Review Article

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

Maria Clareth Mendez Ramos, Jesus Sebastian Martinez Sevilla, Sebastián David Fernández Vertel and Melba Liliana Vertel Morinson*

GI Statistics and Mathematical Modeling Applied to Educational Quality, University of Sucre, Colombia

*Corresponding Author: Melba Liliana Vertel Morinson, GI Statistics and Mathematical Modeling Applied to Educational Quality, University of Sucre, Colombia.

Abstract

In Colombia, several studies have shown that the development and integral well-being of children in rural areas has been marked by the high presence of intestinal parasitosis accompanied by negative repercussions on socioeconomic progress, negative effects on nutritional status and low learning conditions of the child population. Taking into account this context, the present research was born which analyzes the sociodemographic and parasitological factors determining the learning capacity and nutritional status in rural schoolchildren, through descriptive-explanatory research which has as an object population of study a rural community of the department of Sucre. The analysis of the information in this study includes techniques that are part of the tools used in data mining, of which decision trees were initially used for the factors included in the study; as a second, Multiple Correspondence Analysis was used as a multivariate description tool of sociodemographic and parasitological profiles depending on nutritional status and learning capacity, Subsequently, Multiple Factor Analysis (AFM) was performed and finally the significant study variables were incorporated into a logistic regression model. All this was materialized through the free statistical software R [1]. As a result, a prevalence of 73.8% of the presence of intestinal parasites in schoolchildren with low weight was evidenced, likewise it was evidenced that the determining risk factor in nutritional status taking as a reference the Body Mass Index-BMI was the deworming variable, while in the case of academic performance in Mathematics it was evidenced that it is determined by age and deworming in schoolchildren.

Keywords: Parasitosis; Learning Ability; Nutritional Status

Introduction

Intestinal parasitosis is currently composed of an endemic health problem with serious repercussions on the socioeconomic development, physical and intellectual state of the child population [2,3]. In this sense, this problem occurs in several developing countries, where there is a great predominance of intestinal parasites associated with the nutritional status mostly visible in some rural areas, due to the low sanitary conditions and quality of life in these regions [4,5].

In Colombia, several studies have shown that anemia, parasitosis and nutrition status influence school performance and also conditions growth, development and determine the optimal realization of the genetic potential of each human being, however, depending on the region there are several factors that condition these problems [6-8]. According to the above, there is a need for a comprehensive study in the rural school population, which answers the following problem question: What are the sociodemographic and parasitological factors determining the learning capacity and nutritional status of rural schoolchildren?

Methodology

- **Type and approach of the research:** This study consists of descriptive-explanatory research with a qualitative approach, which was carried out in a total of 60 schoolchildren from a rural community in the municipality of Los Palmitos-Sucre in Colombia, the analysis was made from a total of 6 groups of variables as shown in table 1.

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

16

Variable	Guy	Class	Codification of the variable
Socio-demographic variables			
Age (E)	Qualitative	Nominal	From 8 to 10 years (E8-10); From 11 to 13 years (E11-13); From 14 to 16 years old (E14-16)
Sex (GEN)	Qualitative	Nominal	Male (M); Female (F)
Family Structure (PE)	Qualitative	Nominal	Extensive (Ext); Nuclear (NUC); Single parent (Mono); Recomposed (Rec)
Variables housing characteristics			
Housing equipment (MV)	Qualitative	Nominal	Artisan (Art); Materials Both (Mam); Prefabricated (Pre)
Housing flat (PV)	Qualitative	Nominal	Tile (Bal); Cement (Ce); Wood (Mad); Earth (Ti)
Water service (SA)	Qualitative	Nominal	Yes (Yes); No
Gas service (SG)	Qualitative	Nominal	Yes (Yes); No
Electricity service (SL)	Qualitative	Nominal	Yes (Yes); No
Sewerage service (SAL)	Qualitative	Nominal	Yes (Yes); No
Garbage Disposal (DB)	Qualitative	Nominal	Open Field (CA); Garbage Truck (CB); Burning (Que)
Disposal of excreta s (DE)	Qualitative	Nominal	Outdoors (Ali); Bathroom (Ba); Septic tank (Fsep)
Animals in housing (AV)	Qualitative	Nominal	Yes (Yes); No
Variables water management and consumption and hygiene habits			
Water Storage (AA)	Qualitative	Nominal	Pool (Alber); Tank (Tan)
Water Treatment (TA)	Qualitative	Nominal	Yes (Yes); No
Wash Fruits and Vegetables (LFV)	Qualitative	Nominal	Yes (Yes); No
Hand Washing Before Eating (LMAC)	Qualitative	Nominal	Yes (Yes); No
Hand Washing After Defecating (LMDD)	Qualitative	Nominal	Yes (Yes); No
Walk barefoot (AD)	Qualitative	Nominal	Yes (Yes); No
Play with Earth (JT)	Qualitative	Nominal	Yes (Yes); No
Desparasitosis (DP)	Qualitative	Nominal	Yes (Yes); No
Variables parasitológicas			
Intestinal parasites (PI)	Qualitative	Nominal	Yes (Yes); No
Parasitosis classes (CP)	Qualitative	Nominal	Monoparasitosis (Mono); Polyparasitosis (Poli); No aplica (Na)
Presence of Endolimax Nana	Qualitative	Nominal	Yes (Yes); No
Presence of Entamoeba Histolíca (EH)	Qualitative	Nominal	Yes (Yes); No
Presence of Entamoeba Coli (EC)	Qualitative	Nominal	Yes (Yes); No
Presence of Giardia sp (Gsp)	Qualitative	Nominal	Yes (Yes); No
Presence of Balstocytis Hominis (BH)	Qualitative	Nominal	Yes (Yes); No
Learning ability			
Academic Performance in Mathematics (RM)	Qualitative	Nominal	5.0 - 4.6 Superior Performance (DS); 4.5 -4.0 High Performance (DA); 3.9 - 3.0 Basic Performance (Dbas); 2.9 - 0.0 Low Performance (DB)
Academic Performance in Humanities (HR)	Qualitative	Nominal	5.0 - 4.6 Superior Performance (DS); 4.5 -4.0 High Performance (DA); 3.9 - 3.0 Basic Performance (Dbas); 2.9 - 0.0 Low Performance (DB)
Nutritional status			
Body Mass Index (BMI)	Qualitative	Nominal	Under 18.5: Underweight (Bp); 18.5 – 24.9: Normal (N);
Hemoglobina (HGB)	Qualitative	Nominal	Under 12.0: Low (B); 12.0 - 18.0: Normal (N); Greater than 18.0: High (A)
Hematocritos (HCT)	Qualitative	Nominal	Under 37.0: Low (B); 37.0 - 54.0: Normal (N); Greater than 54.0: High (A)

Table 1: Coding of variables under study.

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

17

To carry out this research, a parasitological and nutritional evaluation was carried out in schoolchildren who are studying grades from fourth to sixth grades of Basic in a rural community.

- **Phases of the research:** initially blood and stool samples were taken for subsequent analysis by direct examination, with prior signature of the informed consent of their attendants. The nutritional status was determined by means of a screening of anthropometric indices and the learning capacity was obtained through the academic performance in the areas of Humanities and Mathematics of the same provided by the educational institutions. Additionally, a survey was carried out to know the sociodemographic conditions of the students, water management, consumption habits and hygiene.

For the analysis of the data, techniques were used that are part of the tools used in data mining, first measurement decision trees were made taking into account the total study variable, later Multiple Correspondence Analysis was performed to relate the sociodemographic and parasitological profiles with the nutritional status and learning capacity and then Multiple Factor Analysis was performed taking into account the groups of variables and finally a logistic regression analysis is presented to evaluate the associations between potential risk factors and nutritional status and learning capacity. All this using the Data Mining Tool Orange and the R-Project software [1] with the ADE4 packages [9,10] and FactoClass (Pardo and Del Campo, 2007) and Xtable [11].

Results and Discussion

At this stage, decision trees were initially implemented taking into account the study variables, which are presented in the form of nary trees where from a parent or root node n number of children emerges (see graph N°1), these were implemented in the Data Mining Tool Orange tool.

The most relevant sociodemographic and parasitological factors in terms of nutritional status seen from BMI for the 60 individuals evaluated, refer to family structure, performance in Mathematics, deworming, hematocrit levels, deposition of garbage, presence and/or absence of Endolimax Nana and Entamoeba Coli, presence and/or absence of intestinal parasites, housing material and deposition of excreta. In this sense particularly the probability of presenting low BMI is higher when the family structure is single-parent or recomposed, there is performance in basic mathematics, children are dewormed every 6 months or annually, there is the presence of intestinal parasites, normal levels of hematocrits and when the housing material is prefabricated and deposit the excreta in septic or outdoor pools [12].
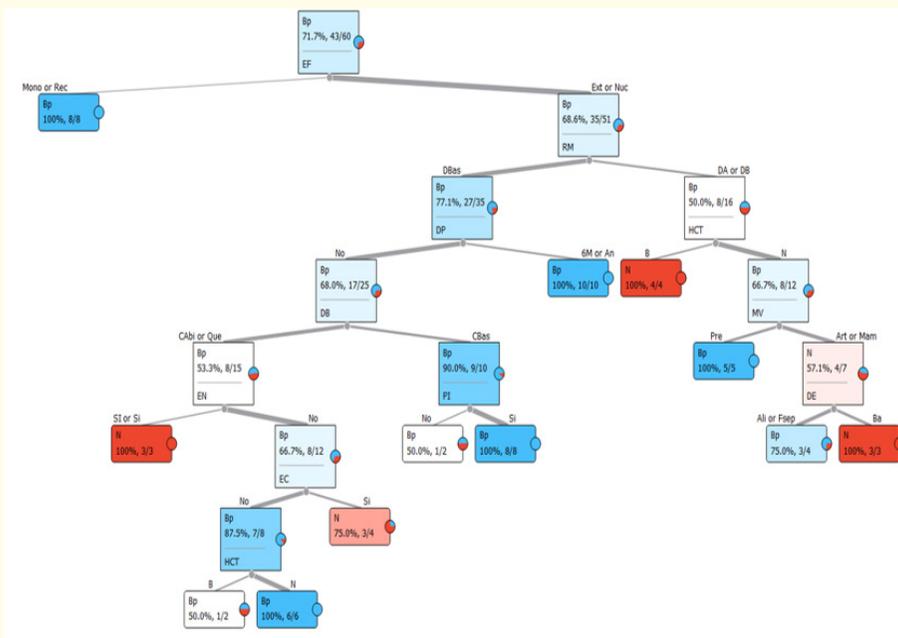

Figure 1: Decision tree sociodemographic and parasitological factors seen from the nutritional status.

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

18

Sociodemographic profile Vs nutritional status-learning capacity

The Multiple Correspondence Analysis-ACM- used to know the sociodemographic profile of the schoolchildren took into account the available information on nutritional status and learning capacity.

Graph N°2 presents the factorial close-up of the MCA of variables such as age, sex and family structure (sociodemographic factor); BMI, hemoglobin and hematocrit level (nutritional factor) and academic performance in Mathematics and Humanities (learning capacity factor) of schoolchildren. The first two axes explain 32.1% of the initial information. The presence of two groups is evident in the population of schoolchildren under study; First group: Low BMI, HGB and HCT at normal levels, academic performance in Humanities and Mathematics at the basic level, single-parent and/or recomposed family structure, male gender, age in the range of 8-10 years. Second group: located in the lower left part of the plane, although a little more dispersed than the first evidence, BMI at optimal levels, associated with high academic performance in the areas evaluated, for the female gender, 11-13 years of age, it is also important to highlight in this group also the association with low levels of HGB and HCT, this may be due to the physical changes to which schoolchildren are exposed specifically at the stage of puberty [13].

Figure 2: P. sociodemographic Vs nutritional status

On the other hand, the factorial foreground of the ACM in graph N°3 shows association between housing conditions, nutritional variables and learning capacity, the first two axes explain 22.7% of the total inertia, where like the previous plane predominates the association between schoolchildren with BMI in normal category, with low levels of HGB and HCT, academic performance in Humanities and Mathematics at the high-higher level, prefabricated material housing, which have light and sewerage services.

Importantly, the evident association between low performance in cognitive areas evaluated and homes made with artisanal material, with absence of drinking water service; it can be an indication of problems related to poor performance and the consumption of non-potable water [14].

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making
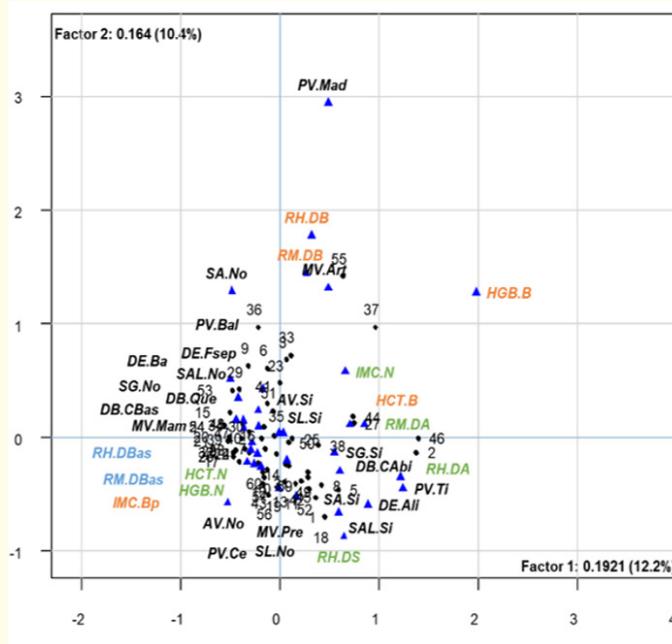
19

Figure 3: P. sociodemographic Vs learning ability.

Parasitological profile vs nutritional status-learning capacity

In order to continue with the analysis of the information in this section, an ACM prepared to know the parasitological profile of schoolchildren is presented, taking into account the variables related to nutritional status and learning capacity.

In the ACM factorial foreground (Graph N°4) for parasitological, nutritional and learning variables, it shows that the first 2 axes explain 31.3% of the total information, it is evident, academic performance at high-higher level (Mathematics and Humanities), normal BMI and low levels of HGB and HCT, which in turn are related to schoolchildren who do not present intestinal parasites. While another group of schoolchildren have parasites and/or intestinal poly parasites with low BMI, likewise, low levels of performance in areas evaluated is related to the presence of the parasite Endolimax Nana [14].

In the factorial foreground of the ACM (graph N°5) as a complement to the information previously presented includes variables related to: water management, consumption habits and hygiene, in which it is evident, association between low BMI, normal HGB and HCT levels, academic performance at basic levels, pool water storage with unhealthy habits such as never having been dewormed, playing with soil, consuming untreated water, walking barefoot, not washing fruits-vegetables before consuming and not washing hands before eating and after defecating. While schoolchildren with normal BMI are associated with high-higher levels in assessed areas and healthy habits.

Global analysis of sociodemographic and parasitological profile vs nutritional status-learning capacity

When performing the global analysis of multiple factor analysis with sociodemographic and parasitological variables. it can be evidenced that the first two factorial axes represent 46.2% of the behavior of the 60 individuals included in the study. From this it is evident that on the factorial plane (Graph N°6), as on the first axis, the groups of socioeconomic variables, characteristics of housing and habits, consumption and hygiene are positioned, while in the case of the second factor it is the group of parasitological variables that have the greatest implication on this axis.
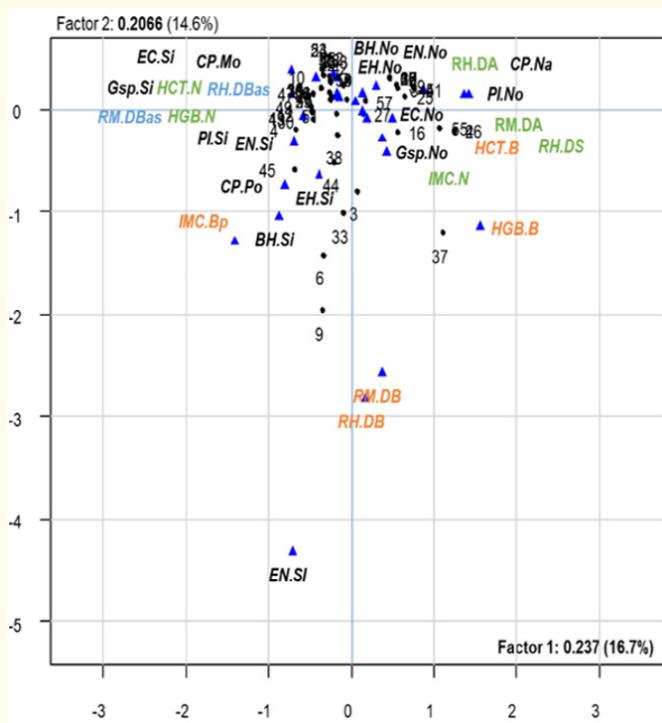
Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

20

Figure 4: P. parasitological Vs nutritional status.
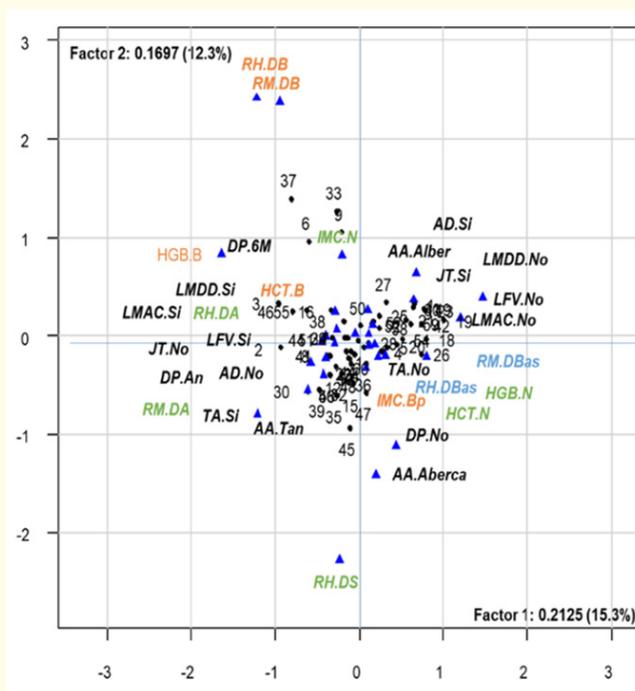
Figure 5: P. parasitological Vs learning ability.

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

21
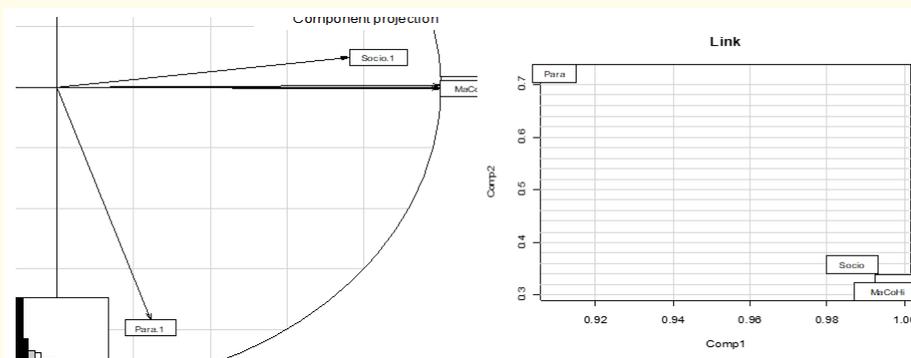
Figure 6: Global analysis of sociodemographic and parasitological factors.

Additionally, a binary logistic regression model was carried out in order to identify determining factors in learning capacity and nutritional status, taking into account the background of the study and in order to test some hypotheses, the variables of gender, drinking water service, presence of intestinal parasites and deworming were chosen in the case of BMI, while in the case of academic performance, the variables of gender, age, presence of parasites and deworming were taken [3-5].

The model is binary logistic regression:

$$P[Y = 1 \mid \mathbf{X}] = \frac{1}{1 + \exp(-\boldsymbol{\beta}^T \mathbf{X})}$$

In this sense, it was evidenced that the risk factor of BMI was the deworming variable (P-value = 0.0447) which evidenced statistical significance and implies that the habit of deworming children every six months or annually influences the nutritional status of the same. While in the case of performance in the area of Mathematics it was evidenced that the variables that help explain the behavior of learning capacity in this area are age (P-value = 0.0156) and also deworming (P-value = 0.0442) [15,16].

Parameters	East	E. E	O. R	Wald Chi	P-value
Gender	-0.60	0.69	0.55	0.75	0.3761
Worming	1.56	0.66	4.76	3.24	0.0447
Intestinal Parasites	0.57	0.87	1.76	0.75	0.3880
Drinking water service	0.44	0,83	1.55	0.28	0.5991

Table 2: Logistic regression model nutritional status (BMI).

Parameters	East	E. E	O. R	Wald Chi	P-value
Constant	0.78	1.55	2.19	0.26	0.6129
Gender	-0.37	1.08	0.69	0.12	0.7299
Age	-2.92	1.48	0.05	3.91	0.0156
Intestinal Parasites	-2.07	1.39	0.13	2.22	0.1364
Worming	-19.75	5206.38	0.0002	0.00014	0.0442

Table 3: Logistic regression model Performance in Mathematics.

Sociodemographic and Parasitological Factors Determining Learning Capacity and Nutritional Status in Rural Schoolchildren: Data Mining for Decision Making

22

Conclusions

In conclusion, an analysis of the sociodemographic and parasitological determinants in the learning capacity and nutritional status in rural schoolchildren of a rural community in the department of Sucre is presented, all through tools used in Data Mining. Some preliminary conclusions are presented from this research

- The probability of presenting low BMI is higher when the family structure is single-parent or recomposed, there is the presence of intestinal parasites, normal levels of hematocrits and when the housing material is prefabricated and deposit the excreta in septic or outdoor pools.
- The association between Normal BMI and normal levels of Hemoglobin and Hematocrits was evidenced in the population of schoolchildren studied, specifically in female schoolchildren in the age range of 11 to 13 years.
- It was found by means of the logistic regression model that with respect to nutritional status the variable that statistically was related to BMI underweight and is evidenced as a risk factor was the deworming variable.
- In the case of performance in the areas evaluated, taking as a reference the performance in Mathematics, statistical significance was evidenced with age and deworming.

Bibliography

1. R Development Core Team. "R: A language and environment for statistical computing". *Vienna: Foundation for Stadistical Computing* (2016).

2. Bundy D. "Inmunoepidemiology of intestinal helminthic infections". *Transactions of the Royal Society of Tropical Medicine and Hygiene* (1994): 254-261.

3. Pollitt E. "Malnutrition, bio-social background and development". *Journal of Neuropsychiatry* (1971): 21-37.

4. Rivero M., *et al.* "Environmental and socio-demographic individual, family and neighborhood factors associated with children intestinal parasitoses at Iguazu, in the subtropical northern border of Argentina. Argentina". *PLOS Neglected Tropical Diseases* (2017).

5. Verhagen LM IR. "High Malnutrition Rate in Venezuelan Yanomami Compared to Warao Amerindians and Creoles: Significant Associations WITH Intestinal Parasites and Anemia". *PLoS ONE* (2013).

6. Fajardo L., *et al.* "Relationship between hemoglobin, iron and ferritin levels and academic performance in a school population". *Revista Colombia Médica* 22.3 (1991): 109-114.

7. Daza C. "Child nutrition and school performance". *Colombia Médica* 28.22 (1997): 92-98.

8. Enriquez C and Segura A. "Risk factors associated with poor academic performance in schoolchildren in Bogotá". *Andean Research* 26.15 (2013): 654-666.

9. Husson F and Pages J. "Exploraty Multivariate Analisis by Example Using R". *Taylor and Francis Group* (2011).

10. Chessel D and Dufour A. "ADE4: Analisis of Evironmental Data: Exploratory and Eucluide and method multivariate*". Lyon, Francia: Data Analysis and Graphical Display* (2004).

11. Dah D. "Package 'xtable' (2016).

12. Peña OG. "Incidence of parasitosis in the nutritional status in school children of the Educational Unit "Julio Enrique Fernández". *UniAndes Institutional Repository* (2015).

13. Barón M., *et al.* "Nutritional status of iron and intestinal parasitosis in children from Valencia, Carabobo State, Venezuela". *Venezuelan Annals of Nutrition* 20.1 (2007): 5-11.

14. Aguin V., *et al.* "Prevalence and relationship between gastrointestinal parasitosis and low academic performance in schoolchildren who attend the Bolivarian school of Jayana, Falcon". *CES Public Health Journal* (2011): 125-135.

15. Amato PR. "Children of divorce in the 1990s". *Journal of Family Psychology* (2001): 355-370.

16. Cervini R., *et al.* "Family structure and academic performance in Latin American countries: Serce data". *RMIE* (2014): 569-597.