

Using Mutual Information and Information Gain Ratios on Entrepreneurial Research: An Empirical Case for Greek I.T. Start-Ups

THEOCHARIS STYLIANOS SPYROPOULOS

International Hellenic University, Serres, Greece

Perrotis College, Thessaloniki, Greece

theospyr2@es.ihu.gr

ORCID ID: <https://orcid.org/0000-0001-8534-2955>

CHRISTOS ANDRAS

International Hellenic University, Thessaloniki, Greece

andraschris@gmail.com

ORCID ID: <https://orcid.org/0000-0001-8674-7823>

ARISTOTELIS DIMKOU

Aristotle University of Thessaloniki, Thessaloniki, Greece

arisdimkos@yahoo.gr

PERSEFONI POLYCHRONIDOU

International Hellenic University, Serres, Greece

polychr@es.ihu.gr

ORCID ID: <https://orcid.org/0000-0003-1517-011X>

Izvorni znanstveni rad / *Original scientific paper*

UDK / UDC: 658:[303.4:007](495)=111

Primljeno / Received: 07. srpnja 2023 / July 7th, 2023.

Prihvaćeno za objavu / Accepted for publishing: 20. studenog 2023. / November 20th, 2023.

DOI: 10.15291/oec.4195

Abstract: The majority of Entrepreneurial quantitative research focuses on Correlation Coefficients. However, new statistical analysis based on Entropy, such as Mutual Information and Information Gain Ratios cast a new light on understanding the relationships among variables and offer a view of non-linear relationships. The study examines key entrepreneurial variables using Mutual Information and Information Gain Ratios and compares findings using the same dataset which examined I.T. Greek Start-Ups. Use of Mutual Information and Information Gain ratios reveals much more relationships between the variables examined, in comparison to Pearson Correlation. Furthermore, the study compares results from Pearson Correlation and Mutual Information and Information Gain ratios to draw new conclusions on the perceptions of Greek I.T. start-up founders. The findings indicate that use of Mutual Information reveals a set of factors that contribute to entrepreneurial perception of success which differs significantly from the conclusions based on Correlation Coefficient Analysis. More specifically factors such as Operation Years and Previous Start-Ups play a far more crucial role than B2B and Sales. The study offers an original contribution to entrepreneurial science, introducing the use of entropy-based mathematical ratios, such as Mutual Information and Information Gain in Entrepreneurial Research. The study highlights that use and results derived of such ratios enable researchers to identify more information regarding (non-linear) relationships between variables, compared to Correlation Coefficient methods.

Keywords: Start-Ups, Business Model, Mutual Information, Information Gain, Network Theory, Entropy, Entrepreneurship

JEL classification: M13, M 19, M20, L26, O30, O34

1 Introduction

In the past decade, most of the research on innovation management and start-up entrepreneurship focused on quantitative research, which mainly depended on Correlation analysis. This was expected since before that a number of critical concepts have been defined the previous period, such as the concepts of sustaining and disruptive innovation, the lean start-up methodology, business model innovation, etc.

Considering the above, the present study provides an original contribution to Entrepreneurship science, by examining data collected from I.T. Start-Up Founders in Greece, with the use of Information Gain. This method is aligned with relevant studies in multifactor complex phenomena, as already presented, and therefore provides a new approach in analyzing entrepreneurship and enables researchers to use the methodology suggested to develop a better understanding of entrepreneurial.

Further research on the specific field is recommended to use larger datasets with the use of the proposed methodology in order to evaluate changes of the critical values across countries or in areas with different cultural and economic environments.

The paper is organized as follows: the review of literature is presented in second section; the data are described in section three. In the fourth section the methodology is analyzed, and the fifth section includes the results of Mutual Information and Information gain Ratios are presented and discussion. Finally, and in the last section the conclusion remarks are discussed.

“The (Academy of Management) Review publishes distinguished original manuscripts which (a) move theoretical conceptualization forward in the field of management, and/or (b) indicate new theoretical linkages that have rich potential for theory and research in management, and (c) provide clear implications of theory for problem-solving in administrative and organizational situations.” (Corley et al. 2011, p. 13).

Considering the above criteria the present study provides an original contribution to Entrepreneurship science, by examining data collected from I.T. Start-Up Founders in Greece, with the use of Information Gain. This method is aligned with relevant studies in multifactor complex phenomena, as already presented, and therefore provides a new approach in analyzing entrepreneurship and enables researchers to use the methodology suggested to develop a better understanding of entrepreneurial.

2 Literature Review

Academic literature review provides a wide range of factors that determine the success of start-up companies, with a further focus on ICT Sector. This study provides an in-depth analysis of Greek ICT start-ups by examining a wide range of factors, such as previous experience (professional and entrepreneurial), education, number of founders, company achievements and challenges, competitive advantage, strategic alliances.

Spyropoulos (2020a) examines Greek IT Start-Ups, summarizes previous research (2020b) and provides a framework of the knowledge management factors with a special focus on Greek I.T. Start-Ups. The key success factors, examined were: Gender, Age, Education, Number of Founders, Working Experience, Previous Entrepreneurial Ventures, Years of Start-Up operation, Reasons for Establishing the current Start-Up (Opportunity, Technology, Business Model Innovation and/or Process Innovation), Key Challenges the Start-Up Faces (Improve Product, Get Customers, Secure Funding and/or Other),

Key Achievements (Prototype, Prof of Concept, Sales of 100 thousand euros and/or Funding of 100 thousand euros), Strategic Alliances, Access to Funding, the value that the new solutions offers the end customer (Unclear, Major, Minor, New Approach towards the market), Disruptive vs Sustaining innovation, type of Competition faced (No Competition, Traditional Companies, New Start-Ups and Disruptive Start-Ups), the Competitive Advantage (Technology, Management, Intellectual Property, Business Model Innovation), Openness of the technology used, and type of Innovation (An Improved Version of an existing Product, a distinctive New Product, a new Market Approach or an attempt to create a New Market). Table 1 below includes the key variables examined.

Table 1: Definitions and Metrics of the Variables Examined

Variable	Definition	Metrics
Gender	Male & Female	Male, Female
Age	Years (date of research)	Age Groups
Education	Level of Education and Degree achieved	Degree Achieved
Founders	Founding team	Number of Founders of the Start-Up
Working Experience	Experience as an Employee, before starting a business venture	
Previous Start-Ups (Entrepreneurial Ventures)	Measures the “serial entrepreneur” effect. Experience as an entrepreneur, number of previous businesses established.	Number of companies that the same founder has established in the past.
Years of Start-Up operation	How many years the Start Up Company is active (established).	Number of years; 0 represents that there is no official establishment of the company as a legal entity and works as a team.
Reasons for Current Start-Up	Examine the reasons why the founder started the current business venture	Examined factors business opportunity, technology, business model innovation and Process innovation.
Reasons for Previous Start-Ups	In case of serial entrepreneurs, examines the reasons for establishing previous companies	The same factors are examined, and examines whether a pattern can be explored (founders follow similar reasons for establishing a new business venture)
Start-Ups Surviving Today	In case of serial entrepreneur, a measure of success (establishment of start-ups which survived)	Measures the number of companies a founder has established in the past and are still active
Key Challenges	Key challenge(s) the founders face in their current venture	Main challenges considered Product Improvement, Getting more Customers,

		Secure Funding or Other; more than one option (challenge) can be selected
Key Achievements	Key achievements of the start-up	Prototype Development, Prof of Concept, Sales Value of 100k euros, Funding of 100k euros. More than one option can be selected
Success	Perceived Level of Success by each founder	Levels of success defined
Disruption	Define whether the start-up disrupts the industry or provides a sustaining innovation proposal	Define the type of innovation in terms of disruptive or sustaining innovation
Competition	Type of Competition faced as perceived by founder	The main options are No competition, Competition from traditional companies, Competition from Start-Ups, Competition from Disruptive Start-Ups
Competitive Advantage	Source of Competitive advantages	Main types are management, technology, Business model innovation and Intellectual Property.
Openness	The level and ease of innovation's integration with third party systems and applications	Distinguishes between Close Systems and Data Export, Basic Connectivity and Seamless Integration
Type of Innovation	Defines how innovative a proposal is	An improved product, a new product, a new product/market approach and a new market creation

Source: Authors' elaboration (2022)

3 Research Methodology

Based on the literature review findings, the research questionnaire was comprised of 18 questions. Defining the total population of start-ups is always a near to impossible task, by definition; start-ups can be just teams trying to develop a product or service, before even establishing a company – as a result there are no official records regarding the actual number of start-ups in any given moment, their commitment, their resources or focus. This of course applies to any country. The study focuses on the founders of the I.T. start-ups who participated in the Digital Greece 2018 Event, organized during the 2018 International Trade Fair of Thessaloniki. This was the first time where Greek Start-Ups had the opportunity to participate in an event focused on innovative start-ups in Greece. The specific sample choice satisfied several critical selection criteria – a well-defined sample population and industry section (start-up founders of I.T. companies), and a common basic background of entrepreneurial and managerial education, since founders of all the participating start-ups had participated at least in one start-up accelerator training event. Start-up founders engaged in other sectors (not I.T.) and other organizations (such as participating accelerators, Venture Capital Companies and Public Organizations) were excluded from the research.

The research took place through structured questionnaires, which were distributed to the sample population on site. More specifically 94 questionnaires were distributed, on site, to the founders of the startups, and were collected immediately upon completion; in case of more than one founder available on site, each founder had to fill-in the questionnaire alone, with no contact with any co-founder or any other team member.

The exact same dataset was used in several published research papers (Spyropoulos, 2020a), where the data were encoded and entered into an advanced statistical analysis software (SPSS), which was used in order to analyze the correlation between variables, with the use of Pearson Correlation Coefficient. Findings indicate Statistical Significance for the Correlations. A second paper (Spyropoulos et al, 2021) used Graph Theory to analyze the same set of data, providing additional information from the same dataset with the use of Graph Theory.

The present study uses Mutual Information to identify relationships between variables and the values of Mutual Information Ratios between variables are used to create a Graph Network between these variables, and further results are drawn through the Graph Theory ratios.

4 Mathematical Background – Key Concepts

4.1 The Concept of Mutual Information

“Entropy: Let X be a random variable on a (discrete) space X , and x an element from X . For every positive integer d , we denote by X a d -dimensional random vector $(X_1, \dots, X_d) \in X_d$, and by the letter x an element from X_d . The (Shannon) entropy of a random variable X on a discrete space X is a measure of its uncertainty during an experiment.

The mutual information is a general measure of the dependence between two random variables. It expresses the quantity of information one has obtained on X by observing Y . On a discrete domain, the mutual information of two random variables X and Y is defined as:

Mutual Information:

$$I(X; Y) = \sum_{x \in X, y \in Y} Pr[X = x, Y = y] \cdot \log \left(\frac{Pr[X=x, Y=y]}{Pr[X=x] \cdot Pr[Y=y]} \right) \quad (1)$$

The mutual information can similarly be expressed as the expected value over X of the divergence between the conditional probability $Pr[Y = y|X = x]$ and the marginal probability $Pr[Y = y]$ (Batina et al. 2010, p. 272-273).

The mutual information is always greater than or equal to zero, with equality if X and Y are independent. It is lower than the entropy of either variable, and equality only occurs if one variable is a deterministic function of the other. The higher the mutual information, the stronger the dependency between X and Y . $0 \leq I(X; Y) \leq \min\{H[X], H[Y]\}$ (Batina et al, 2010, p. 273)

“Pearson’s correlation coefficient is a simple measure of dependence between two random variables X and Y . Computing it does not require to know the probability density functions of X and Y , but it can express only the linear dependence between these variables (whereas mutual information is able to detect any kind of dependence). The correlation coefficient satisfies the following inequality: $0 \leq |\rho(X, Y)| \leq 1$, with the upper bound achieved if Y is an affine function of X . The lower bound is achieved if X and Y are independent but the opposite does not hold X and Y can be dependent and have their correlation equal to zero.” (Batina et al. 2010, p. 273-274).

Information Gain provides asymmetric information between two variables defining the additional information that variable 1 offers as a way to reduce uncertainty (and therefore entropy) of variable 2 (Azhagusundari et al, 2013; Ayyappan et al, 2017).

4.2 Mutual Information and Information Gain Ratio in Business Studies

Azhagusundari et al. (2013) examine databases with different types of data (continuous, discrete and symbolic) and process different types of data into numeric valued attributes; this methodology can be used for hundreds of attributes and variables, in order to apply Information Gain analysis. Ayyappan et al. (2017) use Information Gain Ratio to analyze social networks based on academic datasets, concluding that their work “establishes the significance of information gain for selecting the attributes in the context of social networks of academic data” (p. 941). Prasetyio et al. (2021) use Information Gain ratio to evaluate effectiveness of Bank Marketing (based on Portugal’s Banking system). In their study they highlight the fact that there are several marketing strategies (which can be considered as variables) which can determine the final outcome (in the specific research case a successful customer subscription to bank services).

Furthermore, recent research (Spyropoulos et al. 2022) used Mutual Information and Information Gain ratios in a dataset of 130 Start-Up Founders; the dataset was used in previous research (Spyropoulos, 2019) which examined the dataset using Correlation Coefficient (Spearman). The researchers compared the findings between Correlation Coefficient on the one hand and Mutual Information and Information Gain and concluded that “Entropy-based ratios, and more specifically Mutual Information and Information Gain Ratio offer additional information to researchers focusing on entrepreneurial research. More specifically, Mutual Information and Information Gain highlight the existence of non-linear relationships, which cannot be proved statistically important with the use of Spearman Correlation Coefficients. (Spyropoulos et al., 2022, p. 513).

Finally, Sklavounos et al. (2023) also compare findings between Correlation Coefficient and Mutual Information and conclude that Mutual Information provides additional information and insights that were not identifying with the use of Correlation Coefficient.

4.3 The Present Study

The present study uses Mutual Information and Information Gain Ratios to investigate a complex, real life and multifactor problem, such is the (perceived) level of entrepreneurial success on behalf of the Greek I.T. Start-Up founders, and compare related results derived with Correlation Coefficient (Pearson). More specifically the goal of this research is to evaluate whether use of Mutual Information (MI) and Information Gain (IG) Ratios offer additional information in entrepreneurship research, from information derived from Correlation Coefficient, and examine the managerial implications of related findings. For this reason, the present study uses the exact same dataset from a previous published work (Spyropoulos 2020). The research’s dataset consists of 94 questionnaires of Greek ICT Start-Up Founders and the descriptive statistics are available at Spyropoulos (2020, p.39).

Regarding Mutual Information analysis, Table 2 below includes sets of variables where their Mutual Information Ratio is between 0.46 and 0.10. In total there are 12 sets of variables with MI Ratio values between 0.46-0.25, 18 sets of variables with MI Ratio values between 0.249-0.15, 64 sets of variables with MI Ratio values between 0.149-0.1, 186 sets of variables with MI Ratio values between 0.099-0.5, 645 sets of variables with MI Ratio values between 0.499-0.001 and 51 sets of variables with MI Ratio value zero.

Table 2: Mutual Information Values between Variables

Variable 1	Variable 2	Mutual Information
Previous SU	Previous Reasons	0.464
Pr. Surviving	Previous Reasons	0.396
Operation Years	Success	0.387
Previous SU	Pr. Surviving	0.35
Age	Experience	0.34
Operation Years	Total Reasons	0.302
Operation Years	Previous Reasons	0.296
Operation Years	Strategic	0.285
Experience	Operation Years	0.279
Operation Years	Sales 100k	0.272
Education	Operation Years	0.255
Operation Years	Funding	0.25
Founders	Operation Years	0.23
Age	Operation Years	0.226
Previous SU	Operation Years	0.22
Strategic	New Start Ups	0.214
Operation Years	Funds 100k	0.204
Operation Years	Disruption	0.202
Operation Years	Pr. Surviving	0.2
Strategic	Funding	0.183
No Comp.	Traditional	0.182
Funds 100k	Funding	0.178
Operation Years	Openness	0.17
Process Innovation	Total Reasons	0.167
Operation Years	Business Model	0.163
get funding	Funding	0.158
Funding	Disruption	0.157
Pr. Surviving	Success	0.157
Strategic	Openness	0.151
Age	Funding	0.15
Pr. Surviving	Funding	0.149
Operation Years	B2B	0.149
B2B	B2C	0.147
Disruption	Openness	0.147
Education	Experience	0.147
Education	Disruptive SU	0.145
Previous Reasons	Funding	0.143
Funding	Openness	0.142
Business Model	Total Reasons	0.141
Total Reasons	Business Model	0.141
Age	Pr. Surviving	0.141
Total Reasons	Funding	0.14

Experience	Sales 100k	0.14
Founders	Funding	0.138
Operation Years	New Product	0.136
Previous SU	Success	0.136
Technology	Previous Reasons	0.132
Education	Funding	0.132
Funds 100k	Strategic	0.131
Total Reasons	Disruptive SU	0.131
Operation Years	Other	0.13
Experience	Strategic	0.129
Education	Previous Reasons	0.129
Business Model	Funds 100k	0.128
Success	Funding	0.127
Founders	POC	0.125
Previous Reasons	Funds 100k	0.125
Success	Strategic	0.124
Experience	Previous Reasons	0.124
Operation Years	B2B	0.123
Funding	Disruptive SU	0.122
Operation Years	Process Innovation	0.122
Get customers	Other	0.121
Operation Years	Get customers	0.119
Total Reasons	Management	0.119
Operation Years	Get funding	0.118
Experience	Pr. Surviving	0.117
Operation Years	Traditional	0.117
Education	Openness	0.117
Founders	Pr. Surviving	0.117
Unclear	Major	0.117
Age	Success	0.116
Experience	Success	0.115
Founders	Success	0.115
Operation Years	Disruptive SU	0.114
Operation Years	improve product	0.114
Opportunity	Total Reasons	0.113
Previous Reasons	Success	0.112
Operation Years	No Comp.	0.109
Funding	New Start Ups	0.109
Business Model	Business Model	0.108
Education	Strategic	0.107
get customers	POC	0.106
Founders	Funds 100k	0.105
Minor	new approach	0.105
Operation Years	Technology	0.105

improve product	Funding	0.105
Operation Years	New Market creation	0.104
Education	Business Model	0.104
Funds 100k	Disruption	0.103
Operation Years	Management	0.103
Total Reasons	Funds 100k	0.103
Operation Years	New Start Ups	0.102
Operation Years	IP	0.1

Source: Authors' elaboration (Mutual Information Calculations using R Software), 2023

As it becomes evident from the Mutual Information Ratio, the first relationships, counting up to 10% of the information of the variable 2 tell us a different story and reveal relationships that cannot be identified and described as linear relationships. In this case Mutual Information Ration provides 94 values which explain up more than 10% of the value of variable 2, when the value of variable 1 becomes known (and a much higher number of values up to 5%, and again much higher between 5% and zero). This information becomes more critical in comparison to the information provided from Correlation Analysis, which provided just 10 correlations.

First of all, we have much more detailed information. A part of this has to do with the way of the information (Correlation represents one relationship between 2 variables and whether there is evidence of statistical significance, while Mutual Information represents the level of information we can collect for one variable if we know the value of another variable). Each value represents the level of information that becomes known of the variable 2 when the value of the variable 1 is known, and as a result the degree in which one variable affects the other. In addition, Correlation examines linear relationships, which is not the case in Mutual Information Ratio. Finally, we are able to collect information from zero values; while the typical Correlation Analysis statement is whether there is evidence of statistical significance or not (and in case of not there is no further conclusions), zero values of Mutual Information Ratio indicates that the variable 1 has absolutely no impact on the second variable. Zero Mutual Information Values indicates as a result that variable 1 has no impact on variable 2.

On Variables level, Mutual Information provides us more information as well. For example, the number of Start-Ups founded and the Previous Start-Ups that Still Survive provide us information regarding the Reasons for founding the Start –Up.

The number of Years of operation becomes a key success factor (38.7%); this relationship is not included on Table 3; while it makes sense since operations for many years leads founders to the perception that their company is indeed successful (since it is surviving and likely growing). This indicates that there is no linear relationship, which is also reasonable (more years do not have a linear relationship with the level of (perceived) success).

The MI score between Start-Up founded and Start-ups Surviving is also high (35%) which is reasonable (the more start-ups founded the more likely it becomes to have some of them surviving after a while). However, again there is no evidence of correlation significance and linear relationship (while there is a limiting factor, since a surviving company has to be founded, there is not a linear relationship between surviving and founded companies).

Age and Experience also have a high MI score (34%), which is also reasonable, which is also not included on Table 3, thus not a linear relationship. Years of Operations also linked to the various reasons for founding a company, with serial entrepreneurship (reasons for founding previous companies) with the existence of strategic alliances. These relationships can be easily interpreted since many years of

company operations can lead to strategic alliances and strategic alliances can help the survival of a company; the existence of several reasons for founding a company and for past entrepreneurial experience deriving from the existence of reasons to found companies in the past shows determination and a strong tendency towards (serial) entrepreneurship. Again, these relationships are not included in Table 4.

As already discussed, Information Gain Ratio provides insights of the additional information that can be retrieved from one variable through the observation of another variable; and in this case this is an asymmetric ratio, meaning that the information that can be acquired for variable 1 from observation of variable 2 is different from the information that can be acquired for variable 2 from observation of variable 1.

Table 3 indicates the highest values of Information Gain Ratio with values ranged 0.35 – 0.15. Again, a number of one side relationships are revealed (Reasons for founding Start Ups in the past and number of previous Start-Ups founded, Operation Years of the company and facing different challenges (other)). Perceptions regarding competition reveal that if a founder consider he is facing competition from traditional companies (no other start-ups), there is also a high perception that in reality there is no Competition (which can be explained by the superiority or high differentiation of his proposed solution). In total 490 sets of variables have Information Gain Ratio value between 0.350-0.05.

Table 3: Information Gain Ratio

Variable 1	Variable 2	Information Gain Ratio
Previous Reasons	Previous SU	0.35
Operation Years	Other	0.341
Traditional	No Comp.	0.329
Operation Years	Sales 100k	0.325
Get customers	Other	0.316
Operation Years	Total Reasons	0.295
Previous Reasons	Pr. Surviving	0.294
Pr. Surviving	Previous SU	0.264
Previous SU	Pr. Surviving	0.26
Operation Years	Funds 100k	0.253
Previous SU	Previous Reasons	0.25
Education	Other	0.24
Sales 100k	Other	0.237
Major	Unclear	0.232
Operation Years	Disruption	0.229
Funding	Funds 100k	0.221
New approach	Minor	0.215
Pr. Surviving	Previous Reasons	0.214
Operation Years	Success	0.213
Education	Disruptive SU	0.21
Strategic	New Start Ups	0.209
Total Reasons	Process Innovation	0.208
Operation Years	Minor	0.203
Operation Years	B2BC	0.202
Operation Years	No Comp.	0.198

Experience	Age	0.19
Total Reasons	Disruptive SU	0.189
No Comp.	Traditional	0.185
Age	Experience	0.182
Strategic	Other	0.182
Funding	Disruption	0.178
Funding	Disruptive SU	0.177
Funding	Get funding	0.171
Pr. Surviving	Unclear	0.17
Founders	Unclear	0.17
New Start Ups	No Comp.	0.17
Total Reasons	Business Model	0.169
Experience	Sales 100k	0.167
Operation Years	Business Model	0.167
Operation Years	Disruptive SU	0.166
Operation Years	Previous SU	0.166
Openness	Disruption	0.166
Process Innovation	Total Reasons	0.163
Strategic	Funds 100k	0.163
Operation Years	Strategic	0.163
Operation Years	Previous Reasons	0.16
Previous SU	Minor	0.159
Business Model	Funds 100k	0.159
Operation Years	Education	0.157
Previous Reasons	Minor	0.155
Previous Reasons	Funds 100k	0.155
Openness	Minor	0.154
Major	Minor	0.153
Funds 100k	Business Model	0.153
Operation Years	Process Innovation	0.152
Operation Years	New Product	0.152
Success	Unclear	0.151
Operation Years	B2BC	0.15
Operation Years	Experience	0.15

Source: Authors' elaboration (Information Gain Values, using R Software, 2023)

Information Gain ratios have a total of 1981 sets of variables with values from 0.35 to zero, including those listed on Table 3. However due to the lack of space and since the presented data provide sufficient evidence regarding the use of Information Gain as a tool which can be used to provide additional insights to entrepreneurial research, the rest of the data are not presented here.

Another set of 86 variables have zero Information Gain value, thus any additional observation of the first variable will not enable us to collect any information regarding the variable 2.

In both cases Information Gain Ratio provides us guidance to identify variables for more observation and data collection focus, to collect more information and reduce uncertainty of the variable 2. Table 5

below includes a list with zero values of Information Gain Ratio, indicating that further observation or data collection of the variable 1 will not add new information regarding the variable 2.

As discussed, Spyropoulos (2020a), used the same dataset and SPSS software in order analyze the correlation between variables, with the use of Pearson Correlation Coefficient. Findings indicate Statistical Significance for the Correlations presented at Table 4.

Table 4: Statistical Significant Correlations

Variable 1	Variable 2	Correlation	Pearson Value
Success	B2B	Weak	-0,206*
Success	Sales 100K	Weak	0,218*
Education	Disruption	Weak	0,293**
Education	Get Funding as Challenge	Weak	0,282**
Education	Improve Product as Challenge	Weak	0,204*
Number of Founders	IP as Competitive Advantage	Weak	-0,241 *
Number of Founders	New Market Creation	Weak	-0,238*
Experience	Get Funding as Challenge	Weak	0,296**
B2B	Technology as Competitive Advantage	Weak	0,222*
Opportunity based	Find new customers challenge	Weak	0,274**

Note: * Correlation is significant at the 0.05 level (2-tailed)

** Correlation is significant at the 0.01 level (2-tailed).

Source: Spyropoulos (2020a, p. 41)

Table 5 presents indicative sets of variables for which there was no evidence of Correlation Significance, indicating that no further conclusion could be made at that point using correlation.

Table 5: No evidence of Correlation Significance

Variable 1	Variable 2
Success	Gender
Success	Education
Success	No of Founders
Success	Previous Start-Ups Ventures
Success	Opportunity Based Start-Up
Success	Technology Based Start-Up
Success	Business Model Based Start-Up
Success	Proof of Concept
Success	Prototype
Success	Funds 100k
Disruptive Start-Up	Previous Start-Ups
Disruptive Start-Up	Previous Survived Ventures
Education	Opportunity – based
Education	Technology - based
Education	Business Model – based
Education	Get more customers challenge
Education	Major Value to Customer
Education	Minor Value to Customer
Education	Management as Comp. Adv.
Education	IP as Comp. Adv.

Education	Business Model as Comp. Adv.
Age	Success
Age	Management as Comp. Adv.
Age	Business Model as Comp. Adv.
Age	Intellectual Property as Comp. Adv.
Age	Prototype Development
Age	Proof of Concept
Age	Sales 100k
Age	Funds 100k
Number of Founders	Opportunity
Number of Founders	Technology Based
Number of Founders	Business Model
Number of Founders	Sales 100k
Number of Founders	Funding 100k
Number of Founders	B2B
Number of Founders	Management as Comp. Adv.
Number of Founders	Intellectual Property as Comp. Adv.
Number of Founders	Business Model as Comp. Adv.
Number of Founders	Improve Product as Innovation Level
Number of Founders	New Product as Innovation Level
Employee Experience	Previous Start-Ups
Employee Experience	Opportunity Based
Employee Experience	Technology Based
Employee Experience	B2B
Employee Experience	Business Model
Employee Experience	Process Innovation
Employee Experience	Sales 100K
Employee Experience	Funding 100K
Employee Experience	Proof of Concept
Employee Experience	Prototype

Source: Spyropoulos (2020, p. 21) (SPSS Correlation Analysis)

Table 6 presents the Mutual Information Ratio of Success (as a founder’s perception) with each one of the rest variables of the model.

This is consistent with entrepreneurial success been a multi-factor dependent variable. The key variables appear to be considering Funding as a critical challenge (12.7%), having strategic partnerships (12.4%) and securing Funding of a minimum 100 thousand euros (9.7%).

It is reasonable to assume that these connections can be interpreted easily, since founders who consider securing funding as a key challenge will actively seek funding and make all necessary actions to secure it; in addition, the role of strategic alliances is critical for new business ventures. Again, these relationships are not included in Table 6 (Correlation Results).

It is worth mentioning some relationships such as the set between Success and Prototype (7,6%), Success and Funds 100k (9,7%), which reveal a (partial) non-linear relationship between these variables.

Table 6: Mutual Information for Success Related Factors

Variable 1	Variable 2	Mutual Information
Success	Operation Years	0,387
Success	Pr. Surviving	0,157
Success	Previous SU	0,136
Success	Funding	0,127
Success	Strategic	0,124
Success	Age	0,116
Success	Experience	0,115
Success	Founders	0,115
Success	Previous Reasons	0,112
Success	Funds 100k	0,097
Success	Unclear	0,076
Success	Prototype	0,076
Success	Openness	0,062
Success	Traditional	0,061
Success	Education	0,054
Success	B2B	0,05
Success	Sales 100k	0,05
Success	Process Innovation	0,048
Success	Disruption	0,046
Success	Business Model	0,044
Success	New Product	0,042
Success	Management	0,041
Success	Minor	0,041
Success	Improved product	0,037
Success	New Start Ups	0,029
Success	Other	0,029
Success	Get funding	0,028
Success	Technology	0,025
Success	B2B	0,025
Success	Business Model	0,023
Success	No Comp.	0,022
Success	Dirruptive SU	0,02
Success	Improve product	0,018
Success	Technology	0,016
Success	Major	0,016
Success	Opportunity	0,014
Success	New product Approach	0,014
Success	IP	0,01
Success	New approach	0,007
Success	POC	0,005
Success	Gender	0,005
Success	Get customers	0,004
Success	New Market creation	0,003

Source: Authors' elaboration (Mutual Information Calculations using R Software, 2023)

5 Conclusion

It becomes clear from the above findings and analysis that Mutual Information can provide entrepreneurial researchers a far more detailed picture, in comparison to Correlation Coefficient. By definition, Mutual Information reveals relationships that cannot be identified using linear statistical analysis. As a multi factor variable, entrepreneurial success can be difficult to be defined, measured and identify key parameters that will help entrepreneurs boost entrepreneurial success; however Mutual Information analysis provides additional insights, much more different than the classical linear correlation approach to both academics and entrepreneurs.

However, even though Mutual Information can be used to identify a relationship between variables, the nature of this relationship cannot be easily analyzed; no information is provided in a linear form such as analogous or reverse analogous relationship, and further analysis would be required to reach conclusions regarding the nature of the relationship identified, including usage of other statistical and mathematic tools. In addition, Information Gain Ratio reduces entropy and uncertainty and as such may provide additional insights on which variables researchers may focus more, either through observation or through data collection.

The present study with the use of Mutual Information highlights a number of non-linear relations between variables for the Greek I.T. Start-Ups; this reveals a stronger connection between success and other variables which is in fact a different landscape: according to Pearson Correlation coefficient success is related to B2B (in a reverse analogous relationship) and to sales of over 100k; however Mutual Information links success to Operation Years, Previous Start-Ups and Surviving Start-Ups (relating to serial entrepreneurship), Funding, Strategic Alliances and Age, factors which are closely associated with entrepreneurial success according to theory (Aulet 2013, Santisteban et al., 2017).

The above conclusions do not reduce the value of Correlation Coefficient, as a tool for statistical analysis and forecasting; however, forms a basis for the use of Mutual Information and Information Gain Ratio in order to gain additional insights and reveal new relationships in complex and multi-factor phenomena, such as entrepreneurial success.

6 Limitations and future research recommendations

There are several areas for additional research regarding use of entropy ratios, such as Mutual Information and Information Gain Ratios on entrepreneurial theory.

First of all, the sample size has to be increased; the present study includes a sample of 94 founders, and since the exact dataset was used in previously published research, it was a safe starting point for comparing results between Correlation and Entropy ratios. To this respect, future research with larger sample may offer additional insights.

Second, use of Correlation and Entropy ratios for data processing need to be further tested across different models describing entrepreneurship, which may use similar or different variables. Another promising area for future research is to include not only new models of entrepreneurship theory, but to examine how such models work across different business ecosystems.

Finally, it must be mentioned that identifying relationships between variables with the use of entropy networks provides us a better understanding of the phenomenon examined but additional research may be used to extract further information; use of advanced mathematics and statistical tools (cluster analysis, network theory, machine learning etc.) may be used to generate additional information and a basis for forecasting.

References

- Aulet B. (2013) *Disciplined entrepreneurship, 24 steps to a successful start up*. John Wiley & Sons, Inc; 2013.
- Ayyappan G., Nalini C., Kumaravel A., (2017), *Efficient Mining for Social Networks using Information Gain Ration Based on Academic Dataset*, *International Journal of Civil Engineering and Technology*, (IJCIET), Volume 8, Issue 1, January 2017, pp 936-942
- Azhagusundari B., Thanamani A.S. (2013), *Feature Selection based on Information Gain*, *International Journal of Innovative Technology and Exploring Engineering*, (IJITEE), ISSN: 2278-3075, Volume 2, Issue 2, January 2013
- Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.X. and Veyrat-Charvillon, N. (2011). *Mutual information analysis: a comprehensive study*. *Journal of Cryptology*, 24(2), 269-291.
- Corley K.G., Gioia D.A., (2011), *Building Theory about Theory Building: What constitutes a Theoretical Contribution?* *Academy of Management Review*, 2011, Vol 36, No 1, 12-32
- Prasetyo B., Alamsyah, Muslim A.M., Baroroh N., (2021), *Evaluation of feature selection using information gain ration on bank marketing classification using naïve bayes*, *Journal of Physics, Conference Series*, 1918 042153
- Santisteban J., Mauricio D. (2017). *Systematic Literature Review of Critical Success Factors of Information Technology Startups*, *Academy of Entrepreneurial Journal*, 2017, Vol:23 Issue 2
- Sklavounos, N., Rotsios, K., Kartsiotis, G., Spyropoulos, T.S. and Polychronidou, P. (2023). *Exploring foreign partner perceptions in international strategic alliances of Greek SMEs in South East Europe*. 49th Conference of the European International Business Academy (EIBA 2023), December 15-17 2023, ISEG – University of Lisbon, Portugal.
- Spyropoulos, T.S. 2019. *Greek IT Start Ups – An Analysis of Founder’s Perceptions*. *Scientific Bulletin, Economic Sciences*, 18(1). http://economic.upit.ro/RePEc/pdf/2019_1_1.pdf.
- Spyropoulos TS (2020a). *Digital Greek start-ups – An analysis of founder’s perceptions*. *Advances in Management and Informatics*. 2020;(5th Edition). 33-48, available at https://repository.afs.edu.gr/bitstream/6000/356/4/Spyropoulos_Digital-Greek-start-ups_2020.pdf#page=34
- Spyropoulos TS. (2020b) *Knowledge management challenges for start-ups: A framework proposal*. *International Journal of Entrepreneurship and Business Development*. 2020;3(3). P.248-260
- Spyropoulos T.S., Andras C., Polychronidou P., (2022), *An Analysis of Start-Up Founders Perceptions based on Entropy Ratios – Evidence from the Greek IT Market*, *European Research Studies Journal*, Volume XXV, Issue 3, 2022 p.p. 500-516

Korištenje uzajamnih informacija i omjera dobitka informacija u istraživanju poduzetništva: empirijski slučaj grčkih IT startup tvrtki

THEOCHARIS STYLIANOS SPYROPOULOS
Međunarodno helensko sveučilište u Seresu, Grčka
Perrotis koledž, Solun, Grčka
theospyr2@es.ihu.gr
ORCID ID: <https://orcid.org/0000-0001-8534-2955>

CHRISTOS ANDRAS
Međunarodno helensko sveučilište u Solunu, Grčka
andraschris@gmail.com
ORCID ID: <https://orcid.org/0000-0001-8674-7823>

ARISTOTELIS DIMKOU
Sveučilište Aristotel u Solunu, Solun, Grčka
arisdimkos@yahoo.gr

PERSEFONI POLYCHRONIDOU
Međunarodno helensko sveučilište u Seresu, Grčka
polychr@es.ihu.gr
ORCID ID: <https://orcid.org/0000-0003-1517-011X>

Sažetak: Većina poduzetničkih kvantitativnih istraživanja usredotočena je na korelacijske koeficijente. Međutim, nova statistička analiza temeljena na entropiji, kao što su uzajamne informacije i omjeri dobitka informacija, bacaju novo svjetlo na razumijevanje odnosa među varijablama i nude pogled na nelinearne odnose. U radu se ispituju ključne poduzetničke varijable korištenjem omjera međusobnog informiranja i dobivanja informacija i uspoređuju rezultati koristeći isti skup podataka na uzorku novoosnovanih grčkih IT tvrtki. Korištenje omjera međusobnog informiranja i dobitka informacija otkriva više odnosa između ispitivanih varijabli u usporedbi s Pearsonovom korelacijom. Nadalje, u radu se uspoređuju rezultati Pearsonove korelacije i omjera međusobnog informiranja i dobivanja informacija kako bi se izveli novi zaključci o percepcijama osnivača grčkih IT startup tvrtki. Nalazi pokazuju da korištenje uzajamnih informacija otkriva skup čimbenika koji pridonose poduzetničkoj percepciji uspjeha koja se značajno razlikuje od zaključaka temeljenih na analizi koeficijenta korelacije. Preciznije, čimbenici kao što su godine poslovanja i prethodni startupovi igraju daleko važniju ulogu od B2B i prodaje. Studija nudi originalan doprinos poduzetničkoj znanosti, uvodeći upotrebu matematičkih omjera koji se temelje na entropiji, kao što su međusobno informiranje i dobitak informacija u poduzetničkom istraživanju. Studija naglašava da upotreba i rezultati izvedeni iz takvih omjera omogućuju istraživačima identifikaciju većeg broja informacija o (nelinearnim) odnosima između varijabli u usporedbi s metodama koeficijenta korelacije.

Ključne riječi: Startup tvrtke, poslovni model, međusobno informiranje, dobivanje informacija, teorija mreže, entropija, poduzetništvo

JEL klasifikacija: M13, M 19, M20, L26, O30, O34