ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS

SCHOOL OF INFORMATION SCIENCES & TECHNOLOGY

DEPARTMENT OF STATISTICS

MSc IN STATISTICS

Monitoring techniques for Conversion Rate Index

Marios Kolliakos

Supervisor: Stilianos Psarakis (AUEB)

A Thesis submitted to the Department of Statistics of the Athens University of Economics and Business in partial fulfillment of the requirements for the degree of Master of Science in Statistics.

Athens, September 2023

ΣΧΟΛΗ ΕΠΙΣΤΗΜΩΝ & ΤΕΧΝΟΛΟΓΙΑΣ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

ΤΜΗΜΑ ΣΤΑΤΙΣΤΙΚΗΣ

ΜΕΤΑΠΤΥΧΙΑΚΟ

ΠΡΟΓΡΑΜΜΑ ΣΤΗ ΣΤΑΤΙΣΤΙΚΗ

Τεχνικές ελέγχου για τον δείκτη conversion rate

Μάριος Κολλιάκος

Επιβλέπων : Στυλιανός Ψαράκης (ΟΠΑ)

Εργασία που υποβλήθηκε στο Τμήμα Στατιστικής του Οικονομικού Πανεπιστημίου Αθηνών ως μέρος των απαιτήσεων για την απόκτηση Διπλώματος Μεταπτυχιακών Σπουδών στην Στατιστική.

Αθήνα , Σεπτέμβριος 2023

Acknowledgments

I would like to sincerely thank my advisor, Mr. Psarakis, for his excellent cooperation and communication with me. Mr. Psarakis respected the whole concept of being a professor, as he was right next to me and guided me in whatever I needed. Furthermore, I would like to thank my parents for believing in me and supporting me in every moment.

Abstract

Marios Kolliakos

Monitoring techniques for Conversion Rate Index

September 2023

In this thesis, we attempted to present basic marketing concepts and referred to time series theory and statistical quality control theory. These methods were then applied and related to the conversion rate indicators (relevant to the marketing field) of a certain furniture sales company. These data were obtained with appropriate permission from the company in question.

First, a time series analysis model was applied to evaluate the change over time of the available indicators. Next, once the appropriate time series model is applied, the value of the indicator is predicted. Next, quality control analysis methods for autocorrelation data are used to identify random changes and to qualitatively validate these changes. Finally, we attempt to estimate the firm's earnings based on the constructed model

Περίληψη

Μάριος Κολλιάκος

Τεχνικές ελέγχου για τον δείκτη conversion rate

Σεπτέμβριος 2023

Η πτυχιακή παρουσιάζει βασικές έννοιες του μάρκετινγκ και αναφέρεται στη θεωρία των χρονοσειρών και στη θεωρία του στατιστικού ελέγχου ποιότητας. Στη συνέχεια εφαρμόστηκαν οι μέθοδοι αυτές και συσχετίστηκαν με τους δείκτες ρυθμού μετατροπής (σχετικούς με τον τομέα του μάρκετινγκ) μιας εταιρείας πώλησης επίπλων. Τα στοιχεία αυτά αποκτήθηκαν με την κατάλληλη άδεια από την ενδιαφερόμενη εταιρεία.

Αρχικά, εφαρμόστηκε ένα μοντέλο ανάλυσης χρονοσειρών για την αξιολόγηση των μεταβολών των διαθέσιμων δεικτών με την πάροδο του χρόνου. Στη συνέχεια, οι τιμές των δεικτών προβλέφθηκαν αφού είχε εφαρμοστεί το κατάλληλο μοντέλο χρονοσειρών. Στη συνέχεια, χρησιμοποιήθηκαν μέθοδοι ανάλυσης ποιοτικού ελέγχου για δεδομένα αυτοσυσχέτισης για τον εντοπισμό τυχαίων μεταβολών και την ποιοτική επικύρωση των μεταβολών αυτών. Τέλος, τα κέρδη της εταιρείας εκτιμήθηκαν με βάση το κατασκευασμένο υπόδειγμα.

Table of Contents

List of Tables

List of Figures

Chapter 1

1.1 Introduction

When people hear the word statistics, they usually associate it vaguely with professions such as economists, mathematicians, bankers, professors, and casinos. On the other hand, when business people hear the word, they associate it with the manufacture and sale of products, sales, consumer preferences, and attitudes.

Since today is the age of data and information, statistical data and survey methods are considered indispensable and are widely used in various aspects of business. Large companies spend a great deal of time and money searching for and using data in order to compile surveys and reports and draw conclusions about the future of their businesses based on them. This is especially true in the field of advertising and marketing.

Statistical methods have enabled companies to save time and money while at the same time reaping significant profits through targeted advertising. Through statistics, each company can evaluate its investment in advertising its products and services; at the same time, according to the field of statistical research, companies can measure and evaluate customer satisfaction, attitudes, and complaints with a view to self-improvement through the indicators mentioned in the work.

Marketing is based on consumer habits enforced by human psychology. While businesses have the ability to study some basic habits that have worked well for similar businesses in the past, competition has already changed considerably in the consumer sector, and it is not certain that traditional methods can predict consumer reality. Business needs quantitative measurements, which of course fluctuate.

But with the rise of statistical analysis, numbers now have a unique place in the success of digital marketing campaigns. A branch of mathematics called statistics can now provide all the support and knowledge a marketing strategy needs to accurately capture business objectives, results, and detriments.

In the following thesis, we will implement a statistical approach to an indicator called conversion rate. This indicator is one of the most important indicators in data analysis in the field of marketing and gives an overview and clear picture of business development. Conversion rate is defined as a fraction with the number of users who visited the site in the denominator and the number of users who ultimately made a purchase in the numerator.. The analysis will include time series methods, statistical quality control. The time-series sequence is called the sequence in which data are received at successive equal times (discrete sequence). Statistical quality control is the oldest and best-known method of controlling production processes to improve the quality of services and products produced.

The data collected relate to specific dates mentioned in the text and in the research part, and concern well-known furniture companies in the Greek market. More specifically, an attempt is made to relate time series models to statistical quality control and consistency in revenue forecasting. The data were obtained from a furniture company with the formal permission of the marketing director.

.

1.2 Structure

This first chapter is a general introduction to the concept of marketing, especially e-marketing. In addition, this chapter also explains the relationship between marketing and statistics.

Chapter 2 analyzes the basic concepts and fundamental ideas behind statistical quality control methods. In this chapter, you will learn why quality control is important in business and how quality control allows you to determine attributes and detect changes.

Chapter 3 describes the concept of time series analysis. After first analyzing the main axes of a time series, it presents an analysis of the main components and how to build and evaluate a time series model according to a specific methodology.

In Chapter 4, we conduct an extensive literature review on three specific methods (time series analysis, quality control, and regression) that have been implemented in the past with respect to the indicator conversion rate. In particular, we provide a definition of conversion rate and examples of cases in which this indicator has been used.

Chapter 5 introduces the graph theory of quality control theory. The graphs to be analyzed are sewhart, CUSUM, and EWMA. Chapter 6 presents the application of the statistical methods analyzed in the previous chapters to build time series models, apply quality control, and apply regression to firms' earnings.

Chapter 6 presents the application of the statistical methods analyzed in the previous chapters to build time series models, apply quality control, and apply regression to firms' earnings.

Chapter 2: Electronic Marketing

2.1 A bibliographical approach to Electronic Marketing

The expansion of e-marketing is quite related to the expansion of technology. In 1971 Ray Tomlinson started by sending the first email through a new technology platform. This platform enabled users to send and receive files through different machines. (e.g. TheCentreforComputingHistory). The most famous beginning of e-marketing is considered to have taken place in the 1990s. At that time the Archie search engine appeared which was used as an index for FTP sites. Earlier in the 1980s, computers had quite a lot of storage space. As a result, businesses began to use techniques in sales, such as online marketing and database marketing. These databases make it easier to track customer information more effectively, thus changing the relationship between buyers and sellers.

In the 1990s, the term digital marketing was coined. In addition, in the same decade there was a tremendous development of the use of personal computers and customer relationship management (CRM) applications, which became an important part of marketing. As time went on , marketers had to face competition and so they were encouraged to include more and more services in their marketing, sales and services by forcing their marketing, sales and services software to include more and more services. In addition, marketers using such eCRM software are now able to hold vast amounts of electronic customer data. Through this software, companies can keep abreast of customer needs and record their experiences. This led to the first clickable banner ads in 1994, with AT&T's 'You Will' advertising campaign (McCambley, 12/12/2013).

In addition, during the year 2000 and afterwards, the growth of internet users and the presence of the iphone, a problem created in companies and especially in marketing departments, was that consumers were no longer buying without the advice of a customer. For this reason, they needed to identify innovative digital ways to contribute to their marketing (Hart, 2000). The term automated marketing, introduced in 2007, has helped companies to group customers, develop new marketing campaigns and offer targeted information to their customers.

The 2000s saw a dramatic increase in digital marketing when the proliferation of devices with access to digital media led to rapid growth. Those applications that exploded the use of digital marketing were the growth of social media, such as YouTube, LinkedIn and Facebook . Consumers showed a trend of dependence on this new digital world that was evident in all aspects of their daily lives. As a result, consumers developed a desire for a seamless user experience on various company websites to search for information on various products. Consequently, this change in customer behavior has significantly improved the evolution of digital marketing technology. (GLOBAL, 2016).

2.2 Marketing and E-Marketing

It is widely known that there are numerous definitions of the concept of marketing. Marketing is a very broad concept and could not be defined by a single definition. In fact, to date there is no definition that can capture and convey the full breadth of the marketing concept precisely because it is an important and complex business function while at the same time it is a way of thinking, acting and analysing business.

Philip Kotler (2000) gave an interpretation of marketing and states that it is a social sequence and is directly linked to the development of society, its progress and its well-being. He further adds that marketing is the most important factor in economic development because without this philosophy and mindset of finding innovative ways to deliver value to customers, there can be no progress.

Its philosophy includes improving living conditions and at the same time gives a dynamic dimension to economic development. Therefore, Kotler has managed to separate the function of sales from that of marketing , as he states that sales is merely a business activity. More specifically, he argues that sales is concerned with securing orders so that the customer's needs are met. On the other hand, it is anachronistic for marketing to refer only to sales. On the contrary, good marketing eliminates sales efforts, recognizes the needs of customers and aims to satisfy them. These views are shared by Peter Drucker, who argues that marketing is not necessarily about sales, but that sales are the result of marketing. in short, marketing is about creating satisfied customers and maintaining that satisfaction, since they allow it to persist in the marketplace. Therefore, a business must have two main functions: marketing and innovation(Avronitis, 2010).

Internet marketing is also referred to as online marketing, internet marketing and e-marketing and refers to marketing via the internet. In online marketing , for its construction and implementation, electronic data is used and applied. In addition, through online marketing, the promotion, distribution and pricing of this data is also done. The aim is to generate a large number of exchanges, which in turn will be able to meet the objectives of the company and customers (Turban F. &., 2006).

According to Tzortzakis and Tzortzakis (Tzortzakis, 2002, pp. 580-630), they define electronic marketing as "the use, the application of the internet and new technologies and electronic media, for the realization of marketing objectives, as well as for the support of the ideas of modern marketing". It is the process of promoting products and services via the Internet and more specifically, "it is marketing that uses information technology to increase efficiency and transform marketing strategies to create business models that increase value creation for the customer and/or profitability for the company".

Additionally , in online marketing, a marketing mix strategy is a key part of the strategy . The marketing mix is used as a way of achieving the objectives of this strategy. The marketing mix consists of the tools the company uses to achieve its objectives. They are commonly known as the "4Ps" and the tools are considered to be product(product), price (price), distribution (place) and promotion (promotion).(Petrov,2002).The marketing mix is the combination of human resources and other physical resources required to fulfill the firm's programs and objectives in a market". (Tzortzakis, 1996, p. 380).

The internet and online marketing offer new opportunities to change the way business is done. New e-commerce channels offer new opportunities in the mix and represent a new sales channel (e-commerce). In addition, SMEs that have not been able to market their products in other countries have the ability to enter new markets. Finally, communication between businesses and consumers is becoming more direct.

2.3 Some characteristics of Internet Marketing

Accoriding to the paper Vlahopoulos (2003) identified the main characteristics of e-marketing and can be summarized as follows:

-The market is geographically dispersed and the people who make up its potential come from different social, cultural and economic backgrounds.

- Customers choose the time they spend and the information they wish to receive.

- The distribution of goods helps to save costs and mitigate the speed of delivery.

It is worth emphasising that a key feature of e-marketing is the provision of personalised products and services. In the e-marketing sector, there is no mass market of n individuals, but there are n markets, each containing one individual, each containing one individual. In the paper (Pashopoulos A., 2006). states that online marketing is targeted at people with specific characteristics, requirements and consumption habits. Therefore, according to his thesis, 'it is not a mass market of n millions of people, but n millions of individual purchases'. The internet allows companies to give personalised products and information to their customers. This differs from mass marketing, where the same thing is given to many people. With the Internet, companies can use two different ways to do this. One way is through technology-strategy push , where the prospect directs themselves to the product they want to buy. The other way is through the Pull technology-strategy, in which the expression of the desire is preceded by the expression of the desire and then followed by the information that interests him.

2.4 The functions of marketing

Online marketing helps businesses in many ways. It helps them save money and make more sales. It also helps businesses reach customers around the world. Online marketing involves research to find out more about what customers want and what other businesses are doing. This research helps businesses make good decisions and understand what their customers need. Sometimes, businesses ask customers questions in surveys to find out more about their opinion. Still, e-marketing contributes to market segmentation when ,that is, the large market is divided into sub-markets. The latter are distinguished by homogeneous characteristics in their products and the data used are based on statistical methods, on the behaviour, value and potential of each consumer.

These sub-markets are separated according to the characteristics of the groups in order to provide a targeted product to each one.The science of statistics has provided several methods that can be used in this segmentation. A further favourable aspect of electronic marketing is the development of products either through the creation of new products or through the modification of existing products to meet the needs of consumers. An additional function of marketing is the planning of a program ,in order to increase competitiveness and also to advertise companies and their products through a campaign. It is also possible before and after the sale of a product, to support the customer and to properly retain existing customers through the collection of personal information.

For business promotion to be effective, research is needed to identify target groups and to determine the markets to which their businesses will be targeted. The research should include personal and demographic data. After thorough research, the best online marketing strategies and appropriate methods are selected based on defined criteria.

2.5 Using statistics in marketing

Marketing, whether traditional or digital, is based on human psychology. Companies have the ability to follow certain trends and ways of advertising that have worked in the past, but this does not guarantee success as customer preferences are quite variable. But with the development of statistical analysis , numbers and metrics have a dominant role in the success of digital marketing strategies. Statistical studies provide the necessary expertise to ensure that marketing strategies are able to accurately represent and achieve their target audiences and where they are disadvantaged.

2.5.1 Use of Statistics in Paid Advertisements

Plus it is impossible to create a steady clientele without paid advertising. Marketing strategies with integrated paid advertising are able to avoid short-term returns. The means by which paid advertising can occur are social media platforms, search engines and others.

In this category of advertising there is a significant relationship between statistics and paid advertising. A mistake that happens with marketers is that they think that by spending more on advertising, they get higher returns, which is often not the case. Detailed data analysis is considered a necessary strategy to create satisfactory performance even in the case where advertising is considered optimized. Regardless of the platform, extensive insight is required to optimize ads. But let's limit our discussion here to Google ads, because Google ads are as effective and complex as you need them to be.

One way to start a paid advertising is on google.More specifically one can develop the PPC model. Under this technique marketers take the keywords they need and create ads that run on a pay-per-click billing system. This technique gives marketers information about their ideal customers and their demographics. Thus, through statistics, marketers can identify their buying audience and invest more in them.

2.5.2 Use of Statistics in Social Media Marketing

Plus social networking platforms are the main part of advertising campaigns for most B2B and B2C businesses. Of course, through statistical analysis one can study in depth and choose which platform to start advertising on. Just because a platform has a lot of users does not mean that we will have a lot of customers. Example if ibm wants to sell it systems , it can't do it through tik-tok but maybe through LinkedIn. In addition, if someone wants to sell women's clothing, they should invest in instagram.

Once again the use of data analytics is able to model the profile of the buying public and make more targeted advertising. The users of these platforms have quite different interests and demographic characteristics. The recording of web browsing (via cookies) can be very representative of who the best customers may be. Marketing managers together with analysts are able to create different campaigns to different buyer groups based on their demographics, interests and behavior to maximize their conversion into customers. In addition, through these platforms, companies have direct access to feedback and reviews from customers. Thus they can identify any complaints with a view to satisfying them.

2.5.3 Use of statistical data measurements in marketing

A company to achieve high returns, rely on statistical measurements that this time relate to the product itself and not the buying public. Some quantitative metrics are the following, Click Rate, Conversion Rate, Instant Return Rate, Promotion Rate, Unsubscribe Rate. These metrics are quantitative and can be applied statistical analysis techniques to make predictions, causal relationships and more.

2.6 Return on Investment ROI

One way that can be applied to evaluate the performance of investments or to compare the efficiency of different investments is the "return on investment" (ROI).To determine the ROI, the fraction is used with the difference between the profit on investment and the cost of investment as the numerator and the cost of investment as the denominator.

Consequently, the ROI is a measure of how efficiently a company invests its capital to generate profits. In marketing and finance this indicator could be a very popular one because of its flexibility and simplicity. If the indicator is negative or if there is another investment which is capable of avoiding higher profits, the present investment should be discouraged.

The measurement and determination of ROI should be done continuously during the optimization process of each advertising campaign. In this way the advertiser will be able to monitor the investment and the returns generated by each advertisement. In addition, the measurement of this indicator helps to determine the budget of an investment while at the same time providing information on the profit generated by a particular activity as a percentage of the budget. Advertising ROI is calculated as profit deducted from sales divided by advertising expenditure as a percentage. It can be calculated for any advertising campaign.

There are two ways in e-commerce that can track the progress of an advertising movement (Geddes, 2010).One is metrics and the second is adwords citations.

2.7 Metrics

In this category there are quantitative variables or indicators that give us summary data about the advertisement and its image. These metrics are found in Adwords.

Some of the metrics are clicks , click through rate , impressions. Clicks are an important measurement of the performance of an Adwords ad. This indicator shows us how many visitors enter our site. Clearly, however, a lot of clicks does not correlate perfectly with a high roi as many people enter the site out of curiosity without converting into customers.The clicks through rate the high percentage of clicks on ad impressions is able to show that an ad shows what it is looking for and consequently lead to a customer. On the other hand, neither is this correlated with a high roi. If the company wants to achieve recognition for itself or a product, the number of impressions counted in a campaign is a measure of success.

2.8 Adwords information

Adwords reports can be applied to display various data.The following basic data from adwords reports are:

- Performance statistics for specific placements and keywords in specific ad groups

- Custom reports that focus on specific information.

- Simple design for automatic report generation and delivery to multiple recipients.

- Use of reporting templates

2.9 Measuring results by monitoring conversions

The term conversion usually refers to the conversion of a prospective consumer into a customer. For example, if a consumer visiting a landing page takes a purchase action, the landing page is followed by a sales page. This process is defined as a sales conversion. The visitor - prospective customer is converted into a customer. However, not all different businesses have the same advertising objectives, the conversion can take different forms. It can lead to

1. Candidate client passes his personal data to the cell phone to get an offer

2. The visitor can take an archive to get more information about a product or service

3.The visitor can register on the site

4. The visitor to share the page

Google has made sure to provide a service called Conversion Optimizer (google Inc, 2013). This tool shows the number of conversions an ad has. Some categories of conversion provided by this service are as follows

1. Number of conversions. Count the number of clicks that led to a final user action on the website

2.Average conversion value . The average of the total value of the converters divided by their number

3. Conversion Rate . The number of conversions divided by the number of clicks on the ad. This is the main theme of this thesis , and we will analyze this rate with different statistical methods in the next chapters and especially on chapter 7

4. cost/conversion. Indicates the amount of money spent per conversion

5. Transactions . Number of transactions

6. cost/transaction . Identifies the ratio of the total cost of transactions to the number of transactions

7. Conversion monitoring: conversion costs and cost per conversion are important statistics. However, individual conversions can lead to multiple additional transactions as transaction costs, statistically, show how advertisers can acquire customers and provide useful information on the effectiveness of advertising.

However, it seems that the more statistical science emerges, the more the field of marketing science is demanding its techniques. At the very least, statisticians have developed fairly complex algorithms from which they draw conclusions about customer profiles (cluster analysis), conduct fairly sophisticated market research (component analysis), predict sales through mathematical models (time series analysis), and use intelligence techniques and neural networks to identify relationships and predictions among the various variables mentioned above. Time series analysis techniques and quality control charts are discussed below.

2.10 E-Commerce benefits

2.10.1 Positive towards the Consumer

Buying services and products via the World Wide Web comes with a variety of advantages for the consumer. First of all, the consumer can benefit from the possibility to buy at any time, since services and e-shops do not have specific opening hours like physical shops. Still, there is a significant difference regarding the cost of the products sold in e-shops in which it is obviously lower. In fact, this is due to the fact that e-shops are free of operating costs, for example, electricity, clerical costs and space rent. An additional advantage is the possibility of finding products at discounts after selecting from a wide range of products. Therefore, the consumer can search for the product of interest and buy it at a discounted price. A further advantage, is that the internet market is global. In short, it gives the consumer the opportunity to acquire a good from any other country , without having to move there. In fact, the transaction is instant since, approximately, in 3-4 days the product is in the hands of the consumer even if it came from a distant country.

2.10.2 Advantages for the Company

Buying on the World Wide Web offers significant benefits for both consumers and companies operating on the Web. One important benefit for companies is increased competitiveness as a result of expanding their activities in an effort to eliminate competition in local markets. Furthermore, through electronic transactions, important information such as consumer habits, needs and preferences can be collected, which can be used to improve the company's service policy. Also, it is very important that through e-marketing , the supply chain is significantly reduced. More specifically, before the product reaches the consumer, various intermediaries have intervened which is not very beneficial for the company since there is no direct communication with the consumer.

2.10.3 Disadvantages of e-commerce

However, although e-commerce offers many benefits, there are also risks. Firstly, there is the possibility of fraud or theft by people who take advantage of the low cost of entry to the Internet and the ability to remain anonymous in order to profit illegally. It is therefore very important to observe certain security rules, such as the creation of a unique code by the buyer that has never been used before. An additional disadvantage is the financial burden resulting from the shipping costs which are borne by the consumer. Finally, a major problem is that it is not possible to ensure immediacy with the object of purchase as in physical shops. It is therefore a virtual market.

It is very important for modern man to be able to take advantage of the unlimited possibilities offered to him thanks to the development of technology and not to be hindered by the potential risks that may arise. Instead, he can familiarise himself with the idea of e-commerce, saving effort, money and time. At the same time, through online shopping , the consumer performs a little research , gaining knowledge about things he is interested in while having to be careful in his transactions. The truth is that there is no particular way of doing business on the web. Companies are hammering away with advertisements and finding easier and more profitable ways of trading for their own benefit and for the benefit of the modern consumer.

Chapter 3 : Quality Control

3.1 Basic principles of Quality Control

Quality means that a product has optimal characteristics for the consumer. Products include both manufactured products and services. Thus, quality is the extent to which the customer's requirements, i.e. specifications, are met.

In the literature, the term "quality" has been studied by many researchers. The most common definition of quality is "fitness for purpose or use" by Juran (1999). In parallel, Walter Sewhart (1931), the founder of the control chart theory, focused his research on the characteristics of quality. According to him, quality has two characteristics. The first is objective and the second is subjective. The objective characteristics of quality are related to the characteristics of measurement and minimization of variation, while the subjective characteristics are related to the value and aesthetics of the product.

- Customer satisfaction is one of the core values of quality.

- Measurement methods to determine the level of quality

- Standards help to achieve and stabilize the desired level of quality.

It is worth emphasizing that the most important tool for quality control is SPC statistical process control. Montgomery (1997) defines SPC as a powerful set of problem-solving tools for process stability, variation reduction, and capability improvement. Ledolter, & Burrill (1999) also emphasize that SPC can identify and detect anomalous process behavior and also deal closely with variation reduction, as in Montgomery's case. Smith (2004), in contrast, provides a slightly different definition of SPC. Specifically, he states that SPC consists of data that is collected, organized, analyzed, and interpreted to modify the process to a desired level of quality.

According to the above researchers, the main objectives of SPC are to eliminate or reduce variation in all processes.

1) Achieving process stability in order to achieve - and ensure - customer satisfaction.

2) Minimizing production costs. This can be achieved through corrective actions (eliminating process defects).

3) Involving and supporting all employees working in the establishment or organization in decisions and actions related to the process.

4) Involvement of all members of the organization in continuous process improvement.

With the above in mind, SPC is an essential "weapon" for successful implementation and improvement of quality standards and can be applied to any product or service process. The following section summarizes the main SPC tools to address these challenges. In this chapter, control charts are discussed.

Histogram:

Check sheet:

Pareto chart

Cause and effect diagram

Defect concentration diagram

Scatter diagram

Control chart: Control chart is the most important and sophisticated one of the magnificent seven. We state the definition of the control chart in the next section.

In addition, it is noted that the above tools give the ability to control a process. A process is defined as a series of actions or causes that together lead to an outcome (product or service). In other words, a process is the transformation of a set of inputs into a desired output (Thomas et al.).

The concept of output in the above definition refers to the processed input data that is delivered to the customer. From the definition of process, it can be inferred that to ensure customer satisfaction, companies should base and analyze the process of producing their products. Therefore, in order to achieve this goal, it is necessary to first define and identify the input production process. It is also very important to define the purpose of the process. Only then can the inputs be correctly identified and the customer's requirements can be fully satisfied.

Note that the main inputs are generic. Inputs can include, for example, materials, methods/processes, information, people and the environment. It is clear that it is important to get feedback from the customer when creating the necessary inputs and during the implementation of the process.

The main output of the process is the product or service consumed by the customer. Therefore, in order to properly control and monitor the process, process documentation should be collected to reduce process variability.

Process variation refers to the fact that the probability that two units of a product produced in a given production process are identical is practically zero. There are always minor differences between products and services, even if they are very small (Montgomery, 1997; Griffith, 2000). The literature distinguishes two types of process variation: random variation and non-random variation.

Random variation is defined as process variation imposed by random (or common) sources of variation or random factors. These factors are quite difficult to interpret, model, predict, identify, or modify. Random variation is an inherent part of the process. (Banks, 1989).

Non-random variability, on the other hand, occurs at the output of the process. Random variability is usually less than non-random variability and, if it is at a high enough level, indicates poor process performance. However, the causes of these variations can be detected, identified and eliminated because they are due to fundamental quality characteristics.

Hence, according to the above, the lower the variation, the higher the quality achieved by the process. Asserting that the presence of variation is inevitable, manufacturing must control and maintain a constant level of variation. The action of keeping the variation at an acceptable level is called control. Control is a management process that can be accomplished by at least one of the following actions.

i)actual performance is compared with planned performance,

ii) difference between the two is measured

iii). causes contributing to the difference are identified,

iv) corrective action is taken to eliminate or minimize the difference

There are two types of processes associated with control: controlled and uncontrolled processes. Processes in which the only source of variation is randomness are called statistically controlled, and processes in which the source of variation can be identified are called statistically uncontrolled.

There is another important term. Quality characteristics are quality parameters that describe what consumers perceive as quality. An important tool that controls the quality of the process are the Statistical Process Control Charts and they will be mentioned below.

3.2 Statistical Process Control Charts

First of all, a control chart is a time series graph that shows how the quality characteristics of the product under study change over time. To do this, managers take samples at regular intervals and plot the results on a graph (Ledolter & Burrill, 1999; Smith, 2004).

The purpose of control charts is to identify attributable causes, i.e. non-random variations in a process. Control charts have a very important positive property - they facilitate decision-making by workers by identifying the causes of variation and its signs. Thus, control charts statistically control processes. However, it should be noted that control charts only identify relevant causes. Control charts are very important for forensic scientists because they allow them to identify changes and make quick decisions before anything fatal happens.

Control charts are used as an SPC tool in companies that produce goods and services. They are also an important tool in Six Sigma where the DMAIC methodology is applied in the management part; DMAIC stands for Define, Measure, Analyze, Improve and Control steps.

The basic principles of constructing quality control charts are discussed below. The chart below is the simplest chart found in the statistical quality control literature. This control chart is a graphical representation of the quality characteristics obtained during a sample run, together with the sample run time. The graph has three characteristics

1) The center line (CL), which is the mean value of the quality characteristic corresponding to the control condition

2) Upper control limit (UCL), which is the upper limit of the quality characteristic plotted before the process leaves the control condition.

3) Lower control limit (LCL) is the lower limit of the quality attribute plotted before the process exits the control condition.

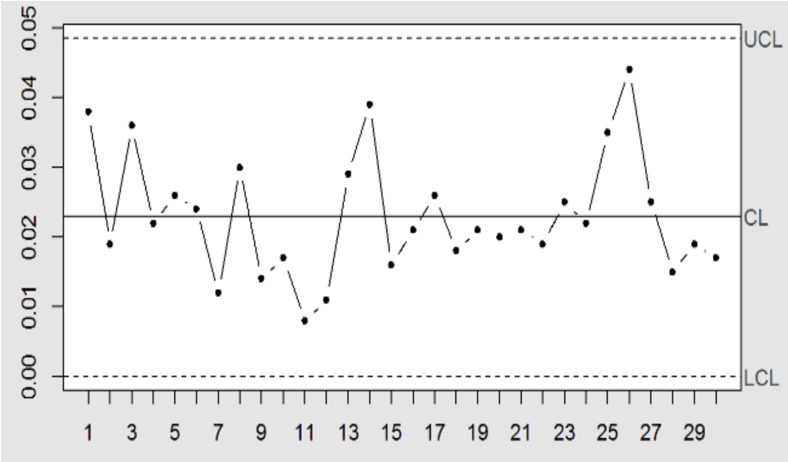

Figure 1 ***A typical control chart for in control processes***

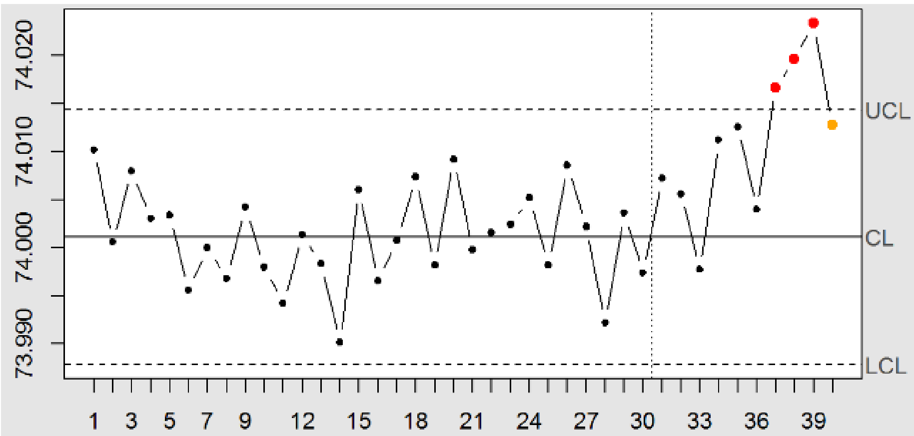

Figure 2 ***A typical control chart for out of control processes.***

The first figure indicates that the process is within the normal range and statistically controlled because there is no pattern of evolution; the second figure indicates that the process is not controlled because many points of the main characteristics of the sample under study are out of the normal range. The challenge, therefore, is to scrutinize, identify and eliminate the reasons why these points are out of control.

A common problem that the quality graph research community wants to address is the estimation of control limits. In other words, "how close or far should the control limits be from the center line?". This is the answer to the question posed.

Researchers strive to minimize all the above risks, so the literature usually uses "three-sigma" to compute control bounds.

This thesis discusses three types of control charts and their relationship with time series. These are the Swehart, cusum and ewma control charts.

But before we finally move on to examine these diagrams, let's look at the benefits they have.

1) The most important benefit is to fulfill the sixth process and improve the process by being able to detect uncontrolled characteristics → investigate causes → make decisions in a timely manner.

2) Control charts can contribute to significant reductions in production costs because they reduce the cost of detecting defective products; Montgomery (1997) states that "without effective process control, it is like paying someone to produce nonconforming products".

3) Control charts provide diagnostic information. In other words, control charts answer the question of whether the presence of defective products is the result of a process defect or a manufacturing defect.

4) Control charts can be used to conduct process capability studies that can serve as the basis for many manufacturing decisions. In other words, control charts can be used as a detection and prediction tool. Similarly, these procedures can be used to determine process parameters such as mean, standard deviation, and coefficient of variation that can be used to determine process capability.

5) The control chart defines a set of methods to be applied by the operator. Operators should follow the guidance of the control chart in improving their processes.

Chapter 4 : Literary review of time series theory

A time series is a sequence of values Y_1 , Y_2 , ...Y_n of a variable Y that varies over time at discrete (equal) time instants $t_1, t_2..., t_n$. Equidistant time instants may be years, months, weeks, days, hours, etc. An example of a time series is:

- The daily prices of a stock exchange share
- The weekly sales of a national newspaper

The characteristic of time series are dependent on whether the values of the variables are used unprocessed (i.e. the original values) or after processing, and are divided respectively into primary and secondary (derived) time series. In addition, time series Y_tA continuous variable, in terms of time, is a variable for which its value is available at any point in time, such as the price of a stock, for example. A discrete variable, in terms of time, is a variable for which its value is not available at any point in time, but only at specific points in time, for example the number of aircraft movements at an airport.

The graphical representation of a time series is done using a time series plot. In this particular plot, the horizontal axis represents time, and the vertical axis represents the values of the variables and their changes over time. It can be mentioned that this plot is a very important tool because it provides a visual/overall picture of the time series and through it one can identify patterns or patterns between data and time, which is of interest to all analysts who study time series.

As an example of a time series series, the graph of the "conversion rate" studied in this thesis is shown below, and it can be easily observed that the "conversion rate" of the time series series series achieves its highest value in the time interval corresponding to the period from week 50 to week 73.

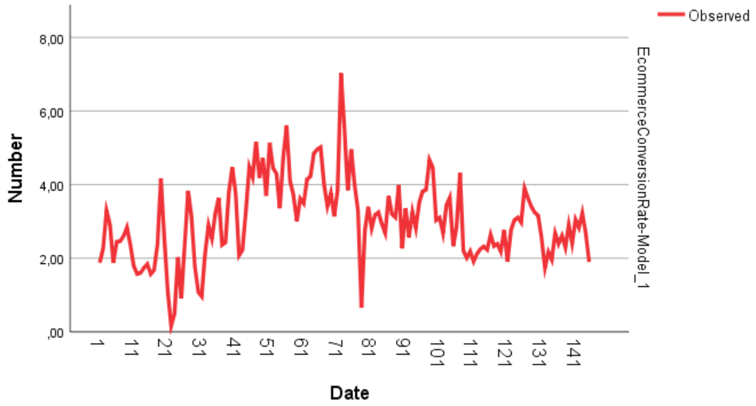

Figure 3 ***A time series plot***

4.1 Time series basic components

Time series of all variables share common analytical characteristics in a number of key components related to how its observations are shaped over time. Identifying, describing and reproducing these common features is of great importance because it makes it possible to describe time series in mathematical models where the ultimate goal is to be able to predict future time series values with sufficient accuracy, consistency and impartiality.

Based on the above, each time series can be analysed in terms of four components, which are described by appropriate statistical methods as indicated by the scientific community. The main components of a time series are described below:

- **(T_t) (trend):** expresses the trend in the values of a time series over a sufficiently long period of time. The possible shapes of the trend in a graph can be upward , or downward. Trends can be described as linear or non-linear.

- **(S_t) (seasonality):** Seasonality in a time series is expressed by repeated/periodic movements in the values of the time series. For example, drug sales increase every winter. This cyclicality is called seasonality. An example of this can be the increase in sales , and specifically the conversion rate of an e-shop every Black Friday.

- **(C_t) (cyclical):** Periodic fluctuations in a time series are represented by repeated movements of the time series' values over a long period of time. If the period of repetition is longer than one year (e.g. 10 years), periodic (as opposed to seasonal) fluctuations occur.

- **(I_t) (irregular patterns):** irregular changes in time series are described by irregular movements of time series values due to random and usually unpredictable factors that cannot be explained many times. The observed values of a time series show abrupt changes that can be easily identified in the time series and do not exhibit regularity or specific recurrence rates. For example, the 20222 earthquake in Turkey in Japan caused a significant increase in sales of building materials due to demand. This example is therefore a phenomenon of irregular movements in the time series, as the phenomenon/event of earthquakes is still unpredictable today.

All the above can be described by the diagrams shown below.

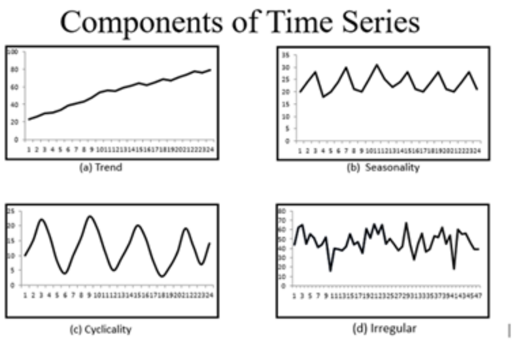

Figure *4 Components of time series*

4.2 Time series components analysis

Time series are appropriately used to predict future values of a set of observations using information from the past. The use of information from past observations of a time series is done by attempting to predict its key elements in the future in order to identify patterns in its behaviour and achieve valid and reliable forecasts. However, in order to achieve a correct prediction of the future values of a time series, its model, which is assumed to have been valid in the past, must also be valid in the future.

For the components of the previous section, it is assumed that there are independent cumulative and/or multiplicative effects on the time series values. Therefore, using a linkage model between them, it is assumed that the dependent variable representing the time series is the sum or a product of the variables T,C,S,I. The basic model of the relationship between the components is as follows:

- **Cumulative model**: $Y_t = T + S + C + I_{tttt}$
- **Multiplicative model:** $Y = T * S * C * I_{ttttt}$

Of course there is also the mixed model in which both operations of addition and multiplication are used, e.g. $Y = T + S * C_{ttttt} + I$. It is therefore considered appropriate for the analyst to choose the correct model , after first evaluating the components, the corresponding statistical results based on the appropriate criteria used in time series analysis.

4.2.1 Classical analysis time series

The classical method of time series analysis requires the identification of the components mentioned and the quantification of their influence on the time series.Time series analysis differs in the way the components are linked. The hierarchical stages of time series analysis according to the classical methodology are listed below:

- Determination of Seasonality.
- Elimination of Epochality.
- Eliminate Voltage.
- Cyclicity determination.
- Elimination of Circularity.

Once the above steps have been carried out, the basic components will have been determined and the time series will be ready to be applied. The steps summarised in this paragraph are set out in detail below.

4.2.3 Identification of Component "Seasonality"

The seasonality of a time series is determined by the following methods:

(i) <u>Percentages in relation to the monthly average</u>

The process of implementing this method includes the following steps:

- Calculation of annual averages.
- Conversion of initial observations into percentages of the previous annual average.
- Calculation of monthly average rate.

The final monthly average rate produced is the seasonality factor (*Grais B.*, 2000).

(ii) <u>Percentages in relation to the monthly trend</u>

The steps to implement the procedure are:

- Locate the least squares line of the trend and calculate monthly values of the variable.
- Converting initial observations to percentages of calculated values (from the 1^{o} step) from the trend equation.
- Calculation of the monthly average rate.

The final monthly average rate produced is the seasonality factor (*Grais B.*, 2000).

4.2.4 Identification of Component "Circularity"

Cyclical fluctuation can be dealt with by the following steps:

- Tackling seasonal fluctuations
- Addressing the trend
- Calculation of moving average terms by eliminating irregular changes.

- Calculation of cyclical fluctuation indices in a similar way to the calculation of seasonality indices

4.3 Forecasting future observations in time series

The main purpose of the time series analysis process is to predict future values through the reliable model that has been identified, but what one should pay attention to is to give importance to finding and selecting the appropriate model in order to reduce the amount of uncertainty. Forecasting methods fall into two main categories:

- **qualitative forecasting methods**: these methods are very effective when there is no historical time series data and depend on the subjectivity of the researcher. These methods are generally less accurate.
- **quantitative forecasting methods**:On the other hand, these methods use historical time series observations, where available. In order to build reliable time series models and make accurate forecasts, historical data should be explored and analysed using appropriate statistical methods (time series analysis).

As mentioned above, qualitative methods are applied when there is no historical data. For example, the supply of a product to the market and the prediction of its sales.

The quantitative methods used when there is historical data , are divided into three branches and can be divided into three branches.

- The first branch includes time series models with forecasting techniques based on the analysis of continuous values of the dependent variable of a time series at regular time intervals.
- The second branch includes models that include dependent and independent variables in econometric models. In other words, in this class of models there are independent variables that are related to the dependent variable.
- In the third class are models that give importance both to the explanatory variables (ARMAX models) related to the modelling of the values of the dependent variable, and to the dependent variable itself.

Where time series may fall short is that the predictions are based only on the variable being studied and more specifically on its past values. Consequently, an important factor for a satisfactory forecast is that the model of the time series under study has not changed significantly. Of course, this is not always the case as the phenomenon described by a time series may be influenced by a number of factors. Perhaps a mixed model consisting of dependent and explanatory variables is the optimal solution for reliable forecasts, as a model that does not take into account the past values of the dependent variable may fail to detect random seasonal variations.

4.4 Modern analysis of time series

In the previous paragraphs, a detailed description of classical methods of time series analysis was given.However, the literature prefers a more modern method of time series analysis called BOX-JENKINS methodology.

The BOX-JENKINS method interprets each time series as a stochastic process using a stochastic model that allows complex models to be fitted to real time-varying data. Therefore, the dependent variables Yt-1, Yt, Yt, Yt+1, ... are correlated, i.e. each time series can be represented as a stochastic model assuming that the observed values of each time series are due to a stochastic process. Thus, assuming that a stochastic process generates the values of the time series, these values are random variables with a common distribution. The methodology is based on the principle of discovering a stochastic model with the same function as the process that generates the actual values of a time series.

The simplest stochastic description of the time series Yt, is to consider the independence of the successive values of Yt-1, Yt with a constant mean m, where m is assumed to be equal to zero:

$$Y_t - Y_{t-1} = e_t \ , \ E(\varepsilon_t) = 0 \text{ and } E(\varepsilon_t \varepsilon_s) = 0 \ , \ t \neq s$$

One of the most important factors examined by the BOX-JENKINS methodology is the concept of stagnation. The time series can be considered as stationary or non-stationary. Below are models of stationary time series:

- Auto-Regressive models of order p, AR(p).
- Moving-Average models of order q , MA(q).
- Mixed class models (p, q), ARMA (p, q).

- Mixed class models (p,d,q) , ARIMA (p,d,q).
- Mixed seasonal models of order (p,d,q) , SARIMA (p, d, q) (p′, d′, q′).

The BOX-JENKINS methodology involves three main steps:

1. Model Recognition searches for and determines the difference order d of the original series required to ensure the stationary ARIMA model (p,d,q). Once the difference order d of the original series corresponding to the stationary model is determined, the lags p and q of AR and MA are searched for and determined.

2. Estimation of the model: In this phase, the parameters of the auto-regressive and moving average of the model selected in the previous step are estimated.

3. Model diagnostic testing:The latter ensures that the theory that the identified and estimated ARIMA model is correctly applied to the actual sample data using appropriate statistical criteria.

4.4.1 Stationarity of Time Series Models

A stochastic process is called stationary when its set of statistical properties is not affected by random changes in the starting point over time. The above expression is a set of statistical properties of n observations with starting point t (Y_t , Y_{t+1} ..., Y_{t+n-1}) are the same as the statistical properties of *n* observations with starting point $t+k$ (Y_{t+k} , Y_{t+k+1} ..., $Y_{t+k+n-1}$).

If the mean and variance are constant over time and the covariance between the values at two time points t, t+k does not depend on time but only on the distance k between these time points, then the time series is called stationary.

The conditions of stability are therefore set out below:

- $E(Y_t) = m_Y$, $\forall t$ (Fixed mean value)
- $Var(Y_t) = \sigma^2_Y$, $\forall t$ (Fixed variance)
- $Cov(Y_t , Y_{t+k}) = c_k$, (Covariance)

Non-stationary time series , contain trends, seasonality and cyclicality and are used on a daily basis in various fields, such as sales of various products. Finding a suitable stochastic model to describe time series data can be done by converting them into stationary time series using first differences or by applying a regression with time as an explanatory variable to remove the trend.In the last chapter, time series will be studied, where the methodology mentioned , the most appropriate one will be studied and will try to interpret and predict conversion rate values.

4.4.2 Autocorrelation function

Autocovariance is defined as the covariance of the same variable at 2 different times, k time points apart, denoted by γ_k

$$\gamma_k = \mathrm{Cov}\,(Y_t\,,\,Y\,)_{t+k}$$

But in the above paragraph Y_t is stationary

$$E(Y_t) = E(Y_{t+k}) = m_Y\,,$$

So the autocovariance takes the form:

$$c_k = \mathrm{Cov}\,(Y_t\,,\,Y_{t+k}) = E(Y_t - m_y)\,(Y_{t+k} - m_y)$$

By definition, the zero lag autocovariance, ($k=0$), is the variance of the variable Y_t denoted by $\gamma_0 = \sigma_Y^2$ while by definition: $\gamma_k = \gamma_{-k}$

We then define the p_k autocorrelation coefficient between two observations Y_t and Y_{t+k} separated by k time periods is defined as follows.

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+k})}}$$

In addition, if there is stationarity then $Var(Y_t) = Var(Y_{t+k}) = \sigma^2{}_Y$, and then O so that the autocorrelation coefficient is given by the formula

$$\rho_k = \frac{Cov(Y_t, Y_{t+k})}{\sqrt{Var(Y_t)}\sqrt{Var(Y_{t+k})}} = \frac{\gamma_0}{\gamma_k}$$

The values of the autocorrelation coefficient ρ_k belong to the interval [-1,1] and $\rho_k = \rho_{-k}$.

In order to obtain information on the stability of the time series, the autocorrelation diagram is constructed, which represents the autocorrelation function (k, ρ_k), $k \geq$

0,. In particular, in this diagram, when the values of the autocorrelation coefficients fall sharply to 0, the time series is said to be stationary as the number of lags k increases, whereas if the autocorrelation coefficients fall slowly to 0, it is said not to be stationary. The statistical significance test of the autocorrelation coefficients p_k , is performed using Bartlett's statistical test with the following assumptions:

$$H_0: \rho_k = 0$$

vs

$$H_1: \rho_k \neq 0$$

The statistical criterion Q is a basic diagnostic test of the appropriateness of the estimated model and tests the hypothesis that jointly a number of coefficients differ from zero or not. The test was introduced by Box and Pierce with the following assumptions:

$$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$$

$H_1: \rho_\iota \neq 0$ let for a value i=1, *2,…, k*

In the case of rejection of the null hypothesis then the hypothesis that the time series is from a random process is rejected.

4.4.3 Partial autocorrelation function

The Partial autocorrelation function is used to control the key characteristics of the time series. The partial autocorrelation coefficient of the obtained values Y_t and Y_{t+k} can be defined defined as the correlation coefficient after removing the effects of all intermediate values Y_{t+1} , Y_{t+2} , …, Y_{t+k-1} .In addition, this coefficient is denoted by φ and has an indirect regression link as the following model can be supported as an example: $Y_t = \varphi\ Y_{12t-1} + \varphi\ Y + e_{22t-2t}$.

Specifically, the first value of φ indicates the time lag of the independent variable Y_{t-1} , while the second value indicates the maximum time lag of the independent variable Y_{t-2} . In this case, when the coefficients of the variables have a lag of 2, i.e. Y_{t-2} , they justify a second order partial autocorrelation coefficient (φ_{22}). In other words, the value φ_{22} indicates the correlation between the variables Y_t and Y_{t-2} , taking into account the intermediate effect of Y_{t-1} .

4.4.4 Self-routing Models of Order p - AR(p)

The autoregressive time series model of order p can be described by the following representation

$Y_t = d + a\, Y_{1t-1} + a\, Y_{2t-2} + \ldots + a\, Y_{pt-p} + e_t$.

In addition, theparameters d, a_1 , a_2 ,..., a_t are constants that can be estimated in a suitable way while e_t is a random variable with E(e_t)=0 and Var(e_t)=σ^2 . This model is identified as an autoregressive model of order p and is denoted by AR(p) . If we use lag operators L then the form of the autoregressive model of order *p* becomes:

$(1-\alpha_1 L - \alpha L_2^2 + \ldots + \alpha L_p^p) Y_t = \varepsilon_t$

Finally, these types of models are stationary if the roots of the polynomial $(1-a_1 L - a L_2^2 + \ldots + a L_p^p$) are outside the unit circle.

4.4.5 Moving Average Models MA(q)

These models of class q mobile media are expressed by the equation

$Y_t = \mu - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_p e_{t-q} + \varepsilon_t$,

where the *parameters* δ, θ_1 , θ_2 , θ ,..., θ_q are constants which can be estimated in a suitable way and e_t is a random variable with E(e_t)=0 and Var(e_t)=σ^2 . In this model case the time series is called the moving average model of order q and is denoted MA(q) This model is represented by a weighted average of the errors ε_t of the *q* previous periods. The moving average model of order q is stationary by definition.

4.5.1 The ARMA Mixed Model (p,q)

It is a coupling between the MA and AR models where its form is

$Y_t = d + a\, Y_{1t-1} + a\, Y_{2t-2} + \ldots + a\, Y_{pt-p} + e_t + m - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \ldots - \theta_p e_{t-q}$

Sequentially, using operators, the mixed model becomes $\boldsymbol{A(L)Y_t = \delta + \Theta(L)e_t}$,

And in this case the ARMA(p,q) model is stationary when the polynomial $A(L)$ has roots outside the unit circle.

4.5.2 ARIMA Mixed First Difference Model (p,d, q)

If a mixed model is constructed where the first differences ($Y_t - Y_{t-1}$) or higher order differences of the dependent variable are applied, then the resulting model will again be a mixed model and will be denoted ARIMA (p,d,q).

There are several criteria for finding the appropriate number of lags and estimation errors to introduce into the model, such as finding the values of p and q so that the series is stationary. This is a necessary condition for the model to be acceptable.

4.6 Evaluation of time series

The objectives for the next chapter are to achieve an analysis of the "Conversion Rate" time series in the following steps:

i. The determination of the time evolution of the conversion rate values and the main components of the time series within the time range of the data

ii. The partial autocorrelation and autocorrelation correlations will be examined in order to apply the appropriate model

iii. The most suitable ARMA or ARIMA model will be selected

iv. The most satisfactory forecast of future values of the indicator.

In order to select the best model, the different models created were compared according to the evaluation criteria provided by the SPSS statistical program. The fit of the models to the pragmatic data was evaluated using the following statistical criteria:

1. Stationary R^2: Determines whether the time series is stationary or not. Desired values should be close to 1.

2. R^2: Identifies the percentage of the variability of the dependent variable that the model manages to explain. The values we want are close to 1

1. Root Mean Squared Error (RMSE):. The lower the value, the better the model that was constructed.
2. Mean Absolute Error (MAE): the lower the value, the better the model that was constructed.
3. Maximum Absolute Percent Error (MaxAPE): the lower the value, the better the model that was constructed.
4. Maximum Absolute Error (MaxAE): the lower the value, the better the model that was constructed.
5. Ljung-Box Q test: shows us how appropriate the model is and we want the p-value to be greater than 0.05.

Chapter 5: Literature review basic quality control charts

5.1 Quality control charts

To begin, as far as the theory of quality control charts is concerned, the literature places them into two categories. The categories are. « Variable control charts» and «Attribute control charts». Let us define that the variables that control charts examine are called quality characteristics, which can be expressed as numbers on a continuous measurement scale (Montgomery, 1997). When a quality characteristic is a variable, it is convenient to control for it on a scale of central tendency and a scale of variance. This type of control chart is called a variable control chart; Besterfield (2004) defines a variable control chart as a way of illustrating the change in central tendency and variance of a particular quality attribute over a series of observations. The types of variable control charts are listed below.

- Shewhart X charts ($\bar{x}$)
- Average and range charts ($\bar{x}$ and R)
- Median and range charts
- Average and standard deviation charts ($\bar{x}$ and S)
- Individual and moving range charts (X and MR)
- Run charts
- s^2 charts

5.1.1 Shewhart X chart

The study begins with the Shewhart control chart. It was first studied by Dr. Walter A. Sewhart in the 1920s and attracted scientific interest in many fields. Subsequently, he introduced statistical control charts, $\bar{x}$, $\bar{x}$ and R, $\bar{x}$ and S , so these control charts are also known as Shewhart control charts.

Shewhart control charts were served as the basis and foundation for other charts that are characterized by simplicity and ease of use.. If $x_1, x_2, x_3, ...$ is a sample of size n, upper control limit (UCL), the center line (CL), and lower control limit (LCL) of the Shewhart X chart is given, respectively in the following:

$$CL = \overline{x} = \frac{x_1 + x_2 + x_3 + x_n}{n}$$

$$UCL = \overline{x} + 3\sigma$$

$$LCL = \overline{x} - 3\sigma$$

where$\overline{x}$ is the mean and σ is the standard deviation of the process. We assume that σ is known or an estimate is available.

Test control limits are used to determine the control limits. Test control limits provide sufficient information to determine if the process is within the control limits selected during the initial observation.

Hence, if all points are within the control boundaries, the process was previously controlled. This proves that test control boundaries can be used to control current and future production.

On the other hand, if any point is outside the test boundaries, it indicates that the process was not controlled in the past. In this case, the causes should be measured and eliminated and once they are eliminated, the control limits should be ignored and recalculated. This process continues until all points are fixed within the constraints. Thus, phase I is called historical data analysis and the control limits obtained from phase I analysis are used in phase II. The control limits are used for current and future monitoring.

Due to their ease of use and lack of need for in-depth statistical methods, sweep charts have been used for several decades. However, despite their satisfactory performance, these charts have several drawbacks. The main disadvantage is that they only consider the last point on the graph and cannot include information about the entire process. In other words, the graphs can reveal large but not small shifts in the parameter under consideration.

However, quality control researchers have developed two other graphs that overcome these shortcomings. These are exponentially weighted moving averages (EWMA) and cumulative sums (CUSUM) . EWMA is suitable for detecting small shifts because it places less emphasis on historical data. However, it is not as responsive to large shifts as the Shugart chart.

5.1.2 Cusum Control charts

The CUSUM control map was proposed by Page in 1954 based on cumulative sums. The advantage of this control map is that even fairly small offsets can be detected quickly, which is not possible with the Shewardt control map. In order to construct this chart we need a continuious sum of deviation from μ_0. If μ_0 are the target value and x_j is sample mean then the cusum chart can be represented as $S_i = \sum_{j=1}^{i}(x_j - \mu_0)$. This procedure is called random walk (Harris, Ross 1991).

5.1.3 CUSUM – chart for samples means (normally distributed) and for individual values

Let values x_j , that they are dependent and they follow the same distribution $N(\mu, \sigma^2)$, as the μ and σare known. Further assume that there are several subgroups with the same population size. Cumulative sum – CUSUM C_n is defined for individual values (n = 1) as:

A) on a base of original scale: $C_n = \sum_{j=1}^{n}(x_j - \mu)$

B) on a base of normal distribution where mean = 0 and standard deviation = 1:

$$U_j = \frac{x_j + u}{\sigma}, S_n = \sum_{3=1}^{n} U_i$$

It is observed that CUSUM C_n is almost the same as CUSUM S_n measured in the units of standard deviation s. So we can write the equation of C_n (Chandra 2001): $C_0 = 0, C_n = C_{n-1} + x_n - \mu,$ and using the same principle for S_n: $S_0 = 0, S_n = S_{n-1} + U_n$

In the case that the μ will be channged from $N(\mu, \sigma^2)$ to $N(\mu + \delta, \sigma^2)$, the population mean μ will face an arbitrary shift of δ.In othe words , this means that the change begins from the point (m, C_m) and increased linearly by slope δ. However, population means shift can be more complicated. The CUSUM control chart can reflect all these changes (Harris, Ross 1991)

5.1.4 CUSUM for sample means $\overline{x}_j$

Unlike individual values, it is assumed here that there is a subset of m observations. Our target is to measure the mean of this subset .Therefore, the standard deviation of the sample is divided by $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{m}}$. Now the change in the mean value due to Δ is not measured in units of σ, but is measured in units of $\sigma_{\overline{x}}$. Thus, from these formulas,x_i values are replaced by $\overline{x_j}$ and the standard deviation of the process is $\sigma_{\overline{x}} = \frac{\sigma}{\sqrt{m}}$(Lu, Reynolds 1999a)

New process mean estimate

If there is a shift, a new process mean may be estimated from the formula:

$$\hat{\mu} = \begin{cases} \mu_0 + K + \frac{C_I^+}{N^+} \; pro \; C_I^+ > H \\ \mu_0 - K + \frac{C_I^-}{N^-} \quad C_I^- < -H \end{cases}$$

where N^+ and N^- are numbers of selected points in a moment (Chambers, Wheeler 1992), when $C_n^+ = 0$, or when$C_n^- = 0$, respectively

But lets see for example small differences between Swehart chart and a Cusum chart.. The two graphs below illustrate the different sensitivity of the graphs to change: observing the second CUSUM graph, it seems to detect deviations from intermediate values at both low and high values. In contrast, the first plot (swehart) does not detect such deviation (Harris, Ross 1991) and only at some higher values (Lu, Reynolds 1999b).

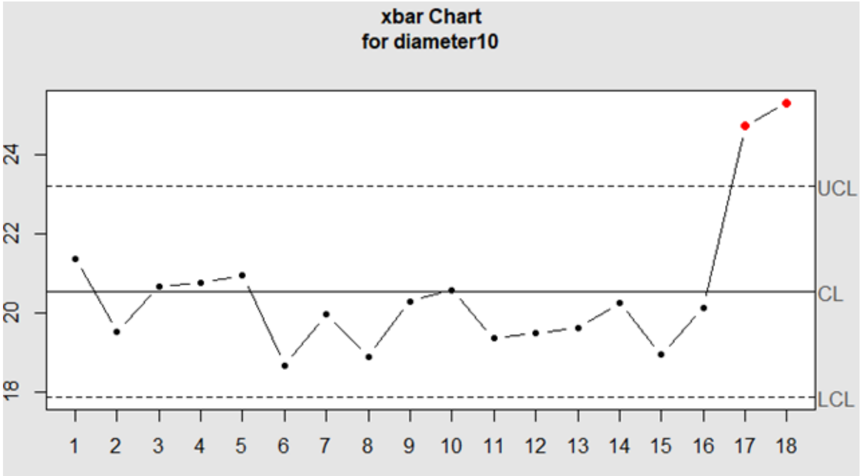

Figure 5 ***Shewhart´s control chart***

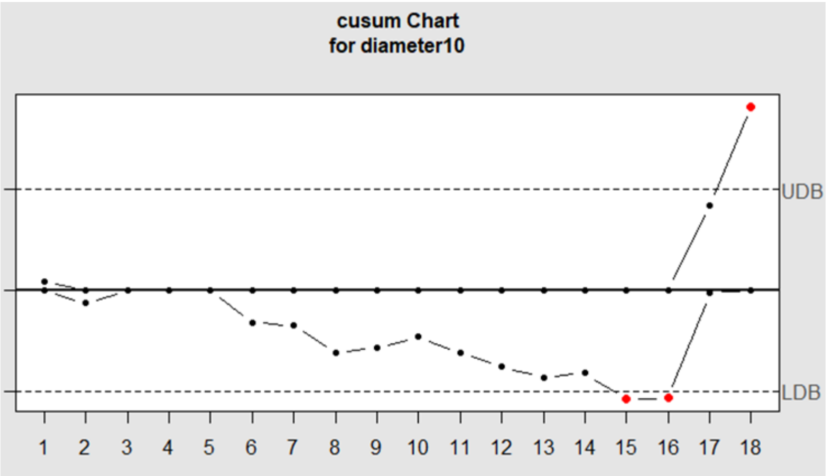

Figure 6 ***CUSUM control chart***

5.1.5 EWMA

This particular classification diagram was stimulated by Roberts (1959) and subsequently examined in more detail by Crowder (1987, 1989) and Lucas and Saccucci (1990). At the outset, it is important to note that these diagrams are based on the quantity

$$Z_t = \lambda X_t + (1-\lambda)X_{t-1}, \quad X_0 = \mu$$

Where X_0 is headstart value . For the purpose of monitoring the average price of the process. let X_t be the value of one observation and μ the average value of the understanding of X_t. Furthermore, the value of the parameter λ ∈(0,1]determines the amount of "importance" given to the value of the latest sample. To plot such a graph,in the case that X_t are independed we need to determine

$$\mu_{Z_t} = \mu_0$$

$$\left(\sigma_{z_t}\right)^2 = \sigma^2\left(\frac{\lambda}{2} - \lambda\right)[1-(1-\lambda)^{2t}]$$

The following shows how to construct a diagram with centerlines and boundaries.

LCL	Center Line	UCL
$\mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$	μ_0	$\mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$

As indicated above, the parameters L and λ must be determined to construct this figure. It is the statistical criterion that determines these parameters; L refers to the distance of the control limit from the centerline; values of 0.05, 0.20, and 0.10 are usually used for λ (Montgomery 2005). These parameters can be chosen based on the chart's ability to detect process changes quickly and avoid false alarms.

It is worth noting that the limits in the above table are called time-varying limits, given that there is a dependence between σ_{z_t} and t. However, as time passes, the values of the control limits tend to become constant, i.e., there tend to be steady-state control limits. These limits are shown below. The good thing about these charts is that they are robust to non-normal data understanding and are therefore more advantageous than Shewhart charts when monitoring a process through a single observation

LCL	Center Line	UCL
$\mu_0 - L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$	μ_0	$\mu_0 + L\sigma\sqrt{\frac{\lambda}{2-\lambda}}$

5.2 Correlated Data and quality control charts

In order to apply someone one of this basic control charts, must first satisfy certain assumptions about the data. Those assumptions are normality and symmetry of the data, constant variance and mean, independence, and absence of autocorrelation (Meloun and Militký, 2006).

The above assumptions apply to many statistical analysis procedures, particularly control charts. If these assumptions are not met, control charts are considered unreliable because they produce erroneous and incorrect conclusions and results. Violation of these assumptions can occur when working with industrial, medical, and chemical data.

Therefore, analysts should verify these assumptions using statistical tests. In addition, it should be noted that autocorrelation of data occurs all the time in chemical, pharmaceutical, food, and metallurgical processes. In addition, in environmentally relevant control processes, data are unevenly distributed and particularly asymmetric.

As noted above, the main condition for applying the chart is that the data be non-self-organizing. A treatment of this problem was proposed by Alwan and Roberts 1998. According to their work, the first step in this type of data is to first fit an ARIMA time series model. Then, after finding and applying the appropriate time series model, the analyst creates residuals. Then, a quality control chart (optional) is applied to the model residuals. The assumptions of data symmetry, normality, and non-self-consistency are satisfied. This theory is applied to the conversion rate index in the last section of this thesis.

Chapter 6: Marketing and conversion rate

6.1 Review of marketing and conversion rate

For some time now and according to Kaufman (1991), more and more businesses, mainly retail, in order to maintain and develop marketing strategies, have developed digital channels. In parallel, online retailers have become highly successful in their visitor traffic. Based on the above, it is predicted that e-commerce sales will increase immediately in the coming years. Retailers are therefore called upon to exploit and manage this development. They need to be invested in resources, functionality to increase the number of potential customers

However, in the media sector, resources, know-how as well as capital are quite limited, which should necessarily lead to investments and analyzes being done in a correct, efficient and original way, as there are no margins for mistakes (Grandon et al. , 2011• Cronin-Gilmore, 2012).In the past, thoughts about such problems in the online retail industry limited SME (small and medium enterprises) participation in e-commerce (Bharadwaj and Soni, 2007), but as online retailing has grown, such concerns have not they are more and more present.

It is widely known that Internet-based technologies are useful as an important tool for efficient sales and marketing (Bell and Loane 2010; Jeansson et al., 2017). An example of change is the Italian fashion industry. This industry is defined as a market where fashion products brought in 2016 revenues worth 90 billion euros and, contrary to the general trends of the sector, has shown positive growth rates in the recent past (source: Mediobanca Study Office) 1.2.

In this market the total of SMEs represents more than half of this market value. The amount of exports (as there was a significant increase of +5.9%) and online sales played an important role in this increase.

The SMEs have discovered this opportunity and are trying in every way to take advantage of it by facing the obstacles that arise. As mentioned above these obstacles can refer to the limited skills and resources available to the SMEs. All of the above reduce the interest to take initiatives on e-commerce. This has resulted in SMEs investing more carefully and purposefully in supporting business intelligence and communication strategies. (Stockdale and Standing, 2006). The development and construction of websites that serve transactional purposes, are more dangerous as they seek more investments (McDowell et al., 2016). Additionally, a problem in investing in online commerce is low conversion rates. The conversion rate will be defined as the number of customers who consume from the site to the total visitors (Ayanso and Yoogalingam, 2009; Viglia 2014). The present fraction will be expressed as a percentage. If this percentage is at low levels, then it does not make sense to invest in the internet.

An easy solution for SMEs is often to focus on large brands such as eBay and Amazon . These websites have a larger footprint on the internet and can bring SMEs in contact with many more customers and on these platforms the conversion rate is too high. However, there is clearly a problem here as well. Independent companies using third-party platforms must pay a commission on sales. In addition, the platforms have the ability to track customer data, with the result that small, medium-sized business customers cannot easily trust them.

Mauri(2003) has previously highlighted the difficulties SMEs face in building loyalty in traditional retail while in parallel data ownership issues can only reinforce such problems in the market on internet.

In addition, analysts are concerned about the appropriateness of using third-party platforms. In particular, analysts argue that these platforms may not support the resale of high value products. This is supported by the fact that consumer empiricism is not optimized as it should be, which is a negative for these consumers. (Joy et al., 2014).. In conclusion, the economies of scope provided by third-party platforms are not so useful for SMEs that have quite limited wealth, production, resources but several competitive products.

The solution to this problem is for the SME to develop its own online trading and trading site. In other words, they can create their own e-shop. However, only a small part of the SMEs that invest in the functionality of electronic exchanges are really developed. This fact has led writers and scientists to study and support that the management of conversion rates is a fundamental factor in the adoption process of digital technology by the media.

Although the conversion rates seem to be quite important for marketing strategies and sales, the research around topics is limited and focused. Mentioned studies have essentially focused on the study of consumer behaviors on the internet (Moe and Fader, 2004; Sismeiro and Bucklin, 2004; Hausman and Siekpe, 2009; Van Slyke et al., 2009; LorenzoRomero et al., 2014). Other research has focused on examining degree of performance that search engine optimization techniques can have if improved (Drèze and Zufryden, 2004; Ghose and Yang, 2009).

The decision for a business to adopt a website for trading is difficult and important, given the limitations of the business. This is because, as the internet develops rapidly, the goals and quality of products on the internet have evolved . Initially, entrepreneurs paid attention to the amount of visitors entering the website. But as the online marketplace matured, the emphasis shifted from website traffic to sales traffic. (Drew, 2003; Jelassi and Leenen, 2003).

In addition, the bibliography followed this evolution. But Nguyen et al. (2016) argue that initial researches focused mainly on using marketing tools in order to positively develop customer satisfaction and service, but there was a lack of research on using consumer service tools to guide consumer behavior, i.e. managing conversion rates and consequently sales.

As websites evolved into sales channels limited research has investigated the factors that positively influenced the intention to purchase. Examples of this type of work include website visits (Moe and Fader, 2004), browsing experience and navigational behaviour (Sismeiro and Bucklin, 2004), web design and language (Hausman and Siekpe, 2009 and website aesthetics and length of exposure (Lorenzo-Romero et al., 2014).

Over time, websites have turned into sales channels. Then they started small surveys which measured the factors that influence consumers to buy a product. Such studies were concerned

1. Study of influence and increase in attendance (Moe and Fader, 2004),

2. How it affects the browsing experience and navigation behavior (Sismeiro and Bucklin, 2004)

3. How the design (Hausman and Siekpe, 2009) and the aesthetics of the website (Lorenzo-Romero et al., 2014) affect the purchase intention.

These studies focus on purchase intent rather than actual sales, with the claim that an increase in purchase intent will increase the conversion rate. More generally, it has been proven that traditional retailing discourages customers from buying once they are in the store. An example in the above theory is by Söderlund et al. (2014), where he reports that 54% leave the retail store without buying anything. On the other hand, on websites, the percentage is between 2-5% (Holzwarth et al., 2006; Sohrabi et al., 2012). The explanations for the above could be the following

1) Visiting retail stores, for entertainment

2) Speed, price comparison, ease of use of online websites

3) Targeted purchase on an online page

4) Reduction of shopping time

But despite the proven usefulness of factors that influence and predict conversion rates, it is somewhat incomplete. Three examples will be studied below where the statistical analysis of conversion rates is examined and carried out.

1) Regression analysis
2) Time series analysis
3) Quality control Charts

6.2 Conversion rate and Regression analysis

Initially, a multiple linear regression will be studied that was carried out in the work of D Di Fatta, D Patton, G Viglia (Di Fatta, D., Patton, D., & Viglia, G. (2018). The data(observations=1.184) was retrieved from 6 different e-commerce media pages registered in the Italian fashion market. An OLS regression, with the stepwise procedure, checks the effect of the independent variables on the conversion rates.

In the above model, the predictive accuracy is quite high as the good fit index in both models is over 60%. According to the model, the disqcounts variable was the one that had the greatest degree of influence on the conversion rate. The variables season, speed of load, luxury contribute to the increase of the conversion rate, while the variable free shipping has a negative effect. The variables free returns and week were statistically insignificant

6.3 Conversion rate and time series analysis

An interesting application of time series analysis to conversion rate indicators appears to have been done by Illya Zinkevich and Tamara Radivilova. In this study, one year of daily data related to time series of conversion rates was collected.

After autocorrelation data were found to exist in this study, three models were constructed, which were then analyzed and compared. The three methods applied were time series analysis, networks of long-shot term memory, and decision tree methods.

To establish the validity of the models, the data were divided into two parts. That is, s is the data used to build the model and m is the data not included in the model. More specifically, for these m data, knowing their true values, the researchers performed a model fit to predict them and check how close they were to the true 7 values.

For the time series models, the MAProcess[0] model was constructed because it had the lowest AIC value among the time series models constructed. In conclusion, the study of conversion rates has shown that time series models are the simplest, fastest, and most convenient forecasting method.

On the other hand, while time series models appear to have a great many positive aspects, they are not applicable in the case of complex or long-term dependencies. Also, decision tree models are easy to learn and understand, but the choice of parameters is inconvenient and they do not work well when training on data with many features.

Finally, neural networks appear to be more accurate, better at prediction, and have smaller errors, but their training time, their cumbersome nature, and the large number of parameters to be estimated make them a more difficult solution.

6.4 Conversion rate and basic Quality Control charts

In the following section, an application of statistical quality control to the conversion rate indicator will be examined from the work of Paul Keller (Keller, Paul. *Statistical process control demystified*. McGraw-Hill Education, 2011) . It is recalled that the conversion rate is obtained by dividing the number of actual sales by the number of potential sales in a given time period.

An example could be given. Let's say an e-shop had, for example, 300 visits and they led to 98 sales, then the conversion rate is 32.7%. Consequently, the question asked here is: "Is this 32.7% better than expected or worse?" . it could be mentioned that there are procedures that are capable of comparing the present conversion rate with one of the past, in order to detect changes.

Additionally a confidence interval for the ratio can be constructed from the range 27.4% to 38.3% (assuming normality), using confidence interval for ratio techniques as described in Keller (2011). As extended with 95% confidence the index falls within the confidence interval and there is no evidence that the sample differs significantly from past experience.

Also a test for two ratios can be used to determine if the September conversion rate is equivalent to the historical conversion rate. For this test, historical data from an e-shop with approximately 9000 potential customers were used, of which 3000 bought in the end. The time period concerns 30 months. The hypothesis test gave a p-value= 0.91, and consequently the hypothesis of equality of proportions cannot be rejected. So the percentages are not statistically significantly different. In other words, there is no evidence that the conversion rates are different. (Note: We would typically look for a p-value of 0.05 or less to indicate that the conversion rates were different).

In this application, the limits of the graph were constructed in accordance with the procedure mentioned in the previous chapter. The upper limit is 041 i.e. 41% while the lower limit is 0.24 i.e. 24%. If the values go out of bounds then some change has occurred or the process is not working well. Consequently, we can see if a strategic marketing positively influenced the process and consequently the change in the chart. Along with this modality it is possible to determine the occurrence of specific causes that bring about changes.

In this way, the responsible marketing, e-shop owners will be able to identify changes, either positive or negative, in their sales, evaluate advertising campaigns and issues related to sales will be handled.

In addition to control limits, the research showed statistical run test rules can also be applied as appropriate to test for unlikely conditions indicative of a special cause. In this way the sensitivity of the control chart is enhanced. Observing each graph, unnatural patterns will be seen within control limits as seen in the application of the research. In the figure below, it seems that from November onwards there is a change in the distribution pattern of the conversion rate. These points were nine in a row and were below the center line of the process. The chance of this happening if the process had not been subject to a special cause of variation is extremely rare (less than 2/10 of 1%). The end result is that there has been a change since January. Also, once we detect such a change, which happened within the limits, then the limits can be adjusted again, as shown in the figure below, so that it is able to control new processes in relation to with sales.

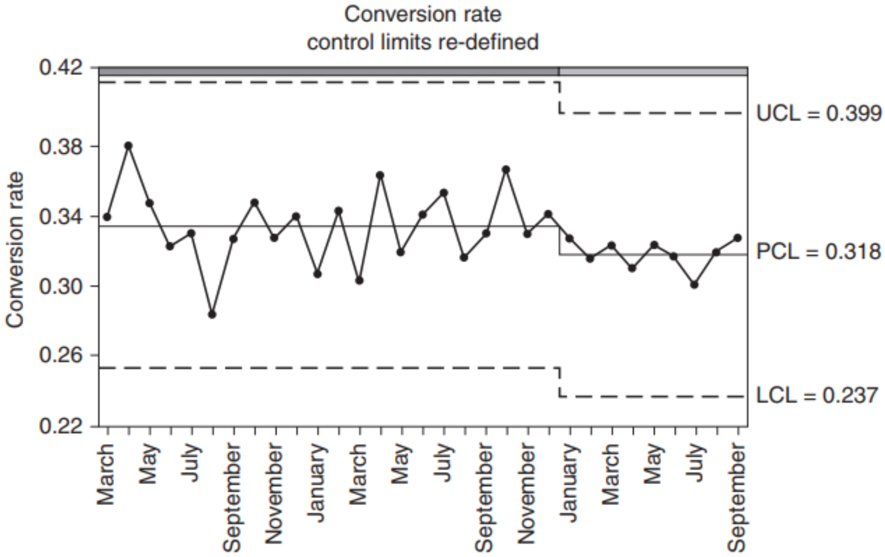

Figure 7 ***control chart for conversion rate***

The main advantage of this procedure is that the control chart detected a drift process that was not detected by classical statistical methods. These classical methods ignore the sequence of process data. This sequence often gives the sales managers very important information about the conversion rate.

In addition, these classical methods are applied to stable populations and not to processes. Process control charts add the element of time to the analysis: This additional dimension takes advantage of the inherent temporal sequence of data produced by a process. For example, even if a histogram is constructed from a properly sampled process, its temporal sequence is lost as the data is collapsed into the cells of the histogram.

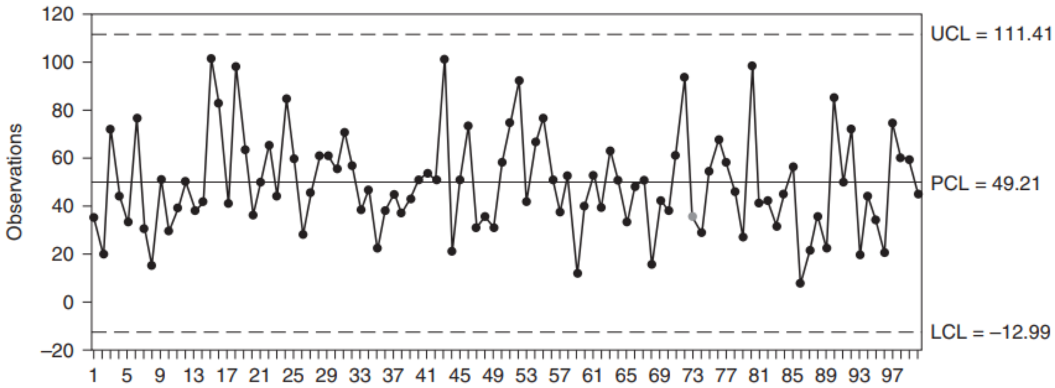

Figure 8 ***process in control***

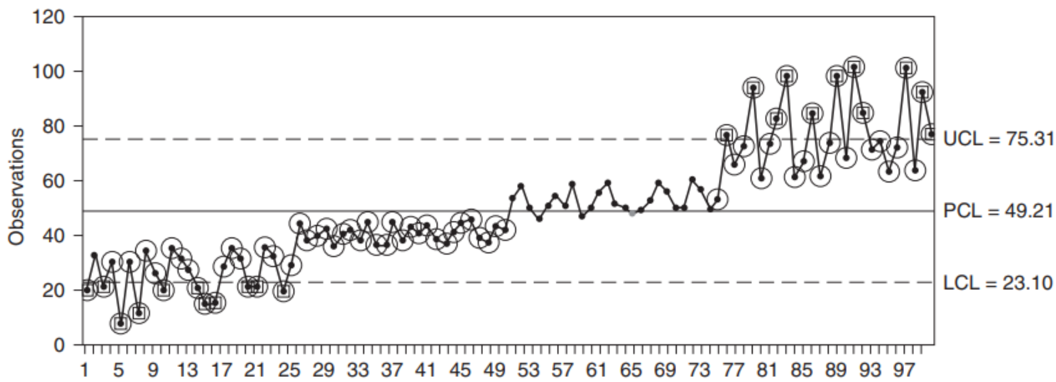

Figure 9 ***process out of control***

There appears to be a strong difference between the two graphs above. The first graph shows a stable process operating within statistical control limits; the second graph shows a process with a presumably positive trend. It is clear that the control process is out of control, and the histogram cannot not indicate this. As a result, the histogram ultimately cannot show such differences or order.

Control charts, on the other hand, can serve these purposes and are the only tool that can predict and evaluate process behavior. Studies have shown that histograms should always be accompanied by control charts. This may not be necessary in special cases where the data are census data or where there is no time series.

Thus, based on the above, a quality control chart is an appropriate tool that can identify difficulties, in which e-business can be improved and, above all, sales can be optimized.

Chapter 7: Application of time series and statistical quality control techniques on conversion rate

7.1 Methodology

In this chapter, the conversion rate is analyzed with two different statistical analysis techniques. The first one will concern the analysis of the index with chronological series. The index for direct and social channels is studied in more detail. In other words, the index will be studied by potential customers who enter the site directly, and potential customers who enter the site from a social network. (facebook, instagram, etc.).

The second way of statistical analysis is using quality control tests. Graphs will be created so that the manager or the person dealing with the marketing of the site can have an overall picture of the evolution of the conversion rate, while at the same time being able to anticipate any drops in order to make strategic changes for the business .

The data refer to a large furniture company based in Greece. The data refer to the conversion rate indicators for those customers who entered the company's site, entered and purchased (they will be called direct customers) and for those who visited the site through a social media application. The data refer to the years 2020 to 2023 and is weekly-. The analyzes were done in the spss statistical package and in the R language. It was proved that the time series conversion rates on both channels were auto-correlated. The test was done on the basis of Ljung -Box test Q where in both cases $p<0.05$ and therefore the tests were statistically significant

The time series models were constructed in R and after the appropriate model is selected for direct and social customers, the residuals of each model will be collected and 3 quality control charts will be applied (in SPSS) for each model. Consequently, 6 graphs will be constructed and compared for reliability and indicator change detection.

7.2 Some descriptive statistics

This section presents interesting descriptive and deductive data. The variables to be analyzed are time spent on the website (in seconds), total number of users per week, number of new registered users, number of transactions, and sales. The above will be analyzed differently for each of the two channels of customer inflow. Initially, in the social scatter plot between duration and conversion rate, there appears to be an exponential relationship. However, with the same variable, but for direct customers, this relationship appears more linear.

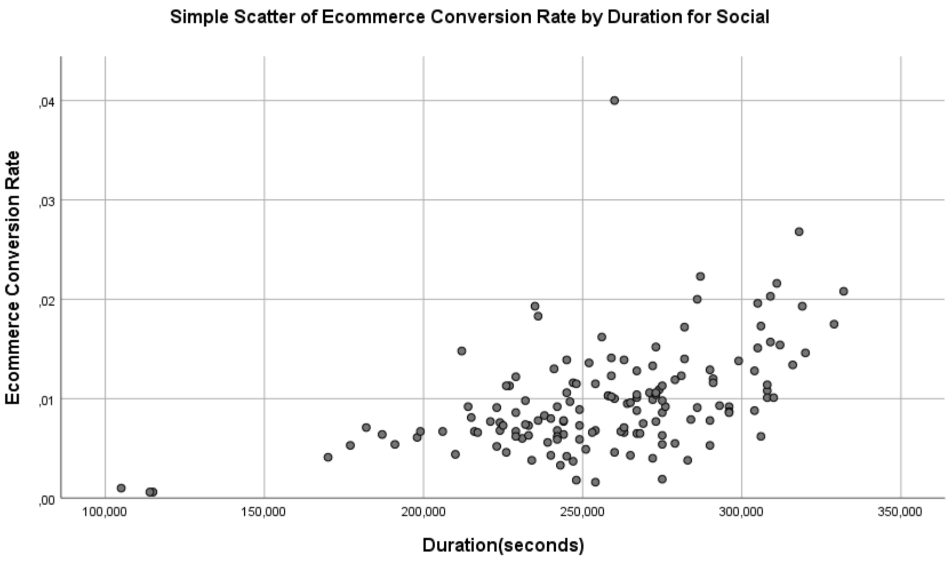

Figure 10 ***scatterplot of duration and conversion rate for social costumers***

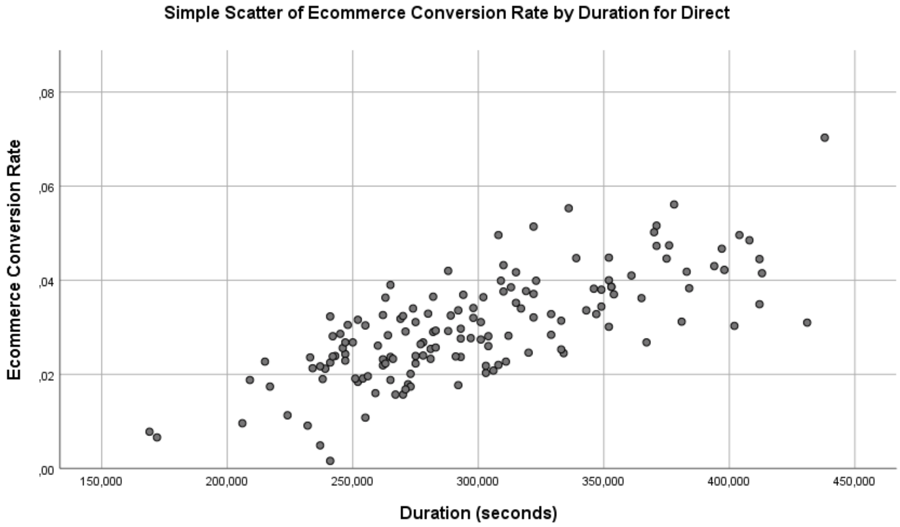

Figure 11 ***scatterplot of duration and conversion rate for direct costumers***

However, the question the company is asking is which channels should be used to invest more money in advertising campaigns. One way to answer this question is to use the Mann-Winey nonparametric test. This test does not care about the normality of the data (which is what is happening in our data) and performs a median test: for all four variables, direct group customers appear to be statistically significantly different from social group customers. Across all variables, direct group customers appear to have higher levels of time, number of users, number of new users, conversion rate, and revenue. As a result, the firm should focus its investments in advertising campaigns targeting direct customers.

Table 1 ***Medians of social and direct costumers with Mann Witney test***

	social	direct	Total	Mann-Witney p-value
Duration (seconds)	259	292	272	,000
Users	1291	1554,5	1424	,000
New Users	848	1339,5	1119	,000
Ecommerce Conversion Rate	0,89%	2,92%	,0173	,000
Transactions	14	71	38	,000
Revenue	1872,015	12636,22	5521,03	,000

7.3.1 Time series analysis–Social conversion rate data

As mentioned above, two time series models were constructed, one for the direct customer channel and one for the social customer channel. The following table shows the various models created for each category, one for each category was ultimately selected. The selection was based on the criteria for choosing the most appropriate model, as described in a previous chapter. The two best-fit models selected are developed and analyzed below.

Table 2 **ARIMA models for both direct and social costumers**

Social	*mape*	R^2	*mae*	*bic*	*rmse*	*dw*
Arima(1,0,3)	63,51	0,15	0,33	-1,47	0,43	>0,05
Arima(1,0,1)	64,63	0,11	0,33	-1,51	0,44	>0,05
Arima(1,0,17)	59,28	0,23	0,31	-0,99	0,44	<0,05
Arima(3,0,3)	63,49	0,17	0,33	-1,41	0,43	>0,05
Arima(2,0,2)	64,79	0,11	0,33	-1,43	0,45	>0,05
Arima(3,0,2)	63,5	0,16	0,33	-1,45	0,43	>0,05
Direct	*mape*	R^2	*mae*	*bic*	*rmse*	*dw*
Arima(1,0,0)	33,03	0,45	0,62	-0,33	0,82	>0,05
Arima(7,0,7)	33,03	0,53	0,57	0,039	0,79	>0,05
Arima(1,0,2)	33,37	0,48	0,6	-0,301	0,804	>0,05
Arima(1,1,2)	32,026	0,47	0,61	-0,29	0,8	>0,05

The main objective was to achieve stationarity in the "conversion rate" time series. So the first thing is to select, the appropriate orders p and q so that the final ARIMA model will be the best model. In this case, the model selected is the ARIMA(1,0,3) model. This model seems to fit the data well as shown by the Ljung-Box Q test (p-value>0.05). In addition, the following table shows the statistical measures of the chronological series.

Table 3 ***goodness of fit***

Model	Number of Predictors	Model Fit statistics: Stationary R-squared	Model Fit statistics: Normalized BIC	Ljung-Box Q(18): Statistics	Ljung-Box Q(18): DF	Ljung-Box Q(18): Sig.	Number of Outliers
Ecommerce Conversion Rate-Model_1	0	,150	-1,476	10,765	14	,704	0

Table 4 ***Statistics metrics of goodness of fit***

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	,150	.	,150	,150	,150	,150	,150	,150	,150	,150	,150
R-squared	,150	.	,150	,150	,150	,150	,150	,150	,150	,150	,150
RMSE	,439	.	,439	,439	,439	,439	,439	,439	,439	,439	,439
MAPE	63,516	.	63,516	63,516	63,516	63,516	63,516	63,516	63,516	63,516	63,516
MaxAPE	1038,579	.	1038,579	1038,579	1038,579	1038,579	1038,579	1038,579	1038,579	1038,579	1038,579
MAE	,333	.	,333	,333	,333	,333	,333	,333	,333	,333	,333
MaxAE	1,451	.	1,451	1,451	1,451	1,451	1,451	1,451	1,451	1,451	1,451
Normalized BIC	-1,476	.	-1,476	-1,476	-1,476	-1,476	-1,476	-1,476	-1,476	-1,476	-1,476

Then the models coefficients values are in the table below. According to this table in MA lag 1 and Lag 2 are not statistically significant while lag 3 of MA and La 1 of AR are statistically significant .

Table 5 ***Coefficients of the optional model***

					Estimate	SE	t	Sig.
Ecommerce Conversion Rate-Model_1	Ecommerce Conversion Rate	No Transformation	Constant		,948	,073	13,007	,000
			AR	Lag 1	,493	,201	2,453	,015
			MA	Lag 1	,265	,201	1,321	,189
				Lag 2	-,027	,094	-,283	,777
				Lag 3	-,273	,094	-2,894	,004

The pacf & acf plot shows the autocorrelation and partial autocorrelation plots of the residuals. However, it is clear that the coefficients are within confidence limits. This fact is a positive diagnostic factor for the model's fit. Thus, on this basis, the model indicates that, to a sufficient degree, the indicator under study can be analyzed.

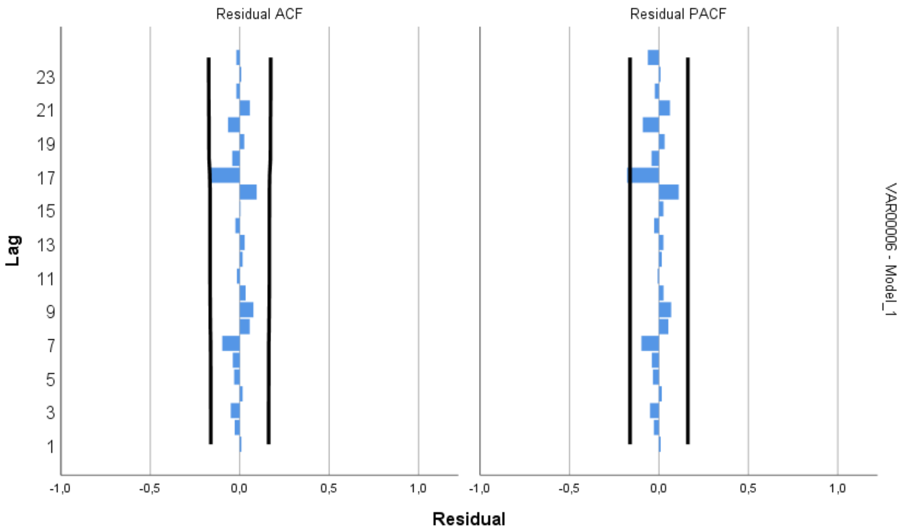

Figure 12 ***PACF & ACF of conversion rate time series model for social chanel***

In addition, a normality test of the model errors was performed. The test was carried out based on the Kolmogorov-Smirnov test. as seen from the table, the errors do not seem to differ from the normal distribution shape as p-value>0.05. As a result of all the above, the conclusion can be drawn that the model is capable of describing and predicting the conversion rate index for the social channel.

Table 6 ***Tests of Normality of residuals***

	Statistic	df	Sig.
Noise residual from time serie model -Model_1	,068	147	,092

a. Lilliefors Significance Correction

And the graph below confirms the degree of adaptation of the model. The red line shows the actual data and the blue line shows the goodness of fit of the model. In other words, it appears that the model adapts to the data quite well and in this way the model is able to make predictions.

The figure shows upper and lower limits of the estimates. The graph then shows the forecasts for the next four weeks. The forecast table shows the values the indicator can take as well as the confidence intervals. However, it is worth pointing out that there is quite a lot of variability at the beginning of the graph, since this time series is also the beginning of the corona various. As a result, there are some large values of the conversion rate, but also some very low values. This could be explained by the fact that while some people purchased online, many did not due to financial constraints.

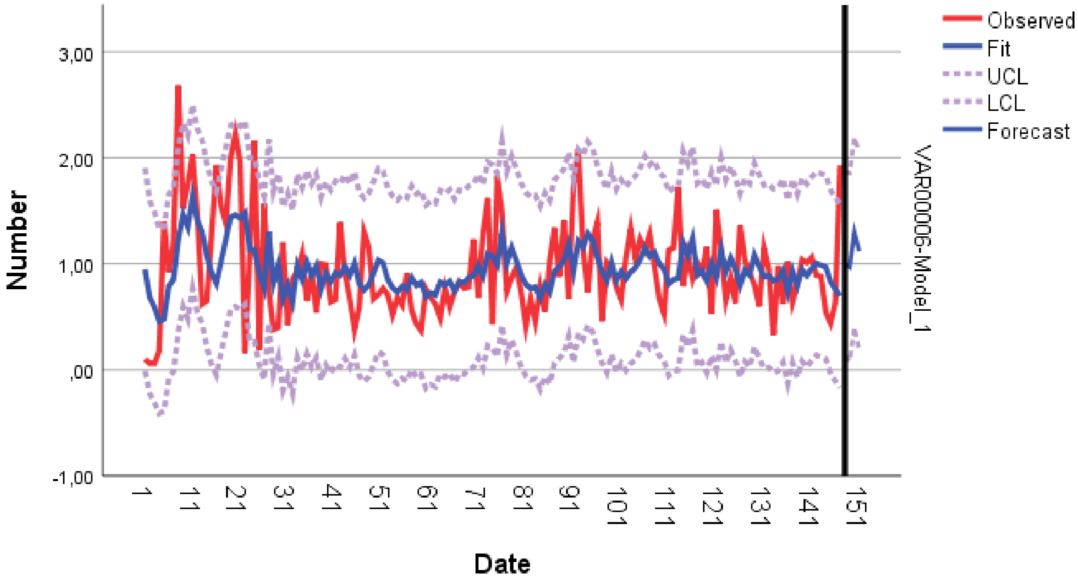

Figure 13 ***Time series plot for fitted model on social costumers***

Below are the predictions that the model manages to make in comparison with the actual prices we have and it seems that maybe the model is good at predicting the conversion rate for social customers.

Table 7 ***Forecasted values of social model***

week	real	forecast
147	0,54	0,67
148	0,44	0,51
149	0,63	0,65
150	1,93	1,5

7.3.2 Time series analysis–direct conversion rate data

The main objective was to achieve stationarity in the "conversion rate" time series. We select the best p and q so that the final ARIMA model could describe the time series in an optimal way. In this case, the model selected is the ARIMA(1,0,0) model. This model seems to fit the data well as shown by the Ljung-Box Q test (p-value>0.05). In addition, the following table shows the statistical measures of the chronological series.

Table 8 ***goodness of fit***

		Model Fit statistics	Ljung-Box Q(18)			
Model	Number of Predictors	Stationary R-squared	Statistics	DF	Sig.	Number of Outliers
Ecommerce Conversion Rate-Model_1	0	,448	25,560	17	,083	0

Table 9 ***Statistics metrics of goodness of fit***

Fit Statistic	Mean	SE	Minimum	Maximum	Percentile 5	10	25	50	75	90	95
Stationary R-squared	,448	.	,448	,448	,448	,448	,448	,448	,448	,448	,448
R-squared	,448	.	,448	,448	,448	,448	,448	,448	,448	,448	,448
RMSE	,820	.	,820	,820	,820	,820	,820	,820	,820	,820	,820
MAPE	33,034	.	33,034	33,034	33,034	33,034	33,034	33,034	33,034	33,034	33,034
MaxAPE	972,182	.	972,182	972,182	972,182	972,182	972,182	972,182	972,182	972,182	972,182
MAE	,620	.	,620	,620	,620	,620	,620	,620	,620	,620	,620
MaxAE	3,460	.	3,460	3,460	3,460	3,460	3,460	3,460	3,460	3,460	3,460
Normalized BIC	-,330	.	-,330	-,330	-,330	-,330	-,330	-,330	-,330	-,330	-,330

Then the coefficient values are n the table below. According to this table of the AR , lag1 is statistical significant. So the model is an AR(1) time series model.

Table 10 ***Coefficients of the optional model***

					Estimate	SE	t	Sig.
Ecommerce Conversion Rate-Model_1	Ecommerce Conversion Rate	No Transformation	Constant		2,955	,207	14,293	,000
			AR	Lag 1	,679	,062	10,963	,000

The pacf & acf plot shows the autocorrelation and partial autocorrelation plots of the residuals. However, it is clear that the coefficients are within confidence limits. This fact is a positive diagnostic factor for the model's fit. Thus, on this basis, the model indicates that, to a sufficient degree, the indicator under study can be analyzed.

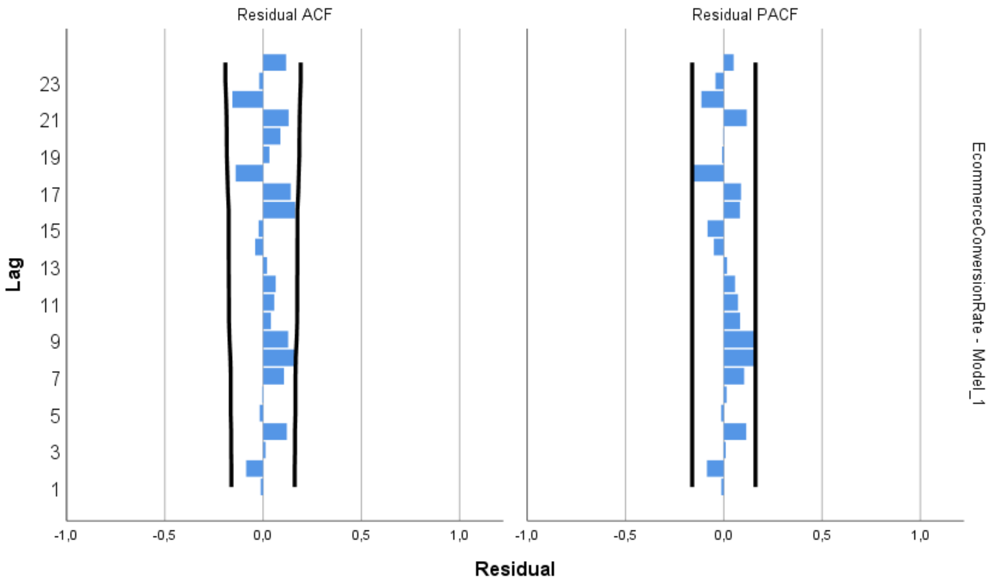

Figure 14 ***PACF & ACF of conversion rate time series model for social chanel***

In addition, a normality test of the model errors was performed. The test was carried out based on the Kolmogorov-Smirnov test. as seen from the table, the errors do not seem to differ from the normal distribution shape as p-value>0.05. As a result of all the above, the conclusion can be drawn that the model is capable of describing and predicting the conversion rate index for the social channel.

Table 11 ***Tests of Normality of residuals***

	Kolmogorov-Smirnov[a]		
	Statistic	df	Sig.
Noise residual from EcommerceConversion-Rate-Model_1	,072	148	,066

a. Lilliefors Significance Correction

And the graph below confirms the degree of adaptation of the model. The red line shows the actual data and the blue line shows he goodness of fit of the model. In other words, it appears that the model adapts to the data quite well and in this way the model is able to make predictions.

The figure shows upper and lower limits of the estimates. The graph then shows the forecasts for the next four weeks. The forecast table shows the values the indicator can take as well as the confidence intervals.

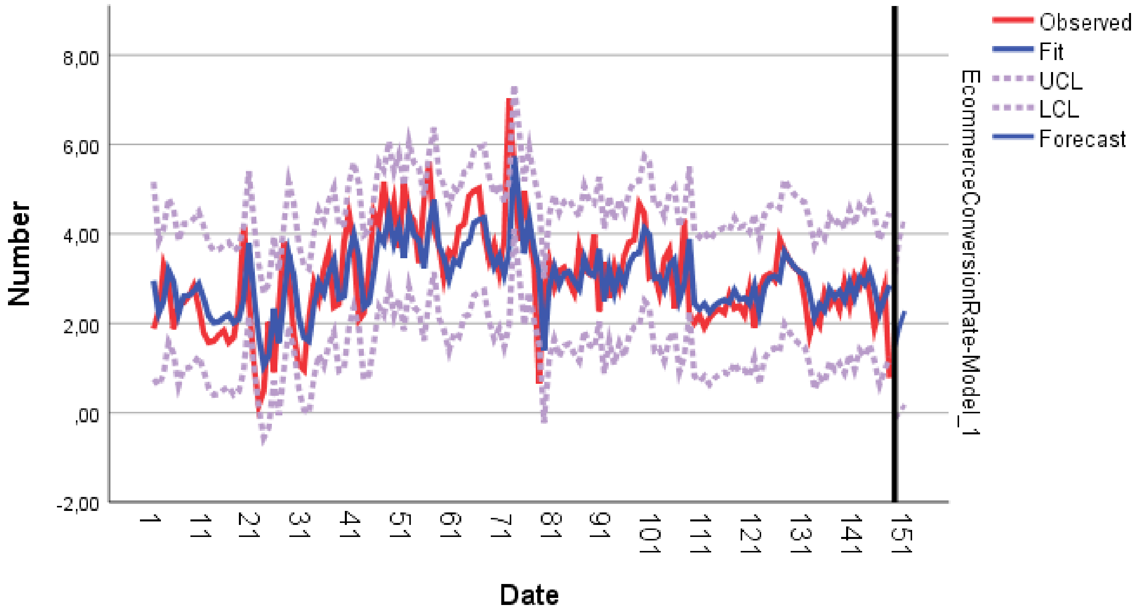

Figure 15 ***Time series plot for fitted model on direct costumers***

Below are the predictions that the model manages to make in comparison with the actual prices we have and it seems that maybe the model is good at predicting the conversion rate for social customers.

Table 12 ***Forecasted values of direct model***

week	real	forecast
146	2,38	2,25
147	2,81	2,49

7.4 Application of statistical quality control charts

Once the appropriate time series models for the direct and social customers have been constructed, the statistical quality control charts will be applied to their residuals, once the data of phase 1 and 2 have been delimited in each type of customer. For the construction of the graphs, the values λ=0.20 were used for the ewma graphs and for the distances of the limits L=2.7 standard deviations were used. However, in order to carry out conclusions from the control charts about the strategies to improve the conversion rate, some rules of interpretation of these charts should be mentioned. The way that will help in the interpretation of the charts is the 4 rules of flows, which were introduced by the Western Electric Company. These rules will help to identify irregular patterns and are known as WESTERN ELECTRIC RULES. The rules are

1. A note out of control

2. Two out of three consecutive points outside the 2 warning control limits (on the same side of the centre line)

3. Four of the five consecutive points to be outside the 1st inner warning control limits (to the same side of the centre line)

4. 8 consecutive points on the same side of the centre line

7.4.1 Quality control charts for data from Direct costumers conversion rate

Initially for direct costumers, phase 1 is defined as the data selected from week 1 to week 40 while phase 2 consists of data from week 41 to the end. The first graph to be examined is the simple Sewhart's common graph. The present company is running an advertising campaign on television mentioning its site during the advertisement . The ad ran from week 47 to week 100. Shewhart's chart identified an out-of-control point where the product is at the top of the ad, the site begins to be known. But while the ad continues the shewhart chart tells us that the process is back in control with the conversion rate indicator returning to low levels. More specifically after week 77 the chart suggests that perhaps the advertising should be stopped and something different should be implemented. Then the CUSUM chart gave 68 points out of control, i.e. as long as the duration of the targeted advertising so it seems that the swehart chart quickly detects huge changes but is not as sensitive to small changes that have occurred with the start of the advertising. The CUSUM chart detects these small changes. In addition, the EWMA chart was also able to detect the change, but at week 72 and afterwards we are told that the control is back within limits. But after week 77 the points may appear to be within boundaries but a non-random pattern appears to be forming, as according to the flow channel number 4, more than 8 points are on one side of the central line. On the other hand, the swehart chart detected a large change in week 77, which was not detected by the other two graphs. Thus confirming the theory that the swehart charts detect sudden large changes more easily than the other two charts. We would still recommend the CUSUM chart as it is the one that showed what happened in reality and indicates how effective the advertising that was done on the part of the company was.

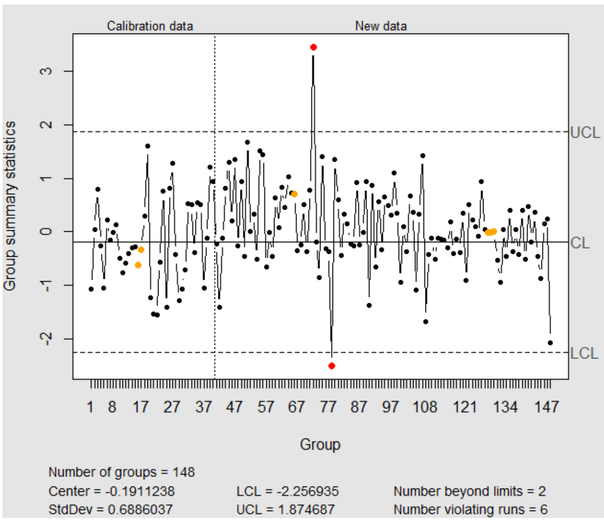

Figure 16 ***Sewhart chart of conversion rate of direct costumers***

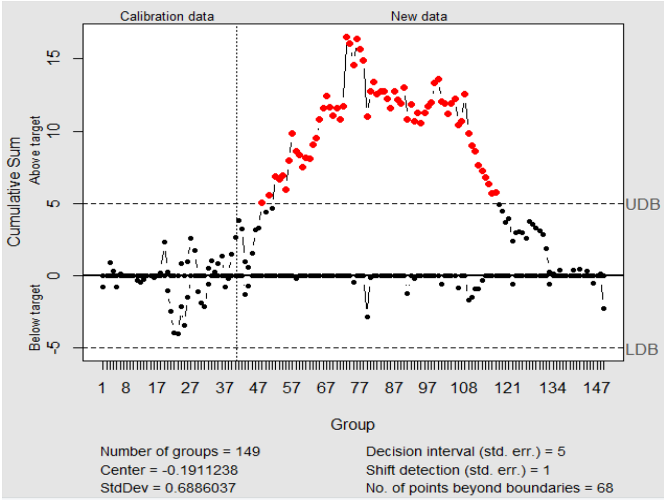

Figure 17 ***CUSUM chart of conversion rate of direct costumers***

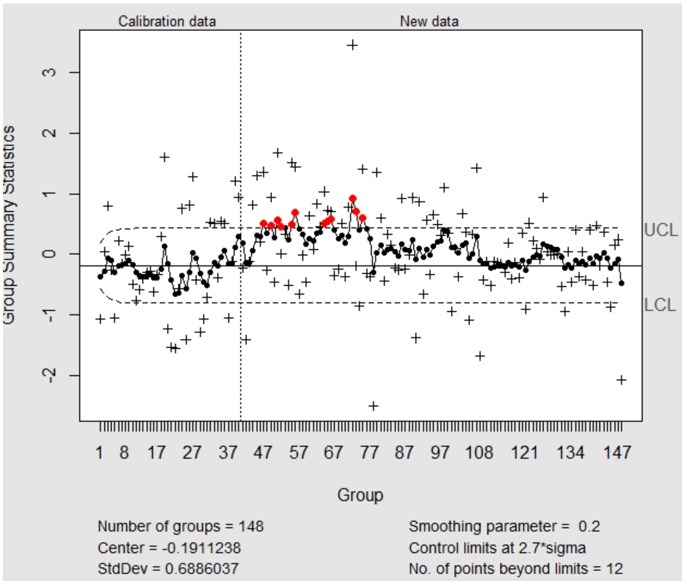

Figure 18 ***EWMA chart of conversion rate of direct costumers***

7.4.2 Quality control charts for data from Social costumers conversion rate

Below we will describe the charts concerning the residuals from the time series model built for customers coming from the social media channel. Initially for direct costumers, phase 1 is defined as the data selected from week 1 to week 16 while phase 2 consists of data from week 17 to the end First of all, it can be observed that in all three charts, from the twentieth week onwards , points are identified out of control. More specifically, it appears that the conversion rate indicator has reached high points. One explanation of the phenomenon is that the company in question carried out advertising through influencers. More specifically, in the period from August to September, the company invested money in people who, through their social networks, promoted the furniture company's products.

In addition, it is worth noting that the company carried out this investment between August and October in week 16 onwards. These months are quite demanding as far as the purchase of furniture is concerned, as young students choose new houses and need new furniture. This investment has been quite successful as all three charts show similar things with some differences. The differences are that the swehart chart identified a negative change and a point off boundaries below the centerline, at 40, which was not identified by the other two charts. In week 18 the EWMA chart gave a point out of control as the index had fallen to low levels and it was the right time to make a breakout investment as it was done. However, if the analyst or the manager observed only the other two charts he might not have been in a position to make a decision to initiate an advertisement.

Additionally in week 111 the company decides to make a 5% discount on its products. This very small change was only detected by EWMA, while the other two charts did not detect it. This discount policy may not be that good as it did not succeed in bringing the index to very high levels. So the company should pay special attention to the influencers as they seem to bring the index to high levels. Finally, it should be pointed out that in this case too the cusum and ewma charts are more sensitive to small changes while the sewhart is more sensitive to large changes. However, ewma seemed to detect and detect that changes were taking place consistently and seemed better.

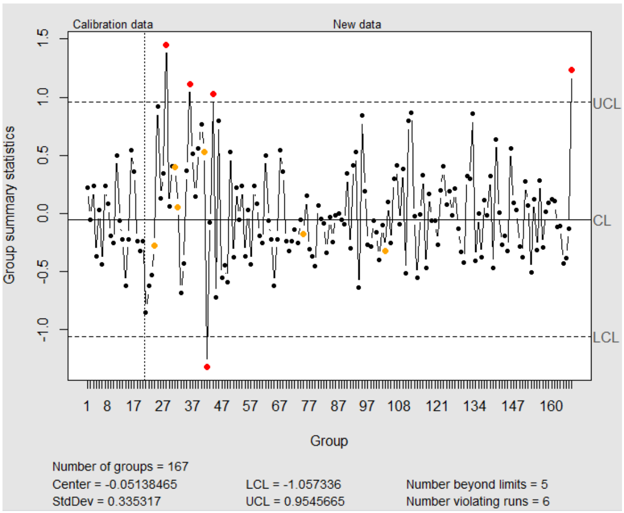

Figure 19 ***Sewhart chart of conversion rate of social costumers***

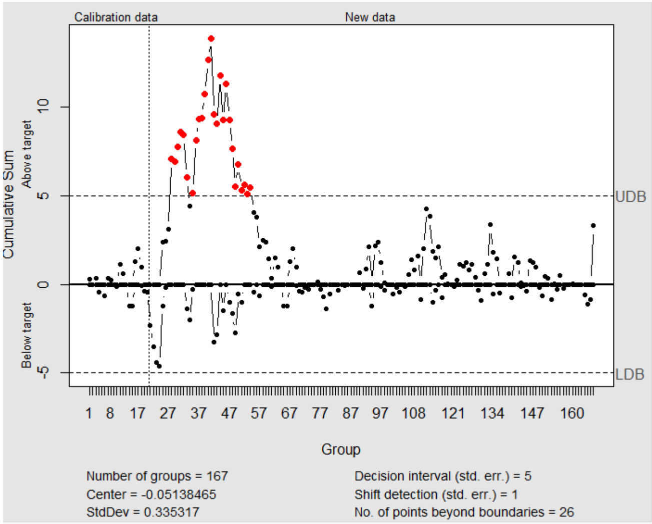

Figure 20 ***CUSUM chart of conversion rate of social costumers***

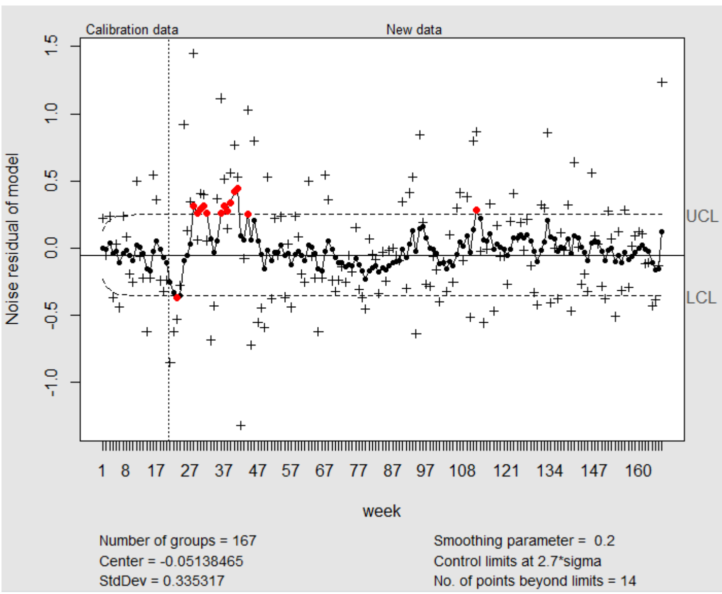

Figure 21 ***EWMA chart of conversion rate of social costumers***

Chapter 8 : Discussion

8.1 Conclusions

In this degree, attempts were made to connect the theory of business marketing with the analytical methods of the statistical field. In more detail, graduates dealt with conversion rates, which are very important indicators that have the ability to provide a complete picture of the business progress of sales. This index is reported as a percentage of those who eventually became buyers of the service than those who visited the online store completely.

After that, an attempt is made to analyze the conversion rate, after the corresponding reference is preceded by the necessary theory. The two methods of analysis were time series analysis and statistical quality control.

The first is data related to the conversion rate of the furniture business for the period 2020-2023. All companies have advertising channels. As a result, the data collected conversion rate data for customers who came to the site (direct costumers) and customers who came from ads on social media (social costumers) showed that the level of conversion rate was lower for direct costumers than for social ones. The same thing happened with the sales. In addition, after the data was tested for autocorrelation, we were able to make 2 time series models for each channel. These models seemed to be able to meet all the necessary requirements and predict conversion rate indicators. So the graduates were able to predict the indicators for each channel for the next 3 weeks.

Then statistical quality control techniques were applied. In particular, since the coverage rate indicator showed autocorrelation, the statistical quality control graph was applied to the residuals of the previously constructed time series. 3 basic charts were built: Swehart, cusum and ewma. The graph in both cases of the channel can detect changes and can be sensitive to small or small changes, as already suggested by the literature.

With the above technology, not only the product marketing department is in a position where you can predict conversion rates, but you can estimate sales. This is very important because the company or site will be able to secure future balances as well as be able to predict future positions in the market pie, and it will be able to arrange orders that should be made to its suppliers. In addition to quality control, managers can oversee the entire process of the business of the product or service produced because the indicators are the result of the entire process, and in conclusion, these changes may not have been detected, studied and improved in time if they were not for statistical quality control charts.

8.2 Limitations of the study - Suggestions for further research

This study deals with data and analysis on e-shops in one industry. Therefore, it is a fairly targeted analysis and cannot be generalized to all companies in all industries. This is because there are no dependencies between the data and direct application of time series applications or control charts may not be necessary. Therefore, analysts should seriously consider the extent to which data do or do not have dependencies. Further research could include a comprehensive study of conversion rates across the industry to understand how firms are addressing their optimization. This necessary control chart would also allow the analyst to study what is qualitatively happening at the macro level of the industry and what is affecting it.

In addition, another drawback is consulting on which channels (and which customer channels) to invest funds in, without considering the accounting aspects. In other words, after conducting the financial analysis of company data necessary to achieve a high ROI, the company must select investments that will truly benefit the company. Thus, in a study such as the one described above, the analyst can use this survey data to study, from a business perspective, the strategies that should be taken to determine if investing in direct customers is truly in the company's best interest. It is possible that an advertising campaign to direct customers will avoid generating more revenue (as analyzed above), but it is not certain whether it will provide more actual benefit.

APPENDIX

Table 13 ***Mann-Witney Ranks***

	Chanel	N	Mean Rank	Sum of Ranks
Duration (seconds)	social	148	115,00	17020,00
	direct	148	182,00	26936,00
	Total	296		
Users	social	148	129,63	19185,50
	direct	148	167,37	24770,50
	Total	296		
New Users	social	148	106,65	15784,50
	direct	148	190,35	28171,50
	Total	296		
Ecommerce Conversion Rate	social	148	81,44	12052,50
	direct	148	215,56	31903,50
	Total	296		
Transactions	social	148	82,06	12145,00
	direct	148	214,94	31811,00
	Total	296		
Revenue	social	148	81,47	12057,00

	direct	148	215,53	31899,00
	Total	296		

Table 14 ***Mann-Witney tests***

	Mann-Whitney U	Wilcoxon W	Z	Asymp. Sig. (2-tailed)
Duration (seconds)	5994,000	17020,000	-6,734	,000
Users	8159,500	19185,500	-3,793	,000
New Users	4758,500	15784,500	-8,412	,000
Ecommerce Conversion Rate	1026,500	12052,500	-13,481	,000
Transactions	1119,000	12145,000	-13,358	,000
Revenue	1031,000	12057,000	-13,474	,000

a. Grouping Variable: Chanel

BIBLIOGRAPHY

Ayanso, A., & Yoogalingam, R. (2009). Profiling retail web site functionalities and conversion rates: A cluster analysis. *International Journal of Electronic Commerce*, *14*(1), 79-114. https://doi.org/10.2753/JEC1086-4415140103

Bell, J., & Loane, S. (2010). ‘New-wave’global firms: Web 2.0 and SME internationalisation. *Journal of Marketing Management*, *26*(3-4), 213-229. https://doi.org/10.1080/02672571003594648

Bharadwaj, P. N., & Soni, R. G. (2007). E-commerce usage and perception of e-commerce issues among small firms: results and implications from an empirical study. *Journal of small business management*, *45*(4), 501-521. DOI: 10.1111/j.1540-627X.2007.00225.x

Burrill, C. W., & Ledolter, J. (1999). Achieving quality through continual improvement. *(No Title)*.

Di Fatta, D., Patton, D., & Viglia, G. (2018). The determinants of conversion rates in SME e-commerce websites. *Journal of Retailing and Consumer Services*, *41*, 161-168. https://doi.org/10.1016/j.jretconser.2017.12.008

Ghose, A., & Yang, S. (2009). An empirical analysis of search engine advertising: Sponsored search in electronic markets. *Management science*, *55*(10), 1605-1622. https://doi.org/10.1287/mnsc.1090.1054

Grandón, E. E., Nasco, S. A., & Mykytyn Jr, P. P. (2011). Comparing theories to explain e-commerce adoption. *Journal of Business research*, *64*(3), 292-298. https://doi.org/10.1016/j.jbusres.2009.11.015

Harris, T. J., & Ross, W. H. (1991). Statistical process control procedures for correlated observations. *The canadian journal of chemical engineering*, *69*(1), 48-57. **https://doi.org/10.1002/cjce.5450690106**

Hausman, A. V., & Siekpe, J. S. (2009). The effect of web interface features on consumer online purchase intentions. *Journal of business research*, *62*(1), 5-13. https://doi.org/10.1016/j.jbusres.2008.01.018

Holzwarth, M., Janiszewski, C., & Neumann, M. M. (2006). The influence of avatars on online consumer shopping behavior. *Journal of marketing*, *70*(4), 19-36. https://doi.org/10.1509/jmkg.70.4.019.

Jelassi, T., & Leenen, S. (2003). An E-Commerce Sales Model for Manufacturing Companies:: A Conceptual Framework and a European Example. *European Management Journal*, *21*(1), 38-47. https://doi.org/10.1016/S0263-2373(02)00151-2

Juran, J. M. (1999). How to think about quality. *JM Juran, AB Godfrey, RE Hoogstoel, and EG, Schilling (Eds.): Quality-Control Handbook. New York: McGraw-Hill*

Keller, P. (2011). *Statistical process control demystified*. McGraw-Hill Education.

Kotler, P., Armstrong, G., Saunders, J., Wong, V., Miquel, S., Bigné, E., & Cámara, D. (2000). *Introducción al marketing*. Pearson Prentice Hall.

Lorenzo-Romero, C., Constantinides, E., & Brünink, L. A. (2014). Co-creation: Customer integration in social media based product and service development. *Procedia-Social and Behavioral Sciences*, *148*, 383-396. https://doi.org/10.1016/j.sbspro.2014.07.057

Lu, C. W., & Reynolds Jr, M. R. (1999). EWMA control charts for monitoring the mean of autocorrelated processes. *Journal of Quality technology*, *31*(2), 166-188. https://doi.org/10.1080/00224065.1999.11979913

Lucas, J. M., & Saccucci, M. S. (1990). Exponentially weighted moving average control schemes: properties and enhancements. *Technometrics*, *32*(1), 1-12. DOI:10.1080/00401706.1990.10484583.

McCambley, J. (2013). The first ever banner ad: why did it work so well. *The Guardian,[online]*, *12*, 12.

McDowell, G., Ford, J., & Jones, J. (2016). Community-level climate change vulnerability research: trends, progress, and future directions. *Environmental Research Letters*, *11*(3), 033001. **DOI** 10.1088/1748-9326/11/3/033001

Meloun, M., & Militký, J. (2006). Compendium of statistical data processing. *Prague: Academia*.

Moe, W. W., & Fader, P. S. (2004). Dynamic conversion behavior at e-commerce sites. *Management Science*, *50*(3), 326-335. https://doi.org/10.1287/mnsc.1040.0153

Montgomery, D. C. (1997). Response surface methods and other approaches to process optimization. *Design and analysis of experiment*, 427-510.

Montgomery, D. C. (2005). Design and analysis of experiments (6th edn.,) john wiley and sons. *New York, NY*.

Neal, A., Griffin, M. A., & Hart, P. M. (2000). The impact of organizational climate on safety climate and individual behavior. *Safety science*, *34*(1-3), 99-109. https://doi.org/10.1016/S0925-7535(00)00008-4

Papamarkos, N., Tzortzakis, J., & Gatos, B. (1996, October). Determination of run-length smoothing values for document segmentation. In *Proceedings of third international conference on electronics, circuits, and systems* (Vol. 2, pp. 684-687). IEEE. doi: 10.1109/ICECS.1996.584454.

Roberts, S. W. (2000). Control chart tests based on geometric moving averages. *Technometrics*, *42*(1), 97-101. DOI: 10.1080/00401706.2000.10485986

Shewhart, W. A., & Deming, W. E. (1986). *Statistical method from the viewpoint of quality control*. Courier Corporation.

Sismeiro, C., & Bucklin, R. E. (2004). Modeling purchase behavior at an e-commerce web site: A task-completion approach. *Journal of marketing research*, *41*(3), 306-323. https://doi.org/10.1509/jmkr.41.3.306.35985.

Söderlund, H., Moscovitch, M., Kumar, N., Daskalakis, Z. J., Flint, A., Herrmann, N., & Levine, B. (2014). Autobiographical episodic memory in major depressive disorder. *Journal of abnormal psychology*, *123*(1), 51. https://doi.org/10.1037/a0035610

Stockdale, R., & Standing, C. (2006). A classification model to support SME e-commerce adoption initiatives. *Journal of small business and enterprise development*, *13*(3), 381-394. ttps://doi.org/10.1108/14626000610680262

Tzortzakis, K., & Tzortzaki, A. (2002). The principles of Marketing a Greek approach. *title in Greek:"Αρχές Μάρκετινγκ-Η ελληνική προσέγγιση*.

Wheeler, D. J., & Chambers, D. S. (1992). Understanding statistical process control. *Knoxville*.

Πασχόπουλος, Α., & Σκαλτσάς, Π. (2006). Ηλεκτρονικό εμπόριο: Επιχειρηματική στρατηγική και marketing στο διαδίκτυο. *Αθήνα: Κλειδάριθμος*.