13th International Conference
EBEEC 2021 - “Economies of the Balkan and Eastern European Countries”

SAMPLING ISSUES AND ECO-NETWORKS ON INNOVATION MANAGEMENT, QUANTITATIVE RESEARCH STUDIES CHALLENGES

T.S. Spyropoulos[1], M. Papageorgiou[2]
[1] *Department of International Business, Perrotis College, Thessaloniki, Greece, hspyro@afs.edu.gr*
[2] *Department of Mathematics, Aristotle University of Thessaloniki, Thessaloniki, Greece, mmpapageor@math.auth.gr*

ABSTRACT
The study examines the implications of Statistical Analysis when used for analyzing complex systems and ecosystems, such as the ones that represent networks of innovation and entrepreneurship. There is a serious concern regarding the effectiveness of using Quantitative Research Tools when analyzing the behavior, decisions, motives and outcomes of special target groups, such as innovators and entrepreneurs, and of course there are critical differences between them as well. Quantitative research depends on a large degree to acceptance of a number of key hypotheses, such as linear relationships, or population distribution in preselected categories, e.g. Gaussian type Curves. More specifically critical decisions related to sample choice, selection of statistical tools and methods which will be used for further analysis and conclusion may lead to less accurate estimations, regarding the population. On the other hand, innovation as a concept, ecosystems of innovation and individual entrepreneurs appear to have certain characteristics that makes it a bit questionable whether their population can actually meet several key criteria or further evidence and further analysis (Theory of Networks and Graph Theory) indicates that a new set of methodologies can provide more accurate results and highlight different aspects of business reality, compared to findings derived from traditional quantitative methods. Therefore the study examines several sampling and methodology questions affecting the results of the traditional quantitative research and hypothesis testing formation and to examine evidence that may question traditional quantitative research methods and their findings in the area of innovation management. Findings will provide researchers, policy makers and members of the innovation ecosystem useful insights for further research and a basis for further decisions and policy formation.

KEYWORDS

Start-Ups, Entrepreneurship, Sampling, Statistical Analysis

JEL CLASSIFICATION CODES

M13, O30, O31

1. INTRODUCTION

The research on innovation and entrepreneurship networks, especially when population consist of startups and their representative founders is not easy to implement due to the heterogeneity of the definition of basic concepts, including, first of all, the definition of a “start-up”. Researchers undertaking startup issues in their research and analysis are based on the well-known definition of Steve Blank, according to which a startup is a “temporary organization searching for a repeatable and scalable business model” [1]. Past research also highlights issues related to start-up definitions, impact on the economy, stage of development, as well as founder’s perceptions [2], [3], [4], success factors [5], business model innovation [6], founder’s experience, education, team sizes and taking into account the way founders had to face the challenges required to scale up their ventures [7]. A large part of entrepreneurial research still focuses on Correlation Coefficient, examining (and assuming) the existence of linear relationships within the data. Ideally the key variables examined address the needs of a theoretical model, and – when statistical analysis is involved – quantitative data are usually collected with the use of (structured) questionnaires.

2. SAMPLING DESIGN AND STATISTICAL THEORY

2.1 Key Problems regarding Sampling Design for Quantitative Research using Questionnaire

Past academic studies suggest [8], [9], [10], [11], [12] that the basic and most important thing in a survey (design) is to define the (target) population with absolute accuracy. This is valid for any survey: the classical survey and even the graphos ones, as it is e.g. web-surveys. The slightest mistake in defining the target population can lead to mistaken and even to ridiculous results. The second step is to find and adopt some sampling framework(s), so that the sampling can take place correctly. Such sampling frames are folders, files, catalogs and all kinds of recording media that contain exactly all the individuals or elements of the Population. Then useful information about the population is collected, regarding its various parameters like size, number of strata and their sizes. First of all the size of the population is determined. Sometimes the size determination is not an easy thing. So we have:

(a) If the population is registered in any recording medium, then one counting procedure is enough especially for not very big sizes.

(b) If the registration takes its place in more than one recording media all the sizes of the subpopulations in the individual media are listed and added (summarization).

(c) For the (more) complicated situations we have to use the appropriate procedure for the case. A very common procedure for complicated kinds of population (e.g. the web or any network but not only) is the one called "capture recapture" (symbol: *c/r*) procedure. We give an example. *Example*: We try to count the "start up" enterprises of the region Central Macedonia, Greece (CM-Gr). For this we think for an official folder of all start enterprises of the area and say that the number of the included start up enterprises is the unknown $N^{(c/r)}$. Also N_1=240 are the start up enterprises in the municipality of Thessaloniki, the capital of the region. A pilot sample of size N_2=140 start up enterprises from all over the region of CM-Gr is taken. Inside the sample m=105 start up enterprises are in Thessaloniki. Then the number of all the start up enterprises of the whole area is given by $N^{(c/r)}= N_1 * N_2/m$=240*140/105=320 start up enterprises. So the population of the startup enterprises in CM-G includes 320 of them.

After the (good) definition of the population a serial number is given to any element of the population. In order to select the sample, a draw is made between the individuals of the population with the help of the serial number of individuals. The serial numbers of the selected elements are drawn by the use of a random number table (RNT) or by the use of some computer program called pseudorandom number generator (PrNG). If the individuals of the population are selected for the sample by the whole population in a uniform manner, then we have simple random sampling (SRS). If selected from different strata of the population we have Stratified Sampling (StS) with its variations. In general there are many techniques and methods of sampling in the literature. Four groups of such methods are roughly distinguished. Each such group has many variations. The four basic groups are see, Cochran (1977), Farmakis (2009, 2015 and 2016):

- Simple random sampling (SRS)
- Stratified sampling (StS)
- Systematic sampling (SyS) and
- Cluster sampling (CluS) with many-many variants.

Determining the sample size is also a very important thing, see Farmakis (2006, 2009, 2015 and 2016), Cochran (1977). For the SRS and if the studied random variables are quantitative ones, then the size of the designed sample depends on:

- The size of the population N (Ascending function of…)
- Squared accuracy d^2 of the estimation of the parameter (Descending function of…)
- Variance of the random variable s^2 (Ascending function of…)
- Significance level α (Descending function of…)

The minimum sufficiently large sample size n_0 is given by the next relation:

$$n_0 = \left[\frac{N}{1+\left(\frac{d}{s\cdot z_{a/2}}\right)^2 \cdot N} \right] + 1 = \left[\left\{ \frac{1}{N} + \frac{1}{\square} \right\}^{-1} \right] + 1, \quad \square = \left(\frac{s\cdot z_{a/2}}{d} \right)^2 ,$$

where [x] is the integral part of x, the maximum integer that does not exceed x.

For the stratified sampling techniques all the above (about the size of the sample) are valid if instead of the variance s^2 the mean variance over all strata is taken, Farmakis (2006, 2009, 2015 and 2016). For the other methods and techniques the decision on the sample size depends on the ever situation and the researcher's opinion. Sometimes a low budget sampling method is applied based on Systematic Sampling. That low budget sampling procedures are non-probability sampling procedures and are in the most of cases based on some kind of additional parallel and independent information, Farmakis (2009,2015 and 2016). In surveys with questionnaire the population is a set with elements people and a questionnaire is prepared to be distributed to the people in the sample. This people will answer to the questionnaire questions with the utmost honesty. Also in many cases pairs of co-submitted questions are used in order to get answers to some "difficult" questions, e.g. "Did you make an honest tax return?", etc. In any sampling survey the problem of bias arises. This is also a major problem in questionnaire surveys. So in any sample survey with a questionnaire, an attempt is made to avoid any kind of bias through the questionnaire. It is noted that even a low percentage of those surveyed and responding to the respective survey is a factor of bias. Investigators' attention is always focused on the risk of bias.

All the above answers will be the data of the survey to be analyzed by the statistician researchers in order to get any output founding. Some statistical (analyzing) packages may be also used. This Statistical Analysis begins with Descriptive Statistics procedures and then various tests are performed to compare means, dispersions, indices, etc. In the next stage, cross-sections and X-square tests are used to reveal dependencies or not between the various random variables involved in the problem we face. The next (third) stage may include very specific Data Analyzes and corresponding conclusions, (ANOVA, Correlation, Regression, etc.). All the founding will be the content of the discussion text and conclusions. This part of the research is usually called "deliverable". The conclusions will be made public, as they will be the subject of any publication. Sometimes the results and the conclusions of the research are not published but remain at the disposal of the client (financier) of the research.

- Relevance of a person to provide an answer – it has to be verifies that the person who fill in a questionnaire expresses the company view; however it may not be uncommon different people in the management or founder's team to have different personal views regarding key areas. This can be even more important in cases with different focus or background of the founders.
- Non- Linearity of relationships – this can be proved a key issue, and statistical tools such as Regression Analysis, Pearson & Spearman Coefficient can be used to identify only linear relationships between variables.
- Sample focuses on the edges of the Gaussian Curve – it is reasonable to assume that entrepreneurs and innovators belong to the extreme parts of a Gaussian Curve, if any; as a result it becomes questionable whether statistical tools designed to analyze context within an Gaussian Curve (and therefore are tools that suggest the existence of a Gaussian Curve and linear relationships) suitable for the task?
- Differences between sample members – do all members of the sample they actually belong to the same Gaussian Curve? E.g. Some entrepreneurs maybe market experts, product design champions or technology experts – it is likely to have different skills, background and education and as a result different tasks within an organization (regardless whether the organization is start-up or not), and as a result, have different priorities and different point of view regarding key variables and metrics that are critical for the organization's performance. A great marketing person may be a bad technical director (e.g. Steve Jobs and Ellon Musk are both innovators, but with a much different focus and approach: as Gates describes: Jobs was a great marketing person and orchestrator with a strong focus on product design, while Musk focuses on product performance and product excellence)
- Past studies [11] highlighted the tendency to avoid to provide information, especially in cases of failure; researchers tend to consider that especially in cases of established companies, companies and managers were particular reluctant in revealing information which may were critical reasons for failure. More recently, start-uppers appeared to be more willing to share their experiences, including their failures, and provide valuable information, however some bias may still exist regarding the extend and the value of information provided.
- Most importantly, Cognitive issues may also be applied; Success of a business venture (and failure as well) is a result which depends on many factors (variables) each one with different significance and impact; since there are actual differences between been able to actually identify some factors; successful start-uppers may attribute their successes to some factors that are not necessarily the ones that led to success or the ones that actually contributed to the success of their venture (and start-uppers who failed may also lack the knowledge of the real reasons of their failure); the best that a researcher can assume is that they respond based on their best of their understanding and as honestly as possible.

2.2 Data and relevant Theory come First, Statistical Theory and Tools Follow

A measure of correlation [12] is a random variable that is used in situations where the data consist of pairs of numbers, such in bivariate data. Suppose a bivariate random sample of size n is represented by (X_1,Y_1), (X_2,Y_2), …., (X_n,Y_n). It can be used (X,Y) when referring to the (X_i,Y_i) in general. That is, the (X_i,Y_i) for i=1,2,….,n have identical bivariate distributions, the same bivariate distribution as (X,Y) has. Examples of bivariate random variables include one where Xi represents the age of a *i*th individual and Y_i represents the weight of him, or where X_i represents a test score of a *i*th individual and Y_i represents her total amount of score. By tradition, a measure of correlation between X and Y should satisfy the following requirements in order to be acceptable. First of all, the measure of correlation should assume only values between -1 and +1. Then, it should be positive and close to +1.0 when the larger values of X and the larger values of Y tend to be paired, and the smaller values of X and Y tend to be paired together too. This tendency is strong and is called a positive correlation between X and Y. Respectively, this measure of correlation should be negative and close to -1.0 when the larger values of X and the smaller values of Y tend to be paired, and vice versa. This tendency is strong too and we can say that X and Y are negatively correlated. On the other hand, if the values of X appear to be randomly paired with the values of Y, the measure of correlation should be enough close to zero. This is the case when X and Y are independent, and possibly some cases where X and Y are not independent. We then say that X and Y are uncorrelated, or have no correlation, or have correlation zero. Pearson's correlation coefficient is the most commonly used measure of correlation. It is denoted by *r* and also is defined as $r = \frac{\sum_{i=1}^{n}(X_i - X)(Y_i - Y)}{[\sum_{i=1}^{n}(X_i - X)^2 \sum_{i=1}^{n}(Y_i - Y)^2]^{\frac{1}{2}}}$. Pearson's r measures the strength of the linear association between X and Y. The meaning of the strength linear association lies in the fact that if a plot of Y versus X gives the points (X,Y) all lie on, or close to, 1.0 then the line is sloping upward, and -1.0 the line is sloping downward. We have [12] also to check the sources of bias if we want to compute a bivariate correlation coefficient. It will be reported the two most important ones: linearity and normality. A researcher must check the relationship between variables. If it is not linear then this model is invalid and he might need to transform the data or use another type of correlation coefficient.

Past studies [11] highlight the importance to look at the data in a graph and after running any other analysis on them. To start a correlation analysis, it should be to look at some scatterplots of variables that the researchers have measured about them. A scatterplot [13] shows the relationship between two quantitative variables measured for the same individuals. The values of one variable appear on the horizontal axis and the values of the other variable appear on the vertical axis. Each individual in the data appears as a point on the graph. As in any graph of data, the researchers have to look for the overall pattern and for striking departures from that pattern. First of all, the direction, form and strength of a relationship can describe this overall pattern of a scatterplot. Direction: The purpose of a scatterplot is to provide a general illustration of the relationship between the two variables. Two variables have a positive association when above-average values of one tend to accompany above-average values of the other, and when below-average values also tend to occur together. Two variables have a negative association when above-average values of one tend to accompany below-average values of the other. Form: A relationship is linear if one variable increases by approximately the same rate as the other variables changes by one unit. In other words, the points on the scatterplot closely resemble a straight line. Other form of relationship between the two variables can be eg. curvilinear, which it is not like a straight line but it has the form of a curve. This is due to the fact that one variable does not increase at a constant rate and may even start decreasing after a certain point. Strength of the relationship: The slope provides information on the strength of the relationship. The slope is 1 when we have the strongest linear relationship. This means that when one variable increases by one, the other variable also increases by the same amount. This line is at a 45 degree angle. In addition, great importance must be given to the strength of the relationship between two variables because is a crucial piece of information. A researcher should not rely on the interpretation only because it is too subjective. More precise evidence is needed, and this evidence is obtained by computing a coefficient that measures the strength of the relationship under investigation. The next thing that is should be under concern is an important kind of departure, the outlier. An outlier is an individual value that falls outside the overall pattern of the relationship.

The measure of correlation [12] may be used with any numeric data and there are no requirements about the scale of measurement or the type of underlying distribution, although it is difficult to define unless the scale of measurement is at least interval. It meets the necessary requirements of an acceptable measure of correlation. However, r is a random variable and, as such, r has a distribution function. Unfortunately, the distribution function of r depends on the bivariate distribution function of

(X,Y). Therefore, r has no value as a test statistic in nonparametric tests or for forming confidence intervals unless, of course, the distribution of (X,Y) is known. In addition to this widely accepted r, many other measures of correlation have been invented that satisfy the preceding requirements for acceptability. Some measures of correlation possess distribution functions that do not depend on the bivariate distribution function of (X,Y) if X and Y are independent and, therefore, they may be used as test statistics in nonparametric tests of independence. The measures of correlation selected for presentation here are functions of only the ranks assigned to the observations. They possess (X,Y) if X and Y are independent and continuous. They may even be used as measures of correlation on certain types of nonnumeric data, if the data meet the ordinal scale of measurement.

When the sample [11] is reasonably large, e.g. 100 observations, it is also possibly large enough for the central limit theorem and the researchers should have not any concerns about normality. However, it would be advisable to use bootstrap function to get robust confidence intervals. The researchers might also consider and use a rank-based method so that they compute the correlation coefficient itself. Spearman’s correlation coefficient ρ (rho) [11] is a non-parametric statistic based on ranked data and the researcher can minimize the effects of extreme scores or the effects of violations of the assumptions so it can be useful. Spearman’s test works by first ranking the data and then applying Pearson’s equation to those ranks. The data [12] may consist of a bivariate random sample of size n, $(X_1, Y_1), (X_2, Y_2), \dots, (X_n, Y_n)$. Let $R(X_i)$ be the rank of X_1 as compared with the other X values, for $i = 1, 2, \dots, n$. That is, $R(X_i) = 1$ if X_1 is the smallest of $X_1, X_2, \dots, X_n$, $R(X_i) = 2$ if X_i is the second smallest, and so on, with rank n being assigned to the largest of the X_i. Similarly, let $R(Y_i)$ equal $1, 2, \dots, or\ n$, depending on the relative magnitude of Y_i, as compared with $Y_1, Y_2, \dots, Y_n$, for each

Kendall’s tau, τ, [11] is another non-parametric correlation and it should be used rather than Spearman’s coefficient because it is useful with small data set with a large number of tied ranks. If the researcher has to rank all of the scores and many scores have the same rank, then Kendall’s tau should be used. Although Spearman’s statistic is the more popular of the two coefficients, it is sometimes suggested that Kendall’s statistic is actually a better estimate of the correlation in the population. More generally, when using correlations as effect sizes the researcher should have in mind either when he reports his own analysis or he interprets others that the choice of correlation coefficient can make a substantial difference to the apparent size of the effect.

Correlation coefficients [13] have a probability (p-value) that shows the probability that the relationship between the two variables is equal to zero (null hypotheses; no relationship). Low p-values may have the strong correlations because the probability that they have no relationship is very low. In addition, in social sciences when p-value is lower than 0.05, correlations are typically considered statistically significant, but the researcher has the liberty to decide the p-value for which he or she will consider the relationship to be significant. It is important to be reported the alpha level, the value of p for which a correlation will be considered statistically significant.

It’s important [11] to remember that correlation coefficients give no indication of the direction of causality. There are two problems:

- The third-variable problem or tertium quid: To recap, in any correlation, causality between two variables cannot be assumed because there may be other measured or unmeasured variables affecting the results.
- Direction of causality: Correlation coefficients say nothing about which variable causes the other to change. Even if we could ignore the third-variable problem, and we could assume that the two correlated variables were the only important ones, the correlation coefficient doesn’t indicate in which direction causality operates. Therefore, correlation coefficients provide evidence of association, not causation.

2.3 Use of Statistical Theory

Pearson Correlation Coefficient appears to be a widely used statistical method for entrepreneurship research, at least for quantitative research. [13] Correlation analysis provides “a quantitative methodology used to determine whether, and to what degree, a relationship exists between two or more variables within a population (or a sample).” [14] Correlation is used in order to investigate the possibility of a linear relationship between two (or more) variables examined.

Further studies [15] examine the effectiveness of Bayesian approach when examining financing markets: "the Bayesian theory has been subjected to both internal and external critiques that have brought substantial revisions of its original formulation. Critics have stressed the paradoxes and anomalies which violate the Bayesian rule of the revision of probability with new information"

Summarizing the conclusions of past research within the Bayesian theory, is highly varied: situations characterized by probabilities with partial ordering; non additive probabilities; intervals of probability; situations of ambiguity of the information à la Ellsberg; elements of arbitrariness, judgements characterised by "vacillation", "indecision", "surprise"; indeterminacy represented by notions of "unreliable", vague and imprecise probabilities. The author concludes that "The original notion of Bayesian rationality in itself was inadequate to explain the actual working of financial markets as it was based on a notion of rationality as internal coherence of the individual. Today as the Bayesian theory is gradually divested of meaning and filled with paradoxes and anomalies, its notion of rationality appears even less enlightening."

Further studies [16] highlight the importance of establishing inclusion and exclusion criteria for sample selection in a quantitative study. "A goal for all researchers regardless of discipline is to yield data findings in support of the predicted correlations or causal relationships as stated in the hypotheses." [17] This means that researchers on the fields of entrepreneurship and e-commerce use specific models and theory to state research hypotheses, then proceed to the actual research phase and – when quantitative research is required, proceed to sample selection, and methods of statistical analysis.

3. CONCLUSION

Application of statistical analysis in Entrepreneurship research sets a number of challenges for the researchers. These challenges have to do with population definition and sampling, as well as the use of traditional statistical methods that require linear relationships (e.g. Correlation Analysis with Pearson Coefficient). A suggestion is the use of several techniques to check for linearity as well as to apply a variety of different, non-linear statistical tools and methodologies.

It is important to refer that correlation requires that both variables be quantitative and describes linear relationships. It does not describe curve relationships, no matter how strong the relationship is. In addition to, correlation is not resistant. Pearson's correlation coefficient is strongly affected by outliers. It is based on means and standard deviations, so it is not robust to outliers; it is strongly affected by extreme observations. These individuals are sometimes referred to as influential observations because they have a strong impact on the correlation coefficient. For these reasons, when describing the relationship between two variables, correlations are just one piece of the puzzle. This information is necessary, but not sufficient. Other analyses should also be conducted to provide more information, such as parametric tests for example t-test and analysis of variance (Anova) and non-parametrics for example X-square test, Mann-Whitney and Kruskal Wallis.

Recent research [18] highlights different statistical findings, using the exact same datasets when using different statistical tools (e.g. graph and network theory tools), and recommend use of a combination of Statistical tools in order for safer statistical findings and related conclusions to be drawn. This is because each combination of Statistical tools acts as a cross-checking mechanism.

It is also a good idea to examine quantitative characteristics in the form of qualitative characteristics to aid in the respective tests, e.g. the X square test for the independence. Thus, the scaling of quantitative characteristics can also be entered from the questionnaire, such as the age with scaling "young", "middle-aged", "elderly". This must be done in accordance with the age limits. Of course, this escalation can also be done at the stage of Statistical processing and analysis, turning the Statistical analysis into a kind of qualitative analysis. It is noted here that the X-square test is a non-parametric (statistical) procedure, free from assumptions about normality, etc.

4. LIMITATIONS OF THE RESEARCH

As the title suggest, the scope of the present paper was to create and establish a reasonable doubt regarding the dependence on statistics on several aspects of the actual world, and more specifically on economic and entrepreneurships fields. More specifically, to highlight areas of the entrepreneurship and economic world where statistical findings and their respective interpretation may lead to misleading conclusions for reasons related to the sample, the phenomenon studied, and the statistical tools used.

There are several reasonable reasons to consider whether quantitative research methods can be used; after there is a closer examination of qualitative variables. Furthermore, use of specific statistical methods and tools may sometimes lead to less reliable conclusions than the ones a researcher would have hoped for.

From this point of view, limitations of the present paper will become evident once scientific research becomes more concerned with the best practices proposed and the conclusions of the present paper, producing more accurate statistical findings and reaching more accurate conclusions.

REFERENCES

[1] Blank, S. (2013). Why the lean start-up changes everything. Harvard Business Review, 91(5), 63–72.
[2] Rostek K., Skala A., (2018), Segmentation Analysis For Polish Digital Startups in Years 2015 and 2016, Studia i Materiay, 1/2018 (26): 55– 67
ISSN 1733-9758, © Wydzia Zarzdzania UW
[3] Spyropoulos, T. S., (2019), "**Start-Up Ecosystems Comparison: MIT and Greece Experiences**", Research Papers Collection, *The Małopolska School of Economics in Tarnów Research Papers Collection, Volume 42, Issue 2, Special Issue: Entrepreneurship Theory & Practice Current Trends and Future Directions, June 2019* DOI: 10.25944 / znmwse.2019.02.4358 http://zn.mwse.edu.pl/en/wp-content/uploads/2019/11/Zeszyty-Naukowe-Vol.-42-4-2019_EN-druk.pdf
[4] Spyropoulos T.S. (2020), "Digital Greek Start-Ups – An Analysis of Founder's Perceptions", Advances in Management and Informatics, Issue 5, 5th Edition, August 2020, https://figshare.cardiffmet.ac.uk/articles/journal_contribution/Advances_in_Management_and_Informatics_Issue_5_Working_Papers_Journal_/13027697
[5] Marom S., Lussier R.N., (2014), A Business Success Versus Failure Prediction Model for Small Businesses in Israel, Business and Economic Research, ISSN 2162-4860 2014, Vol. 4, No. 2
[6] Girotra K., Netessine S., (2014), Four Paths to Business Model Innovation, Harvard Business Review, July – August 2014 Issue, https://hbr.org/2014/07/four-paths-to-business-model-innovation
[7] Øvretveit J., Garofalo L., Mittman B. (2017); Scaling up improvements more quickly and effectively, *International Journal for Quality in Health Care*, Volume 29, Issue 8, 1 December 2017, Pages 1014–1019, https://doi.org/10.1093/intqhc/mzx147
[8] Cochran, W. (1977) "Sampling Techniques" John Wiley &Sons, Inc, New York, Toronto
[7] Farmakis, N. (2006) "A Unique expression for the size of samples in several sampling procedures", *STATISTICS in TRANSITION*, Vol. **7**, No 5, pp. 1031-1043.
[9] Farmakis, N. (2009) "*Survey and ethics*", A & P Christodoulidi, Thessaloniki (in Greek)
[10] Farmakis, N. (2015) "*Sampling and Applications*", https://repository.kallipos.gr/ & www.kallipos.gr (in Greek)
[11] Farmakis, N. (2016) "*Introduction to Sampling*", Kyriakidis Bros Edit. S.A., Thessaloniki (in Greek)
[12] Cooper R. G., (1998) Winning at New Products: Accelerating the Process from Idea to Launch, 2nd Edition,Perseos Books, Reading, Massachusetts, U.S.A.
[13] Field, Andy (2013). Discovering Statistics Using IBM SPSS Statistics (4nd ed.). SAGE Publications Ltd
[14] Conover W.J. (1999). Practical Nonparametric Statistics, Third Edition. John Wiley & Sons, Inc. New York, Chichester, Toronto, singpore
[15] Moore, D. S., Notz, W. I, & Flinger, M. A. (2013). The basic practice of statistics (6th ed.). New York, NY: W. H. Freeman and Company.
[16] Isotalo J., (2014), Basics of Statistics, Available at:http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf
[17] Apuke O.D. (2017), Quantitative Research Methods: A Synopsis Approach, Arabian Journal of Business and Management Review, Vol. 6 (10), 2017, DOI: 10.12816/0040336
[18] Carabelli A. (2002) "Speculation and Reasonableness: a non-Bayesian Theory of Rationality", in S. Dow and J. Hillard, eds, Keynes, Uncertainty and the Global Economy: Beyond Keynes, Elgar, Aldershot, 2002, volumetwo, pp. 165-85
[19] Baxter, P., & Jack, S. (2008). Qualitative Case Study Methodology: Study Design and Implementation for Novice Researchers. The Qualitative Report, 13(4), 544-559. Retrieved from https://nsuworks.nova.edu/tqr/vol13/iss4/2
[20] Picardi C.A., Masick K.D. (2014) Research Methods – Designing and Conducting Research With a Real World Focus, SAGE Publications
[21] Spyropoulos T., Andras C., Dimkou A., (2021 – in press), Application of Graph Theory On EntrepreneurshipResearch– New Findings and a New Methodology Framework