MASTER THESIS

INCUBATION PERFORMANCES AND IMPACT OF START-UP's RESOURCES

STUDY-CASE OF A RESEARCH-BASED INCUBATOR IN NICE.

HUMAN GEOGRAPHY MASTER PROGRAMME 2018/2019

Economic Geography – Regional Development and Policy

Romain MORIN - 6406564

Supervised by Dr Andrea Morrison – Supported by Mark Sanders for IRIS Smart-Cities

This project has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No 774199

Table of contents.

Table of figures.

List of abbreviations.

AgeMean: Mean age of the founding team members.

EdlvlMin: Education level of the least educated member of the founding team.

EExp: Mean entrepreneurial experience (in the founding teams).

FTE: Full-time-equivalent, refers to the national regulatory full-time job working hours.

IC: Information and communication.

ICT: Information and communication technologies.

IExp: Mean industrial experience (in the founding teams).

IncS: Incubation subsidies.

NA: Not applicable, or Not available. Used to mention missing cases or when results cannot be applied to a particular case.

NACE: Statistical Classification of economic Activities in the European Community, from the French « Nomenclature statistique des Activités économiques dans la Communauté Européenne ».

NoTrack: Start-ups that do not fit with any of the five IRIS transition tracks.

PrT: Product type.

RBV: Resource-Based View (Barney, 1991; Penrose, 1959).

ShareK: Share capital.

SIRET: Identification System of Establishment Repertory, from the French « Système d'identification du répertoire des établissements ».

Surv: Survival performance of the start-up standardised by cohort.

TSize: Mean size of the founding team.

Turn: Mean turnover in €/year.

UI: User-innovation.

WS: Women share (in founding teams).

Acknowledgements.

First, I would like to thank my academic supervisor, Dr Andrea Morrison of the Geosciences Faculty at Utrecht University. He helped to schedule my work and structure my steps to conduct rightfully my thesis.

This thesis would not have been possible without my internship supervisor, Dr Mark Sanders. He spent time with me to answer my questions and assist my research and writing. He regularly assessed my work and my progress with detailed and in-depth feedback. He put me back on the right track whenever it was necessary.

More generally, it was an honour for me to take part in IRIS Smart-Cities during these four months. This was an enriching experience for me to work in real research conditions within an engaging project and with a unique database. Particularly, I thank the team based in Nice supervised by Dr Jackie Krafft that hosted me a few days to share about the project. But also Chris Eveleens that guided me in the elaboration of this thesis to guarantee consistency with his PhD on Utrecht's incubators.

I must express all my gratitude to my parents that gave me their support and enabled me to study at Utrecht University in The Netherlands this year. None of what I achieved this year would have been possible without them. Thank you.

Finally, I would also like to give a special thank to Audrey Gombault, a friend that gave me strong support all along with this thesis and assisted me in my uncertain use of R in the beginning.

Romain Morin.

Introduction.

Start-ups have taken an increasingly influential role in public policies as a source of both entrepreneurial vitality and innovation stimulation (Lukeš, Longo, & Zouhar, 2018). Growth of such firms and the study of the very unequal results of support policies have received a lot of interest from academics and is the subject of a lot of empirical research as mentioned by Wright & Stigliani (2013). Since public subsidy policies and private investment has risen through support programmes like incubators, the interest in understanding the underlying process and determinants in the growth of those innovative businesses (i.e start-ups) has also risen (Mason & Brown, 2013; Weiblen & Chesbrough, 2015). Perspectives of enhancing outcomes of incubation, through both innovation and entrepreneurship policies, motivate extensive research to make these policies more effective. All stages of start-ups' development process are therefore scrutinized. This thesis puts the emphasis on capturing the characteristics shared by the most successful start-ups, given the screening policy used by the incubator. This research evaluates the performance of a French incubation programme based on the survival of incubated and non-incubated firms. Then it investigates what start-ups' inherent characteristics have a significant impact on their success, relying on data collected in a French incubator (PACA-Est, Sophia-Antipolis, Nice). This study aims at contributing to the existing literature on incubation assessment as well as the factor of success for start-ups as firms as early-stage firms. Furthermore, it takes part or a larger set of studies on incubators in Europe after research achieved in Utrecht and before an upcoming study in Gothenburg (Sweden). In the end, the contribution of this thesis will also contribute to policy advisory for innovation and entrepreneurial public strategies.

Until now, the literature has been extensive in the study of those innovative businesses. Eveleens, et al (2017) defined start-ups as new business ventures that develop scalable businesses at a small level, embedding themselves in the market they create around it. Research works about start-ups are broad but mainly focuses on the reasons of the inequal success of ventures as well as the performance of support they can benefit from, especially through incubation (Colombo & Delmastro, 2002; Hausberg & Korreck, 2018; Mason & Brown, 2013). Indeed, start-ups are very unequal towards success. Their fragility, due to the uncertainty of demand in the markets they address and their business capabilities, creates deep performance gaps (Levie, 2009; Wright & Stigliani, 2013). However, Mason and Brown (2013) observe that, regardless of their area of business, a high proportion of job creation concentrates within a very limited proportion of ventures. This raises interest in understanding the factors explaining success to improve the support addressed to startups. Such support is complex to define and researchers found that choices made by authorities directly affect the performance of the entrepreneurial ecosystem, employment and innovation (Lukeš et al., 2018).

Public strategies are particularly embodied by incubators, assisting and guiding start-ups in their growth in the early stages of their project. Following the metaphor of the greenhouse, start-ups incubators are directed by "gardeners" and provide "ventures fertilizers" like water and "pampering" to the plants (Phan, Siegel, & Wright, 2005). Phan et al. (2005) describe here an analogy for the support required by early ventures, like financial assistance, facilities and mentoring. The expansion of incubators relies on the assumption that start-ups need to make good decisions which require specific and adequate resources on their path for growth (Wright & Stigliani, 2013). Such resources may be tangible ones that would not be affordable for a single start-up: e.g. reliable facilities like an internet connection or 3D printers. They may also be less concrete, with knowledge sharing and networking that participate in direct and indirect learning processes (Hallen et al. 2016). Both sorts of resources are then supposed to give start-ups the tools to survive and grow faster than non-incubated ones.

Each venture is unique, but support cannot be fully tailored. Approaching this diversity is a challenge for incubators. Hackett & Dilts (2004) stated that incubators try to find the best fit between the incubatees (i.e. incubated firms) and provided services. The assessment of incubator's support and assistance to venture is very well documented but finds opposing findings (Eveleens, 2019). The literature provides a comprehensive study of the different types of support and level of individualization (Hausberg & Korreck, 2018; Scillitoe & Chakrabarti, 2010). At the same time, findings about the role of inherent characteristics of start-ups in the benefits they get from incubation and the success they observe later on, remains quite poor. Some elements like team members' profiles, the knowledge base of the venture or its industrial orientation may lead to different outcomes in terms of performance and survival (Wright & Stigliani, 2013). Inherent characteristics of start-ups belong to the identity of the start-up or its members and cannot be changed, only comprehended in order to adapt and improve support to them. Together with inherent characteristics of start-ups, other factors that may have an impact on their performance constitute "incubatee antecedents" (Hausberg & Korreck, 2018). Initial financial resources of the team before any subsidy, for instance, are part of them. These are referred to as "own resources" by Hausberg and Korreck (2018). Despite the rather rich literature about the outcomes of incubation programmes, the influence of incubatees antecedents towards their success remains understudied. Some studies have investigated separately influences of team members' backgrounds (Baum & Locke, 2004) and financial resources (Lee, Lee, & Pennings, 2001) but comprehensive studies that assess and compare the effects of start-up's antecedents are rare (Wright & Stigliani, 2013). Looking at the process of starting up a venture as the accumulation of firm-specific resources, is reminiscent of the resource-based view of the firm that relates the sustainable comparative advantage of firms to the inimitable resources that make it somehow unique (Penrose, 1959; Barney, 1991). Applied to the context of this study, above-normal returns of part of start-ups would be explained by their outstanding provisions of inimitable resources that differentiate them from their competitors. Human and knowledge capital constitutes part of such rare resources that drives success (Scillitoe & Chakrabarti, 2010). The incubation process is also a substantial resources provider. Through its programme, the incubator provides a large range of services and individualised assistance that might play a powerful differentiating role to incubated start-ups. On the other hand, financial resources represent an imitable factor that does not participate into the creation of a comparative advance for firms (Penrose, 1959; Barney, 1991; Scillitoe & Chakrabarti, 2010).

Consequently, this thesis aims first to evaluate the impact of incubation on our sample. As a major source of support and resources for start-ups, it is interesting to investigate the impact of this process on their success. From a public-policy perspective, incubators today take a major role in innovation and entrepreneurial strategies. The assessment of the outcomes of this strategy seems very necessary. This leads us to our first research question.

Does incubation have a positive impact on start-ups survival?

Subsequently, after evaluating the impact of incubation, we want to understand what else do impact ventures success. In order to capture what matters the most for successful start-ups, we then investigate the effects of inherent start-up characteristics. These characteristics or incubatees antecedents defines the essence of start-ups and are identified as a predictor of growth by the literature (Wright & Stigliani, 2013). They also convey inimitable resources that play a crucial role in the development of ventures. Furthermore, all variables covered by these incubatees antecedents are not equivalent. For instance, influential roles of human or knowledge capital encompass very different implications. Our research question follows these statements to study the performance of start-ups and investigate the impact of their inherent characteristics towards it.

To what extent do start-ups' resources have an impact on their survival and performances?

To answer this question, this thesis relies on extensive data provided by the incubator PACA-EST, based in Nice, as part of a collaboration with IRIS Smart-Cities which support this thesis. IRIS is a European funded project in the HORIZON 2020 programme. It relies on three *Lighthouse cities* in Europe, namely Utrecht (The Netherlands), Nice (France) and Gothenburg (Sweden) and four *Follower* cities. IRIS aims at monitoring, coordinating and scaling local solutions with the support of collaborators and partners in each city to enhance their smartness and sustainability in five *Transition tracks* (respectively *renewable and energy-positive districts*, *energy management and flexibility*, *mobility solutions*, *digital services and monitoring* and finally *citizen engagement*) (IRIS.ue).

As part of this project, IRIS investigates how smart and sustainable city-oriented start-ups may benefit from incubation. The underlying question of this research enquiring the need of specific care for start-ups based on their market orientation. A first study has already been achieved by Eveleens et al. (2019) over data shared by the Climate KIC and Utrecht.Inc partner incubators. Eveleens et al. (2019) concluded that the knowledge-base had no significant impact neither toward the survival nor the performance of start-ups studied in their 259 cases sample. However, the authors did find a significant and positive impact of incubation on survival, growth and investability. Given the focus of the programme, it is now interesting to study whether outcomes are different between cities and across different national or local contexts. The incubator for which I collected data, PACA-EST, is a research-oriented incubator which only focuses on start-ups supported by a research laboratory. This is not the case in Utrecht with the more diverse Utrecht.Inc and Climate KIC, specialized in sustainability-related start-ups. Such singularities support further studies over different incubators.

The present thesis mainly consists of replicating the analysis that was conducted by Eveleens et al. (2019) in Utrecht, adapting this work to the French dataset and the information available in it. For instance, if the French incubator provided us with a 308 cases sample of incubated and non-incubated firms, very little data was collected for non-incubated ones. In fact, we only have their survival up to 2019 and some information from the application. To assess the impact of incubation, however, we need non-incubated firms in the sample as well. Therefore, the thesis investigates the impact of incubation on the survival expectancy of start-ups by using survival analysis with a Cox proportional hazard model. This analysis is followed by the study of three different types of resources and their impact on incubated firm's success indicators with a multilinear regression model.

By answering these research questions, the thesis contributes to the existing literature about start-up's performances in incubators, reviewed by e.g. Hackett & Dilts (2004) and Hausberg & Korreck (2018). The extensive data collected over eighteen years enables this research to contribute to the assessment of incubation impact and the role of knowledge capital, human capital and financial resources in the success of start-ups. This contribution is moreover focused on specific dimensions such as user innovation and smart-city transition tracks. The scientific relevance of the contribution is largely tied to the results found in Utrecht and Gothenburg. In this regard, Utrecht findings are discussed in the margin of findings all the thesis long even if a separate in-depth comparison may be needed. Finally, this thesis contributes to the economic geography literature at a rather micro-level of study. In the meantime, incubators stand as a tool for policymakers, from local to national scales as part of their innovation strategies. Results found may be valuable in adapting incubators to more efficient and successful practices.

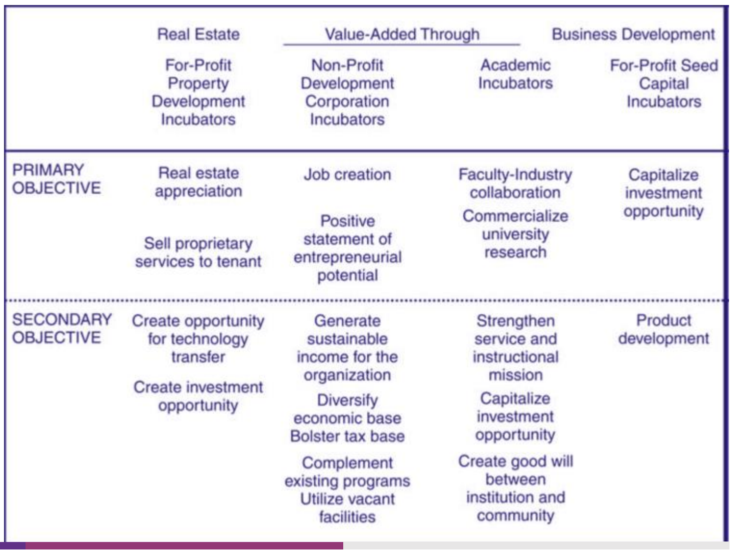
In this thesis, we will first explore the existing theoretical framework of incubation and start-up growth. Particularly, the study conducted by Hausberg & Korreck (2018) offers a comprehensive literature review about incubators. Through their research, they came up with a strong conceptual model that provide good tools for this thesis. Moreover, the research-based view provided a solid framework for the results' interpretation of our study. Second, we detailed the data collection process. This data section also contains a precise description of all variables used in the analysis. These descriptions are key to understand what information variables convey and then interpret the results. It ends with a presentation of the methods that explains and justifies the methodology followed for the analysis. Third, the findings are analysed in the results section. Then, the conclusion section gives insights into the literature contribution of this thesis as well as policy implications. Finally, it ends with the limitations of the study conducted.

Theoretical background.

This thesis takes part in two main theoretical frameworks. First, it is embedded into the incubation literature. The systematic reviews conducted by Hackett & Dilts (2004) and Hausberg & Korreck (2018) offers, for this, rich insights into the research conducted so far. Their findings are crucial for us to investigate the way incubation impact start-ups growth and performances, to then, be able to assess it. Second, entrepreneurship and business growth help us to approach the determinants of firms growth. Within this second theoretical framework, we found that this research particularly fits within the resource-based view developed by Penrose (1959) and Barney (1991).

Diversity of incubation forms and goals.

Incubation programmes are defined differently in the literature. The definition proposed by Phan et al (2005) describes "*property-based organisations with identifiable administrative centres focused on the mission of business acceleration through knowledge agglomeration and resource sharing*". Whether the emphasis is put on provided facilities, sought missions and goals or their institutional structure, incubators can find a broad range of definitions from the broadest to the narrowest (Hausberg & Korreck, 2018). All of them provide a similar set of services and support to start-ups even though final objectives are different (Hausberg & Korreck, 2018). Overall, three main incubator structures have been distinguished, based on their objectives as well as institutional concerns (Hackett & Dilts, 2004). First, "*non-profit development corporation incubators*" that are the most common ones. They focus on job creation and support entrepreneurial potential as primary objectives (Hackett & Dilts, 2004). However, they are also more and more challenged by the emergence of private independent or corporate incubators, that stands as the second type of structures (Hausberg & Korreck, 2018). For-profit incubators aim to capitalize on investment opportunities and usually focus on ICT or high-tech sectors (Allen et McCluskey, 1990; Hausberg & Korreck, 2018). Among these, corporate incubators managed by large companies, serve more individual and strategic objectives in sustaining the innovation of structures that may have lost agility



	Real Estate	Value-Added Through		Business Development
	For-Profit Property Development Incubators	Non-Profit Development Corporation Incubators	Academic Incubators	For-Profit Seed Capital Incubators
PRIMARY OBJECTIVE	Real estate appreciation Sell proprietary services to tenant	Job creation Positive statement of entrepreneurial potential	Faculty-Industry collaboration Commercialize university research	Capitalize investment opportunity
SECONDARY OBJECTIVE	Create opportunity for technology transfer Create investment opportunity	Generate sustainable income for the organization Diversify economic base Bolster tax base Complement existing programs Utilize vacant facilities	Strengthen service and instructional mission Capitalize investment opportunity Create good will between institution and community	Product development

Fig 1. Allen and McCluskey continuum.
Sources: Allen and McCluskey (1990); Hackett & Dilts (2004).

and freshness (Tushman et O'Reilly, 1996). Companies, such as Phillips benefit from encouraging spin-offs and emergence of new businesses as they bring in external ideas and innovation. Leaning on their "half outside-in and inside-out" position, such corporate incubators are designed to meet "*the mother's corporation strategic goals*" (Hausberg & Korreck, 2018; Weiblen & Chesbrough, 2015). On the contrary, private-independent incubators give greater importance to the profitability of their incubatees (Hausberg & Korreck, 2018). Third, academic incubators rely on faculty-industry relations to produce innovation and encourage the commercialization of university research (Allen, D. N., & McCluskey, 1991). Hackett & Dilts (2004) summed up this taxonomy built by Allen, D. N. and McCluskey (1991) in the table presented in figure 1. The academic model matches PACA-EST, the incubator studied in this thesis, the best. It has a strong partnership with the academic ecosystem it belongs to in Nice, Sophia Antipolis. Utrecht.Inc, one of the two incubators studied in Utrecht also fits in the definition of an academic incubator. In such incubators, mostly in line with national or local public strategies, the main success indicator would be job creation (Hausberg & Korreck, 2018). However, even within categories, there still be differences, and particularly in the acceptance, or screening, policies of applicants. Occurring upstream the incubation process, the screening is the first cause of bias in start-ups development between the accepted and the rejected ones, shaping from the beginning the outcomes of the incubator's intervention.

Diversity of screening policies.

To ensure the balance between costs of expenditures and the outcomes of incubation, selection policies towards new incubatees represents a key aspect in the strategy of incubators (Aerts et al., 2007; Patton et al., 2009). Investigating strategies conducted by incubators, Aertz et al. (2007) distinguished four categories of practices over a range of European business incubators. Financial screeners focus on financial resources of applicants while others put the emphasis on the personal characteristics of the founding team such as their entrepreneurial background. While the former strategy remains on the margin, the latter has been found to represent 27% of the studied sample (Aerts et al., 2007; Hausberg & Korreck, 2018). Market screeners are found to be the most common, 61% of incubators evaluate the matching chances between applicants' projects and market factors. Interestingly, although a balanced screening strategy appears to be the most efficient, it is very rare (Aerts et al., 2007). The selection process is the starting point and a key element of the strategy conducted by incubators and policymakers in the support of start-ups. The question of the accepted losses and the elitism degree of selection is core. On the one hand, "picking-the-winners" is seen as a way to select only ventures that are the most likely to succeed (Bergek & Norrman, 2008). On the other hand, the strategy "survival of the fittest" enables to accept a higher number of ventures, letting the market decide which ones fail or to survive. If the latter allows more ventures to get their chance by being incubated, this matches with a global "laissez-faire" strategy that provides lighter support, limiting interventions to basic facilities and on-demand assistance (Bergek & Norrman, 2008; Hausberg & Korreck, 2018). Furthermore, it is interesting here to note that Eveleens (2019) did not find any significant effect of pre-incubation quality of start-ups over their post-incubation growth, survival and investability. Such results give a strong argument against the selection strategy. No matter the strength of the selection process or strategic choices used, the primary objective is always growth and survival of start-ups. Regarding the main goal of public policies, namely job creation, the selection strategy is a crucial element since it's about how to spend public money. A small share of start-ups, called High-Growth Firms (HGF) are very likely to provide most of the job creation while a large majority of start-ups will never have any significant impact on the national or even the local scale (Shane, 2009). In his work, Shane (2009) therefore pleads in favour of a new entrepreneurship policy that would only fund such HGF that can be recognized by their high export-orientation and innovation level. Shane (2009) support her thought by explaining that HGF creates direct as well as indirect jobs by enrolling diverse linked industries in its production process. Although PACA-EST's screening policy also emphasis on most promising projects, it cannot be considered as HGF-focused. Indeed, requirements focus more on the maturity of the project and the strength of the founder's structuration. Second, the innovation potential of the firm is key, which is consistent with its academic orientation. Finally, the venture must be established in the same region than the incubator (incubateurpacaest.org, 2015). In the end, screening policies create biased samples of start-ups that benefit from incubation programmes that are meant to fit the best to them and lead to above-normal returns (Hausberg & Korreck, 2018).

From antecedents to outcomes.

Entrepreneurs' characteristics, the financial resources but also the venture strategy as well as the industry contexts are pinpointed as the best predictors for start-up growth (Wright & Stigliani, 2013). These indicators all qualify as antecedents in the incubation process according to the model proposed by Hausberg & Korreck (2018) replicated in figure 2. Together with the incubator and environmental antecedents, incubatee antecedents are used by incubators to maximize the match between the services and support provided and the needs of selected applicants (Hackett & Dilts, 2004). Incubator antecedents, as detailed in figure 2, correspond to past experience as well as the nature, the strategy

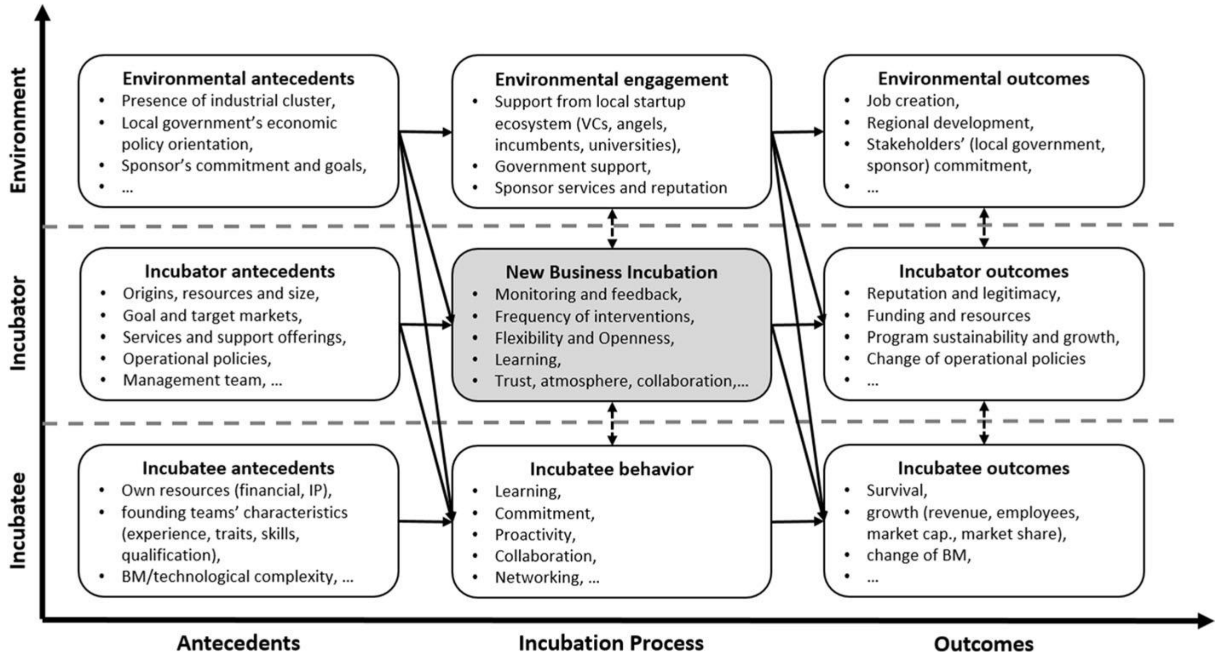


Fig 2. Multi-level antecedents and outcomes of business incubation.
Hausberg et Korreck, 2018.

and characteristics of the incubator while the environmental antecedents refer mostly to the local ecosystem it is embedded in, and national institutional context (Hausberg & Korreck, 2018). Maximisation of the fit between incubatee and incubator after selection is key to forge a strong and adapted relation for the managers with the entrepreneurs they monitor and has proven its importance towards success (Rice, 2002). Rice (2002) found out that the time spent, as well as the diversity of sharing canals used, are positively related to the impact of incubation on start-up's success and survival. In spite of such considerations, incubation is a process that implies self-investments of both parties, meaning that the incubatee behaviour moderates the impact of incubation too (Hausberg & Korreck, 2018). The latter also implies that the founding team of start-ups should also "be aware of their startup's knowledge gaps" in order for them to know well what they need and what part of those needs can be fulfilled by the incubation programmes and their managers. This identification of the start-ups' needs is most crucial when they cannot be fulfilled by the incubatee's team themselves. Therefore, the incubator can assist the firm's team to fix their issues by relying on the network developed by the incubator. This network is, furthermore, often constituted of the local environment ecosystem (Scillitoe & Chakrabarti, 2010). In the case of Sophia-Antipolis in Nice, this ecosystem is very developed.

The outcomes of incubation are also divided into three layers by Hausberg & Korreck (2018), with the incubatee, the incubator and the environment. Unsurprisingly, the incubation outcomes for incubatees are the most studied (Hausberg & Korreck 2018), in order to assess the performance of business incubators in terms of entrepreneurship strategy. Environmental outcomes attract major attention too. Indeed, the literature has a great interest in regional development outcomes such as innovation and job creation. These regional impacts allow us to assess and compare incubation programmes. Still, incubator's performance is complex to measure. Indeed, the different objectives targeted by the diverse forms of incubators create bias in any comparison of the same outcome variables (Bergek & Norrman, 2008). Therefore, it is difficult to generalise the evaluation of one incubator's outcomes to all incubators since they are different and may pursue different objectives. Moreover, the general outcomes of incubation require non-incubated startups in the sample to allow for comparison of programmes while the selection process creates a strong bias in favour of received applicants, supposed to be chosen for their stronger characteristics (Stokan et al., 2015). Hence, results may lead to "overestimation of incubator effectiveness" according to Stokan et al. (2015). Even though Hallen et al. (2016) and (Eveleens, 2019), with the cases of Climate KIC and Utrecht.Inc, found a positive impact incubation, findings remain uncertain. For instance, Colombo and Delmastro (2002) found no significant positive impact of incubation on start-ups. The specific research-orientation of the PACA-EST incubator, as well as the uniqueness of all screening policies, makes it interesting to investigates how incubated start-ups performed in the case of Nice.

The issue of strategic resources.

Accordingly to the conclusions of Wright and Stigliani (2013), to make good decisions, start-ups need good and adequate resources. Resources is a broad term that can refer to tangible resources like furniture, as well as intangible ones like assistance or network connections. The statement made by Wright and Stigliani (2013) relies on the resource-based view (RBV) primarily developed at the end of the 20th century by authors like Jay Barney (1991) but attributed to the earlier works of Edith Penrose (1959). This theory explains that the comparative advantage of firms that allows them to grow faster and perform better than their competitors can't be explained by the neoclassical microeconomics. Barney (1991) cites work by Porter (1990) who explains competitive advantage assuming that all firms have benefited from an equivalent provision of strategic resources and that all resource inequality would be quickly absorbed because of their high mobility. On the contrary, Barney (1991) and Penrose (1959) find that strategic resources are not perfectly mobile and give strong advantages to the firm that has access to them. Those resources are broad and can correspond to various provisions firms can benefit from. According to Barney (1991), firm's resources are "*assets, capabilities, organizational processes, firm attributes information, knowledge etc. controlled by a firm that enables the firm to conceive of and implement strategies that improve its efficiency and effectiveness*". Within the framework of incubation programmes, this theory gives key insights to understand the role they play in the success of start-ups. If Barney et al. (2001) mention that RBV theory is usually applied to large firms, its application to smaller ones like start-ups provides insights to investigate the role of innovation or business capabilities of entrepreneurs with a poorer background and how it affects their chances of success. Key resources and capabilities find their value in their rareness and inimitability (Barney, 1991). What makes the comparative advantage of a firm and enables above-normal returns is both the capabilities and resources that are somehow unique to the firm. Then, their quality, quantity and combination will affect the success of the firm in an inimitable way (Barney, 1991; Barney et al., 2001). Based on this, the support offered by incubators to their ventures plays an equalizing role. Indeed, if all start-ups already own their own combination of resources, they may lack some such as network, business management or marketing capabilities (Scillitoe & Chakrabarti, 2010). But incubation provides equal access to inimitable and rare resources for all incubatees, even though they may be personalised according to each start-up's needs, as well as imitable ones with offices, facilities and furniture. Then, in line with this, we can question what impacts the success of these start-ups that are meant to all obtain an equivalent amount of resources through the incubation. This leads us to formulate a first hypothesis, relying on the overall statement that incubation gives additional inimitable resources to incubated firms compared to others. Such resources are supposed to have a positive impact on survival of those firms (Barney et al., 2001; Penrose, 1959; Pitelis, 2004; Scillitoe & Chakrabarti, 2010) and this is something we want to investigate in Nice.

H1: Incubated firms are more successful than non-incubated ones.

The results that will confirm or not this hypothesis can give a strong argument to the screening policy and the role of the programme as an innovation driver. Incubation is expected to have a positive impact on start-ups' survival since the assistance and services and incubator provides gives start-ups access to resources they may miss otherwise (Barney et al., 2001).

As remaining inequalities, we can still find the initial strategic resources gathered by start-ups. These inequalities also very much overlap with the incubatees antecedents introduced by Hausberg and Korreck

(2018). Human capital represents the first inherent resources of a start-up. It comes with the team members that all benefits from a unique experience and unique capabilities that can differentiate a start-up to another. Furthermore, Harrison et al. (1991) found that complementarity of resources leads to higher performance thanks to the incorporation of new capabilities to the firm (Barney et al., 2001). This implies that the human capital brought by the team members will influence the success of the start-up according to the complementarity of the resources between them as well as with the incubation programme. It leads us to question the impact of the quality and the diversity of the human capital in the success of start-ups within the framework of an incubation programme with the following hypothesis.

H2: A greater and more diverse human capital combination enhances the start-up performance.

Meanwhile, the resources accumulated by a start-up can be observed in the knowledge-base they mobilise in their business activity. All sectors may not need the same amount of resources and some of them might be more difficult to conquer than others. Investigating such issues is very valuable for entrepreneurs, incubators and policymakers to support better some types of ventures and fit better with their needs. In line with this, we question the impact of knowledge capital on the success of a start-up, considering their specific required resources, and end up with the following hypothesis.

H3: The different knowledge capital of start-ups requires unequal amounts and types of inimitable resources, which creates a bias in their respective performances.

Finally, on the contrary to both the former and the latter, financial resources are not inimitable resources, even if they are somehow scarce. However, financial matters are often thought as the be-all and end-all of start-ups support since each new business needs initial investments and money can buy the other resources, at least to the extent that they are imitable. It is then interesting to question its relevance to predict start-up success and compare its significance with those of human and knowledge capital. We thus hypothesize the following.

H4: The level of financial resources of incubatees is positively related to their outcomes.

Based on the literature, these four hypotheses cover the three main types of incubatee resources likely to impact the success of start-ups. It is expected to find some validation or invalidation of our hypothesis to contribute to the literature on the impact of incubatee antecedents and the firm's resources in a start-up's success.

Research design, data and methods.

Research design.

In order to understand the success of incubation for start-ups, this thesis relies heavily on the database provided by the partner incubator PACA-EST in Nice. A cross-sectional research design is adopted here although the data has been collected over 18 years in total, from 2001 to 2019. An observation corresponds to a firm that applied to the incubator since 2001. All firms come with detailed monitoring from the year of application until now (2019). Still, the longitudinal dimension of the database has not been used as much as it could for a range of reasons. First, the available data is very uneven over the years, with many missing cases for most variables, especially before 2010 (not systematic collection of information) and after 2017 (datafiles often not updated yet).

Moreover, the analysis conducted in this research is meant to be as comparable as possible with the (cross-sectional) analysis that was conducted in Utrecht already (Eveleens, 2019). Therefore, Eveleens et al (2019) provide the guideline in the organization of the empirical methods used in this thesis. By all means, since databases differ, variables have been adapted as detailed later on in the variable descriptions. For instance, one main difference is a major lack of data for non-incubated firms. We have survival, but the lack of other variables implies that we are limited in assessing the effect of incubation on the performance of start-ups. The data analysis has been conducted based on the R script already used by Eveleens et al. (2019). In that sense, the thesis is a replication study with new data. This script is made available in the appendix.

In this section, we first discuss the measures of start-up performance. Survival, size, growth and turnover are our four dependent variables. Then I discuss the control and explanatory variables.

Data collection and sample.

The analysis conducted in this thesis relies mainly on incubation data. This data has been provided by the start-up incubator PACA-EST, as a collaborator of the IRIS Smart-Cities project in the “lighthouse-city” of Nice. This partnership allowed us to work on an extensive dataset of three hundred and eight start-ups that applied to enter the incubation programme between 2001 and early 2019. The dataset is distributed between five stages of application. First, rejected or abandoned applications, that represents 92 out of the 308 start-ups. Those start-ups have been either rejected within the application process or abandoned before entering the programme. Second, the application waiting for acceptance or rejection (30/308). Third, start-ups that entered a pre-incubation phase (2/308). Fourth, start-ups that entered the incubation programme and that have not exited already (20/308). And last, the ones that entered and have exited the programme (174/308). Each of these statuses refers to a different stage of the start-up in the incubator and also implies a different amount of information. Hence the data available on these start-ups was very uneven. The most extensive data is provided for the exited start-ups that lived a few years in the programme and even provided additional data after their exit. On the contrary, waiting and rejected/abandoned cases offer the least detail about their team, project, and success.

The dataset is divided into two types of information. On the one hand, the monitoring data corresponds to individual information from each of the application cases with a dedicated folder per start-up. Application procedures as well as team members' information, business plans, expenditure and financial data and incubation forms are contained in them. On the other hand, reporting data separately contained archives and track records about all start-ups in lists and recap tables. These lists and tables are mostly about financial data. In case of missing data for example for share capital or NACE codes, the certified website Societe.com was used to complement the data.

Data extraction was made consistent with the work conducted by Eveleens et al. (2019) on Utrecht.Inc and Climate KIC. Even though both incubators are similar, the data provided by the incubators are significantly different between the two cities. The Nice dataset is richer, but missing cases are more common, especially for time-sensitive variables. Unfortunately, this tends to weaken our analysis. Also, within the framework of his Ph.D., Eveleens allocated significant time to additional systematic data collection from diverse sources that could not be repeated for the Nice sample. Finally, it must be mentioned that most of the start-ups are not firms yet when they enter the incubator in Nice while they mostly are in Utrecht incubators. PACA-EST required applicants to not to be founded yet, with some exceptions in some cases for very young ones. This implies a more mature stage of the projects in Utrecht.Inc and Climate KIC, which may strongly influence survival and our performance measures.

Together with a team based in Nice, we entered the data manually into an Excel sheet. Columns were dedicated to variables that have been selected with lines corresponding to a specific year. Indeed, instead of giving each start-up one line, the choice was made to roll out the lifetime of each start-up from its application year until 2019. When start-ups died, they were still rolled out until 2019, with mention of their failure after it occurred. This enables a more detailed analysis of the evolution of the start-up through time in later work. For this thesis, the data were collapsed into a cross-section in which all ventures represent a single observation, except for the Cox model which needs longitudinal data. This guarantees the least amount of information loss based on the structure of the dataset.

Variables.

Dependent variables.

We are here interested in the performances of the start-ups, that is to say, how they survive, and how they perform. This can be approached by a limited range of variables.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Survival	**Surv**	0	79,1	32,9	16,1	31,2	20,3	31,2	Standardisation of the lifespan by the mean survival expectancy and its standard error, according to the application year.
	Survival (2019)	0,0	1,0	0,8	0,4	1	1	1	Status of the firm in March 2019. Alive: 1, 0 otherwise.
	Lifespan	0,0	18,0	5,3	5,1	4,0	0,0	9,0	Number of active years recorded (from foundation to liquidation/2019).

Fig 3. Descriptive statistics of survival and related variables.
Sources: PACA-EST incubator, Societe.com. 2019.

Firstly, **survival**, which refers to the existence of the start-up up to a given point in time. We define this variable as a dummy variable signifying survival up to the spring of 2019 when our data was collected. Survival has been extracted from the dataset accordingly to the year they were introduced as dead. This is the only dependent variable which covers extensively both incubated and non-incubated firms of the dataset. During the data entry, each start-up has been represented with one line per year since their application year. All lines have been filed by updated data and once a start-up was found dead, its status changed (from Abandoned/Waiting/Pre-incubated/Incubated/Exited) to “Disbanded”. Based on this, we extracted the survival by attributing 0 instead of “Disbanded” and 1 otherwise. Then, grouping the data per ID code, we obtained a table with the status of each start-up for from 2001 to 2019, filling years that precede the venture creation with NAs. Ventures that got acquired by another firm are considered alive. In table 2, the mean value of this variable is 0,8, meaning that around 80% of start-ups in our sample were still alive in 2019. Additionally, because start-ups that applied in 2017 are more likely to be alive in 2019 and observe a longer lifespan, we wanted to distinguish the survival probabilities of firms based on the cohort they belong. To know that, we conducted a survival analysis based on the Cox proportional Hazard model with R and the “coxph”-function. Originally designed for medical purposes, the Cox Hazard model estimates the probability of dying at t time of individuals and allows one to estimate the effect of external parameters in this probability (Cox, 1972). In order to capture the survival probability of a firm given its age and its cohort, we first ran the Cox model, using measures of time, status and a covariate. The time has been approached with the difference between the date associated with the survival observation and the application year, survival is a 0-1 dummy variable and the covariate is the application year (called “App_Year” in the results). We found that the year of the application had a significant and positive impact on the survival chances of firms. This means that firms that applied in the most recent years are found to have a positive impact on the failure probability of the covariate while a be the less likely to last longer than the one that applied earlier in our sample. Indeed, in the Cox model, a negative regression coefficient reflects a positive impact on survival probability. Hazard’s ratio is significant under the threshold of 0.1% and shows a 1.23 coefficient according to the entry year. It means that on average, in our sample, the survival probability drops by 23% between each cohort. Then, we used the *survfit* function to obtain the survival probabilities of every cohort and plotted the results. Looking at the plots of the regressions, it is clear that not all cohorts have the same survival probabilities (fig 4). With no surprise, the youngest cohorts are the ones that are most likely to fail. Years from 2001 to 2004 being underrepresented in the sample, their results depend on limited observations and seem to overperform clearly compared to the others with half chances to survive longer than 10 years. Likewise, there is only one observation that applied in 2019 and apparently failed. Moreover, it seems that the economic crisis of 2007/2008 had no important impact on start-ups development and survival. On average, start-ups get half survival probability (0,5) after five years of life. Exploiting these results, also presented in the table below (fig3), we standardized the survival variable by measuring the effective lifespan of all start-ups (Survival_sum) and standardised it with the mean and standard error corresponding to its entry year, following this standardisation formula:

$$x = \frac{X - \mu}{\sigma}$$

Even though values are no longer meaningful (fig2), it enables us to get a survival variable that now measures the performance of the start-up compared to the mean value of their cohort.

	Records	n.max	n.start	Events	rmean	se(rmean)	Median	0.95LCL	0.95UCL
2001	99	99	99	42	7,762	0,243	16	13	NA
2002	120	120	120	104	6,833	0,260	9	8	11
2003	196	196	196	85	6,999	0,215	15	13	NA
2004	152	152	152	104	6,910	0,238	10	8	12
2005	129	129	129	122	6,233	0,269	7	6	9
2006	110	110	110	100	5,891	0,295	7	5	8
2007	113	113	113	99	5,713	0,302	6	5	8
2008	107	107	107	95	5,412	0,305	6	5	7
2009	176	176	176	164	5,002	0,233	5	4	6
2010	170	170	170	161	4,581	0,225	5	4	5
2011	81	81	81	80	4,012	0,289	4	3	5
2012	80	80	80	80	3,500	0,256	3,5	3	4
2013	136	136	136	130	3,162	0,183	3	3	4
2014	57	57	57	51	2,855	0,295	3	2	4
2015	74	74	74	66	2,420	0,258	2	2	3
2016	61	61	61	61	1,574	0,140	2	1	2
2017	29	29	29	29	1,138	0,144	1	1	NA
2018	18	18	18	18	0,500	0,118	0,5	0	NA
2019	1	1	1	1	0	0	0	NA	NA

Fig 4. Table visualization of survival probabilities according to application year.
Sources: PACA-EST incubator, Societe.com. 2019.

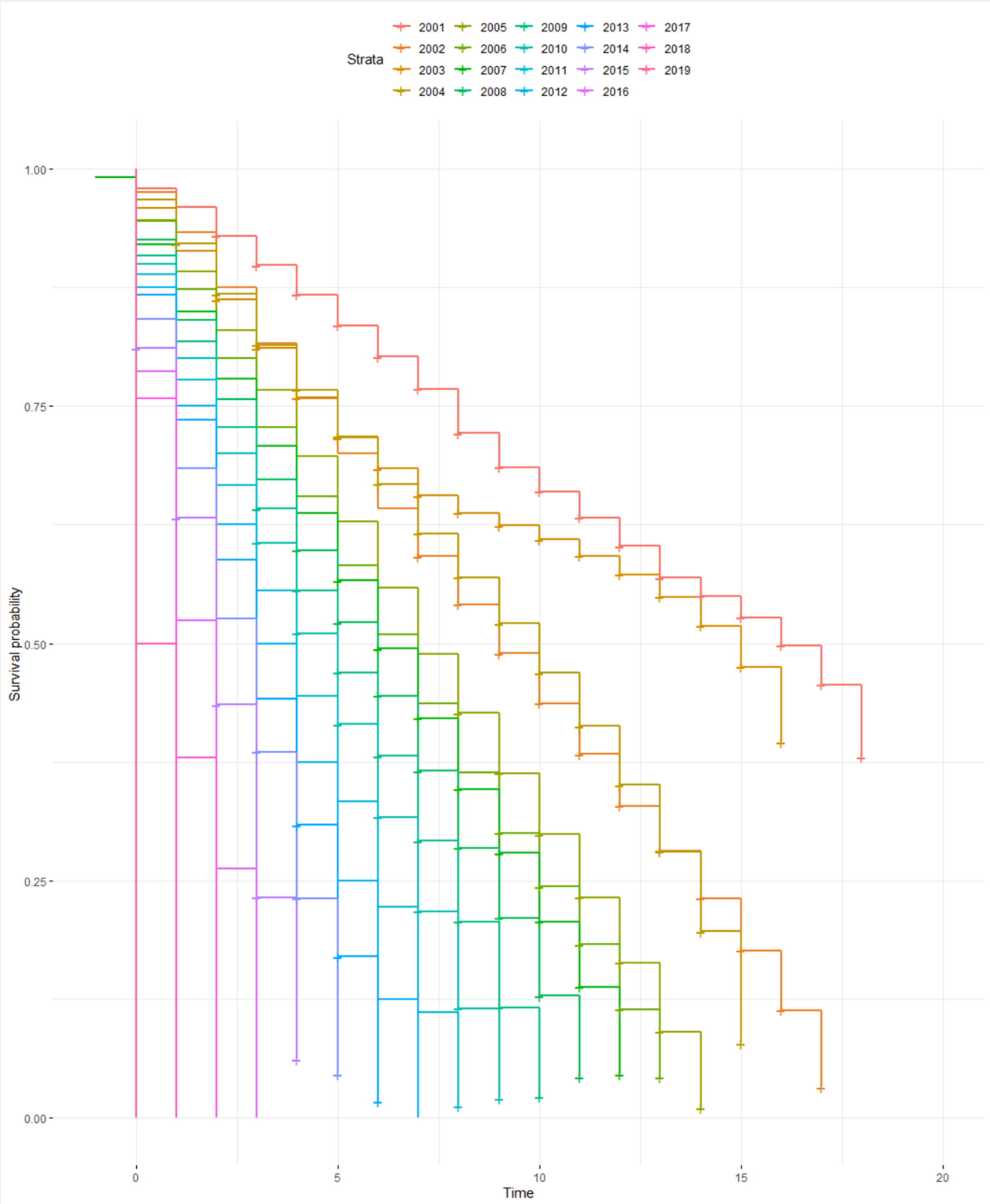


Fig 5. Plot visualization of survival probabilities, from year 1 to year 18, according to application year.
Sources: PACA-EST incubator, Societe.com. 2019.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Employment size	**FTE_mean**	0,8	50,0	6,4	7,1	4,0	2,0	8,2	Average number of people employed annually on a full-time job basis.
	FTE_max	1,0	96,0	9,7	13,2	6,0	2,0	12,3	Maximum number of people employed annually on a full-time job basis.
	FTE_sum	1,0	400	42,3	61,3	19,5	6,6	19,5	Total of people employed annualy on a full-time-job basis.

Fig 6a. Descriptive statistics of employment size and related variables.
Sources: PACA-EST incubator, Societe.com. 2019.

Second, the **employment size** represents a major proxy for the start-up performance since it translates the economic vitality of the firm as well as its commercial expansion. It does not include the founding team and refers to the number of full-time equivalents and not to the absolute number of people employed in the start-up. This variable differs from the definition in the Utrecht database used in Eveleens et al. (2019) which used the number of people employed as the full-time equivalent was not available. Unfortunately, the number of employees has not been recorded in our dataset which means that both studies will differ in this respect. In France, the full-time equivalent is 35 hours per week. This means that two employees working half-time are counted as one single FTE while they may have been counted as two employees in the Utrecht analysis. Furthermore, we faced difficulties to get a consistent measure of employment size since data was not available in the same years for all start-ups. When data is available for a start-up, observation through time rarely starts on their application year and always end before 2019 even when they are still alive because most records were not updated after 2018 yet. Moreover, within this period where employment is recorded, some annual values are missing. As a result, capturing comparable employment data for all start-ups is difficult. The mean value of employment is not the most recent value nor corresponds to a same t moment in the lifetime of ventures, but it can still be analysed as a performance measure as corresponding to its average development level independently to peak or extreme values. We can already observe that, on average, the start-up employment level is little above six employees (6,4), with a few important outliers with very high values (maximum average per year of 50 employees) and half of the start-ups not passing the threshold of four FTE on average per year through their life. Note that the minimum value of employment is never equal to zero for maximum and mean values of employment measurements. This is because it appears that the recording of employment only starts with the first employee of firms. Firms that score zero in employment thus show a missing value in the sample, whereas not all missing values mean zero employees. Then, this result is a little biased by the fact that missing values are numerous and not accounted for. As a complementary measure of employment, we use the sum of employment in each start-up through their life (FTE (sum)) in the table, we note that extremely successful start-ups cumulated up to the equivalent of 400 FTE years since their creation. On the other hand, half of the start-ups that did create jobs cumulated less than 20 FTE years during their life. However, we found that employment has not been recorded for more than half of the start-ups in the sample, either because they did not create any, or because the information is missing. This is the case for all non-incubated firms, as well as for a limited share of incubated ones.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Growth	Growth	0,0	12,0	1,7	2,0	1,0	0,4	2,5	Average annual job creation of the firm.

Fig 6b. Descriptive statistics of growth.
Sources: PACA-EST incubator, Societe.com. 2019.

Third, **growth** measures the evolution of employment in each start-up during its life. Once again, because the data is uneven among the years and the start-ups, the variable has been adapted from the work conducted in Utrecht. Instead of measuring growth between two points in time – between application and the collection of the data – we take the average of a firm's annual job creation over its lifespan. More precisely, each additional employee hired is recorded annually, summed with others and then corrected by the lifespan of the start-up. Therefore, our growth variable cannot be negative since it is not possible to destroy more jobs than have been created before. This way, the variables gains strength with information taking all the lifetime records into account and is less sensitive to the age of start-ups. It is also easier to interpret in terms of participation of each start-up to the local employment than a regular growth rate. It is, however, important to mention that the death of the start-up is not taken into account. Hence the employment growth of the start-up does not reflect its status but only its contribution to employment during its life whether it lasted long or not. Similar to the employment size, growth descriptive statistics show us a mean value (+1,7 job created per start-up in annual average) strongly influenced by very high value (+12 jobs created per start-up in annual average) while half of the sample only performs one job creation per year in average. Note that 25% of the sample creates between 2 and 2,5 job per year on average. On the contrary, the top 25% is very spread, performing between 2,5 and 12 job creation per year on average. Finally, we observe that the lower 25% creates on average 0,4 job per year or less. However, this share may be underestimated here since the missing cases represent 172/308 cases with a high concentration of abandoned as well as failed start-ups.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Turnover	**T_mean**	1000	5 903 636,4	374 260,7	659 512,2	152 348,2	51 875	397 734,4	Average of annual turnovers recorded by the firm (in €).
	T_max	1000	16000 000	884 424,7	2 022 015,2	300 000	78 750	792 500	Maximum annual turnover recorded by the firm (in €).

Fig 6c. Descriptive statistics of turnover and related variables.
Sources: PACA-EST incubator, Societe.com. 2019.

Last, the **turnover** is used for incubated start-ups to approach their vitality and performance during and after the programme. Instead of the investments data used in Utrecht that are not available reliably in this case, we use turnover to introduce another financial performance indicator. The means of all records are used for this. The values of the sample are very uneven, from an annual average of 1000€ up to 5,9M€. In the same way than for employment, turnover's mean is highly influenced by extreme positive values and is twice higher than the median (374 260€/year average against 152 348€/year average). Note that also turnover cannot be negative. Therefore, there is no value below 0 with this variable. Consequently, the distribution is roughly normally distributed but observes a left asymmetry. Seventy-five per cent of the monitored start-ups have an average annual turnover greater than 50 000€ and half of them recorded a turnover higher or equal to 300 000€ at least once. The standard deviation (660 000) is also very high which reflects a high dispersion of the distribution.

Independent variables.

Each of these four dependent variables is designed to approach and assess the performance of the start-ups monitored in our sample. Used through survival analysis and multilinear regressions, we can test their predictability with independent variables. Organised in four categories that fit the hypothesis formulated before.

Incubation.

To assess the performance of the incubator, we want to estimate the impact of the incubation onto the start-ups of our sample. This performance is approached here with the survival dependent variable only. Indeed, this is the only dependent variable whose data is available for both incubated and non-incubated ventures.

The stage variable allows us to evaluate the performance of the incubator. The stage variable is a categorical variable which attributes a number to each stage of incubation start-ups can be embedded in. Start-ups that belongs to stages 2 to 4 are younger than the others on average. Indeed, since that PACA-EST require venture not to be founded on the moment of application, the start-ups that are still in application, pre-incubation or incubation phases are very likely to be between year 0 and year 3 of their life.

Group	Code	Stage	Observations (/308)	Descriptions
Incubation	1	Abandoned/ Rejected	92	Cases where start-up abandoned their application demand to the incubator or that have been rejected.
	2	Waiting	30	Start-ups waiting to know the results of admissibility to the incubator.
	3	Pre-incubated	2	Accepted start-ups that have been accepted but did not enter the incubator yet.
	4	Incubated	20	Start-ups still incubated on March 2019.
	5	Exited	174	Start-ups that entered and left after the end of the incubation phase.

Fig 6d. Descriptive table of the stage variable.
Sources: PACA-EST incubator.

The stage allows us to compare the comparative survival of each of these categories. It is clear that category 1 represents non-incubated firms while 5 reflects incubated ones. For the three other intermediary variables, it is less clear. The waiting list is not really analyzable, they are still in-between and surely very young. About 3 and 4, they remain apart from 5 since they did not fully benefit from incubation yet. Therefore, the focus will be put towards rejected and exited ventures, which also represent the larger share of our sample (266/308).

Human capital.

The Human capital section focuses on information about the resources gathered by members of the founding teams into the start-ups. Human capital is of great interest to explore the implications of the member's characteristics in the success of a venture. Indeed, their uniqueness represents inimitable resources for the firms and this section aims at studying what type of resource is meaningful in firms' success. Data about members contained in this section have been detailed for each of the team members, usually two or three but sometimes up to eight, nine or even ten. To make use of it, the data has been aggregated from one observation per member to a single observation per start-up.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	TSize	1	10	2,7	1,5	3	2	3	Size of the founding team

Fig 6e. Descriptive statistics of the founding teams' size.
Sources: PACA-EST incubator.

The **number of members** in the founding team is used to see whether a limited or a large team is suitable in start-ups project. A higher number of people involved can imply a more diverse or a richer knowledge base which can be beneficial. From the RBV, a larger team implies a greater chance to benefit from inimitable and relevant resources that would be fruitful. However, a big team is also more complex to manage and can also lead to more disagreement. Both the average and a median number of member scores around 3 as standard team size. Half of the sample observes a team between 2 and 3 members.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	WS	0	1	0,13	0,25	0	0	0,25	Share of women in the founding team

Fig 6f. Descriptive statistics of women share in founding teams.
Sources: PACA-EST incubator.

The **gender** dummy of founding members has been aggregated through a share of women within the team, thus, from 0 for 0% to 1 for 100% of women in the team. The variable may be quite trivial and has been found not significant to explain the ex-post success of start-ups already (Banal-Estanol, Macho-Stadler, & Perez-Castrillo, 2016). Nevertheless, considering the very low rate of women (10% on average) in the sample, it may be interesting to see whether the impact on survival or performance is significant or not. We can expect that resources brought by men and women being not equal, more balanced teams could benefit from this diversity. Such a statement is strengthened by the context of the innovation sector that remains male-dominated, and more specifically ICTs, which are overrepresented in PACA-EST. More than half of the teams record no woman, and only 25% of them have a share of females higher than 30%.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	AgeMean	25,5	88	49	9,2	48,5	43	55,6	Average age of the founding team

Fig 6g. Descriptive statistics of the mean age of founding teams.
Sources: PACA-EST incubator.

The **mean age** of the founding team constitutes an interesting variable to approximate the human capital and resources of start-ups. We can question whether a rather senior team is the most suitable to succeed thanks to a greater experience and, to a certain extent, to more knowledge. On the other hand, the younger generation may be more agile or more in line with market expectancy, especially concerning communication technologies. As a control variable for start-ups performance, we use the mean of the team members ages. Mean and median values are similar which traduce a somehow normal distribution around the mean. If maximum (88) and minimum (25,5) values almost meet the limits allowed by the current societal system, we note that half of the start-ups are founded by teams between 43- and 56-years old average.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	EdLvlMax	1	2	1,971	0,17	2	2	2	Maximum education level observed in the founding team
	EdLvlMin	1	2	1,95	0,22	2	2	2	Minimum education level observed in the founding team

Fig 6h. Descriptive statistics of the education level of founding teams.
Sources: PACA-EST incubator.

Then, the **level of education** of the founding team members constitutes a key proxy for the global amount of knowledge, therefore resources gathered in the different projects. It is approached in this research with the minimum and the maximum level of education represented in each team. Therefore, it is possible to know first whether the education level of the team moderates its success, but also, within the team, whether the education level of the most or the least educated member is the most correlated to success. The level of education relies on the highest education level that has been completed according to the French national classification for professional certifications (RNCP) and which is in line with the European system. It goes from level 3, corresponding to the vocational education, to level 8 for PhDs. However, we chose to simplify this classification by giving a score from 1 to 3. Level 1 of education covers levels 3 and 4 of the classification, that is to say, vocational education (3) and baccalaureate level (4). Level 2 of education now refers to higher technician diplomas (5) and bachelor (6) levels. Finally, level 3 of education corresponds to higher academic levels with the master (7) and PhD (8) graduations. Such a classification allows us to capture better the impact on each different types of resources on start-ups performance. If we were interested in both maximum and minimum variables, it appears that the very large majority of teams are founded with level 3 members and almost all teams have a level 3 member. Considering the weak explanatory strength of this variable, this research will focus on the education level of the least educated member. Even though, we observe on the descriptive table (fig 5g) that even the least educated member of teams is, for more than 75% of start-ups, master level and higher. These observations are in line with the research-purpose requirements of the incubator.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	EExp_sum	0	35	3,6	5,9	1	0	4	Sum of the entrepreneurial experience observed in the founding team in years
	EExp_mean	0	13	2,4	2,9	1	1	3	Mean of the entrepreneurial experience observed in the founding team in years
	EExp_max	0	28	3,8	5,1	1	1	5	Maximum entrepreneurial experience observed in the founding team in years
	EExp_min	0	13	1,4	2,2	1	0	1	Minimum entrepreneurial experience observed in the founding team in years

Fig 6i. Descriptive statistics of the entrepreneurial experience of founding teams.
Sources: PACA-EST incubator.

The **entrepreneurial experience** of members reflects the number of years they spent running companies or leading independent projects before their application to the incubator. It is important that only years spent as a founder or co-founder of a project matter here. It is assumed that start-ups lead by people that are experienced in leading positions have better chances of success since it is very likely that they can benefit from similar issues they already dealt with in the past. Experience gives team members a unique background which is valuable for the strategic resources that come with it. The mean value of all members' entrepreneurial experience is used here since its sum would have suffered too much bias from the size of the team. In line with the academic orientation of the incubator, the entrepreneurial experience is rather low, with more than half of the teams that accumulate less than one year of experience and 75% of the team member's with less than 3 years of experience. The most experienced team members have no more than 1 year of experience in half of the cases and more than 5 years in only 25% of the teams. The institutional context matters too. Indeed, the incubator has been created as a response to the call for a project launched by the government in 1999 within the framework of the "Research-innovation" law that aimed at enabling and encouraging academics to found a start-up and register patents. Consequently, teams are mostly constituted of academics.

Group	Variables	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Human Capital	IExp_sum	0	134	18,5	23,2	11	0	28	Sum of the industrial experience observed in the founding team in years
	IExp_mean	0	43	11,5	7,9	10,8	5,5	17	Mean of the industrial experience observed in the founding team in years
	IExp_max	0	43	15,4	9,9	15	7,5	22	Maximum industrial experience observed in the founding team in years
	IExp_min	0	43	8,3	7,9	6	1	14	Minimum industrial experience observed in the founding team in years

Fig 6j. Descriptive statistics of the industrial experience of founding teams.
Sources: PACA-EST incubator.

Finally, **linkages with industry**, as for the entrepreneurial experience, is approached by a number of years spent in the industry – as representative of the private sector –. The industry linkages are useful to measure the role of the experience from the private sector as a resource in terms of network and knowledge, which is very different from the management and market resources brought through the entrepreneurial experience. From the resource-based view perspective, and similarly to the entrepreneurial experience, we expect the industrial experience to be positively related to the success of the ventures. The mean value of the member's industry linkages is used for the analysis for the same reasons than discussed in the case of entrepreneurial experience. With no surprise, on average the industrial experience of founders is much higher than their entrepreneurial experience (very close to the median) of 11 years of background in the private sector. The first quarter of team members are what can be called juniors, with no more than 5,5 years while a large third quarter of them are seniors with more than 17 years of industrial linkages on average.

Knowledge capital.

The knowledge capital section constitutes the core element of this thesis in the understanding of the start-up success in incubators. Patent data, commonly used to approach the knowledge-base in economic geography theories like product and knowledge space, was unfortunately not available for this research (Hidalgo & Haussmann, 2013). Instead, we used a range of other variables that are tightly linked with the knowledge start-ups can benefit from. By using such variables, we can estimate the role of the knowledge-base resources in start-ups performance. The results found out of it would constitute tracks for policy advice.

Group	Code	Observations (/308)	Descriptions
Knowledge Capital	A	2	Agriculture, Forestry and Fishing
	C	23	Manufacturing
	G	7	Wholesale & Retail Trade; Repair of Motor Vehicles
	H	1	Transportation and Storage
	J	91	Information and Communication
	K	3	Financial and Insurance Activities
	M	72	Professional, Scientific and Technical Activities
	N	4	Administrative and Support Service Activities
	P	1	Education
	S	1	Other Service Activities
	NA	103	

Fig 6k. Descriptive statistics of start-ups' industrial sectors.
Sources: PACA-EST incubator, Societe.com. 2019.

First, for all the start-ups that led to an official and registered firm creation, which means having a SIRET code (Identification System of Establishment Repertory), we collected their NAF code from the Societe.com certified website. The NAF code is composed of 732 items and standardized with the international NACE classification and its 615 items. This gives us a very precise proxy of the activity of the start-up as well as the required knowledge for it. However, 732 entries variable is too much in an analysis with only a few hundred firms from a single incubator. Hence, classifications have been converted from level 4 of the classification to level 1, with only 21 entries. In our sample, only 10 out of the 21 entries have been observed. The **industrial sector code** is the first key variable to use in this study and is sharp enough to define a quite precise knowledge base that would be required for each start-up project. We observe on descriptives that both ICT and scientific activities are overrepresented in this sample. The latter covers engineering and medical activities oriented to research and development purpose, consistently with the orientation of the incubator.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Knowledge Capital	PrT	0	1	0,4	0,5	0	0	1	Type of product produced. Hardware = 1, other (software/service) = 0

Fig 6l. Descriptive statistics about the product type of projects.
Sources: PACA-EST incubator.

Secondly, the **product type** is entered as a binary variable and enables to distinguish hardware products (1) from software or service products (0). It is an important variable to test because the process of producing software products or services and a hardware product is absolutely different in the issues that a start-up has to face in its development phase, especially with the needs of manufacturing investments in the case of hardware products (Eveleens, 2019). Moreover, the knowledge type required is different too in terms of management and organization skills. In our sample, about 40% of the start-up conducts non-hardware projects.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Knowledge Capital	NoTrack	0	1	0,87	0,34	1	1	1	Evaluation of the project fit to IRIS tracks. If fits any of the five = 0, otherwise = 1

Fig 6m. Descriptive statistics of the fit of start-ups projects with IRIS tracks.
Sources: PACA-EST incubator, IRIS.ue, 2019.

Lastly, each start-up project has been assessed in comparison to the five tracks for a smarter and more sustainable city, designed within the IRIS project (IRIS.ue). The five tracks are defined as follows: 1- "Renewable and energy-positive districts", 2- "Flexible energy management and storage", 3- "Intelligent mobility solutions, 4- "Digital transformation and services", and 5- "Citizen engagement and co-creation". Those tracks are designed to represent categories of issues to work on in the creation of a smarter and more sustainable city, and we gave each track a binary grade of 1 or 0 according to their fit to one of these tracks. We assessed of the fit of start-ups' projects manually. For each observation, we read descriptions of each project, available in the dataset – then entered in the database –, and evaluated its fit to the five tracks designed by IRIS. We finally attributed a 1-0 separately for each of the tracks. Additionally, a 1-0 grade was given to the **NoTrack** variable with a 1 when the start-up didn't fit in any of the track and 0 if it does. We observe that the share of start-ups that fits at least one of the tracks represents about 13% of the sample.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Knowledge Capital	UI	0	5	0,1	0,7	0	0	0	Level of User Innovation project on a 0-5 scale

Fig 6n. Descriptive statistics of the fit of start-ups projects with user-innovation.
Sources: PACA-EST incubator, Eckinger (2019).

In addition, following works of Eckinger and Sanders (2019) who conducted research about **user innovation** (UI) within the IRIS-project, a user-innovation scaled from 0 to 5 has been built to study the performance of user-innovation start-ups in incubation. Similar to the IRIS tracks method, evaluation of the user-innovation degree of the project has been assessed manually by reading descriptions of the projects in the documents provided by the incubator about start-ups. This assessment relies on the scale built by Eckinger (2019), (see appendix). A cross-checking work has been achieved with Caroline Eckinger to guarantee the proper application and understanding of her work. We found that only 8 start-ups met the requirement to be considered as user-innovation (implying a grade greater of at least 1/5). This can be explained by the research orientation of the incubator PACA-EST which requires the support of a research laboratory to apply, limiting greatly the likelihood to find user-innovation. Because of this limited number of cases, it impossible to use these findings within the framework of multilinear regression, justifying the fact that this variable is abandoned here. The mean of the distribution is about 0,1 and a description per quartile is not sharp enough to observe anything.

Start-up financial information.

The financial information section gathers variables about financial resources start-ups benefited from. They are the most difficult to study due to the fact that they are time-varying and are sometimes confidential or not recorded. However, they are also very interesting to study, differing from human and knowledge capital by the imitability of this type of resource. Indeed, within the framework of the resource-based view, the financial resource does not belong to the inimitable and rare resources that would really matter to create market distinction and comparative advantage. Moreover, financials being a core element of debates about innovation and entrepreneurial policy (Patton et al., 2009), investigation their true impact may participate to capture its real impact on firms' success. Thus, this section aims at evaluating the impact of financials onto start-ups performance. In the end, two main financial variables have been identified to explain the performance of start-ups.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Financial Resources	IncS	0	61742	23 940	17539	23 323	6685	37 332	Amounts of subsidies received by the firm during incubation (in €)

Fig 6o. Descriptive statistics of incubation subsidies.
Sources: PACA-EST incubator.

Hence our first financial variable, namely the amount of financial **subsidy provided by the incubator** during the programme to help start-ups to sustain their needs for growth. This data, in euros per incubation period, has been collected on exit amendments available in the database. The correlation between success and subsidies is interesting even though amounts are also correlated to the time spent by start-ups within the incubator. Furthermore, because non-incubated start-ups had no incubation subsidy, this variable could analyse the effect of money in start-up performances since it would be impossible to distinguish the effect of subsidy amounts from the effect of incubation. To avoid this bias, we muted all records from levels 1 to 3 with NAs and focused on the incubated and exited firms. We also decided to mute the start-ups remaining in incubation at the moment of the data collection. Indeed, start-ups that remain incubated at the moment of the data collection do not have any exit file, making it impossible for us to know about the total amount of subsidy they will get at the end of the programme. For these reasons, the recorded amount for these start-ups is partial and may bias again the analysis with youngest firms having the lowest subsidies. Our final sample is now limited but has a stronger explanatory strength. The sample seems symmetrically distributed around the mean, with an average subsidy of 23 940€. More than 25% of them did not benefit from more than 7000€. At the same time, the largest subsidy granted by the incubator for a venture was slightly over 60 000€.

Group	Variable	Min	Max	Mean	Sd	Med	Q1	Q3	Descriptions
Financial Resources	ShareK	500	2 250 000	109 556	262 664	35 500	10 000	100 000	Share Capital (in €), on the moment of the firm creation.

Fig 6p. Descriptive statistics of start-up's share capital.
Sources: PACA-EST incubator, Societe.com. 2019.

Secondly, we used the **share capital** of the ventures, which corresponds to the total capital brought by members to create the firm. The data has been completed quite extensively relying on the website Societe.com, already used for the NACE codes collection. The share capital corresponds to the capital brought for the creation of firms, which occurs more or less one year after the application date in the cases studied in our sample. For these reasons, the share capital is a good variable to approach the initial input from members into the project and its impact on its success. We can expect that the financial capabilities of the founding team are positively correlated with its chance of success and that the venture gathering the lowest share capital (500€) is very unlikely to perform as good as the one that gathered 2,25 million of euros. The distribution of observations is very spread out, with an important gap between the lower and the top 25%. The first quarter is indeed ten times lower than the third one (10 000€ against 100 000€ for Q3).

Analysis.

Correlations.

Table 7 is a table that shows Spearman coefficients of correlation between each variable. Correlation between variables is coloured according to their values from red for the minimum (-1) to green for the maximum (+1) passing by a neutral white for 0. In order not to skew visually the reading of this table, the diagonal of 1s has been muted into grey. The variables highlighted in white correspond to the dependent variables. Finally, only the variables that are being used in the analysis conducted in this research are represented in this table, leaving additional variables presented in the descriptive table.

First, we observe that the entry year have negative correlations with the incubation subsidies, which means that subsidies tend to be lowered through time since 2001 (-0,63). This is a surprising result considering the fact that the bias of non-exited yet ventures has been fixed. Incubation subsidies are also negatively correlated to the entrepreneurship experience (-0,34). Thus, most experimented start-ups benefit from lower subsidies on average. Second, we note that the mean age of founding teams is negatively correlated to the entry year of start-ups (-0,39). This would mean that last years, founding teams are, on average, becoming younger. The entry year is also negatively correlated with the survival performance of the start-up (-0,55). Such finding was expected since the standardisation created a strictly linear transformation of the entry year for surviving firms, which are, furthermore, more numerous in older cohorts by design. On the other hand, non-surviving firms made correlation drop but not by much. Third, growth appears to be negatively correlated with the NoTrack variable, which would imply that start-ups that follows one of the five IRIS tracks tend to grow more than the rest of the sample. Finally, we observe that the mean turnover and the mean employment are positively correlated, which means that start-ups that perform well in terms of turnover tend to hire more, or, the other way around, start-ups hiring the more people score better in terms of turnover than others.

	Var	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	PrT	1	0,07	-0,05	0,034	0,176	0,093	-0,04	0,216	-0,12	-0,17	0,018	0,222	0,094	-0,06	-0,12	-0,01
2	Stage	0,07	1	-0	NA	0,197	-0,18	0,055	0,026	0,234	-0,05	-0,1	-0,12	0,022	0,285	0,12	-0,24
3	NoTrack	-0,05	-0	1	-0,09	-0,1	-0,13	0,084	0,051	-0,04	0,1	0,085	-0,04	-0,2	0,021	-0,18	-0,28
4	IncS	0,034	NA	-0,09	1	-0,01	-0,63	0,01	0,216	0,112	-0,16	-0,34	0,006	0,28	0,413	0,421	0,321
5	ShareK	0,176	0,197	-0,1	-0,01	1	-0,23	-0,06	0,173	0,235	-0,06	0,049	0,122	0,344	0,124	0,557	0,396
6	Entry	0,093	-0,18	-0,13	-0,63	-0,23	1	-0,02	-0,39	-0,05	0,021	0,128	0,207	-0,17	-0,55	-0,38	0,107
7	WS	-0,04	0,055	0,084	0,01	-0,06	-0,02	1	0,027	0,185	-0,13	-0,05	-0,08	-0,14	-0,03	-0,09	-0,03
8	AgeMean	0,216	0,026	0,051	0,216	0,173	-0,39	0,027	1	-0,07	-0,1	0,214	0,432	0,066	0,285	0,218	0,018
9	TSize	-0,12	0,234	-0,04	0,112	0,235	-0,05	0,185	-0,07	1	0,452	-0,11	0,033	0,189	0,181	0,275	0,162
10	EdlvlMin	-0,07	0,109	0,055	-0,09	-0,05	0,073	-0,04	-0,14	-0,16	1	0,004	-0,24	0,063	-0,15	-0,03	0,03
11	EExp	0,018	-0,1	0,085	-0,34	0,049	0,128	-0,05	0,214	-0,11	0,004	1	0,299	-0,04	0,004	-0,1	-0,07
12	IExp	0,222	-0,12	-0,04	0,006	0,122	0,207	-0,08	0,432	0,033	-0,1	0,299	1	0,168	0	0,145	0,263
13	Turn	0,094	0,022	-0,2	0,28	0,344	-0,17	-0,14	0,066	0,189	0,013	-0,04	0,168	1	0,237	0,73	0,607
14	Surv	-0,06	0,285	0,021	0,413	0,124	-0,55	-0,03	0,285	0,181	-0,07	0,004	0	0,237	1	0,261	-0,2
15	FTE	-0,12	0,12	-0,18	0,421	0,557	-0,38	-0,09	0,218	0,275	-0,14	-0,1	0,145	0,73	0,261	1	0,687
16	Growth	-0,01	-0,24	-0,28	0,321	0,396	0,107	-0,03	0,018	0,162	-0,07	-0,07	0,263	0,607	-0,2	0,687	1

Fig 7. Spearman's correlation coefficients.
Sources: PACA-EST incubator, Societe.com. 2019, IRIS (Iris.eu), 2019.

Analysis methodology.

The test of the hypothesis that will follow for our research is structured in a range of linear and survival hazard regressions based on the nature of our dependent variables. Regressions have been run with the use of the free Software R (Commenges, 2014; R Core Team et al., 2017). Relying on the earlier work of Chris Eveleens (2019) and his R script used on the data for Utrecht's incubators to allow for consistent and comparable studies.

First, to estimate the role of incubation onto survival, we used a survival analysis based on the Cox proportional hazard model. We used the measure of the time since the application date as the "time" argument and the dummy 1-0 survival variable for all years of all start-ups as the status argument. As explained in the data section to explain the time-dependency of survival variable towards the application year, the hazard regressions aim at predicting for each point in time, the probability for the observation to die (or fail in case of start-ups). This enables us to assess the performance of the incubator by taking advantage of the longitudinal dimension of the dataset. The hazard ratios that calculates the model to estimate the chances of survival of a group compared to another takes the following form, with the group A represented with x and a group B represented by x' (Cox, 1972). The hazard ratio $h_k(t)/h_{k'}(t)$ is independent from time t.

$$\frac{h_k(t)}{h_{k'}(t)} = \frac{h_0(t)e^{\sum_{i=1}^{n}\beta x}}{h_0(t)e^{\sum_{i=1}^{n}\beta x'}} = \frac{e^{\sum_{i=1}^{n}\beta x}}{e^{\sum_{i=1}^{n}\beta x'}}$$

Then, we aim at estimating the impact of resources-related parameters that may impact the performance of incubated firms. To estimate these three other groups of variables, we used a multilinear regression based on the standard ordinary least squares model (OLS). Since all the four dependent variables are continuous, all of them has been estimated with this model.

After the regressions of each model, a range of tests has been performed to check the reliability of the results we found. First, multicollinearity has been rejected using the variational inflation factors (VIF). This means that our independent variables are not related to each other. Collinearity reflects the co-evolution of two variables that, together with covers, the entire sample and whose probability to be observed are equal or close to 1. Rejecting the assumption of multicollinearity in our models means that our variables are independent and, thus, estimates are reliable. Although the VIF was slightly higher in the case of NACE industrial codes in our third hypothesis than for the rest of the variables, effective collinearity of variables has been rejected. Also, we checked the null hypothesis of homoscedasticity of our model with the use of the Breusch-Pagan test. Homoscedasticity implies that the variance, so the distribution, of errors is the same for all observations and does not depend on other inherent factors of observations. In our cases, hypothesises of homoscedasticity are not rejected, which means that estimates found are reliable. Also, we checked the distribution of the residuals for each model we tested. Distribution is not linearly distributed. Finally, no reason motivated the removal of any outliers which could have disturbed our analysis.

In the test of our second, third and fourth hypothesis, respectively three, two and two different models have been run. Because of diversely distributed missing values throughout variables, each additional variable limits the strength of the analysis by narrowing the number of cases studied. In order not to miss some significant results and to distinguish better the effect of each variable on the model, they have been added by steps (stepwise method with forward selection).

Results.

The first part of the results on the assessment of the incubation impact on the survival of variables. Relying on these results, we then dive into the results of linear regression to interpret the performance results of incubated firms with the resources-related variables that were introduced in the data section. The strength of the latter results depends on the former ones. Indeed, a significant impact of incubation onto survival implies a bias between incubated and non-incubated firms through the screening policy and the incubation process.

Impact of incubation.

#H1 Impact of incubation on survival.

coxph(formula = Surv(Age, survival) ~ Stage, data = Survivaldf)
n= 1688, number of events= 1589
(832 observations deleted due to missingness)

	coef	exp(coef)	se(coef)	z	Pr(>|z|)
Stage = 2	2.27591	9.73674	0.71356	3.190	0.00143 **
Stage = 3	NA	NA	0.00000	NA	NA
Stage = 4	1.68900	5.41408	0.17855	9.459	< 2e-16 ***
Stage = 5	-0.01624	0.98389	0.08160	-0.199	0.84225

	exp(coef)	exp(-coef)	lower .95	upper .95
Stage = 2	9.7367	0.1027	2.4045	39.427
Stage = 3	NA	NA	NA	NA
Stage = 4	5.4141	0.1847	3.8154	7.683
Stage = 5	0.9839	1.0164	0.8385	1.155

Concordance = 0.52 (se = 0.006)
Likelihood ratio test = 75.43 on 3 df, p=3e-16
Wald test = 117.8 on 3 df, p=<2e-16
Score (logrank) test = 151.4 on 3 df, p=<2e-16

Fig 8. Cox Proportional Hazard model estimating the impact of incubation on survival.
Sources: PACA-EST incubator ; Cox Proportional hazard model, R.

	Records	n.max	n.start	Events	rmean	se(rmean)	Median	0.95LCL	0.95UCL
Stage=1	187	187	187	169	4,762	0,242	4	4	6
Stage=2	2	2	2	2	0,5	0,354	0,5	0	NA
Stage=4	41	41	41	41	1,146	0,145	1	1	1
Stage=5	1458	1458	1458	1377	4,796	0,082	5	4	5

Fig 9. Summary statistics of survival for each stage's value.
Sources: PACA-EST incubator ; Cox Proportional hazard model, R.

On figure 8, the results of the Cox model show no significant effect of the stage variable on survival. We note that if stages 2 and 4 are significant, stage 5 is not found significant. However, the results found for stages 2 and 4 can be self-explained. Indeed, waiting start-ups (stage 3) are all still alive since they are all new ventures, not founded yet, that are applying for the incubator on the moment of the data collection. Stage 4, that refers to start-ups that are still incubated on the moment of the data collection is, also, all alive. What is the most interesting is that exited firms (stage 5) do not significantly survive better than non-incubated ones (stage 1) that are used as the control group in this survival analysis. The p-value of stage 5 is about 0.8, therefore very much higher than any significance level. More precisely, looking at figure 9, we can observe detailed distribution statistics of survival according to each stage value. We see that mean values for stage 1, referring to rejected start-ups, and stage 5, corresponding to exited firms are very similar with respectively 4,762 and 4,796 years of average life expectancy. This difference is slightly more sensible looking at the median value, which can be explained by the lower standard error observed among non-incubated firms. Findings thus show no significant impact of incubation on the survival of start-ups in the sample we studied. The incubated firms do not survive longer than the ones that abandoned their application or that have been rejected by PACA-EST. Such results differ from what has been found by Eveleens (2019) in Utrecht.inc and Climate KIC incubators. Indeed, he found a clear positive impact of these two programmes on start-ups' survival. Consequently, our first hypothesis about the positive impact of incubation on survival is rejected. The absence of a significant impact of incubation also questions the inimitability of resources offered by the incubator to its incubatees. From a resource-based view perspective, it seems that resources gained during the incubation process may not allow start-ups to survive longer than their non-incubated competitors. Thus, these resources might not be valuable enough or can be obtained otherwise by non-incubated firms. Also, the screening policy, the incubator is supposed to select the most promising ventures. Such results tend to show that the strategy operated by PACA-EST is not as performant as it should. However, if the results show no significant impact of incubation, they do not assure that there is no impact at all, but rather that if there is one, it has not been found.

Finally, this implies that there are limited concerns about selection bias in our sample to conduct further analysis.

Performances determinants.

#H2 Impact of human capital.

Table 10 shows the results for the test of the first hypothesis questioning the impact of the human capital on start-ups' success. Consistent with the literature, gender seems to have no significant effect on the performance of our incubated start-ups (Banal-Estanol et al., 2016). Neither turnover, employment nor survival are positively impacted by the share of women in founding teams of our sample. Although, the sample may be too small and men-only to perceive any pattern, causality or correlation.

On the contrary, the average age of team members seems to have a very significant impact on survival. Under the threshold of 0,1%, we observe that the founding team's mean age have a positive impact on the survival with a coefficient included between 0,150 and 0,126 depending on the models tested. Since the survival variable has been standardised, this coefficient is however not intuitively meaningful. Within the framework of the resource-based view, we can say that, on average, the additional resources brought by older team members, positively impact start-ups' success. This impact apparently overcomes the hypothetical impact that youngness could have too.

	Dependent variables:											
	Survival			Employment size			Growth			Turnover		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
WS	-0.201	-0.318	-0.121	-0.975	-0.999	-1.001	-0.475	-0.418	-0.387	0.068	0.081	-0.044
	(0.376)	(0.376)	(0.386)	(0.709)	(0.695)	(0.705)	(0.433)	(0.425)	(0.427)	(1.020)	(1.041)	(1.051)
AgeMemb	0.150***	0.133***	0.126**	-0.016	-0.066	-0.053	0.016	-0.003	0.017	0.015	0.028	0.006
	(0.036)	(0.037)	(0.038)	(0.066)	(0.066)	(0.070)	(0.040)	(0.041)	(0.042)	(0.099)	(0.104)	(0.106)
TSize	1.229	1.186	1.342	4.311***	4.272***	4.046**	1.432*	1.773*	1.303	-0.186	-0.426	-0.861
	(0.674)	(0.676)	(0.722)	(1.157)	(1.165)	(1.274)	(0.707)	(0.713)	(0.771)	(1.814)	(1.841)	(1.993)
Edlvl min = 2		-14.48*	-14.19*		-36.05*	-35.99*		-4.313	-3.975		6.560	4.565
		(6.171)	(7.068)		(17.45)	(17.66)		(10.67)	(10.69)		(17.66)	(22.35)
IExp			0.028			-0.018			0.029			-0.093
			(0.046)			(0.081)			(0.049)			(0.123)
EExp			0.033			0.022			-0.062			0.675*
			(0.115)			(0.209)			(0.127)			(0.326)
Constant	18.851***	34.282***	31.664***	25.465***	63.811**	64.390**	17.404***	20.871	20.744	60.521***	54.548*	53.551
	(3.742)	(7.525)	(8.636)	(6.649)	(19.91)	(20.47)	(4.059)	(12.18)	(12.39)	(10.64)	(22.20)	(27.82)
McFadden R2	0.022	0.045	0.096	0.027	0.047	0.061	0.02	0.04	0.057	0.015	0.023	0.057
Observations	202	198	188	134	132	130	134	132	130	130	129	125
Log Likelihood	-857.893	-837.437	-793.290	-615.950	-603.249	-594.427	-549.819	-538.365	-529.177	-650.699	-645.706	-623.60
Akaike Inf. Crit.	1,723.79	1,684.88	1,600.58	1,239.90	1,216.50	1,202.85	1,107.64	1,086.73	1,072.35	1,309.40	1,301.41	1,261.20
Note:										*p<0.05;	**p<0.01;	***p<0.001

Fig 10. Regression model testing the effect of human capital on start-ups performances.
Sources: PACA-EST incubator.

Then, we can note that the size of teams has been found significant in the prediction of start-ups' employment size, and, to a lower extent, the prediction of employment growth. Under 0,1% to 1% thresholds, one additional member in the founding team would multiply by a factor of 4 the average annual number of employees hired by start-ups. In the meantime, models 7 and 8 shows that larger teams also tend to have greater employment growth than smaller ones. However, the significance of this parameter disappears once industrial and entrepreneurial experiences are introduced in our model. Note that the size of firms and their growth says nothing about their survival probability or their success in terms of revenues.

The education level of the least educated one is found significant with 5% thresholds in the explanation of both the survival and the employment size of start-ups. More precisely, teams whose least educated member has an academic level of education – i.e. between bachelor and PhD – survive shorter and employ fewer people than teams with a member with lower education. From the other point of view, start-ups that have at least one non-academic profile in their team are more likely to live longer and hire more people than others. Moreover, coefficients found are very high, start-ups increasing their survival performance by a factor of 14,19 when having a lower educated member. However, such coefficients can also be caused by the low average value and standard deviation in education. That inflates the coefficient. Such findings show that profiles' diversity benefits to start-ups. Indeed, if most of them are constituted of highly educated people (PhD or master), it appears that the presence of lower educated profiles benefits significantly to their success. Although this tends to show a positive impact of resources diversity in the performances of start-ups, researches conducted by Eveleens (2019) with a study of knowledge novelty and diversity have shown inconclusive results.

Finally, if the industrial linkages of members, measured through their experience in the private sector, has not been found significant to explain any of our dependent variables, this is not the case with the entrepreneurial experience. Indeed, we observe that the entrepreneurial experience has a weak but positive impact on the turnover of firms with a significance threshold of 5%. Furthermore, the explanatory strength of the model increases quite significantly – passing from 0,023 (11) to 0,057 (12) – with the introduction of the variable which shows a true explanatory role of the entrepreneurial experience on turnover's values. Such results imply that the more entrepreneurial experience team members of a start-up have, the better they tend to perform in terms of turnover. However, the turnover being in euros, the effect seems rather limited. This finding meets our expectations and the resource-based view, with a positive, even though limited, impact of rare and inimitable resources on firms' performances. Nevertheless, the mean levels of experience may not be the best proxy to predict start-ups performance. Indeed, similar to the education level, the experience of only one team member in the group might be more influential than the average experience capital of the whole founding team.

Overall, models fail at predicting most of the distribution. Looking at the McFadden R^2S results, we can note that no model is able to predict more than ten per cent of the variation. We can nevertheless conclude that the human resources of start-ups do impact their performances. Variables like the size of the founding team, its average age and entrepreneurial experience have positive impacts on performance inductors. We also note that more than the number of resources available in the teams, their diversity seems to have an impact on start-ups success too. Consistent with the findings of Barney et al. (2001) and Harrison et al. (1991), the complementarity of resources leads to higher performances as shown by the positive effect of a lower educated member in founding teams. However, the absence of the impact of industrial experience is surprising. Within the RVB framework, it would mean that industrial experience does not represent an inimitable enough resource for start-ups.

#H3 Impact of knowledge capital.

Figure 11 shows the results for the test of the third hypothesis questioning the impact of the knowledge capital on the outcomes of incubation over start-ups success.

Results show that the product type is significant in the prediction of the survival and the employment growth of start-ups. Under a significance threshold of 1%, hardware projects appear to last shorter than other types of start-ups (-8,23). Surprisingly, the estimated coefficient of the product-type variable for survival has an opposite sign to the estimated coefficient for growth, which is positive (+7,79). This means that, while the product-type have a negative impact on survival, it impacts positively the employment growth. In other terms, start-ups involved in the production of hardware products grow faster than others in terms of employment but have a lower survival probability. It can be explained by the larger initial capital investments usually required to set up the manufacturing organization (Eveleens, 2019). Then, an intensive and early growth may be required which can, if not controlled properly, lead to failure. Results found in Utrecht's incubator showed a significant negative impact of hardware orientation towards both survival (-0,713) but also on employment growth (-1,655) (Eveleens, 2019).

	Dependent variables:							
	Survival		Employment size		Growth		Turnover	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Prodtype	-2.405	-8.229**	-6.027	0.966	0.278	7.791**	-1.901	-9.758
	(2.520)	(2.957)	(4.357)	(5.369)	(2.483)	(2.951)	(6.365)	(8.397)
NoTrack	-1.150	0.337	0.455	3.637	-11.497**	-9.609**	-0.184	-2.239
	(3.579)	(3.407)	(6.005)	(5.911)	(3.422)	(3.249)	(9.205)	(9.576)
NACE = C		18.765		5.970		13.699		38.083
		(12.414)		(25.472)		(14.003)		(38.321)
NACE = G		11.121		-10.767		11.146		26.828
		(13.331)		(27.488)		(15.111)		(42.880)
NACE = J		8.498		15.460		28.066*		18.749
		(11.926)		(25.061)		(13.777)		(37.884)
NACE = K		25.205		-10.000		29.000		13.500
		(15.169)		(34.604)		(19.023)		(45.092)
NACE = M		15.992		5.065		18.303		23.153
		(11.955)		(24.855)		(13.664)		(37.402)
NACE = N		20.128		-25.517		5.896		43.828
		(15.169)		(30.088)		(16.540)		(42.880)
NACE = S		24.385						
		(20.394)						
Constant	34.644***	24.277	40.775***	26.397	32.058***	4.818	62.456***	44.997
	(3.529)	(12.325)	(5.898)	(25.817)	(3.361)	(14.193)	(9.010)	(39.102)
McFadden R2	0.001	0.113	0.002	0.072	0.01	0.09	0	0.055
Observations	204	185	136	128	136	128	132	125
Log Likelihood	-876.766	-778.189	-631.864	-587.224	-555.371	-510.639	-660.878	-624.444
Akaike Inf. Crit.	1,759.532	1,576.378	1,269.727	1,192.448	1,116.741	1,039.278	1,327.756	1,266.887
Note:						*p<0.05;	**p<0.01;	***p<0.001

Fig 11. Regression model testing the effect of knowledge capital on start-ups performances.
Sources: PACA-EST incubator, Societe.com. 2019, IRIS (Iris.eu), 2019.

Then, we found that start-ups involved in a project fitting IRIS tracks perform better, in terms of employment growth than others. Under a 1% significance threshold, firms that do not meet IRIS tracks' criteria – that score 1 in the NoTrack variable – observe a lower average annual growth, with an estimated coefficient of -9,6. If there is no evidence of better survival, it seems that sustainable and smart-city related start-ups have a more important increase in their employment level. Thus, it seems that the start-up that fits in IRIS tracks hire more than the one from other sectors in Nice. This is interesting since earlier results from Utrecht found no significant effect.

About NACE classification, results are mostly insignificant. The only category that is found significant is the J category. As presented in figure 5k, page 27, the J category refers to information and communication (IC) sector, which is also the most present in this sample, with 91 observations out of 308 (and 205 excluding the missing values). Results show that IC start-ups overperform in terms of employment growth compared to categories A (agriculture, forestry and fishing), H (Transportation and storage) and P (Education). These results are somehow surprising that the J category mostly refers to software and services productions while it has been found that hardware products were also positively correlated to employment growth.

Once again, our models appear to be rather weak when looking at standardized R^2. If the prediction strength increased a little compared to the human capital models, this can be due to a larger number of independent variables in the model. Results show that the industrial categories and the knowledge capital required for each of them differs and thus impact the performance of start-ups. It is made clear that all sectors do not need the same inimitable resources and that some of them are more likely to find them and succeed, especially within incubation. Overall, hardware-focused start-up survives shorter on average. However, together with IC and smart-city related start-ups, hardware-focused ventures observe a higher average employment growth. We can thus conclude that our third hypothesis is verified and that all incubatees may require some specific support to obtain the resources they need.

#H4 Impact of financial resources.

Figure 12 shows the results for the test of the fourth hypothesis questioning the impact of the financial resources on the outcomes of incubation over start-ups success. Financial resources are imitable resources and are not considered as influential explaining the relative outperformance of firms.

	Dependent variable:							
	Survival		Employment size		Growth		Turnover	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Share_K	-0.022	-0.061	0.097	0.109	0.004	0.041	0.0004	0.013
	(0.031)	(0.033)	(0.056)	(0.061)	(0.033)	(0.034)	(0.081)	(0.082)
Inc_sub		0.022		0.031		0.076^{*}		0.151^{*}
		(0.028)		(0.053)		(0.030)		(0.074)
Constant	35.612^{***}	39.135^{***}	32.531^{***}	29.703^{***}	22.510^{***}	12.875^{***}	61.418^{***}	46.787^{***}
	(2.395)	(3.552)	(4.305)	(6.760)	(2.562)	(3.783)	(6.222)	(9.195)
McFadden R2	0.102	0.327	0.025	0.134	0.023	0.145	0.014	0.095
Observations	186	143	133	118	133	118	130	120
Log Likelihood	-787.532	-590.395	-617.052	-548.359	-548.005	-479.858	-651.493	-598.021
Akaike Inf. Crit.	1,579.064	1,186.790	1,238.103	1,102.718	1,100.010	965.715	1,306.986	1,202.042
Note:						$^{*}p<0.05$; $^{**}p<0.01$; $^{***}p<0.001$		

Fig 12. Regression model testing the effect of financial provisions on start-ups performances.
Sources: PACA-EST incubator, Societe.com. 2019.

Results show no significant impact from the share capital on our performance indicators. The insignificance of share capital is rather surprising, meaning that the initial financial resources of start-ups have, in the end, a limited impact on their success. We tested first the share capital alone to test it with the maximum number of observations, but no significant influence has been found in these regressions.

The tests of the incubation subsidies show no significant impact on the survival or the employment size, however, under the 5% significance threshold, we find an impact on both growth and turnover. It is unexpected that the subsidy levels are correlated to performance since the amounts are still on a limited scale and have a limited variance. In this case, it seems more likely that results show more correlation than causality patterns. Also, incubation subsidies may conceal the effects of other related variables such as the length of the incubation. Note that the coefficient found to predict growth is very small reflecting the increase of 0,076 additional FTE per year for each additional euro received by the incubatee during the incubation through subsidies. On the contrary, a 0,151 coefficient to predict turnover is more interesting. Indeed, it means that each additional euro of subsidies converts into 0,151€ of turnover. With an average subsidy amount of about 24 000€, this represents a potential increase of a start-up's mean turnover of about 3 600€ per year, which is meaningful, even though the multiplier returns decreasing outputs.

Finally, these models showed a greater explanatory strength according to the level of their standardised R^2, even though very few results were significant. Overall, we can conclude that financial resources have a limited impact on start-ups' performances. However, if a causality relation is unlikely, it appears that there is still a significant correlation between the incubation subsidies and the turnover for incubated firms. Therefore, results reject our hypothesis of a positive influential role of financial resources on firms' success. However, this comforts the resource-based view which considers financial resources as imitable resources that have no influence on the creation of comparative advantage of firms.

Discussions.

This thesis aimed at investigating to what extent does incubation impact start-ups survival, but also, subsequently, how do the different resources start-ups benefit from having an effect on their overall performance and success. To conduct this study, we relied on data provided by the PACA-EST incubator in Nice over the past eighteen years. At first, we used the Cox proportional hazard model to assess the outcomes of the incubator in terms of survival for the start-ups. Results showed that the impact was insignificant and that start-ups had similar surviving probabilities being incubated or not. This leads to additional questions about the inimitable resources the incubator was expected to bring to its incubatees. Subsequently to these findings, we studied the effects of the resources owned by the start-ups and how they could impact their performances. After differentiating human capital, knowledge capital and financial resources as three different groups of resources we tested them through multilinear regressions. We found that the human resources of the founding teams had a positive impact on performances and that, more interestingly, the diversity of this knowledge additionally to its quantity is significant to explain it. We also found that, depending on their knowledge-base, all start-ups do not perform equally. Relying on the resource-based view, it can be explained by the unequal access to the strategic inimitable resources start-ups may require to succeed. Finally, we found that financial resources were not a good predictor for start-ups performance. Even though, results show a significant correlation between the incubation subsidies and the turnover of incubated firms. This uncovers a perspective for further analyses to investigate what variable can truly explain this relation. These findings may represent some valuable contributions to both the literature and the practice.

Contribution to the literature.

With our findings, this thesis contributes to its theoretical framework. First, it participates to enrich the considerable empirical research on the impact of incubation (Amezcua, Grimes, Bradley, Amezcua, & Bradley, 2013; Cohen, Bingham, & Hallen, 2018; Colombo & Delmastro, 2002; Eveleens, 2019; Lukeš et al., 2018; Peters, Rice, & Sundararajan, 2004; Stokan et al., 2015). Our results do not reject the null hypothesis of the absence of impact of incubation on start-ups' survival. Unlike the conclusions of Stokan et al. (2015), it seems that our results did not overestimate the impact of incubation. The "picking-the-winners" strategy of the incubator, that uncovered Bergek & Norrman (2008) seems not to be systematically verified. This is consistent with what has been found by Eveleens (2019) when comparing the pre-incubation quality assessments of incubated and rejected start-ups in Utrecht.

Furthermore, we tested the impact of diverse types of resources conveyed by incubatees antecedents on their performances (Barney et al., 2001; Hausberg & Korreck, 2018). We found that, consistent with the existing literature, part of them are good predictors of success (Baum & Locke, 2004; Wright & Stigliani, 2013). It is clear from the results that the characteristics of the founding team matters. We found significant impacts from the size, the mean age but also the education level of the members of the founding teams. This support the key role of inimitable resources brought by the entrepreneurs in their firm and how the combination of these resources influence the performances of the start-up to obtain above-normal returns (Barney et al., 2001; Penrose, 1959; Rugman & Verbeke, 2002). From this perspective, our findings tend to show that the resources that start-ups need for growth might not be equally accessible to all types of entrepreneurial projects. For instance, hardware-oriented start-ups survive shorter than others. On the contrary, start-ups involved into smart-city and sustainability that fits the IRIS transition tracks appears to grow more on average in terms of employments. Another conclusion of this thesis is the unclear impact of financial resources on incubated firms' success. The effect of the share capital on our performance indicators has been found not significant. This meets the theory of the RBV stating that financial resources are not inimitable resources and do not participate to outstanding returns of firms (Barney, 1991; Barney et al., 2001; Penrose, 1959; Pitelis, 2004). Meanwhile, confusing results about the impact of incubation subsidies calls for further investigations.

Policy implications.

The results contribute to practice too. In the first place, keeping in mind the limited scale of the research, the absence of a significant impact of the incubator on the success of the incubated firms opens reflection perspectives. The incubator may investigate how it is possible to enhance its action. It can also be a signal addressed to policymakers that support incubators within their innovation and entrepreneurship policy. If the results show no meaningful effect of the incubation at PACA-EST, this is not the case for all incubators, for instance, Climate KIC and Utrecht.inc in the Netherlands. Differences between the findings may lead to a benchmarking to compare PACA-EST practices to observations made elsewhere. This questioning may then improve the outcomes of incubation but also investigate other means in the support given to start-ups.

Second, the results found on the impacts of human and knowledge capital on start-ups' success are valuable for the incubator, and specifically in the design of its screening policy. Indeed, by understanding better what types of founding teams performs the best or least, but also which industrial orientation have more success, incubators can adapt their practices. The first change might come from the screening policy of the incubator which is supposed to target the most promising ventures. Adaptations can also come to the support and services offered to the incubatees to take better into account the different needs of start-ups in terms of resources according to their characteristics.

Third, the absence of significance of the share capital onto the performances of start-ups gives interesting information to incubators, policymakers, and institutions. Indeed, it seems that, at its early stages, start-ups do not perform better when they are better gifted with money. Additionally, the impact of incubation subsidies is significant but very likely as a correlated variable rather than as an explanatory one. Further questioning about the relevance of subsidies with a decreasing output may also be considered. This tends to show to policymakers that financial support may not be the most effective way to assist and help start-ups to perform better.

Finally, we found that start-ups that are committed to sustainability and smart-city related entrepreneurial projects develop themselves more than others in terms of employment. Such findings give evidence IRIS that the smart-city ecosystem is rather favourable to start-ups. They indeed tend to hire more which makes them profitable ventures to invest in for policymakers that want to stimulate innovation and employment.

Limitations.

The assessment of incubation's impact is a complex work that usually involves several bias and limitations (Eveleens, 2019; Hackett & Dilts, 2004; Lukeš et al., 2018; Stokan et al., 2015). The reason is that the sample used to control the effect of incubation is hardly comparable to the incubated one. Indeed, the screening policy is designed to pick the most promising ventures which, therefore, create a sample of start-ups that are somehow the top of the heap. Then, as Stokan et al. (2015) found, the findings are likely to overestimate the effect of incubation. In our case, it can't be the case since we did not find significant results. However, the control group being composed the firms that have been rejected or that abandoned their application, it is not perfectly comparable to the group of incubated firms. Also, the screening policy operated by each incubator being quite unique, the generalizability of the findings appears rather limited.

Second, the data appears to be more or less time-sensitive, depending on the dependent variables. Since our sample is composed of start-ups that have been incubated all over the past eighteen years, observations tend to vary quite significantly according to their age (i.e their year of application). Indeed, because the incubator focuses on non-founded yet ventures only, the entry year of start-ups in our sample is roughly equal to their year of creation. This leads to a bias of observation based on their cohort since the survival performance of younger firms can't possibly be comparable to the oldest ones. Similarly, we can expect employment and turnover variables to be affected. Moreover, because the cohorts were too limited with a single incubator – less than twenty observations per cohort – we could not regress them separately. In the end, we tried to contain this time-dependency by standardising the survival performance and using the mean values of employment and turnover.

Finally, we observed many missing values in the dataset. The poverty of the dataset concerning the non-incubated firm or non-incubated firms yet is normal since no extensive monitoring have been processed by the incubator for these firms. However, we found that even on incubated firms, many observations went missing and particularly in the financial sections that gather the turnover, employment and net operative income figures. This participated in the weakening of this study by narrowing the sampled analysed through the regressions. In the end, most of the models have been tested with between 120 and 160 observations. Only the survival performance benefited from an average of 180-200 observations because of non-incubated observations available for this dependent variable. In the end, the results rely, on average, on half of the 308 start-ups that compose the sample.

References.

Aerts, K., Matthyssens, P., & Vandenbempt, K. (2007). Critical role and screening practices of European business incubators. *Technovation*, *27*(5), 254–267. https://doi.org/10.1016/j.technovation.2006.12.002

Allen, D. N., & McCluskey, R. (1991). Structure, Policy, Services, and Performance in the Business Incubator Industry. *Entrepreneurship Theory and Practice*, *15*(2), 61–77. https://doi.org/10.1177/104225879101500207

Amezcua, A. S., Grimes, M. G., Bradley, S. W., Amezcua, A. S., & Bradley, S. W. (2013). ORGANIZATIONAL SPONSORSHIP AND FOUNDING ENVIRONMENTS : A CONTINGENCY VIEW ON THE SURVIVAL OF BUSINESS-INCUBATED FIRMS , 1994-2007 Source : The Academy of Management Journal , Vol . 56 , No . 6 (December 2013), pp . 1628- Published by : Academy of Manage, *56*(6), 1628–1654.

Banal-Estanol, A., Macho-Stadler, I., & Perez-Castrillo, D. (2016). Key success drivers in public research grants: Funding the seeds of radical innovation in academia? *CESifo Working Paper Series*, (September), 1–36. Retrieved from

http://www.barcelonagse.eu/sites/default/files/working_paper_pdfs/890.pdf%0Ahttps://papers.ssrn.com/sol3/papers.cfm?abstract_id=2774648

Barney, J. (1991). Barnay, 1991.pdf. *Journal of Management*, *17*(1), 99–120. https://doi.org/10.1177/014920639101700108

Barney, J., Wright, M., & Ketchen Jr, D. J. (2001). The RBV of the firm: Ten years after 1991. *Journal of Management*, *27*, 625–641.

Baum, J. R., & Locke, E. A. (2004). The relationship of entrepreneurial traits, skill, and motivation to subsequent venture growth. *Journal of Applied Psychology*, *89*(4), 587–598. https://doi.org/10.1037/0021-9010.89.4.587

Bergek, A., & Norrman, C. (2008). Linköping University Post Print Incubator best practice : A framework Incubator best practise : A framework. *Technovation*, *1–2*(28), 20–28. https://doi.org/10.1093/geront/gns105

Cohen, S. L., Bingham, C. B., & Hallen, B. L. (2018). The Role of Accelerator Designs in Mitigating Bounded Rationality in New Ventures. *Administrative Science Quarterly*. https://doi.org/10.1177/0001839218782131

Colombo, M. G., & Delmastro, M. (2002). How effective are technology incubators? *Research Policy*, *31*(7), 1103–1122. https://doi.org/10.1016/s0048-7333(01)00178-0

Commenges, H. (2014). *R et espace*. (Framabook), Ed.). Paris: CC By-SA 3.0.

Cox, D. R. (1972). Regression Models and Life-Tables Author (s): D . R . Cox Source : Journal of the

Royal Statistical Society . Series B (Methodological), Vol . 34 , No . 2 Published by : Blackwell Publishing for the Royal Statistical Society Stable URL : http://www.js. *Journal of the Royal Statistical Society.*, *34*(2), 187–220.

Eckinger, C., & Sanders, M. (2019). *User Innovation : Why is Business Incubation not a topic ?* Utrecht University.

Eveleens, C. P. (2019). *3 . The impact of incubation on start-up performance : the moderating effect of the start-up ' s knowledge base*. Utrecht University.

Eveleens, C. P., van Rijnsoever, F. J., & Niesten, E. M. M. I. (2017). *How network-based incubation helps start-up performance: a systematic review against the background of management theories*. *Journal of Technology Transfer* (Vol. 42). Springer US. https://doi.org/10.1007/s10961-016-9510-7

Hackett, S. M., & Dilts, D. M. (2004). A Systematic Review of Business Incubation Research. *The Journal of Technology Transfer*, *29*(1), 55–82. https://doi.org/10.1023/b:jott.0000011181.11952.0f

Hallen, B. L., Bingham, C. B., & Cohen, S. L. (2016). th , 2016.

Harrison, S; Hitt, M; Hoskisson, R; Ireland, R. . (1991). Synergies and Post-Acquisition Performance : Differences versus Similarities in Resource Allocations. *Journal of Management*, *17*(1), 173–190.

Hausberg, J. P., & Korreck, S. (2018). Business incubators and accelerators: a co-citation analysis-based, systematic literature review. *Journal of Technology Transfer*, 1–26. https://doi.org/10.1007/s10961-018-9651-y

Hidalgo, C., & Haussmann, R. (2013). *Atlas of economic complexity - Mapping paths to prosperity*. Puritan Press.

Lee, C., Lee, K., & Pennings, J. M. (2001). Internal capabilities, external networks, and performance: A study on technology-based ventures. *Strategic Management Journal*, *22*(6–7), 615–640. https://doi.org/10.1002/smj.181

Levie, J. (2009). A Final Assessment of Stages Theory : Introducing a Dynamic States Approach to Entrepreneurship, (44), 317–350. https://doi.org/https://doi.org/10.1111%2Fj.1540-6520.2010.00377.x

Lukeš, M., Longo, M. C., & Zouhar, J. (2018). Do business incubators really enhance entrepreneurial growth? Evidence from a large sample of innovative Italian start-ups. *Technovation*, (July). https://doi.org/10.1016/j.technovation.2018.07.008

Mason, C., & Brown, R. (2013). Creating good public policy to support high-growth firms. *Small Business Economics*, *40*(2), 211–225. https://doi.org/10.1007/s11187-011-9369-9

Patton, D., Warren, L., & Bream, D. (2009). Elements that underpin high-tech business incubation processes. *Journal of Technology Transfer*, *34*(6), 621–636. https://doi.org/10.1007/s10961-009-9105-7

Penrose, E. (1959). *The Theory of the Growth of the Firm*. (Christos Pitelis, Ed.) (reprint). Oxford University press.

Peters, L., Rice, M., & Sundararajan, M. (2004). The Role of Incubators in the Entrepreneurial Process. *The Journal of Technology Transfer*, *29*(1), 83–91. https://doi.org/10.1023/b:jott.0000011182.82350.df

Phan, P. H., Siegel, D. S., & Wright, M. (2005). Science parks and incubators: Observations, synthesis and future research. *Journal of Business Venturing*, *20*(2), 165–182. https://doi.org/10.1016/j.jbusvent.2003.12.001

Pitelis, C. N. (2004). Edith Penrose and the resource-based view of (international) business strategy. *International Business Review*, *13*(4), 523–532. https://doi.org/10.1016/j.ibusrev.2004.04.002

Porter, M. E. (1990). Competitive Advantage, Agglomeration Economies, and Regional Policy. *International Regional Science Review*, *19*(1996), 85–94.

R Core Team, Team, R. C., & others. (2017). R: A language and environment for statistical computing, *3*. https://doi.org/ISBN 3-900051-07-0, URL http://www.R-project.org/

Rice, M. P. (2002). Co-production of business assistance in business incubators. *Journal of Business Venturing*, *17*(2), 163–187. Retrieved from https://doi.org/10.1016/S0883-9026(00)00055-0

Rugman, A. M., & Verbeke, A. (2002). Edith Penrose's contribution to the resource-based view of strategic management. *Strategic Management Journal*, *23*(8), 769–780. https://doi.org/10.1002/smj.240

Scillitoe, J. L., & Chakrabarti, A. K. (2010). The role of incubator interactions in assisting new ventures. *Technovation*, *30*(3), 155–167. https://doi.org/10.1016/j.technovation.2009.12.002

Shane, S. (2009). Why encouraging more people to become entrepreneurs is bad public policy. *Small Business Economics*, *33*(2), 141–149. https://doi.org/10.1007/s11187-009-9215-5

Stokan, E., Thompson, L., & Mahu, R. J. (2015). Testing the Differential Effect of Business Incubators on Firm Growth. *Economic Development Quarterly*, *29*(4), 317–327. https://doi.org/10.1177/0891242415597065

Tushman et O'Reilly. (1996). Ambidextrous Organizations: Managing evolutionary and revolutionary change. *American Journal of Bioethics*, *16*(2), 15–17. https://doi.org/10.1080/15265161.2015.1120813

Weiblen, T., & Chesbrough, H. W. (2015). Engaging with Startups to Enhance Corporate Innovation. *California Management Review*, *57*(2), 66–90. https://doi.org/10.1525/cmr.2015.57.2.66

Wright, M., & Stigliani, I. (2013). Entrepreneurship and growth. *International Small Business Journal*, *31*(1), 3–22. https://doi.org/10.1177/0266242612467359

Appendix.

UI Scale.

Item	**Value *v***
User innovation identification	
1. Individual consumer / firm / community user innovator	0 **|** 1
2. Modification / new creation	0 **|** 1
3. For personal / in-house use	0 **|** 1
User innovation intensification	
a. Tailor-made / customized	0 **|** 1
b. High expected benefit / value	0 **|** 1
c. Satisfies / better suits own needs	0 **|** 1
d. New to market (novelty)	0 **|** 1
Note:	0 = false **|** 1 = true

$$UI\ Index = (v\text{Item }1 * v\text{Item }2 * v\text{Item }3) * (1 + v\text{Item }a + v\text{Item }b + v\text{Item }c + v\text{Item }d) \quad (1)$$

Graphic 1 User innovation index

0	1	2	3	4	5
Non-UI	*Very weak*	*Weak*	*Moderate*	*Strong*	*Very strong*

Fig 13. User-innovation 0-5 scale.
Sources: Eckinger and Sanders (2019).

R script.

```
# IMPORT STARTUP DATA  ####
#load packages
library(fields)
library(tm)
library(tidyr)
library(dplyr)

#load primary data set
setwd("C:/Users/Romai/OneDrive/Documents/Études/Master 1 - Utrecht
University/Master Thesis/IRIS Data/Nice")
df <- data.frame(read.csv2("V11.4.csv", stringsAsFactors = F))

## CLEAN DATAFRAME ####

#Delete "G" values that could disturb the analysis
df<-sapply(df,as.character)
df <- ifelse(df=="G",NA,df)
#Change "C" to "D", they will both be analysed as failures
df <- ifelse(df=="C","D",df)
#Correction of the survival variable to take failure years into
account
df<-data.frame((df))

#The existing Survival variable has not been completed time-
sensitively,
#we create a new one based on the SIRET code which has been replaced
by "D" for Disbanded on the liquidation year
df <- mutate (df, survival=SIRET)
#The new survival variable is mutated to a domi 1/0 variable
df = mutate(df, survival = ifelse(
  survival =="D", 0,
  ifelse (survival!="D",1,NA)))
df$Survival<-NULL #the former Survival variable is deleted in order
not  to confuse ourselves

write.csv(df,file = "V11_5.csv")

df <- read.csv2("V11_5.csv", stringsAsFactors = F, sep = ",")

df$X<-NULL #Unuseful colomun
names(df)[128]="NetOperativeIncome" #Rename the NA. column.
names(df)[130]="FTE" #Rename the NA..1 column.
df$FTE<-as.numeric(df$FTE)
df$Turnover<-as.numeric(df$Turnover)
df$survival<-as.numeric(df$survival)
df$NetOperativeIncome<-as.numeric(df$NetOperativeIncome)

#creation a FTE growth column
df$FTE_growth <- with(df, ave(FTE, ID_Code,
                              FUN=function(x) c(NA, diff(x)) ))

#creation a Turnover growth column
df$Turnover_growth <- with(df, ave(Turnover, ID_Code,
```

```
                                FUN=function(x) c(NA, diff(x)) ))

      # To convert data stored in lines into columns we use means per lines
(per year) for all time-sensitive variables
      df_Turnover = df %>%  group_by(ID_Code, Year) %>% summarise(MT =
mean(Turnover, na.rm = T)) %>% spread(Year, MT) #For turnover
      df_Turnover_growth = df %>% group_by(ID_Code, Year) %>%
summarise(MTurnover_growth = mean(Turnover_growth, na.rm = T))  %>%
spread(Year, MTurnover_growth) #For FTE_growth
      df_surv = df %>% group_by(ID_Code, Year) %>% summarise(MS =
mean(survival, na.rm = T))  %>% spread(Year, MS) #For survival
      df_FTE = df %>% group_by(ID_Code, Year) %>% summarise(MFTE =
mean(FTE, na.rm = T))  %>% spread(Year, MFTE) #For FTE
      df_FTE_growth = df %>% group_by(ID_Code, Year) %>%
summarise(MFTE_growth = mean(FTE_growth, na.rm = T))  %>% spread(Year,
MFTE_growth) #For FTE_growth
      df_netoperativeincome = df %>% group_by(ID_Code, Year) %>%
summarise(MNOI = mean(as.numeric(NetOperativeIncome), na.rm = T))  %>%
spread(Year, MNOI) #For Net Operative Income

      #There is a column #NA that is due to some data which is not tied
with a specific year. To facilitate the analysis, they will be merged with
2019 since they never overlap
      df_Turnover$`2019` <- coalesce(df_Turnover$`<NA>`,df_Turnover$`2019`)
      df_Turnover$`<NA>`<- NULL

      df_Turnover_growth$`2019` <-
coalesce(df_Turnover_growth$`<NA>`,df_Turnover_growth$`2019`)
      df_Turnover_growth$`<NA>`<- NULL

      df_FTE$`2019` <- coalesce(df_FTE$`<NA>`,df_FTE$`2019`)
      df_FTE$`<NA>`<- NULL

      df_FTE_growth$`2019` <-
coalesce(df_FTE_growth$`<NA>`,df_FTE_growth$`2019`)
      df_FTE_growth$`<NA>`<- NULL

      df_surv$`2019` <- coalesce(df_surv$`<NA>`,df_surv$`2019`)
      df_surv$`<NA>`<- NULL

      df_netoperativeincome$`2019` <-
coalesce(df_netoperativeincome$`<NA>`,df_netoperativeincome$`2019`)
      df_netoperativeincome$`<NA>`<- NULL



      # The created colomns are added to the existing original table in a
new dataframe
      newdf = df %>%
        left_join(df_Turnover, by = c("ID_Code", "ID_Code")) %>%
        left_join(df_surv, by = c("ID_Code", "ID_Code")) %>%
        left_join(df_FTE, by = c("ID_Code", "ID_Code")) %>%
        left_join(df_netoperativeincome, by = c("ID_Code", "ID_Code"))

      rm(df_FTE)
      rm(df_surv)
      rm(df_Turnover)
```

```
rm(df_Turnover_growth)
rm(df_netoperativeincome)

# New added columns in 2008.x format are renamed
names(newdf)[135]="Turnover_2001"
names(newdf)[136]="Turnover_2002"
names(newdf)[137]="Turnover_2003"
names(newdf)[138]="Turnover_2004"
names(newdf)[139]="Turnover_2005"
names(newdf)[140]="Turnover_2006"
names(newdf)[141]="Turnover_2007"
names(newdf)[142]="Turnover_2008"
names(newdf)[143]="Turnover_2009"
names(newdf)[144]="Turnover_2010"
names(newdf)[145]="Turnover_2011"
names(newdf)[146]="Turnover_2012"
names(newdf)[147]="Turnover_2013"
names(newdf)[148]="Turnover_2014"
names(newdf)[149]="Turnover_2015"
names(newdf)[150]="Turnover_2016"
names(newdf)[151]="Turnover_2017"
names(newdf)[152]="Turnover_2018"
names(newdf)[153]="Turnover_2019"

names(newdf)[154]="survival_2001"
names(newdf)[155]="survival_2002"
names(newdf)[156]="survival_2003"
names(newdf)[157]="survival_2004"
names(newdf)[158]="survival_2005"
names(newdf)[159]="survival_2006"
names(newdf)[160]="survival_2007"
names(newdf)[161]="survival_2008"
names(newdf)[162]="survival_2009"
names(newdf)[163]="survival_2010"
names(newdf)[164]="survival_2011"
names(newdf)[165]="survival_2012"
names(newdf)[166]="survival_2013"
names(newdf)[167]="survival_2014"
names(newdf)[168]="survival_2015"
names(newdf)[169]="survival_2016"
names(newdf)[170]="survival_2017"
names(newdf)[171]="survival_2018"
names(newdf)[172]="survival_2019"

names(newdf)[173]="FTE_2001"
names(newdf)[174]="FTE_2002"
names(newdf)[175]="FTE_2003"
names(newdf)[176]="FTE_2004"
names(newdf)[177]="FTE_2005"
names(newdf)[178]="FTE_2006"
names(newdf)[179]="FTE_2007"
names(newdf)[180]="FTE_2008"
names(newdf)[181]="FTE_2009"
names(newdf)[182]="FTE_2010"
```

```
names(newdf)[183]="FTE_2011"
names(newdf)[184]="FTE_2012"
names(newdf)[185]="FTE_2013"
names(newdf)[186]="FTE_2014"
names(newdf)[187]="FTE_2015"
names(newdf)[188]="FTE_2016"
names(newdf)[189]="FTE_2017"
names(newdf)[190]="FTE_2018"
names(newdf)[191]="FTE_2019"

names(newdf)[192]="NOI_2001"
names(newdf)[193]="NOI_2002"
names(newdf)[194]="NOI_2003"
names(newdf)[195]="NOI_2004"
names(newdf)[196]="NOI_2005"
names(newdf)[197]="NOI_2006"
names(newdf)[198]="NOI_2007"
names(newdf)[199]="NOI_2008"
names(newdf)[200]="NOI_2009"
names(newdf)[201]="NOI_2010"
names(newdf)[202]="NOI_2011"
names(newdf)[203]="NOI_2012"
names(newdf)[204]="NOI_2013"
names(newdf)[205]="NOI_2014"
names(newdf)[206]="NOI_2015"
names(newdf)[207]="NOI_2016"
names(newdf)[208]="NOI_2017"
names(newdf)[209]="NOI_2018"
names(newdf)[210]="NOI_2019"


# We call %not_in% what is the contrary to %in%
`%not_in%` <- purrr::negate(`%in%`)

# We create an empty dataframe for finalisation
finaldf <- structure(list(character()), class = "data.frame")

# With i between 1 and the final line number of the dataframe :
for(i in 1:nrow(newdf)){
  # if the ID_Code is not in the ID_Code colomn of the final
dataframe :
  if (newdf[i, "ID_Code"] %not_in% finaldf$ID_Code){
    print(newdf[i, "ID_Code"])
    # We add the line to the final dataframe
    finaldf <- rbind(finaldf,newdf[i,])
  }
}

rm(newdf)

#Change the "NaN" values that appeared with the analysis to "NA"
values to mute them for the analysis
finaldf<-sapply(finaldf,as.character)
finaldf <- ifelse(finaldf=="NaN",NA,finaldf)
finaldf<-data.frame((finaldf))
```

```
#Now the "Year", "FTE", "NetOperativeIncome", "Turnover" and
"survival" variables is no longer needed nor meaningful since the lines are
no longer time-sensitive
finaldf$Year<-NULL
finaldf$FTE<-NULL
finaldf$NetOperativeIncome<-NULL
finaldf$Turnover<-NULL
finaldf$survival<-NULL

write.csv(finaldf, file = "V11_6.csv")

#converttion of required colomns as numeric
finaldf$Founding_year<-
as.numeric(levels(finaldf$Founding_year))[as.integer(finaldf$Founding_year)
]
finaldf$Entry_year<-
as.numeric(levels(finaldf$Entry_year))[as.integer(finaldf$Entry_year)]
finaldf$Exit_year<-
as.numeric(levels(finaldf$Exit_year))[as.integer(finaldf$Exit_year)]

#Convert to numeric values when required
finaldf$Birthdate_1<-
as.numeric(levels(finaldf$Birthdate_1))[as.integer(finaldf$Birthdate_1)]
finaldf$Birthdate_2<-
as.numeric(levels(finaldf$Birthdate_2))[as.integer(finaldf$Birthdate_2)]
finaldf$Birthdate_3<-
as.numeric(levels(finaldf$Birthdate_3))[as.integer(finaldf$Birthdate_3)]
finaldf$Birthdate_4<-
as.numeric(levels(finaldf$Birthdate_4))[as.integer(finaldf$Birthdate_4)]
finaldf$Birthdate_5<-
as.numeric(levels(finaldf$Birthdate_5))[as.integer(finaldf$Birthdate_5)]
finaldf$Birthdate_6<-
as.numeric(levels(finaldf$Birthdate_6))[as.integer(finaldf$Birthdate_6)]
finaldf$Birthdate_7<-
as.numeric(levels(finaldf$Birthdate_7))[as.integer(finaldf$Birthdate_7)]
finaldf$Birthdate_8<-
as.numeric(levels(finaldf$Birthdate_8))[as.integer(finaldf$Birthdate_8)]
finaldf$Birthdate_9<-
as.numeric(levels(finaldf$Birthdate_9))[as.integer(finaldf$Birthdate_9)]
finaldf$Birthdate_10<-
as.numeric(levels(finaldf$Birthdate_10))[as.integer(finaldf$Birthdate_10)]

finaldf$Gender_1<-
as.numeric(levels(finaldf$Gender_1))[as.integer(finaldf$Gender_1)]
finaldf$Gender_2<-
as.numeric(levels(finaldf$Gender_2))[as.integer(finaldf$Gender_2)]
finaldf$Gender_3<-
as.numeric(levels(finaldf$Gender_3))[as.integer(finaldf$Gender_3)]
finaldf$Gender_4<-
as.numeric(levels(finaldf$Gender_4))[as.integer(finaldf$Gender_4)]
finaldf$Gender_5<-
as.numeric(levels(finaldf$Gender_5))[as.integer(finaldf$Gender_5)]
finaldf$Gender_6<-
as.numeric(levels(finaldf$Gender_6))[as.integer(finaldf$Gender_6)]
finaldf$Gender_7<-
as.numeric(levels(finaldf$Gender_7))[as.integer(finaldf$Gender_7)]
```

```
      finaldf$Gender_8<-
as.numeric(levels(finaldf$Gender_8))[as.integer(finaldf$Gender_8)]
      finaldf$Gender_9<-
as.numeric(levels(finaldf$Gender_9))[as.integer(finaldf$Gender_9)]
      finaldf$Gender_10<-
as.numeric(levels(finaldf$Gender_10))[as.integer(finaldf$Gender_10)]

      finaldf$StudyField_1<-
as.numeric(levels(finaldf$StudyField_1))[as.integer(finaldf$StudyField_1)]
      finaldf$StudyField_2<-
as.numeric(levels(finaldf$StudyField_2))[as.integer(finaldf$StudyField_2)]
      finaldf$StudyField_3<-
as.numeric(levels(finaldf$StudyField_3))[as.integer(finaldf$StudyField_3)]
      finaldf$StudyField_4<-
as.numeric(levels(finaldf$StudyField_4))[as.integer(finaldf$StudyField_4)]
      finaldf$StudyField_5<-
as.numeric(levels(finaldf$StudyField_5))[as.integer(finaldf$StudyField_5)]
      finaldf$StudyField_6<-
as.numeric(levels(finaldf$StudyField_6))[as.integer(finaldf$StudyField_6)]
      finaldf$StudyField_7<-
as.numeric(levels(finaldf$StudyField_7))[as.integer(finaldf$StudyField_7)]
      finaldf$StudyField_8<-
as.numeric(levels(finaldf$StudyField_8))[as.integer(finaldf$StudyField_8)]
      finaldf$StudyField_9<-
as.numeric(levels(finaldf$StudyField_9))[as.integer(finaldf$StudyField_9)]
      finaldf$StudyField_10<-
as.numeric(levels(finaldf$StudyField_10))[as.integer(finaldf$StudyField_10)
]

      finaldf$Edlvl_1<-
as.numeric(levels(finaldf$Edlvl_1))[as.integer(finaldf$Edlvl_1)]
      finaldf$Edlvl_2<-
as.numeric(levels(finaldf$Edlvl_2))[as.integer(finaldf$Edlvl_2)]
      finaldf$Edlvl_3<-
as.numeric(levels(finaldf$Edlvl_3))[as.integer(finaldf$Edlvl_3)]
      finaldf$Edlvl_4<-
as.numeric(levels(finaldf$Edlvl_4))[as.integer(finaldf$Edlvl_4)]
      finaldf$Edlvl_5<-
as.numeric(levels(finaldf$Edlvl_5))[as.integer(finaldf$Edlvl_5)]
      finaldf$Edlvl_6<-
as.numeric(levels(finaldf$Edlvl_6))[as.integer(finaldf$Edlvl_6)]
      finaldf$Edlvl_7<-
as.numeric(levels(finaldf$Edlvl_7))[as.integer(finaldf$Edlvl_7)]
      finaldf$Edlvl_8<-
as.numeric(levels(finaldf$Edlvl_8))[as.integer(finaldf$Edlvl_8)]
      finaldf$Edlvl_9<-
as.numeric(levels(finaldf$Edlvl_9))[as.integer(finaldf$Edlvl_9)]
      finaldf$Edlvl_10<-
as.numeric(levels(finaldf$Edlvl_10))[as.integer(finaldf$Edlvl_10)]

      finaldf$Profession_1<-
as.numeric(levels(finaldf$Profession_1))[as.integer(finaldf$Profession_1)]
      finaldf$Profession_2<-
as.numeric(levels(finaldf$Profession_2))[as.integer(finaldf$Profession_2)]
      finaldf$Profession_3<-
as.numeric(levels(finaldf$Profession_3))[as.integer(finaldf$Profession_3)]
      finaldf$Profession_4<-
as.numeric(levels(finaldf$Profession_4))[as.integer(finaldf$Profession_4)]
```

```
      finaldf$Profession_5<-
as.numeric(levels(finaldf$Profession_5))[as.integer(finaldf$Profession_5)]
      finaldf$Profession_6<-
as.numeric(levels(finaldf$Profession_6))[as.integer(finaldf$Profession_6)]
      finaldf$Profession_7<-
as.numeric(levels(finaldf$Profession_7))[as.integer(finaldf$Profession_7)]
      finaldf$Profession_8<-
as.numeric(levels(finaldf$Profession_8))[as.integer(finaldf$Profession_8)]
      finaldf$Profession_9<-
as.numeric(levels(finaldf$Profession_9))[as.integer(finaldf$Profession_9)]
      finaldf$Profession_10<-
as.numeric(levels(finaldf$Profession_10))[as.integer(finaldf$Profession_10)
]

      finaldf$EntrepExp_1<-
as.numeric(levels(finaldf$EntrepExp_1))[as.integer(finaldf$EntrepExp_1)]
      finaldf$EntrepExp_2<-
as.numeric(levels(finaldf$EntrepExp_2))[as.integer(finaldf$EntrepExp_2)]
      finaldf$EntrepExp_3<-
as.numeric(levels(finaldf$EntrepExp_3))[as.integer(finaldf$EntrepExp_3)]
      finaldf$EntrepExp_4<-
as.numeric(levels(finaldf$EntrepExp_4))[as.integer(finaldf$EntrepExp_4)]
      finaldf$EntrepExp_5<-
as.numeric(levels(finaldf$EntrepExp_5))[as.integer(finaldf$EntrepExp_5)]
      finaldf$EntrepExp_6<-
as.numeric(levels(finaldf$EntrepExp_6))[as.integer(finaldf$EntrepExp_6)]
      finaldf$EntrepExp_7<-
as.numeric(levels(finaldf$EntrepExp_7))[as.integer(finaldf$EntrepExp_7)]
      finaldf$EntrepExp_8<-
as.numeric(levels(finaldf$EntrepExp_8))[as.integer(finaldf$EntrepExp_8)]
      finaldf$EntrepExp_9<-
as.numeric(levels(finaldf$EntrepExp_9))[as.integer(finaldf$EntrepExp_9)]
      finaldf$EntrepExp_10<-
as.numeric(levels(finaldf$EntrepExp_10))[as.integer(finaldf$EntrepExp_10)]

      finaldf$IndusExp_1<-
as.numeric(levels(finaldf$IndusExp_1))[as.integer(finaldf$IndusExp_1)]
      finaldf$IndusExp_2<-
as.numeric(levels(finaldf$IndusExp_2))[as.integer(finaldf$IndusExp_2)]
      finaldf$IndusExp_3<-
as.numeric(levels(finaldf$IndusExp_3))[as.integer(finaldf$IndusExp_3)]
      finaldf$IndusExp_4<-
as.numeric(levels(finaldf$IndusExp_4))[as.integer(finaldf$IndusExp_4)]
      finaldf$IndusExp_5<-
as.numeric(levels(finaldf$IndusExp_5))[as.integer(finaldf$IndusExp_5)]
      finaldf$IndusExp_6<-
as.numeric(levels(finaldf$IndusExp_6))[as.integer(finaldf$IndusExp_6)]
      finaldf$IndusExp_7<-
as.numeric(levels(finaldf$IndusExp_7))[as.integer(finaldf$IndusExp_7)]
      finaldf$IndusExp_8<-
as.numeric(levels(finaldf$IndusExp_8))[as.integer(finaldf$IndusExp_8)]
      finaldf$IndusExp_9<-
as.numeric(levels(finaldf$IndusExp_9))[as.integer(finaldf$IndusExp_9)]
      finaldf$IndusExp_10<-
as.numeric(levels(finaldf$IndusExp_10))[as.integer(finaldf$IndusExp_10)]

      # CALCULATE CONTROL VARIABLES ####

      #clean age-related variables
```

```
	finaldf$years_since_creation <- 2019 - finaldf$Founding_year
#calculate age
	finaldf$age_when_creation <- finaldf$Founding_year -
finaldf$Entry_year #split in a) age when application and...
	finaldf$years_since_application <- 2019 - finaldf$Entry_year #b)
years since application
	finaldf$Inc_period <- finaldf$Exit_year-finaldf$Entry_year


	finaldf$Founding_year <- NULL #remove excess columns
	finaldf$Exit_year <- NULL  #remove excess columns

	#Convert the stage from text to numeric categories
	finaldf$Stage <- as.character(finaldf$Stage)
	finaldf$Stage[finaldf$Stage %in% c("Rejected/Abandonned")]<-1
	finaldf$Stage[finaldf$Stage %in% c("Waiting")]<-2
	finaldf$Stage[finaldf$Stage %in% c("Pre-Incubated")]<-3
	finaldf$Stage[finaldf$Stage %in% c("INCUBATED","Incubated")]<-4
	finaldf$Stage[finaldf$Stage %in% c("Exited")]<-5
	finaldf$Stage <- as.numeric(finaldf$Stage)

	#Cut the NACE classification to facilitate the analysis (from 615 to
21 entries)
	finaldf$NACE <- as.character(finaldf$NACE)
	finaldf$NACE <- substr(finaldf$NACE,1,nchar(finaldf$NACE)-3)
	finaldf$NACE <- as.numeric(finaldf$NACE)

	finaldf$NACE[finaldf$NACE %in% c(01:03)]<-"A"
	finaldf$NACE[finaldf$NACE %in% c(05:09)]<-"B"
	finaldf$NACE[finaldf$NACE %in% c(10:33)]<-"C"
	finaldf$NACE[finaldf$NACE %in% c(35)]<-"D"
	finaldf$NACE[finaldf$NACE %in% c(36:39)]<-"E"
	finaldf$NACE[finaldf$NACE %in% c(41:43)]<-"F"
	finaldf$NACE[finaldf$NACE %in% c(45:47)]<-"G"
	finaldf$NACE[finaldf$NACE %in% c(49:53)]<-"H"
	finaldf$NACE[finaldf$NACE %in% c(55:56)]<-"I"
	finaldf$NACE[finaldf$NACE %in% c(58:63)]<-"J"
	finaldf$NACE[finaldf$NACE %in% c(64:66)]<-"K"
	finaldf$NACE[finaldf$NACE %in% c(68)]<-"L"
	finaldf$NACE[finaldf$NACE %in% c(69:75)]<-"M"
	finaldf$NACE[finaldf$NACE %in% c(77:82)]<-"N"
	finaldf$NACE[finaldf$NACE %in% c(84)]<-"O"
	finaldf$NACE[finaldf$NACE %in% c(85)]<-"P"
	finaldf$NACE[finaldf$NACE %in% c(86:88)]<-"Q"
	finaldf$NACE[finaldf$NACE %in% c(90:93)]<-"R"
	finaldf$NACE[finaldf$NACE %in% c(93:96)]<-"S"
	finaldf$NACE[finaldf$NACE %in% c(97:98)]<-"T"
	finaldf$NACE[finaldf$NACE %in% c(99)]<-"U"

	finaldf$NACE<-as.character(finaldf$NACE)

	#Delete the columns that won't be used
	finaldf[,3:5]<-NULL
	finaldf[,4:24]<-NULL
	finaldf[,44:54]<-NULL
```

```
finaldf[,74:79]<-NULL

#clean entrepreneurial experience variable

finaldf$age_memb1 <- 2019-finaldf$Birthdate_1 #age team members
finaldf$age_memb2 <- 2019-finaldf$Birthdate_2
finaldf$age_memb3 <- 2019-finaldf$Birthdate_3
finaldf$age_memb4 <- 2019-finaldf$Birthdate_4
finaldf$age_memb5 <- 2019-finaldf$Birthdate_5
finaldf$age_memb6 <- 2019-finaldf$Birthdate_6
finaldf$age_memb7 <- 2019-finaldf$Birthdate_7
finaldf$age_memb8 <- 2019-finaldf$Birthdate_8
finaldf$age_memb9 <- 2019-finaldf$Birthdate_9
finaldf$age_memb10 <- 2019-finaldf$Birthdate_10

finaldf$Birthdate_1<-NULL #remove excessive columns
finaldf$Birthdate_2<-NULL
finaldf$Birthdate_3<-NULL
finaldf$Birthdate_4<-NULL
finaldf$Birthdate_5<-NULL
finaldf$Birthdate_6<-NULL
finaldf$Birthdate_7<-NULL
finaldf$Birthdate_8<-NULL
finaldf$Birthdate_9<-NULL
finaldf$Birthdate_10<-NULL

#creation of parallel dataframes to calculate means/sum/Min/Max
Gender <-finaldf[4:13]
Gender$MembCount <- (10-(rowSums(is.na(Gender) | Gender == "" |
Gender == " "))) #Here, use of the Gender Variable to count the number of
members in each founding team
Gender$MembCount[Gender$MembCount==0] <- NA #The 0s must be replaced
by NAs cause start-ups without members doesn't make any sense.
Gender$Gender_Wshare<-rowMeans(Gender[1:10], na.rm = TRUE)
Gender<-round(Gender,3)

Edlvl <-finaldf[24:33]
Edlvl$Edlvl_mean<-rowMeans(Edlvl, na.rm = TRUE)
Edlvl$Edlvl_max<-do.call(pmax, c(Edlvl, na.rm = TRUE))
Edlvl$Edlvl_min<-do.call(pmin, c(Edlvl, na.rm = TRUE))
Edlvl<-round(Edlvl,3)

EntrepExp <-finaldf[44:53]
EntrepExp$EntrepExp_mean<-rowMeans(EntrepExp, na.rm = TRUE)
EntrepExp$EntrepExp_max<-do.call(pmax, c(EntrepExp, na.rm = TRUE))
EntrepExp$EntrepExp_min<-do.call(pmin, c(EntrepExp, na.rm = TRUE))
EntrepExp$EntrepExp_sum<-rowSums(EntrepExp[1:10], na.rm = TRUE)
EntrepExp<-round(EntrepExp,3)

IndusExp <-finaldf[54:63]
IndusExp$IndusExp_mean<-rowMeans(IndusExp, na.rm = TRUE)
IndusExp$IndusExp_max<-do.call(pmax, c(IndusExp, na.rm = TRUE))
IndusExp$IndusExp_min<-do.call(pmin, c(IndusExp, na.rm = TRUE))
IndusExp$IndusExp_sum<-rowSums(IndusExp[1:10], na.rm = TRUE)
IndusExp<-round(IndusExp,3)
```

```
Turnover <-finaldf[77:95]
Turnover <- as.data.frame.matrix(Turnover)
for (i in 1:ncol(Turnover)){
  Turnover = lapply(Turnover,as.numeric)
}
Turnover<-as.data.frame(Turnover)
Turnover$Turnover_mean<-rowMeans(Turnover, na.rm = TRUE)
Turnover$Turnover_max<-do.call(pmax, c(Turnover, na.rm = TRUE))
Turnover$Turnover_min<-do.call(pmin, c(Turnover, na.rm = TRUE))
Turnover<-round(Turnover,3)

survival<-finaldf[96:114]
survival <- as.data.frame.matrix(survival)
for (i in 1:ncol(survival)){
  survival = lapply(survival,as.numeric)
}
survival<-as.data.frame(survival)
survival$survival_sum<-rowSums(survival, na.rm = TRUE)
for (i in 1:nrow(survival)){
  if (is.na(survival$survival_2019[i]) == TRUE){
    survival$survival_sum[i] <- NA
  }
}

FTE <-finaldf[115:133]
FTE <- as.data.frame.matrix(FTE)
for (i in 1:ncol(FTE)){
  FTE = lapply(FTE,as.numeric)
}
FTE<-as.data.frame(FTE)
FTE$FTE_mean<-rowMeans(FTE, na.rm = TRUE)
FTE$FTE_max<-do.call(pmax, c(FTE, na.rm = TRUE))
FTE$FTE_min<-do.call(pmin, c(FTE, na.rm = TRUE))
FTE$FTE_sum<-rowSums(FTE[1:19], na.rm = TRUE)
FTE<-round(FTE,3)

for (i in 1:nrow(FTE)) {
  FTE$First[i]<- as.numeric(min(which(!is.na(FTE[i,]))),na.rm=FALSE)
}

for (i in 1:nrow(FTE)) {
  FTE$First[i]<- as.numeric(FTE[i,FTE$First[i]])
}
#Here we get the first non-NA value to add it later to the FTE_growth
that doesn't take into account the first observation


NOI <-finaldf[134:152]
NOI <- as.data.frame.matrix(NOI)
for (i in 1:ncol(NOI)){
  NOI = lapply(NOI,as.numeric)
}
NOI<-as.data.frame(NOI)
NOI$NOI_mean<-rowMeans(NOI, na.rm = TRUE)
```

```
NOI$NOI_max<-do.call(pmax, c(NOI, na.rm = TRUE))
NOI$NOI_min<-do.call(pmin, c(NOI, na.rm = TRUE))
NOI<-round(NOI,3)

AgeMemb <-finaldf[157:166]
AgeMemb$AgeMemb_mean<-rowMeans(AgeMemb, na.rm = TRUE)
AgeMemb$AgeMemb_max<-do.call(pmax, c(AgeMemb, na.rm = TRUE))
AgeMemb$AgeMemb_min<-do.call(pmin, c(AgeMemb, na.rm = TRUE))
AgeMemb<-round(AgeMemb,3)

#These new tables are added to the existing one
finaldf[4:13]<-NULL #To avoid double columns, we delete the former
ones before joining the new tables
finaldf[14:23]<-NULL
finaldf[24:33]<-NULL
finaldf[24:33]<-NULL
finaldf[37:55]<-NULL
finaldf[37:55]<-NULL
finaldf[37:55]<-NULL
finaldf[37:55]<-NULL
finaldf[41:50]<-NULL

#Replacing 0 by NAs for non-incubated firms for IncSub to avoid bias
between incubated/non-incubated firms
for (i in 1:nrow(finaldf)) {
  if (finaldf$Stage[i]== 1){
    finaldf$Incub_sub[i]<-NA
  }
  if (finaldf$Stage[i]== 2){
    finaldf$Incub_sub[i]<-NA
  }
  if (finaldf$Stage[i]== 3){
    finaldf$Incub_sub[i]<-NA
  }
  if (finaldf$Stage[i]== 4){
    finaldf$Incub_sub[i]<-NA
  }
}


#finaldf$ID_Code<-as.numeric(finaldf$ID_Code)
df_FTE_growth$ID_Code<-as.factor(df_FTE_growth$ID_Code)

finaldf<-cbind(finaldf,Gender)
finaldf<-cbind(finaldf,AgeMemb)
finaldf<-cbind(finaldf,Edlvl)
finaldf<-cbind(finaldf,EntrepExp)
finaldf<-cbind(finaldf,IndusExp)
finaldf<-cbind(finaldf,survival)
finaldf<-cbind(finaldf,FTE)
finaldf = finaldf %>%
  left_join(df_FTE_growth, by = "ID_Code")
finaldf$FTE_AvRate<-rowMeans(finaldf[150:169], na.rm = TRUE)
#Creation of a first dependant variable "growth" based on FTE
finaldf[151:169]<-NULL
```

```
finaldf<-cbind(finaldf,Turnover)
finaldf<-cbind(finaldf,NOI)

for (i in 1:nrow(finaldf)){
  if (is.na(finaldf$FTE_min[i]) == TRUE){
    finaldf$FTE_AvRate[i] <- NA
    finaldf$FTE_sum[i] <- NA
  }
}


rm(Gender)
rm(AgeMemb)
rm(Edlvl)
rm(EntrepExp)
rm(IndusExp)
rm(survival)
rm(FTE)
rm(df_FTE_growth)
rm(Turnover)
rm(NOI)
finaldf[35:36]<-NULL
finaldf$First<-NULL

#Reduction of the diploma classification from 6 to 3 categories

finaldf<- finaldf %>%
  mutate(Edlvl_max = case_when(Edlvl_max == 3 ~ 1,#Vocational
education
                               Edlvl_max == 4 ~ 1,#Baccalaureate
                               Edlvl_max == 5 ~  1,#Higher Technician
diploma
                               Edlvl_max == 6 ~  2,#Bachelor
                               Edlvl_max == 7 ~ 2,#Master
                               Edlvl_max == 8 ~ 2,#PhD
                               TRUE ~ NA_real_))

finaldf<- finaldf %>%
  mutate(Edlvl_min = case_when(Edlvl_min == 3 ~ 1,
                               Edlvl_min == 4 ~ 1,
                               Edlvl_min == 5 ~  2,
                               Edlvl_min == 6 ~  2,
                               Edlvl_min == 7 ~ 2,
                               Edlvl_min == 8 ~ 2,
                               TRUE ~ NA_real_))

#Survival's Hazard cox model

library(survival)
library(survminer)

Survivaldf<-df
Survivaldf[4:6]<-NULL
Survivaldf[5:128]<-NULL
Survivaldf[6:7]<-NULL
```

```
Survivaldf$Age<-(Survivaldf$Year-Survivaldf$Entry_year)


res.cox <- coxph(Surv(Age, survival) ~ Entry_year, data = Survivaldf)
res.cox
summary(res.cox)


ggsurvplot(survfit(res.cox, data=Survivaldf), palette = "#2E9FDF",
           ggtheme = theme_minimal())

fit <- survfit(Surv(Age, survival) ~ Entry_year, data = Survivaldf)

ggsurvplot(fit,data=Survivaldf,
legend.labs=c(2001,2002,2003,2004,2005,2006,2007,2008,2009,2010,2011,2012,2
013,2014,2015,2016,2017,2018,2019),
           ggtheme = theme_minimal())

fit
summary(fit)
summary(fit)$table
d<-summary(fit)$table
write.csv2(d,file = "sum_surv.csv")

#Standartisation of firm's survival performance based on the
expectancy estimated by Cox's model
finaldf$ExpectLife <- finaldf$Entry_year
finaldf<- finaldf %>%
  mutate(ExpectLife = case_when(ExpectLife == 2001 ~ 7.762134,
                               ExpectLife ==2002 ~  6.832619,
                               ExpectLife ==2003 ~ 6.998683,
                               ExpectLife ==2004 ~ 6.910128,
                               ExpectLife ==2005 ~ 6.232558,
                               ExpectLife ==2006 ~ 5.890909,
                               ExpectLife ==2007 ~ 5.713298,
                               ExpectLife ==2008 ~ 5.411591,
                               ExpectLife ==2009 ~ 5.001617,
                               ExpectLife ==2010 ~ 4.581357,
                               ExpectLife ==2011 ~ 4.012346,
                               ExpectLife ==2012 ~ 3.500000,
                               ExpectLife ==2013 ~ 3.161985,
                               ExpectLife ==2014 ~ 2.855439,
                               ExpectLife ==2015 ~ 2.420393,
                               ExpectLife ==2016 ~ 1.573770,
                               ExpectLife ==2017 ~ 1.137931,
                               ExpectLife ==2018 ~ 0.500000,
                               ExpectLife ==2019 ~ 0,
                         TRUE ~ NA_real_))

for (i in 1:nrow(finaldf)) {
  finaldf$SurvPerf[i]<-finaldf$survival_sum[i]-finaldf$ExpectLife[i]
}

finaldf$Surv_SE <- finaldf$Entry_year
```

```
finaldf<- finaldf %>%
  mutate(Surv_SE = case_when(Surv_SE == 2001 ~ 0.2433560,
                             Surv_SE ==2002 ~  0.2601362,
                             Surv_SE ==2003 ~ 0.2150579,
                             Surv_SE ==2004 ~ 0.2378249,
                             Surv_SE ==2005 ~ 0.2686495,
                             Surv_SE ==2006 ~ 0.2952364,
                             Surv_SE ==2007 ~ 0.3019228,
                             Surv_SE ==2008 ~ 0.3046623,
                             Surv_SE ==2009 ~ 0.2327854,
                             Surv_SE ==2010 ~ 0.2249301,
                             Surv_SE ==2011 ~ 0.2892653,
                             Surv_SE ==2012 ~ 0.2561738,
                             Surv_SE ==2013 ~ 0.1828031,
                             Surv_SE ==2014 ~ 0.2953487,
                             Surv_SE ==2015 ~ 0.2581040,
                             Surv_SE ==2016 ~ 0.1399870,
                             Surv_SE ==2017 ~ 0.1440384,
                             Surv_SE ==2018 ~ 0.1178511,
                             Surv_SE ==2019 ~ 1,
                                TRUE ~ NA_real_))
for (i in 1:nrow(finaldf)) {
  finaldf$SurvPerf[i]<-finaldf$survival_sum[i]/finaldf$Surv_SE[i]
}

finaldf$ExpectLife<-NULL
finaldf$Surv_SE<-NULL


#Change (again) the "NaN" values that appeared with the analysis to
"NA" values to mute them
finaldf<-sapply(finaldf,as.character)
finaldf <- ifelse(finaldf=="NaN",NA,finaldf)
finaldf<-data.frame((finaldf))

colnames(finaldf)[34] <- "Share_K" #To prevent confusion, "Capital
social" is renamed "Share Capital"

# CORRELATIONS AND DESCRIPTIVES ####

#select columns for correlation table
vars <-
finaldf[c("Prodtype","Stage","UI","NoTrack","Incub_sub","Share_K","years_si
nce_creation",
                   "age_when_creation","years_since_application",
"Entry_year","Inc_period","Gender_Wshare","AgeMemb_mean","MembCount",

"Edlvl_max","Edlvl_min","EntrepExp_mean","EntrepExp_max","EntrepExp_min","I
ndusExp_mean",

"IndusExp_max","IndusExp_min","Turnover_mean","Turnover_max","Turnover_min"
,"survival_2019","SurvPerf",
                   "FTE_sum","FTE_mean","FTE_max","FTE_min","FTE_AvRate")]


vars <- as.data.frame.matrix(vars)
```

```
for (i in 1:ncol(vars)){
  vars = lapply(vars,as.numeric)
}
vars<-as.data.frame(vars)



cors <- as.data.frame(round(cor(x = vars, method = "spearman", use =
"pairwise.complete.obs"), 3))
#View(cors)
write.csv2(cors, paste(Sys.Date(),"_cors.csv"), row.names = FALSE)

desc <- data.frame(min = round(apply(vars, MARGIN = 2, function(x)
min(x, na.rm = T)),3),
                   max = round(apply(vars, MARGIN = 2, function(x)
max(x, na.rm = T)),3),
                   mean = round(apply(vars, MARGIN = 2, function(x)
mean(x, na.rm = T)),3),
                   sd = round(apply(vars, MARGIN = 2, function(x)
sd(x, na.rm = T)),3),
                   median = round(apply(vars, MARGIN = 2, function(x)
median(x, na.rm = T)),3),
                   quantiles = t(round(apply(vars, MARGIN = 2,
function(x) quantile(x, na.rm = T)),3)),
                   missing = round(apply(vars, MARGIN = 2,
function(x) missing(x)),3))
write.csv2(desc, file = paste(Sys.Date(), "desc.csv", sep = "_"),
row.names = T)
rm(vars)



# RUN MODELS ####


# PREPARING MODELS AND PACKAGES####
library(stargazer)
library(psych)
library(car)
library(mitml)
library(ordinal)
library(MASS)
library(pscl)
library(brant)
library(lmtest)

##CHECK THE IMPACT OF APPLICATION YEAR ON DEPENDENT VARIABLE

#survival
mylm_surv1 <- with(finaldf, glm(as.numeric(SurvPerf) ~ Entry_year,
family = "gaussian"))

#Size
mylm_size1 <- with(finaldf, glm(as.numeric(FTE_mean) ~  Entry_year,
family = "gaussian"))
```

```
#growth
mylm_growth1 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
Entry_year, family = "gaussian"))

#Turnover
mylm_turn1 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
Entry_year, family = "gaussian"))


stargazer(mylm_surv1, mylm_size1, mylm_growth1, mylm_turn1,
          star.char = c("*", "**", "***"),
          star.cutoffs = c(0.05, 0.01, 0.001),
          type="text")

summary(mylm_surv1)
summary(mylm_size1)
summary(mylm_growth1)
summary(mylm_turn1)

vif(mylm_surv1)
vif(mylm_size1)
vif(mylm_growth1)
vif(mylm_turn1)

#check residuals
plot(mylm_surv1)
plot(mylm_size1)
plot(mylm_growth1)
plot(mylm_turn1)

R2s <- round(c(pR2(mylm_surv1)[[4]],
               pR2(mylm_size1)[[4]],
               pR2(mylm_growth1)[[4]],
               pR2(mylm_turn1)[[4]]), 3)

stargazer(mylm_surv1, mylm_size1, mylm_growth1, mylm_turn1,
          star.char = c("*", "**", "***"),
          star.cutoffs = c(0.05, 0.01, 0.001),
          add.lines = list(c("McFadden R2",R2s)),
          type="text",out="Stargazer_YEARS.htm")

# test for homoscedacity
bptest(mylm_surv1)
bptest(mylm_size1)
bptest(mylm_growth1)
bptest(mylm_turn1)


## HYPOTHESIS 1 ####INCUBATION ASSESSMENT #SURVIVAL ANALYSIS

Survivaldf$Stage <- as.character(Survivaldf$Stage)
Survivaldf$Stage[Survivaldf$Stage %in% c("Rejected/Abandonned")]<-1
Survivaldf$Stage[Survivaldf$Stage %in% c("Waiting")]<-2
Survivaldf$Stage[Survivaldf$Stage %in% c("Pre-Incubated")]<-3
Survivaldf$Stage[Survivaldf$Stage %in% c("INCUBATED","Incubated")]<-4
```

```
	Survivaldf$Stage[Survivaldf$Stage %in% c("Exited")]<-5
	Survivaldf$Stage <- as.numeric(Survivaldf$Stage)


	res.cox <- coxph(Surv(Age, survival) ~ as.factor(Stage), data =
Survivaldf)
	res.cox
	summary(res.cox)

	fit2 <- survfit(Surv(Age, survival) ~ Stage, data = Survivaldf)

	fit2
	summary(fit2)
	summary(fit2)$table
	dBis<-summary(fit2)$table
	write.csv2(dBis,file = "Surv_H1.csv")


	## MODELS TESTING HYPOTHESIS 2 ####Human Capital

	#survival
	mylm_surv1 <- with(finaldf, glm(as.numeric(SurvPerf) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount), family = "gaussian"))
	mylm_surv2 <- with(finaldf, glm(as.numeric(SurvPerf) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min, family = "gaussian"))
	mylm_surv3 <- with(finaldf, glm(as.numeric(SurvPerf) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min + as.numeric(IndusExp_mean) +
as.numeric(EntrepExp_mean), family = "gaussian"))

	#Size
	mylm_size1 <- with(finaldf, glm(as.integer(FTE_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount), family = "gaussian"))
	mylm_size2 <- with(finaldf, glm(as.integer(FTE_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min, family = "gaussian"))
	mylm_size3 <- with(finaldf, glm(as.integer(FTE_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min + as.numeric(IndusExp_mean) +
as.numeric(EntrepExp_mean), family = "gaussian"))

	#growth
	mylm_growth1 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount), family = "gaussian"))
	mylm_growth2 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min, family = "gaussian"))
	mylm_growth3 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min + as.numeric(IndusExp_mean) +
as.numeric(EntrepExp_mean), family = "gaussian"))

	#Turnover
```

```
    mylm_turn1 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount), family = "gaussian"))
    mylm_turn2 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min, family = "gaussian" ))
    mylm_turn3 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
as.numeric(Gender_Wshare) + as.numeric(AgeMemb_mean) +
as.integer(MembCount) + Edlvl_min + as.numeric(IndusExp_mean) +
as.numeric(EntrepExp_mean), family = "gaussian"))


    stargazer(mylm_surv1, mylm_surv2, mylm_surv3, mylm_size1, mylm_size2,
mylm_size3, mylm_growth1, mylm_growth2, mylm_growth3, mylm_turn1,
mylm_turn2, mylm_turn3,
              star.cutoffs = c(0.05, 0.01, 0.001),
              type="text")

    summary(mylm_surv3)
    summary(mylm_size3)
    summary(mylm_growth3)
    summary(mylm_turn3)

    vif(mylm_surv3)
    vif(mylm_size3)
    vif(mylm_growth3)
    vif(mylm_turn3)

    #check residuals
    plot(mylm_surv3)
    plot(mylm_size3)
    plot(mylm_growth3)
    plot(mylm_turn3)

    R2s <- round(c(pR2(mylm_surv1)[[4]], pR2(mylm_surv2)[[4]],
pR2(mylm_surv3)[[4]],
                   pR2(mylm_size1)[[4]], pR2(mylm_size2)[[4]],
pR2(mylm_size3)[[4]],
                   pR2(mylm_growth1)[[4]], pR2(mylm_growth2)[[4]],
pR2(mylm_growth3)[[4]],
                   pR2(mylm_turn1)[[4]], pR2(mylm_turn2)[[4]],
pR2(mylm_turn3)[[4]]), 3)

    stargazer(mylm_surv1, mylm_surv2, mylm_surv3, mylm_size1, mylm_size2,
mylm_size3, mylm_growth1, mylm_growth2, mylm_growth3, mylm_turn1,
mylm_turn2, mylm_turn3,
              star.char = c("*", "**", "***"),
               star.cutoffs = c(0.05, 0.01, 0.001),
               add.lines = list(c("McFadden R2",R2s)),
               type="text",out="Stargazer_H2.htm")


    # test for homoscedacity
    bptest(mylm_surv3)
    bptest(mylm_size3)
    bptest(mylm_growth3)
    bptest(mylm_turn3)
```

```
## MODELS TESTING HYPOTHESIS 3 ####Knowledge Capital
#survival
mylm_surv1 <- with(finaldf, glm(as.numeric(SurvPerf) ~  Prodtype + NoTrack, family = "gaussian"))
mylm_surv2 <- with(finaldf, glm(as.numeric(SurvPerf) ~  Prodtype + NoTrack + NACE, family = "gaussian"))

#Size
mylm_size1 <- with(finaldf, glm(as.integer(FTE_mean) ~  Prodtype + NoTrack, family = "gaussian"))
mylm_size2 <- with(finaldf, glm(as.integer(FTE_mean) ~  Prodtype + NoTrack + NACE, family = "gaussian"))

#growth
mylm_growth1 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~  Prodtype + NoTrack, family = "gaussian"))
mylm_growth2 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~  Prodtype + NoTrack + NACE, family = "gaussian"))

#Turnover
mylm_turn1 <- with(finaldf, glm(as.numeric(Turnover_mean) ~  Prodtype + NoTrack, family = "gaussian"))
mylm_turn2 <- with(finaldf, glm(as.numeric(Turnover_mean) ~  Prodtype + NoTrack + NACE, family = "gaussian"))



stargazer(mylm_surv1, mylm_surv2, mylm_size1, mylm_size2, mylm_growth1, mylm_growth2, mylm_turn1, mylm_turn2,
           star.char = c("*", "**", "***"),
           star.cutoffs = c(0.05, 0.01, 0.001),
           type="text")

summary(mylm_surv2)
summary(mylm_size2)
summary(mylm_growth2)
summary(mylm_turn2)

vif(mylm_surv2)
vif(mylm_size2)
vif(mylm_growth2)
vif(mylm_turn2)

#check residuals
plot(mylm_surv2)
plot(mylm_size2)
plot(mylm_growth2)
plot(mylm_turn2)

R2s <- round(c(pR2(mylm_surv1)[[4]], pR2(mylm_surv2)[[4]],
               pR2(mylm_size1)[[4]], pR2(mylm_size2)[[4]],
               pR2(mylm_growth1)[[4]], pR2(mylm_growth2)[[4]],
               pR2(mylm_turn1)[[4]], pR2(mylm_turn2)[[4]]), 3)
```

```
	# stargazer(mylm_surv1, mylm_surv2, mylm_size1, mylm_size2,
mylm_growth1, mylm_growth2, mylm_turn1, mylm_turn2,
	#           star.char = c("*", "**", "***"),
	#           star.cutoffs = c(0.05, 0.01, 0.001),
	#           add.lines = list(c("McFadden R2",R2s)),
	#           type="text",out="Stargazer_H3.htm")


	# test for homoscedacity
	bptest(mylm_size2)
	bptest(mylm_growth2)
	bptest(mylm_turn2)

	## MODELS TESTING HYPOTHESIS 4 ####Financial capital
	#survival
	mylm_surv1 <- with(finaldf, glm(as.numeric(SurvPerf) ~
as.numeric(Share_K), family = "gaussian"))
	mylm_surv2 <- with(finaldf, glm(as.numeric(SurvPerf) ~
as.numeric(Share_K) + as.numeric(Incub_sub), family = "gaussian"))

	#Size
	mylm_size1 <- with(finaldf, glm(as.integer(FTE_mean) ~
as.numeric(Share_K), family = "gaussian"))
	mylm_size2 <- with(finaldf, glm(as.integer(FTE_mean) ~
as.numeric(Share_K) + as.numeric(Incub_sub), family = "gaussian"))

	#growth
	mylm_growth1 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
as.numeric(Share_K), family = "gaussian"))
	mylm_growth2 <- with(finaldf, glm(as.numeric(FTE_AvRate) ~
as.numeric(Share_K) + as.numeric(Incub_sub), family = "gaussian"))

	#Turnover
	mylm_turn1 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
as.numeric(Share_K), family = "gaussian"))
	mylm_turn2 <- with(finaldf, glm(as.numeric(Turnover_mean) ~
as.numeric(Share_K) + as.numeric(Incub_sub), family = "gaussian"))



	stargazer(mylm_surv1, mylm_surv2, mylm_size1, mylm_size2,
mylm_growth1, mylm_growth2, mylm_turn1, mylm_turn2,
	          star.char = c("*", "**", "***"),
	          star.cutoffs = c(0.05, 0.01, 0.001),
	          type="text")

	summary(mylm_surv2)
	summary(mylm_size2)
	summary(mylm_growth2)
	summary(mylm_turn2)

	vif(mylm_surv2)
	vif(mylm_size2)
	vif(mylm_growth2)
	vif(mylm_turn2)
```

```
#check residuals
plot(mylm_surv2)
plot(mylm_size2)
plot(mylm_growth2)
plot(mylm_turn2)

R2s <- round(c(pR2(mylm_surv1)[[4]], pR2(mylm_surv2)[[4]],
               pR2(mylm_size1)[[4]], pR2(mylm_size2)[[4]],
               pR2(mylm_growth1)[[4]], pR2(mylm_growth2)[[4]],
               pR2(mylm_turn1)[[4]], pR2(mylm_turn2)[[4]]), 3)

stargazer(mylm_surv1, mylm_surv2, mylm_size1, mylm_size2,
mylm_growth1, mylm_growth2, mylm_turn1, mylm_turn2,
           star.char = c("*", "**", "***"),
           star.cutoffs = c(0.05, 0.01, 0.001),
           add.lines = list(c("McFadden R2",R2s)),
           type="text",out="Stargazer_H4.htm")

# test for homoscedacity
bptest(mylm_size2)
bptest(mylm_growth2)
bptest(mylm_turn2)
```