RES MILITARIS
REVUE EUROPEENNE D ETUDES EUROPEAN JOURNAL OF MILITARY STUDIES
Social Science Journal

Relying on statistical modeling to identify factors affecting women's practice of sports in Algeria

By

Hachem Amel
University Oran 2 Mohamed Ben Ahmed, Algeria
Email: Hachem.amel@univ-oran2.dz
ORCID ID: https://orcid.org/0000-0003-1197-329X

Bakhtaoui Assia
University Oran 2 Mohamed Ben Ahmed, Algeria
Email: Bakhtaoui.assia@univ-oran2.dz
ORCID iD Link (https://orcid.org/)

Received: 22/08/2023; Accepted:21/12/2023

Abstract

This research study aims to identify the factors affecting Algerian women in practicing sports through the use of one of the statistical modeling techniques, represented by binary logistic regression, and relying on the data of the sixth multiple indicator cluster survey. Moreover, this research shows that the place of residence and the educational level of women have a :positive impact on women taking part in sports, while it was found that the age of women has no effect on their practice of sports.

Keywords: Demographics, Practicing, Social factors, Sport, Women.

Introduction

Sport is one of the most important elements of life today, as doctors recommend it to their patients and emphasize its practice to maintain their health and the safety of their bodies. The practice of sports whether it is for recreational, competitive or health purposes is part of the overall education of the individual (Turki, 1982). It is an effective and necessary tool in the physical, psychological and social upbringing of the individual (Khattab, 1965). Hence, it has become necessary to urge people to exercise in their daily lives. Therefore, there is no difference between the sexes. Sports are essential for both, and females need to take part in sport because they are less mobile. However, it does not depend on whether a person is male or female, but there are many social, economic and cultural factors that directly and indirectly affect the women's practice of sports.

In addition, this research paper is to shed light on the most important socio demographic factors affecting women's practice of sports in Algeria, through the use of bilateral logistical analysis, and the database of the sixth multiple indicator cluster survey (2018-2019). Logistical analysis is one of the most important tests that address qualitative phenomena, and allows the identification of a set of variables called independent variables which explain the qualitative variable, and represents the dependent variable of the study.

. In recent years, sports for women have spread in the Algerian society for various purposes, including health, recreation, competitiveness and aesthetics. However, despite the high rates of women practicing sports, the practice remains weak compared to what is found in Morocco and Tunisia (MICS6, Far Morocco and MICS6 Tunisia).

The demand or reluctance of women to play sports depends on many and varied factors that may be social (residential environment, educational level, environment), economic (standard of living), cultural (cultural level, media), or demographic (age, place of residence, individual and marital status).

Thus, this study seeks to answer this key question:

- What is the socio demographic factors that affect Algerian women's practice of sports?

Through this question, a set of sub- questions are raised:

- What are the most important social factors affecting women's practice of sports in Algeria?
- Are there statistically significant differences between married and unmarried women in terms of practicing sport?
- Are there any statistically significant differences between working and non-working women?
- Are there any statistically significant differences in favor of women under 25 years of age?
- Are there any statistically significant differences according to women's educational level?

The basic hypothesis of this study is that there are sets of socio demographic factors that directly or indirectly hinder Algerian women's practice of sports. From this hypothesis a set of sub hypotheses:

- Hypothesis 1: Statistically significant differences between married and unmarried women in terms of practicing sport
- Hypothesis 2: There are statistically significant differences between working and non-working women with regard to the practice of sport and more particularly Aerobic exercise.
- Hypothesis 3: There are statistically significant differences according to their academic level.
- Hypothesis 4: There are statistically significant differences according to the age of the woman.

This study aims to identify the most important socio demographic factors affecting the practice of sports among women in Algeria, through:

1- Disclosure of differences between married and unmarried women.
2- Disclosure of differences between working and non-working women
3. Detect differences between women by educational level and age.
4- Building a mathematical model to explain the relationship between taking part in sports and socio demographic factors, which helps in the understanding the obstacles facing women to practice sports.

This research helps us to be familiar with the most important sociodemographic factors that affect the women’s practice of sports, and helps to find solutions for those who do not practice sports, as the latter is crucial in order to preserve women's health.

Methods

This study relies on the method of statistical analysis by building a binary logistic regression model, in order to know the characteristics of women who exercise through the data of the sixth cluster survey and test the relationship between exercise and sociodemographic variables. To apply this analysis, we have taken as an error rate $\alpha = 5\%$, through which the decision is taken.

The study sample included 35111 women between 15-49 years old, of whom 3712 took part in sports (10.60%). The MICS6 cluster survey includes 5 questions of a questionnaire which was administered to the participants. This study relied on the database of the family file since it is the only one that includes information on women who practice Sports, and the file of women of childbearing age. Thus, the sample included 3712 women between 15-49 years old.

This study is based on two type of variables.

The Dependent variable:
In this research, the dependent variable is whether women play sports or do not play in the sense of "exercise". It is a qualitative variable of duality; that is to say, it contains only two answers or two qualities, either the woman is practicing sports and we symbolized this with the symbol »1«, or she is not practicing sports and we referred to it with the symbol »0«.

independent variables:
According to the source we are relying on in this study; that is to say, the 6 MICS cluster survey, we extracted a set of variables that have a relationship with the dependent variable under study, some of which are quantitative, such as age, and qualitative, such as the index of wealth, gender, place of residence, civil status, individual status and geographical region.

Building the wealth index:
This indicator is measured by household belongings such as electricity, radio, refrigerator, air conditioner, telephone, car etc..., the quality of the wall, floor and ceiling, the presence of drinking water supplies and the type of sewage. Moreover, this indicator was created using these variables and is found in the sixth cluster survey database. It includes 5 characteristics, which are as the followings: very poor - poor - medium - rich - very rich. In addition, the wealth index is a variable that expresses the differentiation of individuals, their behavior and their attitudes. In general, this indicator is measured by the income, but in the sixth survey database, there is no information on household income. In this study, we grouped the five characteristics of the wealth index, and limited them to three characteristics only such poor, average, and rich in order to facilitate the analysis.

Table 1: *Distribution of the sample by women's exercise and wealth index*

	Wealth Index Women's Exercise	
	Absolute number	**%**
Poor	886	23,87
Average	620	16,70
Rich	2206	59,43
Total	3712	100

Source: *MICS6*

2.2- Data Quality Assessment

This step is necessary and important before starting any analysis in order to assess the quality of the data. The latter allows us to estimate the internal and external coherence of the information. Therefore, the reliability of the results of statistical analysis requires that the data must be of high quality, which means that the non-answer rate must be less than 10%. To assess the quality of the data in this study, we calculated the non-answer ratio for the different study variables. The results were very satisfying.

Table 2: *Distribution of the response rate for the different variables of the study*

Variables	Declarative values	Answer rate %	Missing value
Practice sports	3712	100	00
Place of residence	3712	100	00
Age	3712	100	00
Academic Level	3707	99,87	5
Wealth Index	3712	100	00
Individual status in the last month	3712	100	00
Geographic Area	3712	100	00
Family status	3712	100	00

Source: *calculated from MICS6 data*

From table 2, it is clear that the percentage of the missing values is less than 1%, and therefore logistical analysis can be applied to these data as required by this analysis.

2.3 - The principle of logistic regression:

The aim of the use of this type of analysis is to study the relationship between the independent explanatory variables and the two-value dependent variables of whether women play sports or not, through the MICS 6 database and thus to build a mathematical model that explains this relationship.

The Concept of the logistic regression:

Regression in general is a model that analyzes and explains the relationships between a dependent variable and a set of independent variables that are explained by linking them to a linear mathematical equation (linear regression) and may be nonlinear (nonlinear regression) (Rico, 2009). After determining the form of the relationship, we estimate the parameters of the model that express the extent to which the dependent variable is affected by the independent variables for the purpose of interpretation or prediction according to the nature of the study.

Therefore, Logistic regression is a type of regression in which the dependent variable is qualitative, and may take two values (binary logistic regression) and more than two values (multiple logistic regression (Tuffery, 2010). It should be noted that in logistic regression, our goal is not to explain the change in the values of the dependent variable, but to explain the probability of occurrence and non-occurrence of the phenomenon under study.

Results

Sample characteristics:

Table 3 shows that the percentage of women who exercise is slightly over 10%, and is therefore very low. The results of the study showed that there are statistically significant differences attributed to women between 15 and 24 years ($p< 0.05$). The percentage of exercise among women between 25 years and above is between 4 and 6 percent, while this percentage exceeds 25% among women between 15 and 24 years (25.07%), as they are more likely to exercise compared to other groups.

Table 3: *Distribution of the sample by age and exercise*

Age	Status: Practice sports Number	%	Status: Does not practice sports Number	%	Total	%
15-24	2506	25,07	7491	74,93	9997	28,47
25-34	586	5,37	10330	94,63	10916	31,09
35-49	620	4,37	13578	95,63	14198	40,44
Total	3712	10.57	31399	89.43	35111	100.00

Source: *calculated from MICS6 data*

Table 4: *Distribution of Women by Exercise and Marital Status*

Marital Status	Does Not Practice Sports Number	%	Practice Sports Number	%	Total Number	%
Unmarried	2908	19.77	11803	80.23	14711	41.9
Married or ever Married	804	3.94	19596	96.06	20400	58.1
Total	3712	10.57	31399	89.43	35111	100

Source: *calculated from MICS6 data*

Table 4 shows that the exercise of unmarried women is much higher (19.77%) than by married or previously married women, i.e. divorced and widowed women (3.94%). This is explained by the fact that unmarried women are more interested in their body and fitness.

The data showed that there are statistically significant differences by marital status attributed to single women ($p< 0.05$).

3.2 Building a two-answer logistic regression model:

As mentioned above, we can use logistic regression to study the relationship between the interpreted variables and the two-value dependent variables by building a model that explains the relationship between sociodemographic variables and the variable of women's exercise. Before presenting the results of the logistic regression model, we present the coding of the study variables.

Variables coding:

We used symbols for variables in order to reduce the size of the model tables and thus code for variables as follows:

Y : symbolizes the dependent variable represented by women practicing sports. It is a nominal variable consisting of two adjectives (exercise or not exercise).

X1: represents the age group variable and is a categorical variable and takes three characteristics: 15-24 years, 25-34 years, 35-49 years.

X2 : represents the civil status variable. It is a nominal variant, and takes two adjectives: married and unmarried.

X3 : Represents the place of residence variable. It is a nominal variable that takes on the adjectives urban and rural.

X4: represents the variable of the geographical area, which is also a nominal variable and takes 4 adjectives: center, south, east and west.

X5 : Wealth Index, a nominal ordinal variable that takes three characteristics, rich, poor, and average.

X6 : Individual state variable. A nominal variable that takes two adjectives: working and not working.

X7: The educational level variable takes 5 characteristics: without level, primary, intermediate, secondary and university.

Table 5 *shows the symbols of the dependent variable, where women play sports with 1 and no exercise with 0.*

Dependent variable	symbol
Exercise (practice sports)	1
Do not exercise (Do not practice sports)	0

Source: *calculated from MICS6 data*

3.3- Results of the binary logistic regression model:

The results of the model are in two parts. The first part includes results that help to evaluate the model statistically in order to see how strong the model is in order to explain the phenomenon under study. The second and most important part will be devoted to the results that help us explain the relationship between the dependent variable and the independent variables.

3.3.1- Estimation of the bilateral logistic regression model:

The Overall evaluation of the model:

The results of a model with only a fixed limit (i.e. no interpreted variables) are compared with the model with an interpreted (independent) variable, often to see the explanatory power of a model.

Table 6: *Frequencies and the value of the maximum potential function of the model without interpreted variables*

iterations steps	-2-log-likelihood The value of the greatest possibility function	Parameters of the constant
1	987,751	0.087
2	951,232	0.096
3	896,003	0.098

Source: *calculated from MICS6 data*

This table represents the value of the maximum potential function in the case of the model that contains the fixed limit only, i.e. without the explained variables, where the value of the maximum potential function stabilized at the value of 896,003 at the third stage or step. This value is compared with the model that includes the explanatory variables, if the introduction of explanatory variables leads to a decrease in this value, we say that the explanatory variables affect the dependent variable.

Thus, we move on the examination of the regression equation table after introducing the independent variables.

Table 7: *Frequencies and value of the probability function of the model containing the interpreted variables.*

iterations steps	-2log like lihood	Factor							
		Constant	X1	X2	X3	X4	X5	X6	X7
11	598,785	1.003-	0,645	0,145	0,164	0,215	0,187	0,856	0,077
22	511,986	1,125-	0,787	0,156	0,223	0,253	0,168	1,111	0,093
33	496.356	-1,365	0,848	0,164	0,319	0,321	0,213	1,145	1,042
44	489.963	-1,522	0,890	0,166	0,355	0,347	0,256	1,186	1,088
55	489.526	-1,594	0,892	0,167	0,361	0,422	0,259	1,911	1,097

Source: *calculated from MICS6 data*

After the introduction of the explained variables represented in the age group, civil status, place of residence, geographical area, wealth index, individual status and educational level, we found that they are all statistically significant, and this means that they will improve the model's ability to interpret and predict the dependent variable, which is women's practice of sports.

Therefore, the method of presenting the parameters of the logistic regression model is an iterative method in the sense that the calculation is repeated until the parameter values converge at a specific error allowed and at the smallest value of the maximum potential functions. From this table, the following can be observed:

First: The parameters of the model were stabilized at the fifth step, where we notice the values of this step are the smallest value compared to the values of the previous steps.

Second: The estimation process stopped at the smallest value of the probability function (489.526).

We also noticed that this value is much smaller than the value of the maximum potential function in the case of the model involving the fixed term only (896,003), which indicates a relationship between the dependent variable and the variables explained on the one hand. However, these explanatory variables introduced into the model contributed to improve its ability to interpret and predict the dependent variable on the other hand.

Table 8: *Conciliation quality coefficients.*

-2 log of likelihood	R-two de Cox	Snell R-two of Nagelkerke
489.526	0,487	0,511

Source: *calculated from MICS6 data*

This table represents the coefficients that express the explanatory power of the model. We note that the value of these coefficients is average, but they are considered acceptable in such models.

Table 9: *The Khi 2 test to test the significance of the model with degrees of freedom and their statistical significance is injected into the model.*

		Khi-deux	Ddl	Sig.
Step1	Step	663.673	7	,000
	Block	663.673	7	,000
	Model	663.673	7	,000

Source*: calculated from MICS6 data*

Table 9 tests the significance of the model using the Ki-squared test to test the moral, where:

Null hypothesis H0: The model is insignificant.
Hypothesis alternative H1: The model is moral.

We find that the chi-square value of the table is 663.673, and the p-value of the Ki squared statistic is 0.000, which is therefore significant at the level of statistical significance (0.05). Thus, we reject the null hypothesis and accept the alternative hypothesis, that is, the model is significant and represents data well.

Table 10: *Results of the Hosmer - Lemeshow test for model matching quality (Hosmer - Lemeshow).*

Step	chi squared	DDL	level significance sig.
1	93,421	8	0,067

Source*: calculated from MICS6 data*

The table above shows the results of the Hosmer and Lemeshow test to ensure the quality of the model matching using a statistic to square to test the significant difference between the observed values and the expected values, and thus test the estimated model from which the expected observations were calculated as shown in the following table, Table 10, as we have already mentioned, expresses the quality of a model matching so that:

Null hypothesis H_0: The model represents the data well.
Alternative hypothesis H_1: The model does not represent the data well.

Referring to Table 10 for the quality of model matchmaking, we find that the chi-squared value is 93,421, and the p-value of the chi-squared statistic is 0.067, which is therefore insignificant at the level of statistical significance 0.05 and degrees of freedom 8. Thus, we accept the null hypothesis which represents the data well.

Table 11: *Shows the observed and expected values of the Hosmer and lemeshow Hosmer test.*

		Don't practice sports		Practice sports		Total
		viewing	Expected	viewing	Expected	
Steps 1	1	989	991,26	116	126,11	1105
	2	976	988,14	129	176,16	1105
	3	943	976,25	137	211,43	1080
	4	911	945,66	176	259,78	1087
	5	896	911,43	236	361,02	1132
	6	888	903,76	274	401,85	1162
	7	795	896,12	296	479,61	1091
	8	751	896,45	302	502,64	1053
	9	726	845,76	326	569,85	1052
	10	676	797,01	376	641,21	1052

Source: *calculated from MICS6 data*

Table 11 represents the distribution of observed values in the sense of realism and expected values; that is to say, predicted by the model. These results show that there are differences between the two values.

Table 12: *Shows the parameters of regression, standard error and Wald statistic with degrees of freedom and significance of the model in the presence of explained variables.*

Interpreted variables	Logit value	Standard error	Wald	Degree of freedom	calculated level of significance	The exp(B)	95% confidence interval	
							Lowest	The above
Age Group (Age)	-2,036	0,012	296,56	1	0,000	4,123	0,889	1,007
Individual status	1,786	0,041	195,23	1	0,00	2,008	0,806	1,009
Place of residence	2,012	0,035	124,56	1	0,00	1,817	1,128	1,421
Wealth Index	0,685	0,021	143,625	1	0,00	0,789	1,072	1,356
Educational or academic level	2,315	0,019	301,34	1	0,00	1,512	1,026	1,044
Civil status	1,950	0,011	99,912	1	0,00	2,987	0,939	1,045
Geographic region	0,151	0,017	81,101	1	0,00	0,927	0,137	0,165
Fixed	-4,523	0,055	1001,212	1	0,00	0,007	0,975	1,014

Source: *calculated from MICS6 data*

Table 12 shows the parameters of the optimal model obtained in the fifth cycle and its capabilities, in addition to the standard error and Wald statistic for each parameter with the number of degrees of freedom and their statistical significance.

Through the results recorded in the table, it was clear that all variables were significant. We see a high Wald value for variables. Furthermore, we find that all variables have a positive impact on women's practice of sports, but to varying degrees, except for the age variable, which has a negative impact, meaning that the older the woman, the less she practices sports.

Thus, the binary logistic regression equation of the model can be written as follows:

log $\frac{Pi}{1-Pi}$=- 4,523 + 1,786 Individual status + 2,012 Place of residence +0,685Wealth index +1,950Civil status + 2,315Educational level -2,036Age+0,151Geographical area.

The variables can be arranged according to the degree of their impact on the dependent variable according to the following table:

Table 13: *Order of variables*

Rank	1	2	3	4	5	6	7
Variable	Educational level	Age group (age)	Place of residence	Civil status	Individual status	Wealth index	Geographical region

Source: *calculated from MICS6 data*

It is clearly shown in this table that the educational level is more influential on the dependent variable, followed by the age variable and then after the other variables.

8.3.2 –The Interpretation of the odds ratio (Exp(B):

For the educational level: This variable ranked first in terms of the impact on the variable dependent on women's practice of sports. Women with a high level of education are 1,512 times more likely to play sports than other women (other levels of education).

For age groups: Women under the age of 25 are four times more likely to practice sports (4,123) than other groups (25 to 49 years old).

Place of residence: Women in urban areas are almost twice as likely (71.81) to play sports than rural women.

For civil status: Unmarried women are almost three times more likely (2,987) to practice sports than married, divorced or widowed women.

Percentage Individual case: Working women are twice as likely to exercise than non-working women.

Percentage Individual case: Working women are twice as likely to exercise than non-working women.

For the wealth index: a rich woman is less than one time (0.789) more likely than a non-rich woman to play sports.

Geographical area: Women in the north are almost 0.927 times more likely to exercise than women in the southern and interior regions.

Table 14: *Correct classification of the model.*

	Projected	Expectations: Do you exercise		
		Do not exercise	exercise	the correct percentages
Step 1	Do not exercise	29965	1432	95,43
	Exercise	3211	501	13,50
	total percentage			86,78

Source: *calculated from MICS6 data*

Table 14 represents the correct classification ratio for exercise and non-exercise. The results showed that the model was able to correctly classify 95% of women who do not exercise, while only 13.5% of women who exercise were properly classified. The overall correct rating ratio is estimated at 87%, which is considered a good percentage and indicates that the model has a good ability to predict the dependent variable.

Discussion

The difference between women under the age of 25 and those aged between 25 and 49 in the practice of sport is due to the interest of this group in fitness compared to the rest of the groups on the one hand and their possession of free time, since at this age most of them are unmarried. They therefore have no family responsibilities, so they can find time to exercise. As for the women of the rest of the groups with the responsibility of the family and children, they do not find time for that, especially those who work.

For some women, sports is crucial for them not only for fitness, but also for entertainment and for meeting friends. However, Married, divorced or widowed women generally pay less attention to their physical fitness, and secondly, the responsibilities of the home, children and life circumstances do not allow them to find time to exercise.

The age variable has a negative impact on the practice of sport among Algerian women, which means that the older the woman, the less sport she practices. Unfortunately, in our society rarely does women practice sport after the birth of their first child. Family obligations, lack of time are the reasons that lead women not to practice sport.

Age is not the only variable that has an influence on the practice of sport among the Algerian family, but the level of education, the place of residence, the individual, marital status and the geographical area are the main factors that push the woman to practice sport.

Conclusion

The practice of sports for both sexes and women in particular is one of the basic elements to maintain the health of the individual on the one hand and gain psychological, physical and aesthetic comfort on the other hand. Exercise has become one of the ways doctors recommend their patients to treat a wide range of diseases even before resorting to take medications in many cases.

Generally speaking, the use of logistic regression has proven that women's educational level is one of the most important factors that affect their practice of sports, followed by age, place of residence, civil and individual status, standard of living and finally geographical area. The results also showed that age has a negative impact, i.e. the higher the age of the woman, the less she practices sports.

Therefore, strategies must be developed and aimed at motivating women of all ages and different levels to practice sports of all kinds, as such behaviors would contribute directly or indirectly to reduce the spread of several diseases such as obesity, high blood pressure, strokes, diabetes and other diseases may be due to the less practice of sports, and the matter does not stop at this, but sports contribute to the entertainment of women and improve their psychological state.

The bibliography

Ghanem Adnan and Al-Ja Awni Farid Khalil (2001), 'The use of dual-response logistic regression technique in the study of the most important economic and social determinants of family income adequacy: an applied study on a random sample of households in Damascus Governorate', *Damascus Journal of Economic and Legal Sciences*, Volume 27, First Issue.

Mustafa Rebhi Elayyan et al. (2008),' *Methods of Scientific Research'*, 1st Edition, Dar Safa, Amman, Jordan.

Mohammed Mahmoud Al-Gohari (2009), *Foundations of Social Research*, 1st Edition, Dar Al-Masirah, Amman, Jordan.

Turki Rabeh (1982). 'Educational views of university students. and teachers. and students of the Teacher Training Institute. and those engaged in education and education'. Diwan of University Publications. Algeria.

Khattab Adel Mohammed (1965), 'Physical Education for Social Work', Dar Al-Nahda Al-Arabiya.

Mahdi Muhammad Al-Qassas (2007), Principles of Statistics and Social Measurement, a course issued by the Faculty of Arts, Mansoura University, Dar Al-Multaqiyat, Lectures and Articles, Egypt.

MSPRH (2020); Rapport MICS 6(2018-2019)

Rico Racotomolala (2009), Pratique de la régression logistique. Université Lumière Lyon.

TUFFERY Stéphane (2012), 'Data Mining et statistique décisionnelle'. Edition TECHNIP. Paris.