Exploring the Factors Influencing School Dropout: A Logit Model Analysis

Noaman LAKCHOUCH, Lamarti Sefian MOHAMMED, Mustapha KHALFOUNI
Abdelmalek Essaadi University, Tetouan, Morocco

***Abstract*—School dropout negatively impacts a country's development index, with numerous factors contributing to this complex phenomenon. To investigate the factors associated with school dropout in a specific region of Morocco, a cross-sectional study was conducted, encompassing a weighted sample of 274 junior secondary education students. Data collection was facilitated through a questionnaire administered to school directors. The data processing involved two main stages: preparation and modeling. The modeling phase employed a binary logistic regression model, focusing on the student's dropout status as the dependent variable. The study's findings highlighted several significant factors associated with school dropout: academic performance (as indicated by exam grades), the student's age and gender, and the availability of school transportation services that encourage students to continue their studies. Additionally, while class size also played a significant role, its impact was deemed less critical compared to the other factors identified. These results underscore that school dropout is influenced by a multitude of factors, suggesting the need for targeted interventions to prevent dropout and foster academic success, particularly among female students.**

Keywords*—*School dropout; development index; Morocco; cross-sectional study; logistic regression; academic performance; exam grades; age; gender; school transportation; academic success

I. Introduction

School dropout is a phenomenon that affects many countries around the world. It is defined as the interruption of studies before reaching higher education levels ([1], p. 26). This complex and multifactorial phenomenon is influenced by a set of factors such as poverty, learning difficulties, disinterest in school, school violence, discrimination, distance from school, among others [4]. School dropout has significant consequences for both individuals and society, leading to difficulties in professional integration, health problems, challenges in social and cultural development, and negatively impacting economic growth and social development [4]. According to the data from the Ministry of National Education, the school dropout rate in Morocco has slightly decreased over the past three years. In 2020 to 2021, it was 5.3%, with more than 331,000 students, girls and boys, leaving school. In 2021 to 2022, this rate remained at 5.3%, with 334,664 dropouts recorded. In 2022 to 2023, it slightly decreased to 5.2%, with 327,271 dropouts recorded [2]. This trend is also observed in secondary education in Morocco, where the dropout rate decreased from 2.9% in 2020-2021, with over 114000 students, to 2.8% in 2022 to 2023, with 112,925 dropouts recorded. For a geographical area located in the north of Morocco, there is a slight decrease in the dropout rate among secondary school students, dropping from 3.0% in 2020 to 2021, to 2.8% in 2022 to 2023, with a decrease of one-tenth of a percentage point (0.1%) per year [2]. Despite these encouraging results, many efforts are still needed to improve student retention in the education system. In fact, 90.4% of students enrolled in the first year of primary school reach the end of primary school (with repetition), and 71% of them manage to complete secondary education, compared to 90% in the United States, for example [3]. To better understand the causes of this phenomenon, we rely on personal data from 300 secondary school students, collected through a questionnaire distributed in 10 establishments in the study area. In doing so, we aim to assess the risk of school dropout and identify the influence of students' personal and academic characteristics on their intentions to leave school. These two objectives are achieved by answering the following question: What are the personal and socio-economic factors contributing to school dropout in secondary education?

We conducted our study in several stages: a literature review allowed us to examine the main previous studies. In a section dedicated to tools and methods, we used a logistic model to identify the factors significantly influencing the decision to drop out of school before completing the cycle. In a section devoted to presenting the results of our statistical modeling, we presented the findings of our study followed by a discussion. In conclusion, we provided the final remarks of our study.

II. Literature Review

School dropout is a major issue with significant consequences for individuals, families, and societies. In this study, it is defined as the interruption of education before obtaining a secondary school diploma in Morocco, more specifically within the study area during the 2022 to 2023 academic year. At the global level, the phenomenon remains alarming. According to a UNESCO report (2020), 262 million school-aged children and adolescents were out of school, of which 123 million were girls [4].

Numerous international and national studies have examined the factors contributing to dropout. For instance, [5] highlighted the strong link between school failure and social exclusion. His comparative study between France and the United Kingdom showed that school exclusion is closely related to social exclusion, each reinforcing the other. Similarly, [6] emphasized the multifactorial nature of the phenomenon, pointing to economic, family, school-related, and personal determinants.

In Croatia, [7] identified four categories of dropouts: low-performing students, "silent" disengaged dropouts, maladjusted dropouts with emotional difficulties, and "stressed"

dropouts overwhelmed by academic demands. This typology illustrates the diversity of dropout pathways and the interdependence of risk factors.

In the United States, [8], through a longitudinal study of 6,330 ninth-grade students, demonstrated the importance of both individual and contextual characteristics. His findings showed that gender (boys being more at risk), academic performance, part-time employment, family background, and the socio-economic environment of schools were significant determinants of dropout. Similarly, [9] found that academic difficulties, low self-efficacy, and behavioral issues significantly increase the risk of dropout, especially when combined with contextual factors such as poverty and low parental education.

From a gender perspective, [10] studied early school leaving among girls in Spain, showing that stereotypes, discrimination, violence, and the lack of institutional or family support play a critical role. This highlights how gender acts as a transversal factor that amplifies the effects of other social and school-related determinants.

In Morocco, several studies have also examined this phenomenon. In [11], the authors found that male students from disadvantaged families and those with grade repetitions were more likely to leave school. Similarly, [12], [13] categorized the determinants of dropout into individual, socio-economic, educational, and family-related factors. They noted that socio-economic reasons tend to affect boys more, whereas educational difficulties are more significant for girls. More recent research [14] has further stressed the importance of institutional factors (such as school quality and discipline) and socio-economic conditions (family income, rural environment).

Taken together, these studies confirm that school dropout is a complex and multidimensional phenomenon, shaped by the interaction of individual factors (academic performance, age, gender, risky behaviors, and mental health), family factors (family structure, parenting practices, parental education, socio-economic background), and school-related or social factors (class size, school climate, educational policies, and socio-economic environment). However, several limitations can be identified. Few Moroccan studies employ robust quantitative methods capable of evaluating the relative effects of different variables. Support services such as school transport, canteens, or boarding facilities are seldom analyzed, even though they are critical for keeping students in school, particularly in rural areas. The gender dimension, while mentioned, is often addressed superficially, without deeper exploration of the mechanisms making girls more vulnerable. Finally, many studies are constrained by small samples or rely exclusively on qualitative approaches, which makes it difficult to generalize the findings.

It is within this context that the present study positions itself. By applying a binary logit model to a weighted sample of junior secondary school students in northern Morocco, it provides a rigorous statistical measure of the determinants of dropout. It also highlights the crucial role of school transportation, a factor rarely addressed in previous studies, confirms the differentiated impact of gender, and offers recent empirical evidence to inform educational policies aimed at preventing dropout in Morocco.

III. Materials and Methods

A. Sources of Data

In collaboration with the educational authorities of the study area, a simple random survey (SRS) was conducted to select ten junior secondary schools. Principals from these selected schools were then requested to complete a questionnaire designed to collect data on students who either dropped out or completed their secondary education during the 2022 to 2023 academic year. This effort resulted in the collection of 274 records, representing individual students or observations.

The questionnaire was divided into two primary sections. The first section gathered details about the participating schools, including their location, staff composition, student enrollment numbers, and infrastructure facilities. The second section, designed as a table, collected individualized information on the students selected for the study. This data encompassed sociodemographic aspects such as age, gender, family background, and academic history (including dates of enrollment and withdrawal, instances of grade repetition, and average academic scores). Additionally, it collected information on participation in social assistance programs like boarding facilities and school transportation services. A set of variables was then derived from this collected information to streamline the statistical analysis process, which are outlined in Table I.

B. Sample Weighting (Data Preprocessing)

The student selection process was implemented through a mixed-method survey approach that used both stratified and cluster sampling techniques. Initially, ten schools were randomly chosen, each with an equal chance of selection from a total of 94 schools.

Subsequently, within each of these schools, every student who dropped out during the academic year 2022 to 23 was included in the study. Additionally, a comparable number of third-year students were randomly selected. Thus, the calculation of individual weights for the students proceeded are as described:

Computation of Sampling Weights [15], [16]:

If the student is abandoned:

$$\omega_i = \frac{1}{prob(i \in S_0)}$$

Here, S_0 denotes the sample of selected institutions (schools).

The idea is that the probability of selecting a dropout student depends solely on the probability that their institution was chosen in the sample. Thus, each dropout student observed in the sample represents the inverse of this probability.

For example:

10 schools are chosen out of 94 $\rightarrow$ probability $= \frac{10}{94}$.

Therefore, the weight is $(\frac{94}{10}) = 9.4$.

This means that each dropout student represents 9.4 students in the population.

TABLE I. DESCRIPTION OF VARIABLES

Variable	Description
Statut	A binary categorical variable takes the value 1 if the student leaves school and 0 if not
Std.Tchr	Supervision rate, this is the number of students per teacher (StudentTeacher)
Std.Admin	Number of students per administrator (students/administrator)
Std.Serv	The number of students per support service member
Acad.Rat	The proportion of academy staff to the total number of teachers
Std.Cl	Average number of students per class (rate=students/class)
Std.Cl.Rm	Average number of students per classroom (rate=students/classroom)
Std.L	The students' academic level
Space.H	Number of classrooms at the school
Sprt.F	The existence or not of the sports field at the school (1 if yes 0 if no)
Water	The existence of the water area at the school (1 if yes, 0 if no)
Electr	The existence of the electricity area of school (1 if yes, 0 if no)
Sanit.Serv	The existence of sanitation services (1 if yes, 0 if no)
Age	Age of student
Gender	Gender of student
Rep.Nb	Number of repeats of the student since his/her first access to school
Score	Average mark of the student in three subjects (Arabic and French language, Maths)
Dist.S.H	The distance covered by the student to arrive at the school
Partl.Sit	Parental situation (two-parent (2) or one-parent (1) household)
Luch.Sch	The student is a beneficiary of the school lunch program (1 if yes, 0 if no).
Std.Internat	A derived categorical variable that takes the value 2 for a student who is a beneficiary of the boarding school, 1 if the boarding school exists but the student is not a resident and vice versa, and 0 if there is no boarding school and not a beneficiary.
Trspt.B	The student benefits from the school transportation program (1 if yes, 0 if no)
Tayssir.B	the student benefits from the Teyssier program (1 if yes, 0 if no)

If the student has completed their course

$$\omega_i = \frac{1}{prob(i \in S_0)} \times \frac{1}{prob(i \in S_E)}$$

where, S_E is the sample of students selected within the institution.

Therefore, the probability of selecting a dropout student is simply the probability of choosing their institution; consequently, their probability of belonging to the sample is the same as that of selecting the institution.

In practice, if a student has dropped out, their weight is $\omega_i = (\frac{10}{94})^{-1} = \frac{94}{10} = 9.4$, meaning each selected dropout student in the sample represents 9.4 students. If the student has completed their course, we have $\omega_i = \frac{94}{10} \times \frac{N_e}{n_e}$, where N_e is the number of students in the 3rd year of junior secondary education in institution e and n_e is the number of selected students from the same institution.

C. Methods

1) Model specification: Our aim is to predict whether a student is likely to drop out of school by examining both their personal characteristics and the features of their school environment. Thus, school dropout is viewed as a binary outcome, where a student either discontinues their education prematurely or completes their secondary schooling. This scenario satisfies the prerequisites for binary classification, enabling the use of the binary logistic regression model [17].

Binary logistic regression (LR) is a statistical technique used to forecast the likelihood of a binary outcome, based on a set of independent variables. This outcome is binary on nature, signifying it can assume one of two possible values, such as "yes" or "no", or "success" or "failure" [18].

The LR model is presented as a logistic function, which is a sigmoid function that takes a real number as input and outputs a value between 0 and 1. The logistic function is defined as:

$$Logistic(x) = \frac{1}{1 + e^{-k(x+x_0)}}$$

where, x is a real number. The logistic function (see Fig. 1) has the properties illustrated in the following graph:

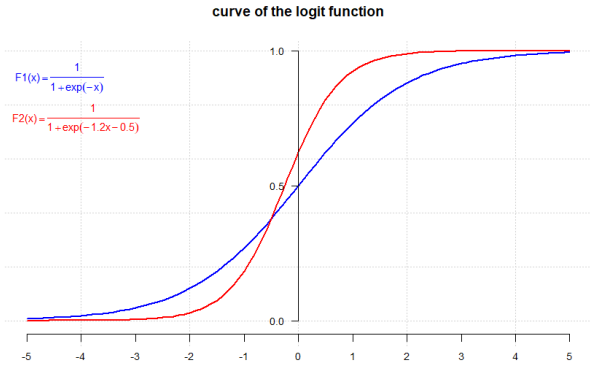

Fig. 1. Graphical representation of the logistics function.

The logistic function is used in the logistic regression model to transform the linear combination of the predictor variables into a probability (Boateng, et al. 2019) [17].

The logistic regression model is defined as:

$$P(Y = 1|X) = Logistic(\beta_0 + \beta_1 X1 + \beta_2 X2 + ... + \beta_3 Xp)$$

where, Y is the binary outcome variable, X1, X2, ..., Xp are the predictor variables, β_0 is the intercept, and β_1, β_2, ..., β_p are the coefficients.

Here is an example of a logistic regression model that predicts the probability of a student passing an exam based on their GPA and the number of hours they studied and gender [Eq. (1)].

$$P(\text{Pass}|\text{GPA, Hours Studied, Gender}) = Logit(\beta_0 + \beta_1 \text{Score} + \beta_2 \text{Hours Studied} + \beta_3 \text{Gender}) \quad (1)$$

Score: The student's academic performance, Hours Studied: The number of hours they studied, Gender: The student's gender, β_0: The intercept and β_1, β_2, and β_3: The coefficients.

The logistic regression model can be used to predict the probability of a student passing the exam for any given Score, number of hours studied and Gender. The model can also be used to identify the factors that are most important for predicting whether a student will pass the exam [19].

2) Model selection: According to the analysis of the AIC index, we can eliminate a certain number of variables and retain the following variables: Std.Cl, Std.Srv, Age.Std, Gender, Rep.Nb, Score.Av, Std.L, and Trspt.B. The AIC value decreases from 63.76 for a saturated (complete) model to 61.57 for a model containing only the mentioned variables [Eq. (2)].

3) Model Validation: The VIF [1] analysis allows us to conclude that there is no collinearity among the explanatory variables (all VIFs are below 5).

Furthermore, the ROC curve (see Fig. 2) analysis on the test data confirms that this model is excellent for predicting school dropout in a student, with an AUC of 0.87.

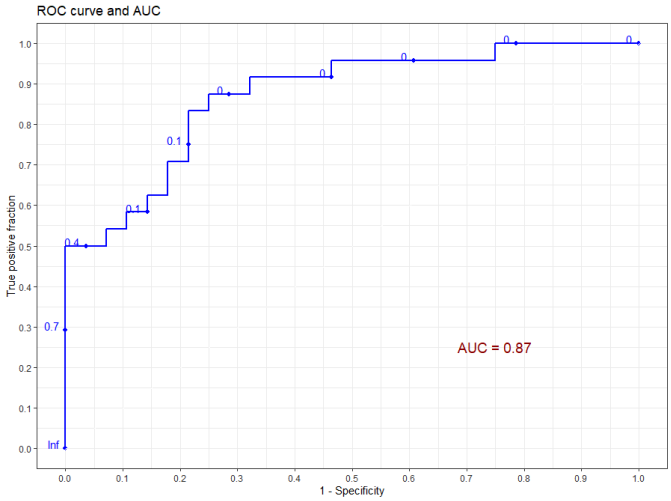

Fig. 2. ROC curve.

Still, on the test data, we perform the Hosmer-Lemeshow test [2] [19]. The test allows us to say that the model is adequate $(Statistic = 8.736, df = 8, p - value = 0.365)$.

[1]The Variance Inflation Factor (VIF) is a measure used to detect the presence of multicollinearity in the model such as:

$VIF = 1$: Indicates that there is no correlation between the independent variable in question and the other independent variables in the model. In other words, there is no multicollinearity.

$1 < VIF < 5$: VIF values in this range are generally considered acceptable. There is some correlation, but it is not high enough to cause serious multicollinearity problems.

$VIF \geq 5$: A VIF of 5 or more indicates multicollinearity that may be concerning and requires attention. This suggests that the independent variable is strongly correlated with other explanatory variables in the model.

[2]The Hosmer-Lemeshow test is a statistical test used to evaluate the goodness of fit for logistic regression models. Specifically, it tests whether the observed event rates match expected event rates in subgroups of the model population. The null hypothesis for the Hosmer-Lemeshow test is that there is no difference between the observed and expected frequencies of the outcome variable

4) Model estimation: After selecting and validating the model on the test data, we can keep the following logistic regression model:

$$P(\text{Statut} = 1|X) = \text{Logit}(\beta_0 + \beta_1 \text{Std.Cl} + \beta_2 \text{Std.Srv} + \beta_3 \text{Age.Std} + \beta_4 \text{Gender} + \beta_5 \text{Rep.Nb} + \beta_6 \text{Score.Av} + \beta_7 \text{Std.L} + \beta_8 \text{Trspt.B}) \quad (2)$$

TABLE II. THE ESTIMATION OF THE FINAL LOGIT MODEL

Characteristic	Beta[1]	Sig.[2]	OR[3]	95% CI (OR)	Sig.	VIF[4]
Std.Cl	0.203	0.062	1.23	0.99–1.52	0.065	4.7
Std.Srv	0.004	0.054	1.00	1.00–1.01	0.054	3.8
Age.Std	0.497	0.019	1.64	1.09–2.48	0.018	3.2
Gender					0.011	1.3
Girl	Ref[5]					
Boy	-0.117	0.012	0.33	0.14–0.78	0.012	
Rep.Nb	0.561	0.023	1.75	1.08–2.84	0.001	2.6
Score.Av	-0.601		0.56	0.45–0.68	0.001	1.7
Std.L	-1.400	0.001	0.25	0.10–0.59	0.001	1.8
Trspt.B		0.019			0.019	1.9
No	Ref					
Yes	-1.432	0.020	0.24	0.07–0.80	0.020	

IV. RESULTS

The results from Table II allows us to conclude that the number of students per service, the age and gender of the student, the number of grade repetitions, the average score in the main subjects (Math, French, and Arabic), the student's academic level, and the use of school transportation are factors that significantly influence school dropout differently. Thus, the average number of students per class has a positive influence $(beta = 0.213, p - value = 0.061)$, as well as the number of students per service and age, which also have positive effects on dropout. Specifically, an additional year in age increases the risk of leaving school by 1.43 times $(OR = 1.43, p - value = 0.042)$, and each additional repetition of a grade increases the chance of leaving school by 1.75 times $(OR = 2.13, p - value < 0.001)$. On the other hand, some factors have negative and significant effects, meaning they decrease the risk of dropping out of school. For instance, being a boy makes the chance of not dropping out of school 1/0.38=2.6 times higher than for a girl, indicating that girls are more likely to leave school than boys. School dropout is influenced by various factors, but some are more important than others. The average score in main subjects is the most crucial factor, with a weight of 5.78. The student's academic level is also important, with a weight of 3.19. Age, gender, the existence of school transportation services, and the number of grade repetitions have moderate importance, with weights ranging from 2.33 to 2.54. The number of students per class does not have a significant impact on the risk of school dropout (see Table III).

V. DISCUSSION AND IMPLICATIONS

As mentioned earlier, the objective of this analysis was to identify the factors that significantly influence school dropout

TABLE III. THE IMPORTANCE OF FACTORS

Factor (Variable)	Weight
Std.Cl	1.845
Std.Srv	1.929
Age.Std	2.372
Gender	2.545
Rep.Nb	2.289
Score.Av	5.788
Std.L	3.191
Trspt.B	2.338

in junior secondary education. Based on a weighted sample of 274 observations, our findings confirm several results from previous studies while also offering new insights specific to the Moroccan context.

Regarding student characteristics, academic performance emerges as the most critical determinant, consistent with the findings of [8], [9]. Students with poor grades or repeated years face a substantially higher risk of dropping out, highlighting the importance of academic support and remedial programs. Age and gender also play a role, though to a lesser extent. In line with [10], our study shows that girls are more likely to drop out than boys when academic performance is weak, suggesting that gender-related vulnerabilities may exacerbate the effect of poor performance. This underscores the need for gender-sensitive policies aimed at keeping girls in school, particularly in rural and disadvantaged contexts.

With respect to school-related factors, our results provide a more nuanced perspective. While smaller class sizes are associated with lower dropout risk, their effect appears modest compared to student performance and support services. This aligns with the mixed evidence reported in earlier research [6]. More importantly, our analysis identifies school transportation as a significant factor in student retention—an aspect that is often overlooked in international studies but is particularly relevant in Morocco, where geographical distance and lack of infrastructure remain barriers to education. By contrast, the provision of boarding schools with catering services did not show a significant impact, suggesting that not all forms of institutional support contribute equally to reducing dropout.

Our analysis also reveals certain limitations. First, the study relied on cross-sectional survey data, which restricts the ability to capture the dynamics of dropout over time. Second, due to institutional constraints, family-related variables such as parental education, household income, and family structure could not be fully incorporated, even though prior research suggests that they are highly relevant [11], [12]. Addressing these limitations in future studies would allow for a more comprehensive understanding of the dropout phenomenon.

Future research should therefore build on this work by integrating family and community-level characteristics, employing longitudinal data to assess causal relationships, and exploring interaction effects between academic performance, gender, and socio-economic status. Moreover, additional attention could be given for evaluating the effectiveness of targeted interventions, such as transportation subsidies, remedial classes, or gender-sensitive programs, to determine which policies are most effective in reducing dropout rates in the Moroccan context.

VI. CONCLUSION

In this study, we analyzed the determinants of school dropout using weighted survey data from junior secondary education students in Morocco. The results show that school dropout is primarily driven by academic performance, with interacting effects of age and gender. In particular, girls with weak academic achievement are more likely to leave school prematurely. Moreover, our analysis highlights the importance of social support services, such as school transportation, which significantly reduce dropout risk by improving accessibility and retention. These findings provide valuable insights for policymakers seeking to design effective interventions against school dropout.

Our results suggest that educational support programs are essential to help students enhance their academic performance, with particular attention to vulnerable groups such as girls and students in rural areas. At the same time, logistical support measures, notably transportation, should be prioritized as effective levers to ensure equity in access to education.

This study nevertheless presents some limitations. It relies on cross-sectional survey data, which, although weighted to reflect the population, cannot capture the longitudinal dynamics of dropout nor the full range of family-level factors such as parental education, income, and household structure. Consequently, the results should be interpreted with caution and cannot be generalized to the entire population without further validation.

Future research should extend this work by incorporating longitudinal data, exploring family and community-level determinants, and assessing the effectiveness of targeted interventions. Furthermore, emerging methodologies such as machine learning and deep learning models offer promising avenues to predict dropout risk with greater precision, identify complex patterns of interactions between variables, and provide early warning systems for students at risk. Combining statistical modeling with artificial intelligence approaches could therefore enhance both the accuracy of dropout prediction and the design of preventive strategies tailored to the Moroccan educational context.

ACKNOWLEDGMENT

The authors would like to thank the school directors and students who participated in the survey, as well as the educational authorities who facilitated access to the data. Special thanks are also extended to colleagues and reviewers for their constructive comments, which helped improve the quality of this work. Any remaining errors are solely the responsibility of the authors.

REFERENCES

[1] La Bouture, Les lycéens décrocheurs : de l'impasse aux chemins de traverse, Chronique sociale, Librairie Mollat, 15 rue Vital-Carles,33 080 Bordeaux Cedex, 1998.

[2] MEN, Annuaire statistique de l'éducation, Ministère de l'Education Nationale, 2022.

[3] W.Bank, Lower secondary completion rate, World Bank, Website, 2022. https://data.worldbank.org/indicator/SE.SEC.CMPT.LO.ZS?locations=MA

[4] UNESCO, Rapport mondial sur l'éducation 2020 : par amour de l'apprentissage, Paris, FRANCE, 2020.

[5] Danielle Zay, Prévenir l'exclusion scolaire et sociale des jeunes. Une approche franco-britannique, Paris, PUF, 2005.

[6] Muller F., & Rocher G., L'abandon scolaire : un phénomène complexe et multifactoriel, Éducation et formation, 77(1), 1-16, 2014.

[7] Ogresta J., & Rezo I., Why do we drop out? Typology of dropping out of high school, Youth & Society, 53(6), 934-954, 2021.

[8] Zvoch, K., Freshman year dropouts: Interactions between student and school characteristics and student dropout status, Journal of Education for Students Placed at Risk, 11(1), 97-117, 2006.

[9] Gençtanır Ö., & Ercan, E., Prediction of school dropout among Turkish high school students: A multilevel analysis, Educational Studies, 58(1), 1-17, 2022.

[10] Rodríguez-Gómez, M. J., School dropout for girls: An intersectional perspective, Gender & Education, 35(2), 202-218, 2023.

[11] BIYOUDA S, et al., Les déterminants sociodémographiques et scolaires du risque de décrochage scolaire chez des collégiens marocains, Revue Française d'Economie et de Gestion, 2(3), 252-266, 2021.

[12] Ait Hammou M., & Bettach A., Causes de l'abandon scolaire au Maroc : une étude exploratoire, Revue Marocaine des Sciences de l'Education, 28(1), 117-134, 2021.

[13] Ait Hammou M., & Bettach A., L'abandon scolaire au Maroc : analyse des facteurs socio-économiques et institutionnels, Economie et Humanisme, (535), 55-64, 2022.

[14] Vinciguerra, A. et al., Les déterminants du décrochage dans l'enseignement secondaire : une revue de littérature, Revue française de pédagogie, (202), 111-130, 2020.

[15] Lohr, S. L., *Sampling: Design and Analysis*, 2nd ed., Chapman and Hall/CRC, New York, 2019.

[16] Bethlehem, J., *Applied Survey Methods: A Statistical Perspective*, Wiley, Hoboken, NJ, 2009.

[17] Boateng, E. and Abaye, D., A Review of the Logistic Regression Model with Emphasis on Medical Research, Journal of Data Analysis and Information Processing, (7), 190-207, 2019.

[18] Hyeoun-Ae Park, An Introduction to Logistic Regression: From Basic Concepts to Interpretation with Particular Attention to Nursing Domain, Journal of Korean Academy of Nursing, 43(2), 154-164, 2013.

[19] Hosmer D.W. and Lemeshow S., Applied Logistic Regression, John Wiley & Sons, New York, 2nd ed., 2000.