ISSN: 2527-3159 (print) 2527-3167 (online)



Bootstrap Logistic Regression On Determining Factors Affecting The Level Of Entrepreneurial Capability

Muhammad Ridho¹ and Dodi Devianto²

Andalas University, dho22.mhd@gmail.com¹ Andalas University, ddevianto@sci.unand.ac.id²

doi: https://doi.org/10.15642/mantik.2019.5.1.10-18

Abstrak: Tujuan dari penelitian ini adalah untuk mengetahui faktor-faktor yang mempengaruhi tingkat kemampuan wirausaha pada daerah wisata di Nagari Salayo Sumatera Barat. Dalam hal ini, tingkat kemampuan wirausaha menjadi variabel respon dengan skala ordinal yang terdiri dari empat kategori, yaitu rendah, sedang, tinggi dan sangat tinggi. Sedangkan, variabel prediktor terdiri dari empat faktor sosio-demografi yaitu jenis kelamin, tingkat pendidikan, kelompok usia dan pekerjaan serta lima variabel motivasi wirausaha. Untuk menentukan variabel prediktor yang mempengaruhi variabel respon digunakan regresi logistik ordinal dengan pendugaan Bootstrap. Hasil dari penelitian ini menunjukkan terdapat dua variabel prediktor yang secara signifikan mempengaruhi variabel respon, yaitu motif wirausaha dan motif sosial dengan *hit ratio* sebesar 61,667. Sehingga dapat disimpulkan bahwa model yang terbentuk dapat digunakan untuk menentukan tingkat kemampuan wirausaha pada daerah wisata.

Kata Kunci: regresi logistik ordinal, pendekatan Bootstrap, kemampuan wirausaha

Abstract: The purpose of this study is to determine the factors that affect the level of entrepreneurial capability in tourism of rural area in Nagari Salayo of West Sumatra. The level of entrepreneurial capability is the response variable in this study with an ordinal scale consisting of four categories; they are lower, middle, high, or very high. Whereas the predictor variables consist of 4 sociodemographic factor variables, they are gender, education level, age group, and occupation, and also five entrepreneurial motivation variables. To determine the predictor variables that are significantly affecting response variables, an ordinal logistic regression with a bootstrap estimation is executed. The study's result shows two predictor variables that affect the response variable significantly; they are the entrepreneurial motive and social motive with the hit ratio of 61,667%. With that result, the model formed by bootstrapping logistic regression can determine the level of entrepreneurial capability in tourism of the rural area.

Keywords: ordinal logistic regression, bootstrap estimation, entrepreneurial capability

1. Introduction

Tourism in Indonesia is currently growing rapidly, along with the increase in visits by domestic and foreign tourists. This extraordinary tourism potential which owned by Indonesia can be a mainstay to raise society's lives. Cultural village is one type of business that immediately captivates many regions, one of which is Solok regency. Solok as one of the tourist destinations in West Sumatra, Indonesia seeks to encourage increased tourist visits to this area by developing a cultural or traditional village program in the rural areas of Nagari Salayo at West Sumatra Province.

The entrepreneur in tourism certainly needs serious attention so that they can successfully invite tourists to come. The arrival of tourists can grow society' economic, which is marked by the growing number of business communities. There is some evidence about regional development being community entrepreneurs with research surveys on tourism motivation [1] entrepreneurial motivation [2], travel motivation and value [3], SME performance [4] and also investigations about the effects of motivation on entrepreneurial capability.

The entrepreneurial ability influences the entrepreneurial process [5]. This means that the success rate of an entrepreneurial is directly proportional to the level of the entrepreneur's ability to run a business. The entrepreneurs who have a high level of achievement of initial motivation also have high entrepreneurial abilities [6]. There are specific factors that influence entrepreneurial capability, which is solved using panel Bootstrap analysis [5], exploratory factor analysis and confirmatory factor analysis [7] Logistic regression biner [8], logistic regression ordinal [9].

This study examines factors which affect the level of entrepreneurial capability in tourist areas in Nagari Salayo West Sumatera, which are grouped into four levels. Ordinal regression analysis with Bootstrap estimation is used to predict socio-demographic factors and entrepreneurial motives that significantly affect the level of entrepreneurial capability. The confident interval of Bootstrap estimation is used to test the accuration of ordinal regression. The results of this study can be considered in the development of creative economics and local resources to support the sustainability of tourism village programs in West Sumatra.

2. Literature review

A data analysis used to find the relationship between the response variable (Y) with one or more predictor variables (X) is called the regression method [10]. This method aims to get a suitable model to describe the relationship between response variables and set of predictor variables in the most simple and best way. Meanwhile, logistic regression is a regression used to analyze the relationship between nominal or ordinal scale with two or more categories variables with a set of continuous or categorical predictor variables [11].

Logistic regression equation used from the approximate form of probability function $\pi(x) = E(Y|x)$ as the following equation [10]

$$\pi(x) = \frac{\exp(\alpha + \beta_1 x)}{1 + \exp(\alpha + \beta_1 x)},$$
(1)

Then to simplify equation (1), a logit transform is performed as:

$$g(x) = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \alpha + \beta_1 x \tag{2}$$

Furthermore, an ordinal logistic regression is a regression analysis used to describe the relationship between response variables with predictor variables, where the response variable is ordinal scale with more than two categories. The cumulative opportunity $P(Y \le r|x)$ is defined as:

$$P(y \le r | x) = \pi_r = \frac{\exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)}{1 + \exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)},$$
(3)

Where x is the vector value of the predictor variable with t categories [11]. The estimation of the regression parameter is done by describing it using the logit transformation of $P(Y \le r|x)$.

Logit P (Y \le r|x) = ln
$$\left(\frac{P(Y \le r|x)}{1 - P(Y \le r|x)} \right)$$
 (4)

By substituting equation (3) to equation (4), obtained

Logit P (Y
$$\leq r | x) = \alpha_r + \sum_{k=1}^t \beta_k x_k$$
 (5)

If there are *r* response categories, then the cumulative opportunities of the response in the equation below:

$$P(Y \le 1|x) = \pi_1(x) = \frac{\exp(\alpha_1 + \sum_{k=1}^t \beta_k x_k)}{1 + \exp(\alpha_1 + \sum_{k=1}^t \beta_k x_k)}$$
(6)

$$P(Y \le 2|x) = \pi_2(x) = \frac{\exp(\alpha_2 + \sum_{k=1}^{t} \beta_k x_k)}{1 + \exp(\alpha_2 + \sum_{k=1}^{t} \beta_k x_k)}$$
(7)

$$P(Y \le r | x) = \pi_r(x) = \frac{\exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)}{1 + \exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)}$$
(8)

Based on the cumulative opportunities in the above equation, we obtain the opportunity for each category of the response as follows

$$P(Y = 1|x) = \frac{\exp(\alpha_1 + \sum_{k=1}^{t} \beta_k x_k)}{1 + \exp(\alpha_1 + \sum_{k=1}^{t} \beta_k x_k)}$$
(9)

$$P(Y = 2|x) = \frac{\exp(\alpha_2 + \sum_{k=1}^{t} \beta_k x_k)}{1 + \exp(\alpha_2 + \sum_{k=1}^{t} \beta_k x_k)} - \frac{\exp(\alpha_1 + \sum_{k=1}^{t} \beta_k x_k)}{1 + \exp(\alpha_1 + \sum_{k=1}^{t} \beta_k x_k)}$$
(10)
:

$$P(Y = r|x) = \frac{\exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)}{1 + \exp(\alpha_r + \sum_{k=1}^t \beta_k x_k)} - \frac{\exp(\alpha_{r-1} + \sum_{k=1}^t \beta_k x_k)}{1 + \exp(\alpha_{r-1} + \sum_{k=1}^t \beta_k x_k)}$$
(11)

The estimation of ordinal logistic regression model parameters is used the Maximum Likelihood Estimator (MLE). In the MLE method, maximizing a joint probability density function is performed to estimate the regression parameter, which is also called the likelihood function.

In ordinal logistic regression analysis, the response of each observation is assumed to be spread according to the multinomial distribution. If the response variable Y~Multinomial $(y_1, y_2, \dots, y_r; p_1, p_2, \dots, p_r)$, then the function likelihood for the response y is

$$L(\alpha,\beta) = \prod_{i=1}^{n} [\pi_1(x_i)^{y_{1i}} \pi_2(x_i)^{y_{2i}} \cdots \pi_r(x_i)^{y_{ri}}]$$
(12)

By doing natural logarithm of the equation (12), we obtain a log-likelihood function as follows

$$\ln \left(L\left(\alpha,\beta\right) \right) = \sum_{i=1}^{n} \{ y_{1i} \ln \left[\pi_{1}(x_{i})\right] + y_{2i} \ln \left[\pi_{2}(x_{i})\right] + \dots + y_{ri} \ln \left[\pi_{r}(x_{i})\right] \}$$
(13)

A maximum ln-likelihood can be obtained by differentiating L (α,β) against α and β and by equating to zero. The first derivative solution of the ln-likelihood function is a nonlinear function so that a numerical method is needed to obtain the parameter estimation, one of them is Newton-Raphson method [11].

3. Data and Research Methods

The data used in this study come from the questionnaire that was distributed to 60 beginner entrepreneurs. The questionnaire measures for motivation adopted from [12] and the capability of entrepreneurship measurement are extracted from [13]. There are two variables that are used, they are response variables (Y) and predictor variables (X). The level of entrepreneurial capability becomes the response variable that divided into four categories, they are Y = 1 if the entrepreneurial capability is low, Y = 2 if the entrepreneurial capability is middle, Y = 3 if the entrepreneurial capability is high, or Y = 4 if the entrepreneurial capability is very high whereas the predictor variables consisted of 4 variables derived from socio-demographic factors and five variables of entrepreneurial motivation.

The predictor variables are (1) Gender (X1): Entrepreneurs' genders are divided into two categories, X1 = 1 if the entrepreneur is male or X1 = 2 if the entrepreneur is female; (2) Education Level (X2): Entrepreneurs' education levels are divided into 3 categories, X2 = 1 if the entrepreneur does not attend/primary education graduate, X2 = 2 if the entrepreneur attends medium school graduate, or X2 = 3 if the entrepreneur is a high school graduated; (3) Age group (X3): Entrepreneurs' age groups are divided into 5 categories, X3 = 1 if the entrepreneur is 20-30 years old, X3 = 2 if the entrepreneur is 30-40 years old, X3 = 3 if the entrepreneur is 40-50 years old, X3 = 4 if the entrepreneur is 50-60 years old, or X3 = 5 if the entrepreneur is more than 60 years old; (4) Occupation (X4) Entrepreneurs' occupations are divided into 3 categories, X4 = 1 if the entrepreneur is an enterpriser, X4 = 2 if the entrepreneur is a farmer, or X4 = 3 if the entrepreneur is not an enterpriser or a farmer; (5) Entrepreneurial motive (X5); Working motive (X6); (7) Social motive (X7); (8) Individual motive (X8); and (9) Economy motive (X9).

The data analysis steps undertaken in this study are as follows:

- 1) Providing the statistics for the level of entrepreneurial capability variable and predictor variables.
- 2) Establishing the best-guessed model of ordinal logistic regression by involving all predictor variables that affect the level of entrepreneurial capability.
- 3) Testing the overall parameter significances of the ordinal logistic regression model simultaneously

This test aims to investigate the significance of the coefficient β of the response variable simultaneously. The hypothesis of this test is:

H₀:
$$\beta_1 = \beta_2 = \cdots = \beta_r = 0$$

H₁: there is
$$\beta_k \neq 0$$
; $k = 1, 2, ..., n$

The statistic test used is a G-test or Likelihood Ratio Test.

$$G = -2\ln\left[\frac{L_0}{L_M}\right],\tag{14}$$

Where L_0 is a likelihood of the model containing constants only and L_M is the likelihood of the model containing the variables. H₀ is rejected at a significant level α if $G > \chi^2_{\alpha,df}$ with degrees of freedom is *r* or p-value < α . [10]

4) Conducting a partial test between predictor variables and response variables This test is done if the simultaneous test occurs rejection H_0 and the purpose of this test is to determine predictor variables that have a significant effect on the response variable. The tested hypothesis is:

$$\mathbf{H}_0: \, \mathbf{\beta}_k = \mathbf{0}$$

H₁:
$$\beta_k \neq 0$$

This test is performed for each k = 1, 2, ..., r. The statistic test used is Wald's test.

$$\mathbf{W}_{k} = \frac{\beta_{k}}{SE\left(\hat{\beta}_{k}\right)},\tag{15}$$

Where *SE* (β_k) is the standard error of the regression coefficient of *k* and β_k is the expected regression coefficient of *k*. H₀ is rejected at a significant level of α if W_k > $\chi^2_{\alpha,r}$ or p-value < α . [10]

- 5) Perform bootstrap estimate of standard error
 - The algorithm of bootstrap estimate according to [14] is
 - a) Select *B* independent bootstrap samples x^{*1} , x^{*2} , ..., x^{*B} , each consisting of *n* data values drawn with replacement from *x*,
 - b) Evaluate the bootstrap replication corresponding to each bootstrap sample,

$$\hat{\theta}^*(b) = s(x^{*b}), \qquad (16)$$

where b = 1, 2, ..., B

c) Estimate the standard error $se_F(\theta)$ by the sample standard deviation of the *B* replications

$$s\hat{e}_{B} = \{\frac{1}{B-1}\sum_{b=1}^{B} [\hat{\theta}^{*}(b) - \hat{\theta}^{*}(\cdot)]^{2}\}^{\frac{1}{2}},$$
(17)

where
$$\hat{\theta}^*(\cdot) = B^{-1} \sum_{b=1}^{B} \hat{\theta}^*(b)$$

- 6) Creating the best ordinal logistic regression models that loading the significant variables in the partial test only.
- 7) Finding and interpreting the odds ratio for each significant predictor variable.
- 8) Conducting Test Accuracy Model
 - This test is done to determine the measure of data prediction accuracy to know how accurate the model that is formed. The accuracy of the model formed is determined by a measure called the hit ratio, which is defined as follows:

hit ratio =
$$\frac{\text{many objects that are classified correctly}}{\text{total of observations}} \times 100\%$$
 (18)

4. Results and Discussion

In this study, we use the ordinal logistic regression analysis with bootstrap estimation. The response variable is the level of entrepreneurial capability that divided into four categories, namely low, middle, high, or very high. The data used are 60 responses on questionnaires filled by beginner entrepreneurs in tourism of rural area in Nagari Salayo of West Sumatra.

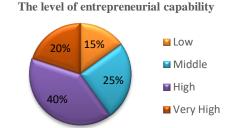


Figure 1. The Level of Entrepreneurial Capability

The graphic in Figure 1 shows that most beginner entrepreneurs' capability has been high. Characteristics of beginner entrepreneurs can be known through descriptive statistics as follows. These analytics are used to find a simple view of the data or the beginner entrepreneurs from all sides.

Variable		Sum	Percentage
Gender	male	24	40,000%
	female	36	60,000%
Education Level	not attend/primary school graduate	2	3,333%
	medium school graduate	38	63,333%
	high school graduate	20	33,333%
Age Group	< 30	8	13,333%
	31 - 40	14	23,333%
	41 - 50	13	21,667%
	51 - 60	14	23,333%
	> 60	11	18,333%
Occupation	entrepriser	15	25,000%
	farmer	15	25,000%
	other	30	50,000%

Table 1. Characteristics of beginner entrepreneurs

The characteristics of beginner entrepreneurs are shown in Table 1. The proportion of female beginner entrepreneurs is more than male beginner entrepreneurs with 60,000%. Meanwhile, the beginner entrepreneurs with medium education have the highest percentage than other education levels, with 63,333%. In contrast, the beginner entrepreneurs' age groups are almost the same, and half of the beginner entrepreneurs are not an enterpriser or a farmer.

After knowing the characteristics of beginner entrepreneurs, the next step is establishing the best-guessed model of ordinal logistic regression by involving all predictor variables and testing the overall parameter significances of the ordinal logistic regression model simultaneously to find out the presence or absence of predictor variables that affect the response variables significantly, the data are tested simultaneously using the G statistic test or Likelihood Ratio Test.

Table 2. Model Fitting Information	1
------------------------------------	---

Model	-2 Log Likelihood	Chi-Square	Df	Sig,
Intercept Only	158,345			
Final	82,795	75,550	14	0,000

From Table 2, we obtain that the test value $G > \chi^2_{0.05, 14}$ and p-value < 0.05. The rejection of H₀ happens because the G test is bigger than χ^2 table. This means that the

response variable is significantly affected at least by one predictor variable. The test is continued with the partial test by using the Wald test because H_0 is rejected.

Variable	Estimate	Wald	Df	Sig,
Entrepreneurial Motive (X ₅)	0,288	4,453	1	0,035
Social Motive (X ₇)	0,527	21,882	1	0,000

Table 3. Parameter Estimates

The significance value of all predictor variables is less than α . Therefore, we can conclude that the variable entrepreneurial motive (X₅) and Social Motive (X₇) affect the response variable.

To increase the confidence interval of the model, Bootstrap estimate of the standard error is performed. Using B = 200 bootstrap sample, each consists of n = 60 observations, the ordinal logistic regression with Bootstrap estimation of standard error approach yield as the following result:

			95% Confidence Interval (CI)			
Predictor	Estimate	Std	Ordinal Logistic		Bootstrap Logistic	
Predictor	Estimate	Error	Regression		Regression	
			Lower	Upper	Lower	Upper
Constant (1)	12,530	3,668	7,128	17,933	7,114	23,101
Constant (2)	15,017	4,196	9,111	20,923	9,509	26,360
Constant (3)	18,161	4,519	11,705	24,616	12,466	31,461
Entrepreneurial Motive (X ₅)	0,288	0,134	0,021	0,556	0,019	0,630
Social Motive (X ₇)	0,527	0,179	0,306	0,747	0,302	0,983

Table 4. Bootstrap for Parameter Estimates

The confidence interval of each predictor variables in Table 4 show increasing. The confidence interval length for an entrepreneurial motive variable increase from 0,535 to 0,611, and a social motive variable, the confidence interval increase from 0,441 to 0,681. The best ordinal logistic regression models:

$$\pi_1(x) = 12,530 + 0,288x_5 + 0,527x_7 \tag{19}$$

$$\pi_2(x) = 15,017 + 0,288x_5 + 0,527x_7 \tag{20}$$

$$\pi_3(x) = 18,161 + 0,288x_5 + 0,527x_7 \tag{21}$$

The next step is to interpret how the significant variables affect the level of entrepreneurial capability. The value of odds ratio can be used to interpret the effect. The value can be seen in the following table:

Table 5. Value Odds Ratio

Variable	Estimate	Exp (β)	
Entrepreneurial Motive (X5)	0,288	1,334	
Social Motive (X7)	0,527	1,694	

Base on Table 5, the odds ratio for entrepreneurial motive is 1,334. This means that each increase of 1 unit entrepreneurial motive will increase the risk of 1,334 times the entrepreneurial capability will be very high. On the other hand, the risk posed by an increase of 1 unit variable social motive is 1,694 times the entrepreneurial capability will be very high because the value of the odds ratio is 1,694.

The last step is to calculate the exact value of the classification between the actual value and the predicted value obtained from the model that has been formed. The model predictions on response categorize eight low entrepreneurial capability, three middle entrepreneurial capability, 17 high entrepreneurial capability, and nine very high entrepreneurial capability correctly. For more details, this classification mistake can be seen in the following table:

		Classification			
Observation	Low	Middle	High	Very High	accuracy percentage
Low	8	1	0	0	88,889%
Middle	1	3	11	0	20,000%
High	2	5	17	0	70,833%
Very High	0	0	3	9	75,000%
Overall percentage	5,369%	2,013%	11,409%	6,040%	61,667%

Table 6. Accuracy of Classification

From Table 6, the accuracy of the classification of the model is 61,667%. It means the model can predict the data of 61,667% correctly, which means that the resulting model is good enough and feasible used to predict the response variable

5. Conclusion

This study explores the development of entrepreneurial capability in tourism of rural area based on local resources and creative economy in Nagari Salayo of West Sumatra. More than half of new entrepreneurs involved in this study are female with 60,000%, and most of them are medium school graduate. While the new entrepreneurs' age is averaged, and half of them are not an enterprise or a farmer.

This study's results show two variables positively have a significant effect on the level of entrepreneurial capability; they are an entrepreneurial motive and social motive. Meanwhile, the variables, gender, education level, age group, occupation, working motive, individual motive, and economy motive do not significantly influence the level of entrepreneurial capability. The variable social motive gives the highest effect on the level of entrepreneurial capability with odd ratio 1,694. This means that each increase of 1 unit of social motive will increase the risk 1,694 times the capability of new entrepreneurs will be very high. While the variable entrepreneurial motive gives 1,334 times affect the level of entrepreneurial capability. Higher variable social and entrepreneurial motive will lead to higher entrepreneurial capability level.

Based on the hit ratio, the model with the significant predictor variable gives 61,667% model accuracy. Thus, it can be concluded that the model that is formed is feasibly used to determine the factors that affect the level of entrepreneurial capability in tourism of rural area in Nagari Salayo of West Sumatra.

References

- M. Li, H. Zhang, and L. A. Cai, "A subcultural analysis of tourism motivations," *Journal of Hospitality & Tourism Research*, vol. 40, no. 1, pp. 85–113, 2013, https://doi.org/10.1177/1096348013491601.
- [2] S. Shane, E. A. Locke, and C. J. Collins, "Entrepreneurial motivation," *Human Resource Management Review*, vol. 13, no. 2, pp. 257–279, 2012.
- [3] A. Plangmarn, B. G. Mujtaba, and M. Pirani, "Cultural value and travel motivation of European tourists," *Journal of Applied Business Research*, vol. 28, no. 6, pp. 1295–1304, 2012, https://doi.org/10.19030/jabr.v28i6.7344.

- [4] F. W. Swierczek, and T. H. Thai, "Motivation, entrepreneurship and the performance of SME's in Vietnam," *Journal of Enterprising Culture*, Vol. 11, no. 1, pp. 47–68, 2003, https://doi.org/10.1142/S0218495803000044.
- [5] M. N. I. Afzal, S. A. Siddiqui, K. M. Mansur, and R. S. Sulong, "An empirical investigation of factors affecting entrepreneurial capability (EC) environment in ASEAN-5 economies," *Asian Academy of Management Journal*, vol. 23 no. 2, pp. 25–44, 2018, https://doi.org/10.21315/aamj2018.23.2.2.
- [6] N. Rametse, T. M. Nganunu, M. J. Ding, and P. Arenius, "Entrepreneurial motivations and capabilities of migrant entrepreneurs in Australia," *International Migration*, vol. 56, no. 4, pp. 217–234, 2018, https://doi.org/10.1111/imig.12452.
- [7] R. W. Lv, C. Lai, and J. Liu, "Entrepreneurial capability scale and new venture performance: The moderating role of entrepreneurship education," *National Social Science Foundation of China*, 2015.
- [8] C. Petchprapunkul, N. Koojaroenprasit, R. Pitipanya, and S. Pleehajinda, "Using logistic regression to identify the entrepreneurial capability of thai agricultural cooperatives members on their farm management," *Cultural and Religious Studies*, vol 4, no. 11, pp. 668–676, 2016, https://doi.org/10.17265/2328-2177/2016.11.002.
- [9] M. J. Silva, and J. Leitão, "What Determines the Entrepreneurial Innovative Capability of Portuguese Industrial Firms?," *Munich Personal RePEc Archive Paper*, no. 5216, 2007.
- [10] D. W. Hosmer and S. Lemeshow, Applied Logistic Regression, New York: John Wiley & Sons, Inc., 2000.
- [11] A. Agresti, *Categorical Data Analysis*, New York: John Wiley & Sons Inc, 1990.
- [12] V. Vijaya and T. J. Kamalanabhan, "A scale to assess entrepreneurial motivation," *Journal of Entrepreneurship*, vol. 7, no. 2, pp. 183–197, 1998, https://doi.org/10.1177/097135579800700204.
- [13] J. A. M. Leon, and M. Gergievski, *Psychology of entrepreneurship: Research and education*, Madrid: Universidad Nacional de Education a Distancia, 2007.
- [14] B. Efron and R. J. Tibshirani, *An Introduction of the Bootstrap*, New York: Chapman and Hall/CRC, 1993.