López Puga, Jorge; García García, Juan
A Comparative Study on Entrepreneurial Attitudes Modeled with Logistic Regression and Bayes Nets
Universidad Complutense de Madrid
Madrid, España

Available in: http://www.redalyc.org/articulo.oa?id=17224489026

The Spanish Journal of Psychology,
ISSN (Printed Version): 1138-7416
psvjour@sis.ucm.es
Universidad Complutense de Madrid
España
Entrepreneurship research is receiving increasing attention in our context, as entrepreneurs are key social agents involved in economic development. We compare the success of the dichotomic logistic regression model and the Bayes simple classifier to predict entrepreneurship, after manipulating the percentage of missing data and the level of categorization in predictors. A sample of undergraduate university students (N = 1230) completed five scales (motivation, attitude towards business creation, obstacles, deficiencies, and training needs) and we found that each of them predicted different aspects of the tendency to business creation. Additionally, our results show that the receiver operating characteristic (ROC) curve is affected by the rate of missing data in both techniques, but logistic regression seems to be more vulnerable when faced with missing data, whereas Bayes nets underperform slightly when categorization has been manipulated. Our study sheds light on the potential entrepreneur profile and we propose to use Bayesian networks as an additional alternative to overcome the weaknesses of logistic regression when missing data are present in applied research.

Keywords: entrepreneurship, logistic regression, Bayes nets.
The study of entrepreneurs has challenged the social sciences for the past decades, just as subatomic particles have confused physicists, because, as noted by Rogoff and Lee (1996), “its impact is observable, but the thing itself seems ephemeral and invisible” (p. 1). Despite several decades of enthusiastic research of the topic, a precise and broadly accepted definition of entrepreneurs, how they develop or what effects they have on the business network has not been established. However, the nature of the entrepreneur seems to be crucial to promote the processes of economic and social development in contemporary societies. As noted by Shapero (1985) almost three decades ago, the phenomenon of the “entrepreneurship provides communities with the diversity and dynamism that not only assures continuous development, but also an environment in which personal freedom and individual rights can flourish” (p. 5).

In this work, we shall try to identify the elements associated with attitudes towards business creation, using two statistical techniques: logistic regression and Bayes nets. The contributions of this work can therefore be interpreted in two senses. On the one hand, we shall deepen our substantive knowledge of attitudes towards business creation and, regarding the methodological aspect, we shall compare the ability of two statistical techniques to predict entrepreneurship. With regard to our first goal, we note that the study of entrepreneurship has recently become particularly relevant, in view of the socioeconomic consequences involved (Corman, Lussier, & Nolan, 1996). More specifically, the study of entrepreneurship or entrepreneur tendencies in the university is relevant from a practical and theoretical viewpoint (Cano, García, & Gea, 2003; Ruiz, Rojas, & Suárez, 2008). From the theoretical viewpoint, the study of entrepreneurship in university students may shed light on the nature of the potential entrepreneur (Fuller-Love, 2006; Huefner, Hunt, & Robinson, 1996; Krueger & Brazeal, 1994; López, García, Cano, Gea, & De la Fuente, 2010); whereas from a practical viewpoint, intervention programs could be implemented to optimize the possibilities of success in businesses created within the universities. With regard to our second goal, the comparison of the techniques that are the target of the present investigation has not been extensively studied in our context, and it is appropriate to determine their performance conditions in order to provide decision criteria for investigations, either theoretical or applied (López & García, 2011).

**Entrepreneurial Activity**

Scientific interest in the tendency to create businesses goes back to the 18th and 19th centuries, although there are many gaps and there is some controversy about its conceptualization (i.e., Cano et al., 2003; Rogoff & Lee, 1996). In general, one could say that the concept of entrepreneur or entrepreneurship originates in France. In fact, the English word coined to refer to this phenomenon (entrepreneur) comes from French. However, as noted by Hébert and Link (1989), the taxonomy of the theories referring to the entrepreneurial phenomenon originate with the Irish economist Richard Cantillon, who spent a large part of his life in France. Cantillon gained most of his reputation thanks to his work *Essai sur la nature du commerce en général*, which remained unknown for almost a century and was accidently discovered by William Stanley Jevons (Hayek, 1985). The economic model proposed by Cantillon identifies three main agents that form a relevance hierarchy: (a) the *landowners*, who are economically independent; (b) the entrepreneurs, who fool each other in risky market exchanges with the intention of gaining benefit; and (c) the *hirelings*, who are persuaded by economic funding to carry out unattractive or unpleasant work and who avoid active decision-making in order to ensure contractual guarantees of stable income. Although Cantillon proposed that the top of the hierarchy belonged to the landowners, the entrepreneurs played an essential role in economic development. For him, entrepreneurs are economic agents who gain benefits from business judgments under uncertainty and not by means of revenue or salaries. Later, other authors have followed this idea of considering the entrepreneur as a social being who defies risk and uncertainty (i.e., Hayek, 1985; Hébert & Link, 1989; Polopolus & Emerson, 1991; Samuelson, 1970).

Many definitions have been proposed to clarify the concept of entrepreneur, which have shed light on this object of study. Nevertheless, the purpose and content of the definitions of entrepreneur are very heterogeneous. For example, from a historical-synthetic viewpoint, Hébert and Link (1989) indicate that the entrepreneur is “someone who specializes in taking responsibility for and making judgmental decisions that affect the location, the form, and the use of goods, resources, or institutions” (p. 39); whereas from a business viewpoint (Genesca & Veciana, 1984; Veciana, 1989), it has been noted that an entrepreneur is “the person who has created an ongoing business which survives over the years” (pp. 147 and 19, respectively). Likewise, Steensma, Marino, and Weaver (2000) consider that the entrepreneur holds more than 50% of the property of the business and has managed it actively. Gartner (1988), focusing on the entrepreneur’s function, defined him as “a role that individuals undertake to create organizations” (p. 30); whereas from a practical viewpoint centered on the operationalization of the phenomenon, Huefner et al. (1996) defined him as “those who said they were entrepreneurs and had owned and managed one or more businesses” (p. 62). In any event, the figure of the entrepreneur corresponds to a person who seeks benefit in return for his entrepreneur activity, in contrast to the director or executive, who seeks a salary. Or, in more general terms, the entrepreneur is the person or group that seeks to exploit an economic
opportunity (McKenzie, Ugbah, & Smothers, 2007). Entrepreneurs embark upon new activities, whereas the goal of managers or executives is to keep a previously established business running. As stated by Samuelson (1970), the most characteristic traits of entrepreneurs are having long-term vision, being original, being brave, and introducing, rather than discovering, new products or services.

The trait model is probably the most successful within the study of attitudes towards business creation and the one that has produced the most works. From this perspective, people who create a business are characterized by a series of personality traits or a fixed state of concrete existence and, despite the criticisms (i.e., Bird, 1988; Gartner, 1985, 1988), this model has been the most successful theoretical psychosocial perspective until the mid 1980s to study the entrepreneur’s figure (i.e., Fuller-Love, 2006; McKenzie et al., 2007; Thompson, 2004). For example, McClelland (1955, 1961) noted that certain psychological traits such as achievement motivation, understood as the desire to attain a goal or objective, are determinants of the economic dynamics of countries. Traits like achievement motivation, intelligence, uncertainty tolerance, optimism, self-efficacy, or locus of control have been associated with entrepreneurs (i.e., Gottfredson, 1998; Huefner et al., 1996; Sánchez, 2003; Stanworth, Stanworth, Granger, & Blyth, 1989, Veciana, 1989). However, according to Robinson, Stimpson, Huefner, and Hunt (1991), this model of entrepreneurship has been gradually losing its strength, for at least four reasons. Firstly, because methodologies based on personality traits were not adapted to the specific context of business creation because they used generic scales. On the other hand, neither were personality theories adapted for use in the sphere of business creation. Thirdly, there were problems with the convergent validity indexes because different instruments that measured the same construct had low correlations with each other. Lastly, the emphasis during that epoch on the interactive theories to explain human behavior within the psychosocial sphere weakened the trait model in business creation.

A new perspective that originated within the context of psychosocial research has overcome the weaknesses of the trait model, considering entrepreneur activity a conscious and intentional process mediated by personal factors that interact with the social environment (i.e., Bird, 1988; Krueger & Brazeal, 1994; Krueger & Carsrud, 1993; Krueger, Reilly, & Carsrud, 2000; Liñán, Battistelli, & Moriano, 2008; Moriano, Gómez, Laguna, & Roznowsky, 2008). As noted by Licht and Siegel (2006), the origin of this perspective is found in the pioneer work of Shapero and Sokol (1982), which emphasized that the values system fostered by society conditions business creation culturally. For Shapero and Sokol (1982), human behavior is guided by some kind of inertia that keeps people functioning normally. However, at certain times of the life span, this balance is upset or skewed by factors that force the individual to move towards, and choose between, different behavioral alternatives (Shapero, 1975). In Shapero and Sokol’s model of the “entrepreneurial event,” the chosen course of action depends on the relative credibility of the alternative behaviors as well as on some predisposition to act. This credibility of the creation of a business will depend on its desirability and viability (Shapero & Sokol, 1982). About a decade later, Krueger and Carsrud (1993) and Krueger and Brazeal (1994) merged the entrepreneur event model with the theory of reasoned action (Ajzen & Fishbein, 1980; Fishbein & Ajzen, 1975) to explain both business creation and entrepreneurship within organizations. Their contribution was to consider the phenomenon of business creation an intentional process mediated by the subjective perceptions of the potential entrepreneur. Within this context, the credibility of creating a business will depend on two aspects: (a) the attitude towards the behavior and the social norms (perceived desirability), and (b) the perception of control over the initiative, or self-efficacy (perceived viability). These elements, in interaction with the proclivity to act and the influence of some triggering event, would condition entrepreneurship (Krueger & Brazeal, 1994).

**Logistic Regression versus Bayes Nets**

Both Bayes nets and logistic regression are statistical techniques aimed at modeling probabilistic relations between variables. Nevertheless, there are formal differences between them that make them qualitatively different, depending on the application context. The origin of logistic regression lies in the decade of the 1960s and it emerged as an alternative to the *ordinary least squares* (OLS) estimation procedure that was traditionally used in the linear regression model (Long, 1997; Ortega & Cayuela, 2002). The goal pursued was to estimate the probability of occurrence of an event as a function of a series of independent variables (Walker & Duncan, 1967). It became popular during the 1970s when it began to be used in computer programs for statistical calculus (Peng & So, 2002). Bayes nets (also known as probabilistic causal networks, expert Bayes systems, expert probabilistic systems, causal networks, belief networks, or influence diagrams) are statistical tools that belong to a set of techniques aimed at graphic modeling (Martinez & Rodriguez, 2003) and are part of the family of *highly structured stochastic systems* (Cowell, Dawid, Lauritzen, & Spiegelhalter, 1999). Following Kadie, Hovel, and Horvitz (2001), one could say that a Bayes net is a set of variables, a graphic structure connecting these variables, and a set of distributions of conditional probability. The net represents the uncertainty associated with each variable by means of probabilities, and this uncertainty is susceptible to modification based on observations (or evidence) of the model, using Bayes’ theorem. Although Bayes nets have been used in psychology as formal normative models to model psychological processes such as learning-general
reasoning and, more particularly, causal reasoning (i.e., Conati, Gertner, & VanLehn, 2002; Conejo et al., 2001; Conejo, Millán, Pérez-de-la-Cruz, & Trela, 2004; Gopnik & Schulz, 2004; Gopnik, et al., 2004; Gopnik, Sobel, Schulz, & Glymour, 2001; Glymour, 2001, 2003; Juraůsky, 1996; Narayan & Juraůsky, 1998, 2002; Krynski & Tenenbaum, 2007; López & García, 2009; Martin & VanLehn, 1995; Mislevy & Gitomer, 1996), they have not been extensively used as data analysis tools in the context of psychology (López, García, De la Fuente, & De la Fuente, 2007).

Likewise, the analytical applications in our near context have been related to the study of attitudes (i.e., García, López, Cano, Gea, & De la Fuente, 2006; López, 2009; López & García, 2007; López, Ramírez & Casado, 2012; López et al., 2010).

Sample size is possibly one of the most problematic aspects for logistic regression (i.e., Bull, Mak, & Greenwood, 2002; Conato, Peduzzi, Holford, & Feinstein, 1995; DeMaris, 2002; Firth, 1993; Harrel, Lee, Matchar, & Reichert, 1985; Irala, Fernández-Crehuet, & Serrano, 1997; Jovel, 1995; King & Zeng, 2001a, 2001b; Lilienfeld & Pyne, 1984; Long, 1997; Ortega & Cayuela, 2002; Peduzzi, Concato, Feinstein, & Holford, 1995; Peduzzi, Conchat, Kemper, Holford, & Feinstein, 1996; Silva & Barroso, 2004), as a large quantity of observations is needed for the estimations of the models to be valid and reliable. Nevertheless, diverse strategies were developed to make logistic regression work as expected. For example, Whittermore (1981) and Hsieh (1989) published a series of sampling tables for simple and multiple logistic regression with which to estimate the necessary sample size in order to obtain valid and reliable estimations without jeopardizing the statistical power. In addition, procedures to estimate optimum sample size, taking into account the prevalence rate of the independent variables, the statistical power, as well as their level of measurement have been developed (Hsieh, Bloch, & Larsen, 1998). From an analytical viewpoint, exact logistic regression (Hirji, Mehta, & Patel, 1987; King & Ryan, 2002; Mehta & Patel, 1995), based on conditional calculus and permutations, has been developed.

Reade and Kupper (1995) indicated that some variables are discretized, which can affect the results of the logistic regression analyses in the social sciences particularly. Although the categorization of continuous variables has the advantage of avoiding distributional assumptions about the variables, it allows simpler interpretations of the results, it provides higher robustness if the model was not well specified or if there were extreme data (Cusmille & Bangdiwala, 2000; Zhao & Kolonel, 1992), it also implies a loss of important information produced by a series of biases in the statistical results. More specifically, it has been observed that the estimations of the measures of association of the variables can vary depending on the cutoff point used to make the classification, the statistical power decreases, and Type I error can increase when testing the hypothesis (Altman, Lausen, Sauerbrei, & Schumacher, 1994; Chuang, 1997; Flegal, Keyl, & Nieto, 1991; Fung & Howe, 1984; Ragland, 1992; Reade & Kupper, 1995).

Some works have suggested that models based on a network structure could have advantages versus logistic regression in terms of prediction (i.e., Ankarali, Canan, Akkus, Bugdayci, & Ali, 2007; Bartfay, Mackillop, & Pater, 2006; Eftekhar, Mohammad, Ardebeli, Ghodsi, & Ketabchi, 2005; Finch & Schneider, 2007; García, López, De la Fuente, Cano, & Gea, 2007; Jaimes, Farbierz, Alvarez, & Martinez, 2005; Kumar, Rao, & Soni, 1995; Lee, Abbott, & Johantgen; 2005; López, Ruiz-Ruano, & García, 2008; Terrin, Schmid, Griffith, D’Agostino, & Selker, 2003). Nevertheless, there is no robust evidence that shows a better performance of the techniques based on network models versus logistic regression. According to the bibliography reviewed, there are few works that compare Bayes nets to logistic regression.

**Goals and hypotheses**

In accordance with the above, the goal of this work is to generate models that predict the tendency to create businesses in potential entrepreneurs (Krueger & Brazeal, 1994, Thompson, 2004) in terms of their general desirability (Veciana, Aponte, & Urbano, 2005), the time frame to initiate the business creation, and the entrepreneurs’ behavioral predisposition as a function of their attitude towards business creation, their perceived preparation, the perceived obstacles (or viability), their motivation, and the perceived deficiencies. All these predictor variables were considered relevant in the literature on business creation (i.e., Cano et al., 2003; García, et al., 2006, 2007; García, Cano, & Gea, 2005; Genesca & Capelleras, 2004; Genesca & Veciana, 1984; Gómez, Mira, & Martínez, 2007; HayGroup & SAP AG, 2003; Rogoff & Lee, 1996; Sánchez, 2003; Veciana, 1989). In the literature reviewed, we found no works that address this problem the same way we have addressed it in this work, although it is reasonable to think that the general desirability of creating a business will be directly related to motivation, whereas the behavioral dimension will be inversely related to the obstacles, as suggested in the work of Miller, Bell, Palmer, and Gonzalez (2009). Miller et al. (2009), examining students registered in courses on business creation and students not registered in this kind of training programs, found a negative relation between the level of barriers perceived by the students and the entrepreneurial intentions, although they found a positive relation between entrepreneurial intention and the awareness of entrepreneurial challenges. The attitude towards business creation and perceived viability have also been indicated as predictors of the intention to create a business (Liñán, Rodriguez-Cochoard, & Rueda-Cantuche, 2011). Othman and Ishak (2009) observed that entrepreneurship related to
locus of control, the monetary value of work, and autonomy were important predictors of the decision of choosing an entrepreneurial career. It has also been reported that when a person perceives the creation of a business as socially acceptable, this perception exerts a positive effect on the intention to create a business, independently of the person’s age (Kautonen, Tornikoski, & Kibler, 2011). Moreover, Fini, Grimaldi, Marzocchi, and Sobero (2012) also found that the entrepreneurial corporate intention depends on the attitude towards entrepreneurial behavior, and that this intention is conditioned by specific situational motivations as well as by the entrepreneur’s perceived skill.

In a sample of Catalan university students, Veciana et al. (2005) found that the men scored higher in perceived desirability and intention of creating a business. They also found that, in this same sample, there was a relation between the family profile of the potential entrepreneur and the intention to create a new company. However, these relations did not occur in a sample of students from Puerto Rico. Harris and Gibson (2008) also found that the men displayed more attitudes of personal control in the business and more innovative attitudes in the new business in comparison to the women. These authors also observed that prior family experience in business creation had a considerable effect on entrepreneurship.

We shall compare the predictive performance of the logistic regression and the Bayes nets to predict these components of the tendency towards business creation, manipulating the number of missing cases as well as the degree of categorization in the predictor variables. For this purpose, the binary logistic regression model and the Bayes naive Classifier (Martínez & Rodríguez, 2003) or Bayes simple classifier (Domingos & Pazzni, 1996) will be used. As the rate of missing cases is related to the sample size (because logistic regression does not use a case for estimation when it has a missing value in the vector case), and considering the problems presented by logistic regression when categorizing the variables of the model, we expect to find that the Bayes simple classifier will show higher rates of predictive classification.

Method

Participants

A sample of 1230 university students participated in the study. The participants were selected through block sampling with proportional affixation by sex and degree program (sampling error ± 3%, 95 CI%, z = 1.96 and p = q = .5) from the first and last courses of all the degree programs that are taught in the Universidad de Almería. Of the sample, 426 (34.6%) participants were men and 797 (64.8%) were women, and their age ranged between 17 and 56 years ($M = 22.45, SD = 4.46$). Of the participants, 48.2% ($n = 557$) studied a career in the sphere of human or juridical sciences, 26.26% ($n = 323$) studied careers within the business field, and 27.9% ($n = 344$) studied technical degrees.

Operationalization of the variables

To collect the data, we used a 3-page, A4 format, questionnaire. On the back of the front cover were questions about the sociodemographic data. On the next page were various questions about the intention to create a business. The first three questions of this second page are the variables used as outcome variables in the statistical models. Each one of these variables was considered (one by one) an outcome variable in the logistic regression models and a class or divergent node in the Bayes net models. The questions were: Do you consider it worthwhile to create your own business? ($N_{YES} = 913$), Have you given serious thought to creating your own business, as a real short-/medium-term option? ($N_{YES} = 463$), and Have you ever initiated actions aimed at creating your own business? ($N_{YES} = 109$). Thus, the first question refers to the desirability of business creation (DES), the second question is worded in temporal terms about the short- and medium-term tendency of setting up a business (TEM), and the third question refers to the behavioral dimension (BEH). These three questions were responded by the participants in positive or negative terms.

After the questions about business creation, we presented five scales whose scores were used as predictors or independent variables in the models of logistic regression and as divergence variables in the Bayes net models. In all of them, the items were rated on a 5-point Likert-type scale (except for the Attitudes towards Business Creation Scale, which was rated on a 4-point scale), and they were corrected by dividing the sum of all the items by the number of items of each scale after reversing the items that were negatively worded. First was the Perceived Training Deficiencies scale (12 items, $\alpha = .89$, $M = 2.38$, $SD = 0.66$) [i.e., administrative formalities, marketing, human resources], Perceived Preparation (4 items, $\alpha = .85$, $M = 2.08$, $SD = 0.76$) [i.e., knowledge, experience], Motivation to create a Business (11 items, $\alpha = .88$, $M = 3.26$, $SD = 0.57$) [i.e., finding a job, economic independence] and Perceived Obstacles in the Creation of a Business (17 items, $\alpha = .83$, $M = 3.17$, $SD = 0.58$) [i.e., having to work too many hours, lack of experience], where all the items were positively expressed. On the third page was the “Actitudes hacia la Creación de Empresas” (ACEMP, in English, the Attitudes towards Business Creation Scale; Cano et al., 2003; García et al., 2005, 2006, 2007; López et al., 2010). The ACEMP scale has 29 items referring to the frequency with which certain behaviors or thoughts are carried out, of which 13 items are reverse-scored (with regard to a positive interpretation); for the rest, a higher score indicates a higher
of the variable. We built 24 models, of which 12 were recategorized in two, three, five, and ten equal sized class intervals (Domingos & Pazzni, 1996), based on the range of the variable. We built 24 models, of which 12 were

<table>
<thead>
<tr>
<th>Variable</th>
<th>Deficiencies</th>
<th>Obstacles</th>
<th>Motivation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Preparation</td>
<td>.236**</td>
<td>-.167**</td>
<td>.090**</td>
</tr>
<tr>
<td>ACEMP</td>
<td>.302**</td>
<td>-.251**</td>
<td>.237**</td>
</tr>
<tr>
<td>Deficiencies</td>
<td>-.071*</td>
<td>.212**</td>
<td>.264**</td>
</tr>
</tbody>
</table>

*p < .05 (bilateral). **p < .01 (bilateral).

quantity of entrepreneurship (α = .75, M = 2.77, SD = 0.31).

Table 1 shows the correlations among the variables involved in the study.

Procedure

The questionnaires were administered to students in the first and last courses of the careers taught in the University of Almería, at the end of a session in a core course in each study plan. The questionnaire was administered in groups with the teacher’s consent and the voluntary participation of the students. To request the teachers’ collaboration so they would consent to give up part of their class time so their students could complete the questionnaire, we contacted them by e-mail. Before administering the questionnaire, general instructions were given and the participants were asked to carefully read the presentation letter, in which they were informed about the goal of the study, which was on the first page of the form. When collecting the data, the doubts of the participants were clarified by the person who administered the questionnaire. The questionnaire took about 20 to 30 minutes to complete. No reward was given for collaboration, except for thanking the students verbally.

Data analysis

As there were five independent variables that could form part of the best substantive and statistical model, and taking into account the principle of parsimony with regard to the number of variables, we compared all the possible combinations of independent variables that could occur in this situation (all the possible combinations without replacement). Of the n variables taking p variables at a time. In all, we estimated 186 models, one third for each dependent variable. No interaction terms were entered so as not to deal with complexity in this case. Although we could have dealt with this theme, later studies will have to aim their research at solving this topic. In any event, our results can be taken as a conservative indication of the most complex case, which would include interaction terms (Peduzzi et al., 1995).

To study the effect of categorization, all the variables were recategorized in two, five, and ten equal sized class intervals (Domingos & Pazzni, 1996), based on the range of the variable. We built 24 models, of which 12 were Bayes nets (BN) and 12 were logistic regression models (LR). Likewise, for each predicted variable (DES, TEM, and BEH), we estimated four models: one where the predictors were not recoded, and three where they were categorized in two, three, five, and ten categories. To study the effect of the missing values, we conducted a Monte Carlo simulation to generate completely random patterns of missing data (Hair, Anderson, Tatham, & Black, 1998). In total, we generated 24 databases with random patterns of missing data characterized by different percentages of missing cases (0, 5, 10, 15, 20, 25, 50, and 70%); of which one third was used to estimate the models of the desirability component, one third was used to estimate the models of the temporal component, and the rest were used to estimate the models of the behavioral component of entrepreneurship. Figure 1 shows an outline of the design of the data analysis.

The estimation of the logistic regression models was done with the PASW statistical package, version 18.0 (SPSS), in its unconditioned format, without including any interaction term and with a 95% confidence level of the estimations. No method of sequential selection of the independent variables was employed. The parameters of the Bayes nets were estimated with the maximum likelihood method corrected with Laplace’s succession equation (Morales, 2006; Ng & Jordan, 2002) with the Netica program, version 4.02 (Norsys Software). For each model, we estimated their corresponding goodness-of-fit indexes and for all of them, we obtained the values of sensitivity, specificity, proportion of false positives, proportion of false negatives, positive predictive value, negative predictive value, and general proportion of correct classifications.

Predictive efficiency is understood, in general, as the degree to which a classification tool frequently provides correct responses (Greiner, Su, Shen, & Zhou, 2005; Greiner & Zhou, 2002; Shen, Su, Greiner, Musilek, & Cheng, 2003). Nevertheless, the two techniques compared in this work produce different rates of correct classifications depending on the cut-point threshold established for the classification. Thus, we shall use ROC curves to compare the levels of specificity and sensitivity for the diverse cut-points in the probabilities provided by both models (i.e., DeMaris, 2002; Hanley & McNeil, 1982, 1983). The area under the ROC curve (θ) was estimated with the nonparametric technique proposed by Hanley and McNeil (1982). To determine whether there were statistically significant differences between the ROC curves generated by logistic regression and Bayes nets, we used the procedure proposed by Hanley and McNeil (1982), taking into account the correlation between both areas under the curve.

Results

Table 2 shows the logistic parameters associated with each one of the models that obtained the best combination of sensitivity, specificity, proportion of false positives,
Figure 1. Outline of the design.
proportion of false negative, positive predictive value, negative predictive value, and general proportion of correct classifications when studying the conjoint behavior of the two statistical techniques employed. As can be observed, all the predictor variables are positively associated with the tendency to consider the creation of a business desirable, except for perceived obstacles, which is inversely related to this tendency. That is, a higher score in the ACEMP scale, higher perceived preparation, and a higher score in the motivation scale imply a higher probability of considering the creation of a business desirable. The largest parameter of the model is associated with the ACEMP score.

Although we selected three variables to explain the temporal aspect of the tendency towards business creation (TEM), only two of them were statistically significant at the individual level, the score in the ACEMP scale and perceived preparation. Both the ACEMP score and the score in perceived preparation were positively related to the probability of seriously considering the creation of a business at short-/mid-term. Again, the greatest parameter of the model was associated with the ACEMP score, indicating that people who score high in this scale are 1.5 to 6.6 times more likely to create a new business.

With regard to the manipulation of the degree of categorization, in contrast to our initial expectations, and as can be observed in Figure 2, it benefitted logistic regression globally. Although both logistic regression ($\theta = .7992, z = 23.18, p < .001$) and Bayes nets ($\theta = .8365, z = 28.19, p < .001$) generated acceptable areas under the ROC curve, we observed a statistically significant difference of 3.72%, favoring logistic regression ($z = 4.06, p < .001$).

Table 3 shows that the Bayes nets had problems when the predictor variables were recoded to explain the temporal and behavioral dimension of the attitude towards business creation. In these cases, the Bayes nets did not generate a ROC curve that was statistically different from .5, whereas in the case of the logistic regression, the areas under the ROC curve were always statistically different from this reference value. Nevertheless, when the response variable is the desirability of creating a business, the Bayes nets performed as well or better than logistic regression. Thus, when the independent variables were recoded in five or ten categories, the Bayes nets generated areas under the ROC curve that were statistically larger than those generated by logistic regression ($z = 1.93, p < .05$ and $z = 2.00, p < .05$, respectively) In the case of dichotomization, logistic regression obtained a 1.3% predictive advantage over the Bayes nets but this difference was not statistically significant ($z = 1.36, p = .09$).

Table 2

<table>
<thead>
<tr>
<th>OV</th>
<th>PV</th>
<th>$\beta$</th>
<th>SE</th>
<th>Wald ($df=1$)</th>
<th>$p$</th>
<th>$e^\beta$</th>
<th>CI inf</th>
<th>CI sup</th>
</tr>
</thead>
<tbody>
<tr>
<td>DES</td>
<td>A</td>
<td>0.985</td>
<td>0.264</td>
<td>13.956 *</td>
<td>2.677</td>
<td>1.597</td>
<td>4.489</td>
<td></td>
</tr>
<tr>
<td>$\chi^2 = 12.33$</td>
<td>P</td>
<td>0.497</td>
<td>0.111</td>
<td>19.954 *</td>
<td>1.643</td>
<td>1.321</td>
<td>2.043</td>
<td></td>
</tr>
<tr>
<td>$p = .14$</td>
<td>O</td>
<td>-0.596</td>
<td>0.156</td>
<td>14.663 *</td>
<td>0.551</td>
<td>0.406</td>
<td>0.747</td>
<td></td>
</tr>
<tr>
<td>M</td>
<td>1.084</td>
<td>0.153</td>
<td>50.434</td>
<td>*</td>
<td>2.956</td>
<td>2.192</td>
<td>3.986</td>
<td></td>
</tr>
<tr>
<td>Cst.</td>
<td>-4.139</td>
<td>0.913</td>
<td>20.55</td>
<td>*</td>
<td>0.016</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TEM</td>
<td>A</td>
<td>1.499</td>
<td>0.228</td>
<td>43.357 *</td>
<td>4.476</td>
<td>2.865</td>
<td>6.993</td>
<td></td>
</tr>
<tr>
<td>$\chi^2 = 12.32$</td>
<td>D</td>
<td>0.158</td>
<td>0.118</td>
<td>1.813 .178</td>
<td>1.172</td>
<td>0.93</td>
<td>1.476</td>
<td></td>
</tr>
<tr>
<td>$p = .14$</td>
<td>P</td>
<td>0.624</td>
<td>0.102</td>
<td>37.373 *</td>
<td>1.867</td>
<td>1.528</td>
<td>2.281</td>
<td></td>
</tr>
<tr>
<td>Cst.</td>
<td>-6.363</td>
<td>0.641</td>
<td>98.474</td>
<td>*</td>
<td>0.002</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$\chi^2 = 4.02$</td>
<td>P</td>
<td>1.121</td>
<td>0.15</td>
<td>55.781 *</td>
<td>3.069</td>
<td>2.287</td>
<td>4.119</td>
<td></td>
</tr>
<tr>
<td>$p = .85$</td>
<td>O</td>
<td>-0.083</td>
<td>0.203</td>
<td>0.168 .682</td>
<td>0.92</td>
<td>0.618</td>
<td>1.37</td>
<td></td>
</tr>
<tr>
<td>Cst.</td>
<td>-7.994</td>
<td>1.433</td>
<td>31.14</td>
<td>*</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>


*p < .001.
When manipulating the rate of missing values, the Bayes nets showed better global functioning in the three components of the tendency toward business creation. The greatest differences between the areas under the ROC curve were observed in the desirability component (z = 2.93, p = .002), where logistic regression obtained an area of 0.6993 (z = 9.47, p < .001) and the Bayes net an area of 0.7468 (z = 13.35, p < .001). In the model of the temporal dimension of the attitude toward business creation, the differences between the curves was of 4.64% (z = 5.18, p < .001) and logistic regression obtained an area under the ROC curve of 0.6881 (z = 11.44, p < .001) whereas the Bayes net generated a curve with an area of 0.7351 (z = 15.30, p < .001). Lastly, with regard to the behavioral dimension, although the Bayes net had a larger global area under the curve (.6597, z = 4.68, p < .001) than logistic regression (.6440, z = 4.31, p < .001), this difference was not statistically significant (z = 0.59, p = .28). In all cases, the area under the ROC curve was lower in logistic regression than the one generated by the corresponding Bayes net. Nevertheless, no statistically significant differences were

Table 3
Areas under the ROC Curve as a Function of Categorization Level of the Outcome Variable (OV) and the Type of Technique.

<table>
<thead>
<tr>
<th>Technique</th>
<th>OV</th>
<th>NC</th>
<th>θ</th>
<th>SDθ</th>
<th>Z</th>
<th>p</th>
<th>CIinf</th>
<th>CI sup</th>
</tr>
</thead>
<tbody>
<tr>
<td>LR</td>
<td>DES</td>
<td>2</td>
<td>.6673</td>
<td>.0189</td>
<td>8.87</td>
<td>*</td>
<td>.6303</td>
<td>.7042</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>5</td>
<td>.6752</td>
<td>.0186</td>
<td>9.42</td>
<td>*</td>
<td>.6388</td>
<td>.7117</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>10</td>
<td>.6972</td>
<td>.0183</td>
<td>10.8</td>
<td>*</td>
<td>.6614</td>
<td>.733</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>WC</td>
<td>.703</td>
<td>.0182</td>
<td>11.14</td>
<td>*</td>
<td>.6673</td>
<td>.7387</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>2</td>
<td>.6385</td>
<td>.0175</td>
<td>7.9</td>
<td>*</td>
<td>.6041</td>
<td>.6729</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>5</td>
<td>.6753</td>
<td>.0165</td>
<td>10.63</td>
<td>*</td>
<td>.643</td>
<td>.7076</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>10</td>
<td>.6772</td>
<td>.0166</td>
<td>10.66</td>
<td>*</td>
<td>.6446</td>
<td>.7097</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>WC</td>
<td>.688</td>
<td>.0164</td>
<td>11.43</td>
<td>*</td>
<td>.6558</td>
<td>.7202</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>2</td>
<td>.632</td>
<td>.0326</td>
<td>4.05</td>
<td>*</td>
<td>.568</td>
<td>.6959</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>5</td>
<td>.6366</td>
<td>.0335</td>
<td>4.08</td>
<td>*</td>
<td>.571</td>
<td>.7022</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>10</td>
<td>.6492</td>
<td>.0335</td>
<td>4.46</td>
<td>*</td>
<td>.5837</td>
<td>.7148</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>WC</td>
<td>.6412</td>
<td>.0336</td>
<td>4.21</td>
<td>*</td>
<td>.5754</td>
<td>.707</td>
</tr>
<tr>
<td>BN</td>
<td>DES</td>
<td>2</td>
<td>.6549</td>
<td>.0182</td>
<td>8.53</td>
<td>*</td>
<td>.6193</td>
<td>.6905</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>5</td>
<td>.6924</td>
<td>.0175</td>
<td>11.01</td>
<td>*</td>
<td>.6582</td>
<td>.7267</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>10</td>
<td>.7161</td>
<td>.0174</td>
<td>12.43</td>
<td>*</td>
<td>.6821</td>
<td>.7502</td>
</tr>
<tr>
<td></td>
<td>DES</td>
<td>WC</td>
<td>.7549</td>
<td>.0167</td>
<td>15.28</td>
<td>*</td>
<td>.7222</td>
<td>.7876</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>2</td>
<td>.5066</td>
<td>.017</td>
<td>0.39</td>
<td>.349</td>
<td>.4732</td>
<td>.54</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>5</td>
<td>.4947</td>
<td>.017</td>
<td>-0.031</td>
<td>.622</td>
<td>.4614</td>
<td>.5281</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>10</td>
<td>.505</td>
<td>.017</td>
<td>0.3</td>
<td>.383</td>
<td>.4717</td>
<td>.5384</td>
</tr>
<tr>
<td></td>
<td>TEM</td>
<td>WC</td>
<td>.729</td>
<td>.0148</td>
<td>15.47</td>
<td>*</td>
<td>.7</td>
<td>.758</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>2</td>
<td>.5136</td>
<td>.0293</td>
<td>0.46</td>
<td>.322</td>
<td>.4562</td>
<td>.5709</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>5</td>
<td>.5247</td>
<td>.0298</td>
<td>0.83</td>
<td>.204</td>
<td>.4662</td>
<td>.5831</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>10</td>
<td>.5334</td>
<td>.0299</td>
<td>1.12</td>
<td>.132</td>
<td>.4749</td>
<td>.5919</td>
</tr>
<tr>
<td></td>
<td>BEH</td>
<td>WC</td>
<td>.7015</td>
<td>.0314</td>
<td>6.41</td>
<td>*</td>
<td>.6398</td>
<td>.7631</td>
</tr>
</tbody>
</table>


*p < .001.
observed between the curves generated by each technique in the model designed to predict the behavioral component of the attitude towards business creation (Table 4). In the rest of the cases, statistically significant differences were observed between the curves generated by logistic regression and Bayes net, except in the cases that dealt with the model of the desirability component of the entrepreneurial attitude with 20 and 70% of missing cases, where the differences can be considered marginal.

Discussion

The most relevant result we can underline was that the behavioral dimension of the attitude toward business creation was the most difficult to predict correctly from the variables used in this study. In all cases, the area under the ROC curve generated both by logistic regression and by Bayes nets was lower compared with the desirability and temporal dimensions. This difficulty to predict the behavioral component could be due to a synergy derived from the complex relation of the object studied—the relation between the attitudinal components and behavior (i.e., Brehm, Kassin, & Fein, 2005; Deaux, Dane, & Wrightsman, 1993; Feldman, 1998; Franzoi, 2005; Morales, Rebollos, & Moya, 1994)—and to the methodological difficulties derived from the low rate of positive responses in this dependent variable. As noted Phan and Butler (2003), “the relationship between beliefs and attitudes further complicates the linkage between attitudes and strategic choice” (p. 76) and, therefore, we suggest that future works aimed at modeling entrepreneurial potential take into account both beliefs and attitudes about this when using multivariate statistical models, such as structural equation models (i.e., Krueger et al., 2000).

In terms of the variables related to attitude towards business creation, we observed that the score on the ACEMP entrepreneurial attitudes scale was a good predictor of all the dimensions of the tendency towards business creation. Thus, this study could be seen as another proof of the validity of this scale (Cano et al., 2003; García et al., 2005, 2006,

Table 4
Comparison of the Areas under the ROC Curve as a Function of the Percentage of Missing Cases (MC) of the Outcome Variable (OV) and the Type of Technique.

<table>
<thead>
<tr>
<th>OV</th>
<th>MC</th>
<th>( r_p )</th>
<th>( r_\alpha )</th>
<th>( r )</th>
<th>SD</th>
<th>z</th>
<th>( p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>DES</td>
<td>0</td>
<td>.7017</td>
<td>.7661</td>
<td>.69</td>
<td>.01</td>
<td>4.056</td>
<td>*</td>
</tr>
<tr>
<td>DES</td>
<td>5</td>
<td>.6521</td>
<td>.7538</td>
<td>.67</td>
<td>.02</td>
<td>3.115</td>
<td>.001</td>
</tr>
<tr>
<td>DES</td>
<td>10</td>
<td>.7034</td>
<td>.7742</td>
<td>.69</td>
<td>.01</td>
<td>3.664</td>
<td>*</td>
</tr>
<tr>
<td>DES</td>
<td>15</td>
<td>.6425</td>
<td>.7555</td>
<td>.65</td>
<td>.02</td>
<td>3.193</td>
<td>.001</td>
</tr>
<tr>
<td>DES</td>
<td>20</td>
<td>.6724</td>
<td>.7807</td>
<td>.69</td>
<td>.02</td>
<td>1.413</td>
<td>.079</td>
</tr>
<tr>
<td>DES</td>
<td>25</td>
<td>.6939</td>
<td>.7394</td>
<td>.66</td>
<td>.02</td>
<td>3.21</td>
<td>.001</td>
</tr>
<tr>
<td>DES</td>
<td>50</td>
<td>.667</td>
<td>.6888</td>
<td>.63</td>
<td>.02</td>
<td>3.098</td>
<td>.001</td>
</tr>
<tr>
<td>DES</td>
<td>70</td>
<td>.6255</td>
<td>.7671</td>
<td>.65</td>
<td>.03</td>
<td>1.488</td>
<td>.068</td>
</tr>
<tr>
<td>TEM</td>
<td>0</td>
<td>.8745</td>
<td>.8517</td>
<td>.84</td>
<td>.01</td>
<td>5.182</td>
<td>*</td>
</tr>
<tr>
<td>TEM</td>
<td>5</td>
<td>.8818</td>
<td>.8453</td>
<td>.84</td>
<td>.01</td>
<td>4.266</td>
<td>*</td>
</tr>
<tr>
<td>TEM</td>
<td>10</td>
<td>.8812</td>
<td>.859</td>
<td>.84</td>
<td>.01</td>
<td>3.154</td>
<td>.001</td>
</tr>
<tr>
<td>TEM</td>
<td>15</td>
<td>.8608</td>
<td>.8338</td>
<td>.82</td>
<td>.01</td>
<td>3.127</td>
<td>.001</td>
</tr>
<tr>
<td>TEM</td>
<td>20</td>
<td>.8478</td>
<td>.8206</td>
<td>.79</td>
<td>.01</td>
<td>4.345</td>
<td>*</td>
</tr>
<tr>
<td>TEM</td>
<td>25</td>
<td>.8883</td>
<td>.8472</td>
<td>.84</td>
<td>.01</td>
<td>4.265</td>
<td>*</td>
</tr>
<tr>
<td>TEM</td>
<td>50</td>
<td>.8413</td>
<td>.834</td>
<td>.79</td>
<td>.01</td>
<td>3.368</td>
<td>*</td>
</tr>
<tr>
<td>TEM</td>
<td>70</td>
<td>.8316</td>
<td>.7917</td>
<td>.77</td>
<td>.02</td>
<td>2.519</td>
<td>.006</td>
</tr>
<tr>
<td>BEH</td>
<td>0</td>
<td>.7727</td>
<td>.685</td>
<td>.69</td>
<td>.03</td>
<td>0.594</td>
<td>.276</td>
</tr>
<tr>
<td>BEH</td>
<td>5</td>
<td>.7942</td>
<td>.6405</td>
<td>.67</td>
<td>.03</td>
<td>0.9</td>
<td>.184</td>
</tr>
<tr>
<td>BEH</td>
<td>10</td>
<td>.7796</td>
<td>.6743</td>
<td>.69</td>
<td>.03</td>
<td>0.823</td>
<td>.205</td>
</tr>
<tr>
<td>BEH</td>
<td>15</td>
<td>.749</td>
<td>.676</td>
<td>.67</td>
<td>.03</td>
<td>0.318</td>
<td>.375</td>
</tr>
<tr>
<td>BEH</td>
<td>20</td>
<td>.7424</td>
<td>.6908</td>
<td>.67</td>
<td>.03</td>
<td>0.106</td>
<td>.458</td>
</tr>
<tr>
<td>BEH</td>
<td>25</td>
<td>.8376</td>
<td>.6391</td>
<td>.69</td>
<td>.03</td>
<td>0.407</td>
<td>.342</td>
</tr>
<tr>
<td>BEH</td>
<td>50</td>
<td>.6482</td>
<td>.6652</td>
<td>.61</td>
<td>.04</td>
<td>1.179</td>
<td>.119</td>
</tr>
<tr>
<td>BEH</td>
<td>70</td>
<td>.6861</td>
<td>.699</td>
<td>.65</td>
<td>.04</td>
<td>0.372</td>
<td>.355</td>
</tr>
</tbody>
</table>

Note. SD: standard deviation of the difference between curves. \( r_p \): correlation of the responses of the techniques for the cases with positive response. \( r_\alpha \): correlation between the responses of the techniques for the cases with negative response. \( r \): correlation between the areas under the ROC curves. DES: desirability. TEM: temporal component. BEH: behavioral component. *\( p < .001 \).
Moreover, subsequent studies should examine in more detail the relations established between the attitude towards business creation and the ACEMP dimensions. Thus, as Gartner’s (1985, 1988) criticism of the trait model in the study of business creation has been recently rebutted (McKenzie et al., 2007), this scale could be useful in diverse contexts, both theoretical and applied, to identify and describe the profile of the person who sets out to create a new business venture.

Another variable we found that was related to all the dimensions of entrepreneurship was preparation. We also observed that, although in all cases, the relation between the level of preparation and the tendency toward business creation was positive, the behavioral dimension had the closest relation to perceived preparation, whereas desirability and the temporal aspect were at the same level. This result is consistent with prior works that revealed the importance of training in the entrepreneur’s profile (i.e., Genescá & Capelleras, 2004; Genesca & Veciana, 1984; Gómez et al., 2007; HayGroup & SAP AG, 2003; Rogoff & Lee, 1996; Sánchez, 2003; Veciana, 1989). Thus, as training is presented as an essential element in the description of the entrepreneur as well as a determinant trait to predict entrepreneurs’ competitive success, intervention programs aimed at recommending and creating training itineraries for potential entrepreneurs are needed. In the university sphere, for example, the urgency of promoting a favorable context for professional and intellectual development linked to the development of new businesses has been underlined (i.e., Cano et al., 2003; Díaz, 2003; Peñas & Quijano, 2008).

Besides the ACEMP score and the degree of perceived preparation, which are common factors to all three dimensions of entrepreneurship dealt with in this study, we also observed that the desirability of business creation is related to a motivational aspect and to the quantity of obstacles (viability) perceived by the person. Thus, people who are more motivated to create their own business will have a greater desire to create a business, whereas the people who perceive more obstacles will tend less towards business creation. The behavioral component is inversely related to the number of obstacles perceived in business creation. That is, the more obstacles a potential entrepreneurs perceive, the lower will be the behavioral dimension of their entrepreneurial intention. These results are consistent with the model of the entrepreneur event (Krueger & Brazeal, 1994; Krueger & Carsrud, 1993; Shapero & Sokol, 1982) and with the empirical results of Miller et al. (2009). Lastly, we observed that the people with more perceived deficiencies have the highest tendency towards business creation in temporal terms. Although this aspect had the lowest reported reliability, it would be consistent with the study of Rogoff and Lee (1996), which found that a certain type of real entrepreneurs typified as heirs believed they had more knowledge about the management of a business (measured with a 7-point Likert-type item) when, in fact, they had the least knowledge (measured with 19 items of general knowledge about business extracted from the curricular material of business sciences studies) in comparison with the typologies of creator and operator. As our study did not deal with any variable that considered the models of family role in business creation, it would be interesting for future research to consider it—along with the tendency towards the creation of new businesses—to predict the degree of perceived training deficiencies in potential entrepreneurs, taking into account the different entrepreneurial typologies (Gartner, 1985). A limitation of our study related to the above paragraphs is that we did not consider the three dimensions of business creation as variables at different levels of the network that may explain entrepreneurial behavior. Although we could have addressed this issue from a methodological viewpoint, we wished to study the three variables independently as variables explained by the rest of the variables. Future works will have to assess the goodness of fit of the models that include variables at their corresponding level, using Bayes nets.

In general terms, these results can serve to characterize the entrepreneur or, more specifically, the potential entrepreneur, as we worked with a sample of university students (i.e., Huefner et al., 1996; Krueger & Brazeal, 1994; López, 2009; López et al., 2010). This has practical utility because potential entrepreneurs are economically valuable and should therefore be protected or, at least, vocationally guided (i.e., Valls, 1996). In this sense, the results we present can be used institutionally to classify and identify entrepreneurs within the university with the aim of optimizing their probabilities of success in their businesses (Díaz, 2003), because, as noted recently by Morillas (2009), the optimization of possibilities for entrepreneurs stimulates local development programs and is a seedbed for employment. This is even more urgent when underlining the lack or deficiency of certain managing skills in entrepreneurs, which seems to have a drastic effect on the growth and permanence of newly created businesses (Fuller-Love, 2006). In the sphere of stimulation of the entrepreneurial attitude, we should try to increase the desirability and viability of business creation (Krueger & Brazeal, 1994). As noted by Dreisler, Blenker, and Nielsen (2003) in Denmark, this would include, among other aspects, informing potential entrepreneurs, carrying out propagandistic intervention programs, introducing social interventions aimed at modifying attitudes and social norms about entrepreneurs, developing economic incentives or modifying normative and/or legislations. As commented by Lihán et al. (2011), we should promote the development of entrepreneurial skills and competence in primary and/or secondary school, as well as the development of tools that assess the degree to which the social intervention programs aimed at optimizing entrepreneurial potential are successful.

With regard to the statistical aspect of our work, we observed that, when a sample presents many missing cases,
this is problematic for logistic regression, whereas the Bayes net seems to deal with this situation more successfully. Nevertheless, this advantage of Bayes nets to deal efficiently with missing data disappears when the rate of positive responses in the criterion variable is very low (as in our study when predicting the behavioral component of entrepreneurship). The suitability of working with Bayes nets when the samples present missing data was already known (i.e., Heckerman, 1995) but, as shown in this study, it is necessary to perform comparative studies with different techniques to determine the best application conditions of each technique. The development of hybrid parametric estimation procedures from the structural properties of Bayes nets and of logistic regression is a promising advance (Greiner et al., 2005; Greiner & Zhou, 2002; Ross, Wettig, Grünwald, Myllymäki, & Tirri, 2005). Despite the fact that one of the parameters that negatively affects regression analysis is the dichotomization of the predictor variables under certain conditions (i.e., Altman et al., 1994; Ragland, 1992; Reade & Kupper, 1995), larger areas under the ROC curve were observed for logistic regression in general and specific terms when the predictor variables were categorized using arbitrary recoding with intervals of the same width. Nevertheless, these differences, albeit statistically significant, are not very large and could be due to the influence of other factors (such as the type of recoding or the way the variables are distributed). Future research will have to explore these issues in more detail. Another limitation, one of the main ones of our work with regard to the comprehension of the entrepreneur phenomenon, derives from having used potential entrepreneurs instead of real ones. In future investigations, the perceptions of entrepreneurs should be assessed in comparison to those of potential entrepreneurs in order to have another perspective of the phenomenon. Lastly, we propose that the methodology of Bayes nets could be useful to model the processes of business creation from the viewpoint of the psychosocial models of entrepreneurial intention (Krueger et al., 2000; Liñán, et al., 2008; Moriano et al., 2008). Thus, we could have probabilistic models that would represent the process of decision-making that real or potential entrepreneurs carry out during the task of creating their own company.


References


