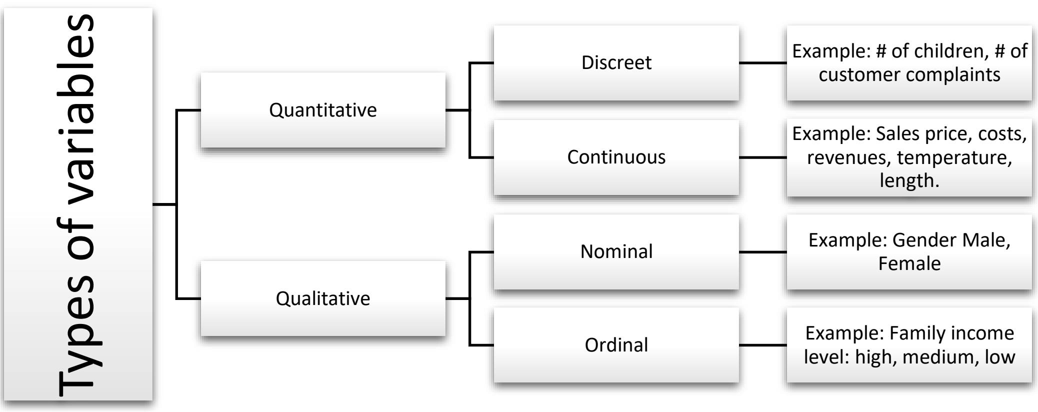
1 Introduction

Industrial engineering is presented as the disciplinary branch that manages to lead organizational changes through processes of innovation, quality, productivity and competitiveness, in such a way that optimized results are obtained in search of the satisfaction of the expectations of a given community, in any field [1]. In order to achieve this objective, the industrial engineer must be able to analyze, interpret, understand, design, program and, in general, take control of the productive systems to establish strategies of continuous improvement in the processes involved in the creation and development of goods and services. In this sense, industrial engineering brings together knowledge and interdisciplinary skills in such a way as to integrate resources in an unbeatable way. Marketing, then, becomes an important part of the study of this discipline, since it allows the engineer to know the characteristics and expectations of its internal and external clients, by means of the use of research tools.

2 Bibliographic Review

Initially it is necessary to establish the differences between the types of variables; quantitative variables are those that can be represented in numerical terms, while qualitative variables are expressions of data that represent an attribute or a category, in turn. Quantitative variables are classified as discrete or continuous, the former being those that can only take integer values, while continuous quantitative variables can take no integer values within an infinite set of data. On the other hand, qualitative variables can be ordinal or categorical when they have a degree of importance in each attribute, while in the nominal variables the order is not relevant, a graphical representation of this classification is presented in the figure below.

Fig. 1: Types of variables



3 Establishment of Techniques

Based on the types of variables to be involved in each study, the appropriate technique and therefore the type of quantitative tool to be used in the data analysis phase should be established; these techniques can be classified into:

3.1 Univariate techniques

These refer to when only one characteristic of individuals in a population is measured (Gender), while bivariate techniques measure two characteristics simultaneously (Gender and age) and multivariate techniques analyse more than two characteristics (Gender, age, educational level) [2]. Univariate analysis is the researcher's first approach since it allows him to explore the initial characteristics of the study variables. The first tool used to describe the characteristics of the variables studied is the frequency tables for qualitative data (categorical) and measures of central trend (mean, median, fashion); dispersion (variance, standard deviation) for quantitative data [3], table 1 shows this relationship.

Table 1. Univariate quantitative tools by variable type

Variable types	Scale	Tools for use
Qualitative (Categorical)	Nominal	Absolute and relative frequency tables (percentages).
	Ordinal	
Cuantitativas	Discrete	Central tendency and dispersion measurements, ranges, interquartile ranges.
	Continuous	

At the graphical level, it is common to use bar charts and pie charts, while for quantitative variables, line charts are the most appropriate, especially if time series are being presented. Another widely used and very useful graph is the BoxPlot diagram, this type of graph shows the distribution of the data by means of the representation of the quartiles, which allows us to observe the range in which the greatest number of observations are found (boxes) and in which interval they are found, in this way we can design strategies and actions for these findings.

3.2 Bivariate techniques

With this set of techniques, it is possible to find relationships between two variables, either by cause and effect, association or dependence; it also allows to determine if the means between two groups differ or are equal and to predict variables. The main tool used to perform the analysis is the contingency table or cross table, which relates the absolute frequencies of both variables. From this configuration it is possible to determine percentages in each category of answers, by rows or columns, based on the total, among other calculations. In the inferential area, different tools are used to compare different hypotheses about the association or relationship between two categorical variables, some of the most relevant of which are listed below.

3.2.1 Chi Square test of independence

When two categorical variables are had it is possible to determine if there exists some type of relation or association between these. The formula for his application in tables of contingency appears later:

$$\chi^2 = \sum_{i=1}^{n} \frac{(fo - fe)^2}{fe}$$

Where:

fo = Observed Frequency
fe = Expected Frequency

It is necessary to take into account that this tool only indicates the association or independence between two categorical variables, it does not inform about the direction, nor the magnitude of this relation; additionally, the value or frequency observed of the cells in the contingency table must be greater than 5 in at least 80% of the cells, so that the test works correctly [4]. Its application can have multiple purposes according to the objectives, hypotheses and questions that are raised, for example, the intensity of flavor of a product and the gender [5].

3.2.2. Linear correlation

To determine the degree of association of two variables it is possible to use three statistical tools that provide information on the magnitude of the relationship between two data sets, these are the Pearson correlation coefficient (r) commonly used for continuous quantitative variables, the Spearman correlation coefficient (Rho) used when categorical ordinal variables are involved and the Cramer V coefficient when nominal variables are used, these coefficients measure the same, these coefficients measure the same:

- Pearson's correlation coefficient: This is the tool that best fits the analysis in the case of two variables analyzed, whether quantitative, discrete or continuous; this coefficient only measures the degree of linear association between two variables and does not establish a cause-and-effect relationship, for this reason, in many cases the relationships may not be linear but curvilinear; can take values between -1, 0 and 1, values close to 0 indicate the existence of little or no relation, values close to -1 indicate an inverse or negative linear relation and values close to 1 indicate a positive direct linear relation, in this way the magnitude and direction of the relation can be determined. A classic example of the relationships described, can be the amount of dollars invested in a marketing campaign for a given product or service (X1) and where you want to see if there was a real impact on sales of that product or service (Y1), you would expect the relationship to be at least direct curvilinear or direct linear, among the most perceptible results. The formula for calculating Pearson's linear correlation for two quantitative variables is shown as follows:

$$r = \frac{N\sum XY - (\sum X)(\sum Y)}{\sqrt{N\sum X^2 - (\sum X)^2} * \sqrt{N\sum Y^2 - (\sum Y)^2}}$$

Where:

X = Independent variable
Y = Dependent variable
N = Number of data

- Spearman correlation coefficient (RHO): With this technique, as with Pearson's R coefficient, it is possible to determine the magnitude and direction of the relationship between two ordinal variables and its result is interpreted in the same way. The mathematical formula associated with the calculation of the RHO coefficient is given by the expression, as can be seen:

$$r_s = 1 - \frac{6\sum d^2}{n(n^2 - 1)}$$

Where:

d = Difference between X and Y statistics
n = Number of data

- Cramer correlation coefficient V: It is pertinent to use this type of tool when the variables analyzed are of nominal type, that is, when the order or importance of the response categories is not relevant. It is used in contingency tables larger than two classes and is interpreted in the same way as the previous correlation coefficients; a limitation is that it is required to calculate the Chi Square value which adds a little more calculation. The formula for this calculation looks like this:

$$V = \sqrt{\frac{\chi^2}{n(k-1)}}$$

Where:

X^2 = Chi Square value
n = number of observations in the table
k = the lowest number of rows or columns

3.2.3. Tests to compare means from two groups

- T Student Test: The purpose of this test is to verify that the average of one group is significantly different from the other. The data used must meet the requirements of: a) Normality in data distribution b) Independence and randomness in sample observations c) Sample sizes greater than $n > 30$ and d) Homocedasticity (homogeneity of variances). This test applies different formulas in the case of two independent samples, for example in the collection of data from men and women in a questionnaire, or if they are paired or related samples when for example before and after data are available in experimental designs.
- Mann Whitney U-Test: This test is considered the non-parametric alternative to the T Student test, i.e. when it is not possible to verify the necessary parametric assumptions it is recommended to use this statistic. As a non-parametric test it is more flexible in terms of the assumptions to be met and, in terms of interpretation, gives the same results as the T Student [7]. The test statistic for the U Mann Whitney, can be seen as follows [7]:

$$U_1 = n_1 n_2 + \frac{n_1(n_1 + 1)}{2} - R_1$$

$$U_2 = n_1 n_2 + \frac{n_2(n_2 + 1)}{2} - R_2$$

Where:

n_1 and n_2: sample size of each group respectively
R_1 and R_2: sum of the ranges of each group respectively

- ANOVA analysis of variance: It is common to present cases where it is necessary to contrast the average of more than two groups, under this assumption it is not convenient to use the T Student test since it would require more effort in the calculation and would increase the probability of falling into type I errors. To avoid these situations, the best option is to apply the ANOVA [8] analysis tool. This parametric test contrasts the null hypothesis that the means of more than two groups are equal or if they differ, where the variable to be studied or factor contains several categories or groups. It is important to note that for the application of this test it is necessary that the assumptions of the parametric tests mentioned in the T Student section of this document are met.

3.2.4. Bivariate forecasting tools

If you know that two variables are significantly related and you want to forecast the value of one of them, called dependent, from the other (independent), it is convenient to use linear regression models as a forecasting tool:

- Simple Linear Regression: It is possible to design a mathematical model that helps to predict a quantitative dependent variable (Y1) from another independent quantitative call (X1) when we know a priori that there is a relationship and that this relationship is linear, the tool used in this case is the Simple Linear Regression. With this tool it will be possible to estimate the value of the dependent variable with an acceptable degree of confidence.
- Binary Logistic Regression: This type of regression is used when the variable to be predicted (Y1) presents two response categories, that is, dichotomous, the model tries to estimate the probability that the event studied will occur according to an independent variable.

3.3. Multivariate Techniques

In this set of tools there centers a great number of technologies of major complexity that contribute more robust information about the investigated phenomena, his intention consists of analyzing simultaneously multiplicity of variables and of diverse types to observe his behavior and to establish conclusions [9]; there will qualify in two groups of agreement to his intention, first, those which end is to describe the characteristics of the phenomenon (descriptive approach) and those who are in use for realizing conjectures and generalizations on the population (approach inferencial), on the other hand, the method that they use also will be a criterion of classification, which can be methods of dependence and interdependence:

3.3.1. Multivariate technique with a dependency approach

- Multiple linear regression: This technique works the same as simple linear regression and its main difference lies in the number of independent variables assumed by the model. By incorporating different variables into the equation, it is possible to determine which of them provide or predict the value of the dependent variable in the best way; in this sense, the analysis tool has greater robustness.
- Multinomial logistic regression: In the case of the multinomial, the dependent variable has more than two categories (polynomial) and the independent variables can be metric or non-metric. As in classical linear regression and logistic models, it is necessary to determine the significance or goodness of fit of the model and the effect of each independent variable (regressors) in order to identify which ones contribute significantly and which ones do not [10].

- MANOVA: Like the ANOVA variance analysis, this technique evaluates the relationship between a set of metric variables, but against different categorical dependent variables, i.e. if multiple ANOVA analyses were performed at one time, similar results would be obtained by performing the MANOVA procedure. The main advantages of this tool lie in the increase of the power, with respect to multiple ANOVAS, detection of multivariate response patterns and reduction of type I errors in the procedure [11]. In the market research sector, applications can be multiple, including assessing the effects of promotional campaigns for two brands of products [12] or measuring the intention to purchase a product or service when there are multiple response variables and independent variables [13].
- Canonical Correlation: This technique is, to a large extent, more robust and more powerful than multiple or simple linear correlation techniques, the reason being that it is useful in situations where there are multiple dependent and independent variables, allowing them to have different scales of measurement (metric and non-metric), this allows the tool to be adapted to multiple real-life cases, specifically in the area of market research [14].
- Discriminant Analysis: From the discriminant analysis it is possible to explain if an individual or object belongs to a pre-established group. Categories of non-metric variables are used as dependent variables; unlike other techniques, independent variables are metrics that result in linear functions of each combination of independent variables that maximize the separation between groups. The difference between this analysis and other classification tools is that the groups know each other beforehand [2].
- Structural equations: This analysis tool allows determining cause and effect relationships between multiple variables, since the computational need for this technique is greater than for other techniques such as regression or factor analysis. To apply this technique, it is necessary to have a theoretical model to guide the estimation process. Based on this model, the dependency relationships and other relevant aspects for the model must be established [15].
- Joint analysis: One of the most relevant features of this tool is its ability to evaluate metric and non-metric variables, facilitating analytical processes with greater flexibility than other procedures. The model also allows an individual analysis of preferences, adding all of these to a generalized model. Using the decompositional technique, it is possible to disaggregate consumer preferences according to the degree of importance shown, contrary to the compositional or classification techniques (discriminant analysis) in which the valuations are used to construct the preferences of a product [16].
- AID analysis (automatic detection of interactions): Like the previous techniques, AID analysis analyses the dependency relationships between variables, the particularity of AID is that the dependent variable must be metric and the independent variables in non-metric scale; the AID performs dichotomous sequential divisions of the independent variables based on the analysis of the variance [10].

3.3.2. Multivariate technique with a focus on interdependence

- Principal Component Analysis (ACP): When you have a large number of variables and you want to synthesize the information in a more manageable way, use the principal component technique. This technique reduces the number of variables (dimension reduction) with the least possible loss of information by generating components or factors that are no more than a linear combination of the original variables [17]. This tool is a fundamental reference when applying grouping and interdependence techniques.
- Correspondence analysis: This analysis seeks to graphically represent in a multidimensional space the relationship that exists between the categories or classes of two nominal (non-metric) variables. With this technique it is possible to visualize the relationship between two variables in a better way than in a contingency table, since sometimes the amount of data in a table makes the analysis difficult [18].
- Cluster or cluster analysis: This is one of the most commonly used techniques in market research because it allows the identification of similar segments [19], it is necessary to follow a methodology that consists of the approach of the problem to be addressed, the choice of a distance measurement, the choice of the clustering procedure, the decision on the number of clusters, the interpretation and description of the clusters and, finally, the evaluation of the validity of the model [20].
- Multidimensional scaling: It is a multivariate interdependence technique that seeks to represent in a geometric space the proximities between a group of objects, it is similar to other interdependent classification analyses, however, its utility lies in the visualization of the results which facilitates its interpretation and analysis.

The different approaches analyzed become tools of great value for management at the business level, since as long as a correct articulation is achieved between the needs of the business areas [21], it will be possible to respond more accurately with innovations and solutions that adjust to what must be inserted in the markets [22], and therefore attend in a more competitive way to each of the instances that are derived from the processes of the current economy [23].

4 Conclusions

A statistical model is not a deterministic model, but rather a model that involves some probabilistic aspects; generally, a model is the basis of the assumptions made by the analyst, since the use of statistical methods cannot generate sufficient information or experimental data to describe the entire population; however, it does facilitate the learning of certain population properties from the analysis of a data set.

Throughout history, statistical models have been developed applicable to each type of variable and its interaction with the other aspects to be studied, in such a way that the generation of analysis results is more and more accurate; the acquisition of knowledge and its adequate handling allows the development of a better analysis of the cases studied in such a way that the prediction of their future behaviour is carried out with a high degree of reliability. In this document, the techniques for quantitative variables, either discrete or continuous, and for qualitative variables, categorized as nominal or ordinary, were observed, showing the conceptual basis of each one and its possible usefulness in relation to market research. From what has been observed, it can be concluded that the adequate and assertive use of these tools helps the industrial engineer to really determine the behavior of the variables to be controlled, developed and continuously improved in search of the satisfaction of a demand that evolves day by day.

References

[1] J. Stark, Product lifecycle management, In *Product Lifecycle Management*, Springer, Vol. 1, 2015, 1 - 29. https://doi.org/10.1007/978-3-319-17440-2_1

[2] M.J. Crowder, Statistical analysis of reliability data, *Routledge*, **1** (2017), 1 - 34.

[3] G. Ferrer, *Investigación Comercial*, 4ª Ed., Esic Editorial, 2016.

[4] J. Cerda and L. Villarroel Del, Interpretation of the Chi-square test (X^2) in pediatric research, *Revista Chilena de Pediatría*, **78** (2007), no. 4, 414-417. https://doi.org/10.4067/s0370-41062007000400010

[5] C. Esquilín and I. Hurtado, *Market Study to Determine Coffee Consumption Preferences*, PhD Thesis, Universidad Autónoma de Occidente, 2013.

[6] R. Martínez, L. Tuya, M. Martínez, A. Pérez and A. Cánovas, The correlation coefficient of the Spearman characterization ranges, *Habanera Magazine of Medical Sciences*, **8** (2009), no. 2.

[7] C. Martínez, *Estadística y Muestreo*, Bogotá: Ecoe Ediciones, 2012.

[8] M. Rubio-Hurtado and V. Berlanga-Silvente, Com aplicar les proves paramètriques bivariades t de Student i ANOVA en SPSS. Cas pràctic, *REIRE. Revista d'Innovació i Recerca en Educació*, **5** (2012) no. 2, 83-100.

[9] M. Quero and K. Inciarte, Classification of multivariate statistical techniques, *Telos*, **14** (2012), no. 2.

[10] V. Fernández and R. Fernández, Multinomial logistic regression, *Cuadernos de la Sociedad Española de Ciencias Forestales*, **18** (2004), 323-327.

[11] R. Warne, A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists, *Practical Assessmen, Research & Evaluation*, **19** (2014), no. 17, 1-10.

[12] A. Tessen, Multivariate analysis of variance (manova) applied to market research, II National Congress of the College of Mathematicians of Peru, 2011.

[13] O. Salem, Applied research in the hamburger market in the city of Guayaquil: Leader and a Follower, *FCSHOPINA*, **61** (2014).

[14] M. Mestre, Application of market research to the analysis of marketing problems, *Icade Magazine. Journal of the Faculties of Law and Economics and Business Administration*, (2012), 339-359.

[15] M. Cupani, Analysis of Structural Equations: concepts, development stages and an application example, *Revista Tesis*, **1** (2012), 186-199.

[16] J. Hurtado, Commercial use of joint analysis in Spain, *Survey Methodology*, **10** (2008), no. 1, 27-38.

[17] M. Terrádez-Gurrea, *Análisis de Componentes Principales*, Spain: UOC, 2012.

[18] M. Aranaz, *SPSS Para Windows*, Mc Graw-Hill Interamericana de España, 2002.

[19] A. Batanero and C. Díaz, Informal approach to hypothesis contrast, *Didactics of Statistics, Probability and Combinatorial*, **2** (2015), 207-2014.

[20] R. Walpole, R. Myers, S. Myers and K. Ye, *Probability and Statistics for Engineering and Science*, 9th edition, Mexico, Pearson Educación, 2012.

[21] H.G.H. Palma, La gestión empresarial, un enfoque del siglo XX, desde las teorías administrativas científica, funcional, burocrática y de relaciones humanas, *Escenarios*, **9** (2011), no. 1, 38-51.

[22] D. Cardona, A.M. Rada and H.G.H. Palma, Creación de empresa como pilar para el desarrollo social e integral de la región caribe: Apuntes críticos, *Saber, Ciencia y Libertad*, **12** (2017), no. 1, 15.

[23] H.G. Hernández, D.A. Cardona and J.L. Del Rio, Direccionamiento Estratégico: Proyección de la Innovación Tecnológica y Gestión Administrativa

en las Pequeñas Empresas, *Información Tecnológica*, **28** (2017), no. 5, 15-22. https://doi.org/10.4067/s0718-07642017000500003

Received: July 24, 2018; Published: August 12, 2018