Europeaan Research Studies Journal
Volume XXV, Issue 3, 2022

pp. 500-516

An Analysis of Start-Up Founders Perceptions Based on Entropy Ratios - Evidence from the Greek IT Market

Submitted 05/09/22, 1st revision 20/09/22, 2nd revision 09/10/22, accepted 30/10/22

Theocharis Stylianos Spyropoulos[1], Christos Andras[2],
Persefoni Polychronidou[3]

Abstract:

Purpose: *The study examines use of Entropy related ratios in Entrepreneurship studies. Entropy ratios, such as Mutual Information (M.I.) and Information Gain (I.G.). More specifically, the study focuses on perceptions of I.T. (Information Technology) Greek Start-up founders with the use of Mutual Information and Information Gain ratios.*
Design Methodology: *The study compares and discusses key findings between conclusions drawn from correlation coefficient and entropy ratios, regarding the managerial and entrepreneurial implications, based on the exact same dataset of previous published reserch. Entropy based ratio focus on probability analysis in order to measure dependencies between variables. While the mutual information is a measure of dependence between variables, which expresses the quantity of information obtained on one variable when the value of another variable is known, the information gain ratio measures the reduction of the entropy and therefore the reduction of uncertainty of one variable that derives from information gained regarding the value of another variable.*
Findings: *The study concludes that Mutual Information and Information Gain ratios offer significant information to entrepreneurial research, identifying non-linear relationships. Correlation Coefficient provides a more limited amount of information.*
Practical Implications: *Use of Entropy ratios will offer additional insights to both researchers and managers, by providing evidence of non-linear relationships.*
Originality/Value: *The research presented here is part of a larger study and further confirms preliminary findings conducted on a smaller sample.*

Keywords: *Innovation Management, Marketing, Start-Ups, Founders, Business Models, Entrepreneurship, Strategy, Greek.*

JEL Classification Codes: *L26, M13, O30, O31, O32, O33.*

Paper type: *Research article.*

[1]*Corresponding author, Department of Economic Sciences, International Hellenic University, Greece/Department of International Business, Perrotis College, Thessaloniki, Greece, email: theospyr2@es.ihu.gr;*
[2]*Department of Industrial Engineering and Management, International Hellenic University, Thessaloniki, Greece, email: andraschris@gmail.com;*
[3]*Department of Economic Sciences, International Hellenic University, Greece, email: polychr@es.ihu.gr;*

1. Introduction

The present study examines the use of Entropy Ratios, and most specifically Mutual Information and Information Gain on entrepreneurial research and compares findings with previous published research. Most of the entrepreneurial research depends on quantitative data to create, revise or evaluate models, usually with entrepreneurial success as a dependant variable and several other variables (supported by entrepreneurial theory and past research) as independent variables. In most of the cases, the independent variables are considered independent between them. Usually, correlation coefficient analysis is used to identify linear relationships within a given dataset.

Recent studies confirm the dependence on analysis in quantitative research: correlation analysis, a quantitative research method, is used for explanatory research, in order to explore the extents to which some variable co-vary (Cresswell, 2008) while further research examines the use of applied research through surveys and collection of large amounts of data (Picardi *et al.,* 2014). Isotalo (2014) provides the framework of Statistical Correlation Coefficient Analysis. 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)." (Apuke, 2017, p. 44).

However, the amount of information confirmed by Correlation Analysis appears to be limited, and use of other statistical and mathematical methods may be able to provide us more information regarding the relationships between the variables included in the entrepreneurial models. This will enable researches to gain additional insights and secure a deeper understanding regarding the relationships between variables examined and the complex entrepreneurial reality.

Boer *et al.* (2015, p. 1242) argue that "A contribution to theory consists of a better or more inclusive explanation of phenomena in the world, often couched in mathematical language."

Regarding theory building and testing result the same authors conclude that "The key difference between building and testing research is then where the theoretical argument should mainly be developed, not how it should be constructed. Hence the suggestions that follow should be equally useful in developing good theoretical arguments for all papers. When done well, the theoretical argument is tied to the task at hand." (Boer *et al.,* 2015, p. 1245). Finally, they conclude regarding theory building and testing: "A good theoretical argument is linked to the data and builds on a small number of existing theories, preferably one or two, to make a coherent argument.

The variables that are measured align with the relationships the theory predicts. Boundary conditions are clearly spelled out and there is a clear path from supporting or rejecting a hypothesis to the theory being used. Finally, a good theoretical argument makes it clear how results could be used to falsify as well as confirm." (Boer *et al.,* p. 1246).

The authors of the present study not only conclude that Mutual Information and Information Gain ratio offer significant information to entrepreneurial research, and highlight potential impacts, such as the use of subsets of datasets to identify more insights on the relationship between variables.

2. Literature Review

Related literature review on entrepreneurial studies has been summarized in recent studies. 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 IT Start-Ups. More specifically the dataset used is the one used in Spyropoulos (2019) where 130 Start-Up Founders responded through a closed questionnaire, and results were drawn using Spearman Correlation Coefficient.

The key issue examined in the present research is the use of Entropy Ratios, in entrepreneurial research, in order to identify and evaluate non-linear relationships. Recent Research (Spyropoulos and Papageorgiou, 2021) discuss the limitations of using solely statistical tools to identify linear relationships while Spyropoulos *et al.* (in press) provide evidence of identifying nonlinear relationships using Mutual Information and Information Gain ratios on entrepreneurial research. The same authors suggested further research, using larger datasets.

As a result, the present study explores relationships between variables on entrepreneurial research using Mutual Information and Information gain ratios, and comparing findings with previous study (Spyropoulos, 2019) which was based on Spearman Correlation Coefficient to examine whether additional information can be drawn.

Recent studies highlight the role of Business Model Innovation. "First, it represents an often underutilized source of future value. Second, competitors might find it more difficult to imitate or replicate an entire novel activity system than a single novel product or process". (Amit *et al.*, 2012, p. 1). Further studies (Zafar *et al.*, 2013) highlight the role of Culture, Genter, Education, Family and self-perception on entrepreneurial success.

Song *et al.* (2008) summarize previous research on entrepreneurships examining factors such as Competition, Business Environment, Product Innovation, Marketing, Industry & Market Experience, Firms age, prior start-up experience, alliances and founding team. Wilde *et al.* (2018) summarize previous literature review regarding the role of Age and Gender in entrepreneurship.

3. Research Approach and Methodology

The study examines 130 questionnaires, with the literature review and descriptive statistics available at Spyropoulos (2019, p. 5-9). The data collected from Greek IT Start Up founders, with the use of closed questionnaires, from September 2018 to March 2019. The dataset was examined with the use of SPSS software and Spearman Correlation, in order to identify relationships between the key entrepreneurial variables, as defined by literature review.

4. Mathematical Background – Entropy, Mutual Information and Information Gain– Key Concepts

Information Gain ratio measures the feature(s) possessing the most information, based on a specific class (Shaltout *et al.*, 2014). Recent studies indicate that I.G. can also be also used in Artificial Intelligence models, such as the tree structure formation (Alhaj *et al.*, 2016). Such an approach enables future research to further use predictive analytics and Artificial Intelligence models in order to analyze entrepreneurship.

5. The Present Study

Regarding the Mutual Information Ratio in the dataset examined, there were 12 pairs of variables with Mutual Information ratio value between 0.4 and 0.2, 42 pairs of variables with Mutual Information ratio value between 0.2 and 0.1, 135 pairs of variables with Mutual Information ratio values between 0.099 and 0.5, 734 pairs of variables with Mutual Information ratio values between 0.49 and 0.001, and 65 pairs of variables with mutual Information value zero.

The first remark at this point is that the information acquired from the fact that there are pairs of variables with M.I. ratio value equals to zero (or extreme low values of 0.01%) is in fact important; this in fact indicates that the two variables are independent, and the researcher can use this information accordingly (e.g., if variables were used as explanatory factors in a multi-factor phenomenon, such as entrepreneurial success, and their M.I. ratio with the dependant variable is zero, they can be effectively removed from the model, since there is no relationship between the variables; in case they share an important share of M.I. values, adjustments may be required).

The first pair of variables with the highest Mutual Information ratio value is Age and Experience (0.409). However, this relationship is not included in Table 4, which includes all Statistical Significant Correlations (Spyropoulos, 2019). The interpretation is that there is a non-linear relationship between the variables Age and Experience; from the business perspective the non-linearity can be explained with the fact that people may change careers, thus actual age does not have a linear relationship with experience in a specific field.

Previous Start-Ups and Previous Surviving Start-Ups have a M.I. ratio value of 0.38 and 0.34 with Previous Reasons variable, indicating that there may be non-linear relationships between the Reasons for Establishing a Start-Up and the mount of companies established and still Surviving; this means that a certain percentage of founders tends to establish new companies for a similar reason (e.g., identify a business opportunity, take advantage of a new technology or a new business model) and that he or she may establish a new company in the future if the same reason appears, therefore several founders who are serial entrepreneurs are likely to establish or develop a certain pattern of behavior and identification of an opportunity to establish a new company. Once again, such relationships have not been identified with the use of Spearman Correlation Coefficient (as included in Table 4).

Table 1 includes the pairs of variables where the value of their Mutual information ranges between 0.4 and 0.1. In addition, the pairs of variables with M.I. ratio values between 0.4 and 0,1 are compared with the information available at Table 4, the existence of Statistically Significant Correlation (Spearman).

Table 1. *Mutual Information Ratio*

Variable 1	**Variable 2**	**Mutual Information Ratio**	**Spearman Correlation**
Age	Experience	0.409	No
Previous SU	Previous Reasons	0.385	No
Pr. Surviving	Previous Reasons	0.341	No
Success	Years	0.28	No
Previous SU	Pr. Surviving	0.246	No
Years	Strategic	0.243	No
Experience	Years	0.242	No
Business Model	Total CompAd	0.221	No
Years	Previous Reasons	0.22	No
No Comp.	Traditional	0.211	No
Years	Sales 100k	0.21	No
Years	Funding	0.2	No
Founders	Years	0.185	No
Previous SU	Years	0.182	No
Strategic	Funding	0.181	No
Education	Years	0.179	No
Management	Total CompAd	0.17	No
Strategic	New Start Ups	0,162	No
Age	Years	0,156	No
Years	Total CompAd	0.155	No
Years	Disruption	0.155	No
Education	Experience	0.154	Yes
Success	Pr. Surviving	0.15	No
Education	Funding	0.147	Yes
Business Model (Opportunity)	Business Model (Competitive Advantage)	0.146	No
Get funding (Challenge)	Funding (seeking/secured)	0.144	No

IP	Total CompAd	0.139	No
Funding	Total CompAd	0.132	No
Years	Minor	0.129	No
Founders	Pr. Surviving	0.128	No
Success	Strategic	0.125	No
Age	Funding	0.123	Yes
Age	Education	0.122	Yes
Funding	Openness	0.12	No
Technology (as Opportunity)	Technology (as Competitive Advantage)	0.12	No
Experience	Sales 100k	0.119	No
Success	Funding	0.116	No
Experience	Previous Reasons	0.115	No
Success	Experience	0.115	Yes
Years	New Product	0.115	No
Funding	Disruption	0.115	No
Technology	Previous Reasons	0.114	No
Years	Management	0.113	No
Years	Openness	0.112	No
B2B	B2C	0.112	No
Experience	Prototype	0.11	No
Years	get funding	0.109	No
Years	Pr. Surviving	0.108	No
Years	Traditional	0.106	No
Years	IP	0.105	No
Years	Process Innovation	0.105	No
Previous Reasons	Funding	0.103	No
Pr. Surviving	Total CompAd	0.102	No
Major	new approach	0.101	No

Source: *Own study.*

The key finding and observation is that from the pairs of variables with the highest Mutual Information value ratio, none of the top 12 (Mutual Information ration between 0.4-0.2) in listed in Table 4, which includes findings of the Spearman Correlation Coefficient. In fact, just the 22^{nd} pair in the list of Table 1 variables pairs, “Education” and “Experience” with Mutual Information ratio value of 0.154 is the first one included in Table 4. In total, only 6 of the 54 pairs of variables listed in Table 1 are also listed in Table 4.

The direct conclusion is that even though Mutual Information analysis provides 54 pairs of variables where the data and observed values relate with a minimum 10% of total data, and (most probable) they relate in a non-linear, straight forward way, Spearman Correlation Coefficient identified only 5 of these relationships as linear and statistically significant.

Therefore the primary key finding is that actually Mutual Information Ratio may provide researchers much more information that Correlation Coefficient. The second conclusion is that search for linear relationships, with the use of Correlation

Coefficient enables researchers usually to identify a limiting number of relationships between variables (and this in terms of the total sample and the whole range of values).

Since Mutual Information reveals information shared between variables, and effectively areas where values of one variable are related with the values of another variable, while these relationships are not identified with the use of Correlation Coefficient in our sample, the logical conclusion is that there are many more non-linear relationships to be identified; however these relationships may not be linear or even of the same direction (analogous or reverse) for all values of the sample. In other words there may be specific ranges of values between variables where there may be even a direct linear relationship (Correlation Coefficient) but this may not apply to the total range of value and our whole sample.

However considering the findings of the present research the following conclusions can be drawn:

First, that use of Mutual Information ratio reveals relationships between variables, that are not identified with some commonly used methods of statistical analysis, such as Correlation Coefficient, and in this case, Spearmen Correlation Coefficient.

Second, and as a result of the previous remarks, entrepreneurship is recommended to be viewed not just by examination of all the available data as a whole, but further analysis across different values of specific variables may reveal more information and provide a much better understanding of the entrepreneurial research.

These observations add much in closing potential gaps between theory and practice and cast new light into entrepreneurial research; analyzing entrepreneurial data from different groups of entrepreneurs is much closer to reality than analyzing all entrepreneurs as a whole. In a similar way academics and marketing professionals use several segmentation methodologies and criteria, entrepreneurs (and start-up founders) are recommended to be analyzed with the use of some segmentation criteria as well.

By considering a Mutual Information ratio value between 10% and 15%, in cases where the same variables have no Correlation Coefficient marks a possible cluster in our data (data subset) where actually (even a linear) relationship may exist. This in turn leads to very interesting conclusions, especially if we consider the large number of variables in entrepreneurial models and the wide range of values between these variables. By segmenting the data (creating clusters), more relationships can be identified and explained, provided much more variable findings to both entrepreneurs and researchers.

Table 2 provides a list of the Mutual Information ration value between each variable and “Success” variable, providing information whether the relationship between the variable examined and "Success" has been proved a statistical important Correlation Coefficient, using the Spearman Correlation Criterion.

Table 2. *Success Mutual Information Ratios*

Variable 1	Variable 2	Mutual Information Ratio	Spearman Correlation
Success	Years	0.28	No
Success	Pr. Surviving	0.15	No
Success	Strategic	0.125	No
Success	Funding	0.116	No
Success	Experience	0.115	Yes
Success	Previous Reasons	0.083	No
Success	Founders	0.08	No
Success	Prototype	0.076	No
Success	Age	0.075	No
Success	Previous SU	0.074	No
Success	Education	0.071	No
Success	Minor	0.064	No
Success	Openness	0.056	No
Success	Unclear	0.054	No
Success	Total CompAd	0.051	No
Success	Disruption	0.044	No
Success	B2B	0.038	Yes
Success	Traditional	0.036	No
Success	Sales 100k	0.034	Yes
Success	Opportunity	0.034	No
Success	Improved product	0.032	No
Success	B2C	0.031	No
Success	get funding	0.03	No
Success	Technology	0.029	No
Success	New Product	0.026	No
Success	POC	0.026	No
Success	New Start Ups	0.025	No
Success	new approach	0,021	No
Success	Business Model	0.019	No
Success	Gender	0.018	No
Success	Major	0.018	No
Success	IP	0.017	No
Success	Technology	0.016	No
Success	Management	0.016	No
Success	No Comp.	0.015	No
Success	Business Model	0.015	No
Success	Other	0.014	No
Success	Dirruptive SU	0.013	No
Success	Process Innovation	0.013	No
Success	Funds 100k	0.012	No
Success	New product Approach	0.005	No
Success	New Market creation	0.004	No
Success	improve product	0.004	No
Success	get customers	0.002	No

Source: *Own study.*

Regarding Success factor, the variables set with the highest scores in Mutual Information Ratio are Success and variables “Years” (28%), “Previous Surviving Start-Ups” (15%), “Strategic” (12.5%) “Funding” (as a challenge) (11.6%) and “Experience” 11.5%. It has to be noted that from these variables “Experience” is the only one included in Table 4, which includes Spearman Correlations (thus implying a linear relationship between the variables).

The relationship between “Success” and “Years” (of Operation) variables can be explained that since the company operates for several years, founders consider it a success; however if after a period of several years growth remains low, founders may change the perception of the level of success. The relationship between the variables “Success” and “Previous Surviving Start-Ups” can be explained since serial entrepreneurs may be more successful in a new venture, and the non-linear relationship may impose some negative impact as well (e.g., lack of focus or time for the founder of the new venture due to commitments to the other companies).

Strategic Partnerships can help, but again the non-linear relationship may suggest that strategic partners offer benefits but may as well impose limitations for a new company (e.g. spin-offs and funded companies have members on their boards which may not share the same philosophy, values or managerial approaches). “Funding” as a challenge may work in a similar way; prepare the company to be ready to accept investor’s funds but may also mislead the company from the market or customer focus.

In most entrepreneurial models, “Success” is the dependant variable and several “independent” variables are considered in order to help entrepreneurs and academics to understand the phenomenon of business success. These variables are then examined and statistically analyzed with the use of various statistical analysis tools.

Again, the findings are very interesting: from the 44 pair of variables examined in Table 2, only 3 pairs of variables are included in Table 4 (Statistically Important Correlation Coefficient). The first 4 pairs with the highest Mutual Information ratio are not present in Table, and the first pair of variables listed in both Tables 1 and 4 is the pair of variables “Success” and “Experience”, listed 5th in Table 1 with Mutual Information ratio value 11.5% (and a very weak relationship).

Even though there are several variables with lower Mutual Information ratio value, the information we may retrieve from further clustering or segmenting the data can be important.

For example, “Success” and “B2B” have a very weak Spearman Value (-0.188) and a relative low Mutual Information ratio value (0.038). This can be interpreted that business that address to the B2B market have a negative impact on success (-0.188) but there were successful companies operating in the B2B market; considering the small number of years of operation for most companies (since they are start-ups) and the fact that a large number of these companies has not been established yet (so actually they cannot issue invoices, have sales and feel successful), the fact remains that some (very few) founders of start-up companies even in B2B sector are actually successful (Spyropoulos, 2019).

And this is exactly one research area (of many potentials) where further study of what the actual few successful founders did differently that can offer useful insights, instead of trying to create generalizations from a the whole samples, reaching conclusions that may be great generalizations but poor practical insights that can promote entrepreneurial success.

Another example comes from "Success" and "Founders". Literature review supports that a team has more possibilities to be successful rather than a single individual (Aulet, 2013; Roberts *et al.*, 2015; Spyropoulos, 2020b). So while the theory supports that in principle a team of founders has more possibilities to be successful, an effort to identify a linear relationship may be in vain.

From a managerial perspective the question is "how many founders should a start-up have" and there cannot be a simple question. Two, three of four founders may appear ideal if their skills are complimentary and share the same philosophy or add some value (network, experience), however in practice if more founders are added the risks and disagreements become disproportional high and unmanageable – so there cannot be a linear relationship between "Success" and number of "Founders" variables.

A very promising analysis are ways to relate successful founders with the specific number of founders team and reach a conclusion in the form that "teams with X founders tend to be the most successful ones". A relevant promising research approach was recently examined analyzing entrepreneurial datasets with the use of Network Theory (Spyropoulos *et al.*, 2021b).

The mutual information expresses the quantity of information one has obtained on X by observing Y. The mutual information of two random variables X and Y is defined as:

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

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).

Information Gain ratio expresses the reduction of entropy between two variables and therefore a reduction in uncertainty.

$$I_{AB} = \frac{S[A] - S[A|B]}{S[A]} = \frac{S[A;B]}{S[A]}$$

Kent (1983) examined the relationships between Correlation Coefficient and Information gain ratio. In addition academic research concludes that "In the case of probabilistic forecasts, Information Gain can give a simple and convincing measure of the accuracy." (Peirolo, 2010, p. 11).

Recent research examines the use of decision trees in machine learning and use Information gain ration to reduce uncertainty "In some real world issues, instances may be ill-known for some factors such as randomness, data incompleteness and even expert's indefinite subjective opinions; however, traditional decision trees can only handle certain samples with precise data. The incompletely observed instances are usually ignored or replaced by a precise one, despite the fact that they may contain useful information" (Gao *et al.*, 2022, p. 1).

Table 3 below includes the pairs of variables with I.G. Ratio value between 0.352 (max) and 0.1. More specifically there are 12 pairs of variables with Information Gain ratio values between 0.35 and 0.2. There are 73 pairs of variables with Information Gain Ratio value between 0.199 and 0.1, 215 pairs of variables with I.G. Ratio values between 0.99 and 0.05, 1561 pairs of variables with I.G. ratio values between 0.499 and 0.001, and 79 pairs of valuables with I.G. ratio value zero.

Regarding the Information Gain ratio, most of the values below are easily interpreted. Founders believing that they face competition only from traditional companies may underestimate competition as well and there seems to be a group of founders that are consistent to the reason why they established companies in the past and whether they survived.

Table 3. *Information Gain Ratio*

Variable 1	**Variable 2**	**Information Gain Ratio**
Traditional	No Comp.	0.352
Previous Reasons	Previous SU	0.274
Years	Sales 100k	0.274
Previous Reasons	Pr. Surviving	0.253
Total CompAd	Business Model	0.227
Age	Experience	0.225
Years	Minor	0.222
Experience	Age	0.221
No Comp.	Traditional	0.216
Total CompAd	Management	0.212
Previous SU	Previous Reasons	0.209
Major	Unclear	0.2
B2B	B2C	0.192
Pr. Surviving	Previous Reasons	0.185
Previous SU	Pr. Surviving	0.183

Education	Other	0.183
Pr. Surviving	Previous SU	0.176
Years	Disruption	0.174
Total CompAd	IP	0.173
Strategic	New Start Ups	0.166
Years	Funds 100k	0.164
Years	Other	0.163
Funding	Funds 100k	0.163
Business Model	Business Model	0.163
Founders	Unclear	0.162
Business Model	Total CompAd	0.16
New Start Ups	No Comp.	0.16
Funding	Unclear	0.159
Funding	Dirruptive SU	0.157
Years	Dirruptive SU	0.156
Experience	Sales 100k	0.156
Years	Success	0.153
Years	B2C	0.151
Business Model	Business Model	0.15
Education	Dirruptive SU	0.147
Years	Strategic	0.144
Traditional	Dirruptive SU	0.141
Years	Management	0.141
Funding	get funding	0.138
Years	Experience	0.133
Years	IP	0.131
Years	Previous SU	0.13
Funding	Other	0.129
Funding	Disruption	0.129
Disruption	Minor	0.128
Years	No Comp.	0.127
Total CompAd	Technology	0.127
Success	Unclear	0.126
Strategic	Unclear	0.125
Technology	Technology	0.125
Years	New Product	0.124
Management	Total CompAd	0.124
Years	Process Innovation	0.122
Technology	Technology	0.121
Years	Previous Reasons	0.119
Strategic	Funds 100k	0.118
Previous Reasons	Technology	0.115
Pr. Surviving	Unclear	0.113
Previous Reasons	Funds 100k	0.113
Age	Unclear	0.113
Years	Total CompAd	0.112
B2C	B2B	0.112
Total CompAd	Process Innovation	0.112
Success	Pr. Surviving	0.111
Success	Minor	0.111
Experience	Prototype	0.111
Prototype	Dirruptive SU	0.11

Years	Traditional	0.109
Funding	Strategic	0.107
Years	Education	0.107
get funding	Sales 100k	0.106
Years	get funding	0.104
Major	Minor	0.104
Funding	IP	0.104
Business Model	Unclear	0.103
Years	New Market creation	0.102
Years	Business Model	0.102
new approach	Major	0.101
Major	new approach	0.101
IP	Total CompAd	0.101
Openness	Disruption	0.101
Dirruptive SU	Funds 100k	0.101
new approach	No Comp.	0.1
Opportunity	Technology	0.1
Total CompAd	Business Model	0.1

Source: *Own study.*

Table 4 below provides a list of the Statistical Significant Correlations of the exact dataset (Spyropoulos, 2019).

Table 4. *Statistical Significant Correlations*

Variable 1	Variable 2	Correlation	Spearman Value
Success	Sales 100k Euros	Very Weak	0.195*
Success	B2B	Very Weak	-0.188*
Success	Experience	Very Weak	0.177*
Age	Education	Weak	0.353**
Experience	Education	Moderate	0.402**
Get Funding as a Challenge	Education	Weak	0.310**
Unclear Value to Customer	Education	Very Weak	-0.19*
Competition from Disruptive Start Ups	Education	Very Weak	0.18*
Education	Combination of Competitive Advantages	Very Weak	0.189*
Education	New Product	Very Weak	0.186*
Previous Surviving Start-Ups	Prototype Achievement	Very Weak	0.196*
Previous Surviving Start-Ups	Funding 100k	Weak	0.222*
Previous Surviving Start-Ups	Major Value to Customer	Very Weak	-0.178*
Previous Surviving Start-Ups	New Product	Weak	-0.223*
Previous Surviving Start-Ups	New Market Creation	Very Weak	0.193*
Gender	Improve Product as a Challenge	Very Weak	0.182*
Gender	Funding 100K	Very Weak	-0.194*

Age	Get Funding as a Challenge	Weak	0.258**
Age	Prototype	Weak	-0.244**
Age	Funding 100k	Very Weak	0.174*
Age	Previous StartUps	Very Weak	0.190*
Opportunity (Reason)	POC	Very Weak	0.175*
Opportunity (Reason)	Management as Competitive Advantage	Very Weak	0.199*
Disruption	Prototype Development	Very Weak	-0.173*
Disruption	Minor Value to Customer	Weak	-0.318**
Sales 100k	Funding 100k	Weak	-0.318**

Note: *Statistical Significance, *Correlation is significant at the 0.05 level (2-tailed).*
*** Correlation is significant at the 0.01 level (2-tailed).*
Source: *Own study.*

6. Conclusions

Results of the present study confirm the conclusions from previous research (Spyropoulos *et al.*, 2022, in press) and confirm 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.

This can be an indication for further qualitative research to test theories and models developed, or for further quantitative analysis, especially focusing on further clustering; relationships that can be non-linear for the whole sample may be more linear across selected values of specific variables. For example, a founder may consider his company successful if the business operates after 5 years and demonstrates some profits, but at 7 years he/she may reconsider due to limited growth or profitability, which may not fit his/her perceptions or ambitions.

Recent research (Spyropoulos *et al.*, 2021b) also conclude further clustering of the data, and use Network Theory to analyze data not as a whole but as separate clusters, based on network centralities. Such an approach, alongside the use of Mutual Information and Information Gain ratios may offer areas of research that offer more specific insights, instead of focusing on the identification of linear relationships, which may fit or even mislead within selected areas of the data or specific value ranges of the datasets.

For example, a linear trend identified through correlation coefficient analysis of the total dataset may not provide meaningful and useful insights which fit the specific challenge or situation faced by entrepreneurs within a subset of the total dataset. Or, in order to reverse and rephrase this implication from a different point of view, identification of generalized insights may be not applicable or even misleading in specific circumstances an entrepreneur faces, which are limited to a narrow area of our data. The risk in this case for entrepreneurial researchers is to offer insights that do not fit into the specific situation an entrepreneur faces, which adds little to entrepreneurial science, thus the risk of offering general advice that does not fit a specific case.

Baxter *et al.* (2008) also examine the practice of setting specific criteria during sample (establishing inclusion and exclusion criteria for sample selection) in a quantitative study; such an approach is further confirmed by the findings of the present study, where high Mutual Information ratio scores may reveal relationships between variables in smaller subsets (or variable value ranges) of the total dataset.

Recent research in the fields of Computer science highlight the risk of Logical Fallacies, (Thorne *et al.*, 2018; Jin *et al.*, 2022) identify the risks of Logical Fallacies (and highlight the use of Computer science generated Reasoners); fallacies of relevance and generalization appear to be related to entrepreneurial research, when trying to identify relationships between variables from a total dataset, however these relationships maybe different within specific subsets of the dataset.

The research concludes that use of Mutual Information and Information Gain ratios offer additional insights for the relationship between the variables of entrepreneurial models. Furthermore, additional insights can be drawn from Network theory.

This sets a number of critical questions and side effects in entrepreneurial research such as:

- Entrepreneurial reality is too complex and relying on identification of linear relationships through correlation coefficient may be at a cost of a deeper understanding of the entrepreneurial reality.
- High M.I. and I.G. ratios when combined with the lack of Correlation Coefficient may suggest that (a) there are non-linear relationships between variables for the total dataset, but there may be selected range values of the variables where there are linear relationships; in this case moving from examining the total data available versus a smaller subset of data, which may fit better to the reality of specific entrepreneurs. (b) High I.G. values suggest that additional observation and data collection of one specific variable may offer us (in comparison) more information regarding the second variable, thus offering a guide for more focused data collection efforts.
- Findings below may suggest that more focused research in data clusters may reveal additional insights, especially regarding the very specific challenges faced by entrepreneurs in real life. Network Theory findings (Spyropoulos et al., 2021b) also points to the same direction.

6.1 Limitations and Future Research Recommendations

Even though the key findings of the research further support the similar conclusions from related studies, there is of course area for further research.

First of all, the existing dataset remains rather low (130 questionnaires, based on Spyropoulos 2019). So, one direction for further research is to work and analyze larger datasets, and ideally working with different sets variables as well. Working with larger datasets and more variables may also provide more opportunities for creating subsets of data or data cluster that can be further analyzed, in order to identify additional insights between variables which may have significant managerial and entrepreneurial implications and provide further insights.

Another insight is to include datasets from different sectors and economic areas, which may highlight critical insights or variables across cultures, economic zones and other environmental, political and cultural factors.

References:

Alhaj, T.A., Siraj, M.M., Zainal, A., Elshoush, H.T., Elhaj, F. 2016. Feature Selection Using Information Gain for Improved Structural-Based Alert Correlation. PLoS ONE 11(11), e0166017. doi:10.1371/journal.pone.0166017.

Amit, R., Zott, C. 2012. Creating Value Through Business Model Innovation, MIT Sloan Management Review, Spring. http://sloanreview.mit.edu/article/creating-value-through-business-model-innovation/.

Apuke, O.D. 2017. Quantitative Research Methods: A Synopsis Approach, Arabian Journal of Business and Management Review, 6(10). DOI: 10.12816/0040336.

Aulet, B. 2013. Disciplined Entrepreneurship, 24 Steps to a Successful Start Up. John Wiley & Sons, Inc.

Batina, L., Gierlichs, B., Prouff, E., Rivain, M., Standaert, F.X., Veyrat-Charvillon, N. 2010. Mutual Information Analysis: a Comprehensive Study. Journal of Cryptology.

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.

Boer, H., Howleg, M. Killduf, L., Pagel, M. Schmenner, R. Voss, C. 2015. Making a meaning contribution to theory. International Journal of Operations & Production Management. DOI: 10.1108/IJOPM-03-2015-0119.

Creswell, J. 2008. Educational research: Planning, conducting, and evaluating quantitative and qualitative research. New Jersey, Pearson, Merrill Prentice Hall.

Gao, K., Wang, Y., Ma, L. 2022. Belief Entropy Tree and Random Forest: Learning from Data with Continuous Attributes and Evidential Labels. Entropy, 24, 605. https://doi.org/10.3390/e24050605.

Jin, Z., Lalwani, A., Vaidhya, T., Shen, X., Ding, Y., Lyu, Z., Sachan, M., Mihalcea, R., Schölkopf, B. 2022. Logical Fallacy Detection, Computer Science. https://arxiv.org/abs/2202.13758,https://doi.org/10.48550/arXiv.2202.13758,

Isotalo, J. 2014. Basics of Statistics. Available at: http://www.mv.helsinki.fi/home/jmisotal/BoS.pdf.

Kent, J.T. 1983. Information gain and a general measure of correlation, Biometrika, Volume 70, Issue 1, 163-173. https://doi.org/10.1093/biomet/70.1.163.

Peirolo, R. 2011. Information gain as a score for probabilistic forecasts, Meteorological Applications, 18, 9-17. DOI: 10.1002/met.188.

Picardi, C.A., Masick, K.D. 2014. Research Methods – Designing and Conducting Research With a Real World Focus. SAGE Publications.

Roberts, E.B., Murray, F., Kim, J.D. 2015. Entrepreneurship and innovation at MIT. Continuing Global Growth and Impact, Martin Trust Center for MIT Entrepreneurship, MIT.

Shaltout, N.A., El-Hefnawi, M., Rafea, A., Moustafa, A. 2014. Information Gain as a feature selection method for the efficient classification of Influenza based on Viral Hosts. Proceedings of the World Congress on Engineering, Vol I, WCE, London, U.K.

Song, M., Podoynitsyna, K., Bij, H.V.D., Halman, J. I.M. 2008. Success Factors in New Ventures: A Meta Analysis. Journal of Product Innovation Management, 25, 7-27.

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, T.S. 2020a. Digital Greek start-ups – An analysis of founder’s perceptions. Advances in Management and Informatics, 5th Edition, 33-48. https://repository.afs.edu.gr/bitstream/6000/356/4/Spyropoulos_Digital-Greek-start-ups_2020.pdf#page=34.

Spyropoulos, T.S. 2020b. Knowledge management challenges for start-ups: A framework proposal. International Journal of Entrepreneurship and Business Development, 3(3).

Spyropoulos, T.S., Papageorgiou, M. 2021a. Sampling Issues and Eco-Networks on Innovation Management. Quantitative Research Studies Challenges, EBEEC 2021, The Economies of the Balcan and the Easter European Countries, Pafos, Cyprus May 14-16, 2021, Proceedings of the 13th International Conference on "Economies of the Balkan and Eastern European Countries".

Spyropoulos, T.S., Andras, C., Dimkou, A. 2021b. Application of Graph Theory on Entrepreneurship Research. EBEEC 2021, The Economies of the Balcan and the Easter European Countries, Pafos, Cyprus May 14-16, 2021, KnE Social Sciences Publications, The 13th International Research Conference Economies of the Balkan and Eastern European Countries (EBEEC) held online on the 14th-16th of May.

Spyropoulos, T.S., Andras, C., Dimkou, A., Polychronidou, P., (in press). Using Mutual Information and Information Gain Ratios on Entrepreneurial Research: An Empirical Case from Greece I.T. Start-Ups.

Thorne, J., Vlachos, A., Christodoulopoulos, C., Mittal, A. 2018. FEVER: a large-scale dataset for fact extraction and VERification. In Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers), 809-819. New Orleans, Louisiana, Association for Computational Linguistics.

Wilde, R.J., Leonard, P. 2018. Youth enterprise: the role of gender and life stage in motivations, aspirations and measures of success. Journal of Education and Work, 31(2), 144-158. DOI: 10.1080/13639080.2017.1421311.

Zafar, S., Khan, I.M. 2013. Examining Factors of Entrepreneurial Success: Culture, Gender, ucation, Family, Self-Perception. Journal of Poverty, Investment and Development - An Open Access International Journal, Vol. 2.