



INTENSIVE USE OF STATISTICAL SOFTWARE AS EDUCATIONAL STRATEGY TO IMPROVE DATA ANALYSIS

Javier Alonso Trujillo¹, Abraham Alonso Ricardez¹, José Armando Iruz Manzano², Naomi Maciel Arvizu Vieyra²

¹ Professor of research methodology, Degree in nursing, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México.

² Undergraduate students in nursing, Facultad de Estudios Superiores Iztacala, Universidad Nacional Autónoma de México.

ABSTRACT

A recent study uncovered errors in data analysis carried out in a significant proportion of research conducted by nursing students. Such errors suggest dubious conclusions that highlight educational strategies aimed at getting students to make correct use of statistical software and good data analysis. We think that using the statistical software intensively improves the data analysis. The objective of study was to demonstrate that an educational strategy (ES) improves the quality of data analysis performed by nursing students. **Materials and methods:** Quasi-Experimental study with repeated measurements alternating to ES. Quality of data analysis had three dimensions; General knowledge on research methodology, Integration of knowledge on research levels and Application of knowledge in data analysis. Statistical analysis plan. Measures of central tendency and variability were calculated. The 95% Confidence Intervals were obtained. Comparisons were made for k-means with one-way ANOVA, and Tukey test was applied as post-hoc. Significance level ≤ 0.05 . **Results:** General knowledge about research methodology. Significant differences were observed between the measurements ($p = 0.000$). Post-hoc test tells us that there are three subsets. Integration of own knowledge of investigative levels, significant differences between measurements were observed ($p = 0.001$). Post-hoc test tells us that these differences allow construction of two subsets. Finally, Application of knowledge in data analysis, no significant differences were observed ($p = 0.157$). **Conclusion:** The ES positively influenced on General knowledge and Integration of knowledge but there were no significant changes in Application of knowledge in analysis of data. Initial basic training on use of statistical software should be included in a subsequent study.

KEY WORDS: Data analysis, Statistics, Teaching, Educational strategy, Nursing, Scientific training.

INTRODUCTION:

The scientific training is educational process performed by a teacher, who through some pedagogical model, encourages students to acquire and develop research skills, according to what is established in the graduation profile of current curriculum. During this process, systematic incorporation of up-to-date knowledge, for example, the analysis of data through use of Information and communication technologies (ICT), represents a factor of first importance, not only for investigations themselves, but also to be in accordance with changes that they are occurring in world of work and knowledge (Coriat B, 1996, Rifkin J, 1996, Ruiz C, 1998).

The relevance of information obtained through data analysis begins to be considered a strategic resource in many of human activities in world. Some authors point out that many professions will disappear in next ten years, leaving their place to those who make use of technology and generate information from large amount of data that will be analyzed by those who know how to do it (Oppenheimer A, 2018).

In this sense, university scientific training, specifically in the nursing undergraduate program of the School of Higher Studies Iztacala, UNAM, represents a particular problem. Recent studies about methodological quality of student research work of last three years, showed serious deficiencies in use of statistical tools for data analysis, as well as ignorance of basic concepts, both in methodological aspects and statistics, which impacts on analysis of inappropriate data, discussions and conclusions without scientific basis and what is more serious, students without quality scientific training, but still, with passing grades (Alonso, Alonso, Ordoñez, and Cruz, 2018).

However, this problem would seem to be completely unknown in the discipline at local level. A probable explanation could be the fact that the failure rates and low academic performance are extremely low, which would seem to indicate that everything works well in the module responsible for learning and application of tools for data analysis. An important fact is that, every semester, one or two students out of approximately 120 that integrate each graduation class undergo an extraordinary examination after failing the Research Methodology module. Some voices have stated that for a teacher who does not master his subject, it is more comfortable to approve the student than to expose their limited knowledge. Because in this investigation our object of study is the analysis of data, this will be understood as that research competence that allows consistent application of statistical tools to data sets that have been obtained with methodological rigor and processed using statistical software with the purpose of obtaining scientific conclusions.

State of the art. The school learning problems are often faced with educational strategies once identified, although the factors that cause them are also analyzed. There are some educational theories that emphasize the student as the central axis of learning strategies, which have their origin in the seventeenth century, and it is until the early twentieth century when the idea that the curriculum and

teaching should focus on in the needs, interests and experiences of the students. This position is represented by the new European school and the progressive and democratic education of John Dewey (Bellocchio, 2009, Diaz, 2006, Tobón, 2005).

Nowadays, changes in educational paradigms, fostered by scientific and technological advances in various fields, point towards a restructuring of the meaning of teaching staff and its necessary constitution as a facilitator of learning processes. In this way, the conception of the teacher as a transmitter and the student as a receiver of knowledge is replaced by the teacher's conception as a guide and guidance that accompanies the student (subject of learning) in the process of not only building knowledge, but also development of skills and values associated with an efficient, ethical and responsible professional performance (Alonso, Valadez, Carrasco, and Guzman, 2016, Flores, 2000, López, 2011).

Philosophy training by competences has two central characteristics:

- The conception of the subject in training as an active subject who knows the competences defined in their graduation profile and the criteria for evaluating them, and who has at their disposal the training offers that allow them to achieve them.
- Creation of an educational environment favorable to development of competences through the use of various teaching aids (Tobón, T S, 2005).

Currently, the learning process is undoubtedly nuanced with various theoretical influences. Doctrines such as behaviorism have been overcome to make way for others that raise the importance of the student, teacher, subject and school, all of them integrated. The experiential learning that links significant experiences of the apprentice with new knowledge and on the other hand the experience of not only knowing but doing (learning by doing), have allowed a significant advance in the strategies of university learning, as some recent works demonstrate (Alonso, Alonso, Zamora, and Medina, 2017, Alonso, et al., 2016).

Justification. The present study was carried out by virtue of the problems reported by Alonso et al., (2018, p.3). They identified that the analysis of data made by nursing students during six research forums presents some deficiencies related to the methodological quality and the use of technology. This affects the results that students presented in their research. Some evidences are the following; A significant fraction of the studies evaluated were methodologically simple, 44.1% univariate, that is, only descriptive statistics were applied, 36.9% used samples ≤ 30 units of study. In the studies that methodologically had a higher degree of complexity, 47.4% omitted the hypothesis tests, 23% made incorrect interpretations of the "p" value, 72.3% performed non-probabilistic sampling. In the last three forums, there was a clear downward trend in the use of hypothesis tests: Forum XX: 34.3%, Forum XXI: 21.9% and Forum XXII: 20.0% (Alonso et al., 2018).

This situation has motivated us to try to reduce the deficiencies observed in the

analysis of data developed by students, experimenting with an educational strategy and demonstrating that it has a positive impact. This experiment is congruent with the ethics of the academic profession, since the principle of beneficence is fulfilled, understood as doing good to young people through teaching both in the intellectual and moral dimension; that is, teaching the best and the best ways (López, 2011, López and Félix, 2010).

The objective was to demonstrate that an educational strategy improves quality of data analysis performed by nursing students.

The variable was divided into three dimensions with sequential logic:

1. General knowledge on research methodology
2. Integration of knowledge on research levels
3. Application of knowledge in data analysis

MATERIALS AND METHODS:

Quasi-Experimental study, longitudinal, prospective and analytical. Research design: Quasi-Experimental with repeated measurements alternated to the intervention, as seen in the diagram.

G n = 8 O1 ES O2 ES O3 ES O4 ES O5

The diagram expresses the following:

Group (G): A group of eight students selected according to criteria, will receive their research methodology classes in which an ES is incorporated.

Educational Strategy (ES): Consists of dividing the course into two parts; 48 hours of theory in the classroom and 144 hours of practice (48 hours of field work and 96 hours of intensive use statistical software). During the theory classes, the study of general knowledge on methodology of research and integration of knowledge of research levels proposed by the Hispanic Society of Scientific Researchers, was carried out. During the 144 hours of practice we proceeded to use a teaching support material that included 32 exercises to apply knowledge in data analysis, emphasizing tasks such as specialized software management, application of hypothesis tests, interpretation of outputs and contextual interpretation from the results. ES included 96 hours of intensive use of statistical software Microsoft Excel and Statistical Package for the Social Sciences (SPSS) and 48 hours to build, design, validate and apply its documentary measurement instrument, activity called field work.

Measurements (O1 to O5): The random variable was measured five times.

Each measurement included the specific measurement of each of three dimensions:

1. General knowledge about research methodology (cognitive skills, concepts, methods and strategies). It refers to the understanding of key concepts used in research, methods and strategies.
2. Integration of knowledge on research levels (cognitive abilities of systematization). It is the process that develops the subject in order to logically link the key concepts with their location in the pyramid of the research levels proposed by the Hispanic Society of Scientific Researchers, and that according to some authors, represent some of the cognitive skills, such as example, the systematization of concepts and ideas.
3. Application of knowledge in data analysis (cognitive skills analysis and procedural skills specialized in the management of statistical software) (Rivera et al., 2009). It corresponds to the phase in which the students demonstrate their procedural skills in research and manage to make use of the statistical tools that will allow them, on the one hand, the expression of the results of the analysis made from a set of data provided to them in the frame of a particular problematic, and on the other hand, the statistical and contextual interpretation of results that the specialized software throws during data analysis developed by them.

The three dimensions in which the random variable was divided are part of the inventory of research competences that has been proposed by Rivera some years ago and by Hispanic Society of Scientific Researchers recently (Rivera et al., 2009; Supo, 2017).

Measuring instrument. A questionnaire was applied that included three dimensions of the random variable. The instrument presented content validity (Lara et al., 2006, Prieto and Herranz, 2010, Rivera et al., 2009, Supo, 2014), construct validity (positive item-total correlations) and Reliability (KR-20 > 0.7). The grades obtained in each measurement were adjusted to the scale established to evaluate the academic performance in educational institution and according to the General Regulations of Exams of UNAM, and this scale goes from 0 to 10 points (Universidad Nacional Autónoma de México, 2015).

Ethical aspects. For the students who participated in the experiment, the confidentiality of their qualifications as well as their anonymity was maintained. It was considered ethical that none of students who participated were excluded from receiving the potential benefit represented by ES, so it was not necessary to ask them if they wished to participate or not in the study (López ZR, 2011). The ethics of the academic profession refers to the application of the principle of beneficence, understood as doing good to young people through teaching both in the intellectual and moral dimension; that is, to teach the best and the best ways (López, 2011, López and Félix, 2010).

Statistical analysis plan. Measures of central tendency and variability were calculated. 95% Confidence Intervals were obtained. Comparisons were made for k-means with one-way ANOVA, and Tukey test was applied as post-hoc. Significance level ≤ 0.05

Statistical software Excel 2013 and SPSS version 22 were used (Landero and González, 2009).

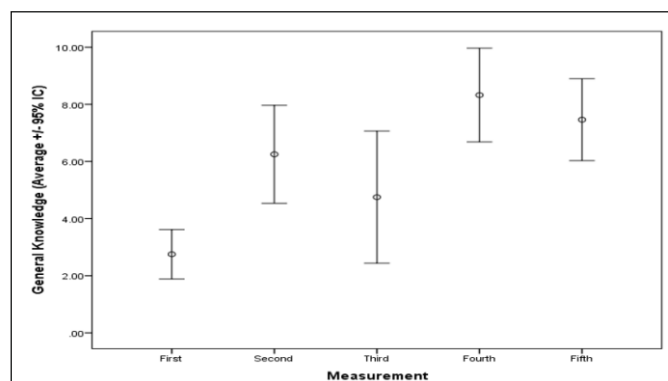
RESULTS:

Starting from the fact of a problem reported by Alonso (2018), in this study an ES was designed and implemented to demonstrate that participating students were able to handle basic concepts of methodology, integrate objectives in pyramid of research levels and apply his knowledge in data analysis.

First, a comparison of five measurements made to "General knowledge on research methodology" dimension is shown. Recall that first measurement was made before implementing the ES, and the rest during the implementation. As can be observed in figure 1, significant differences are observed between measurements ($p = 0.000$). Post-hoc test tells us that there are three subsets. In subset 1 the first and third measurement are located. Third, second and fifth measurements belong to subset 2. Therefore, subset 3 contains second, fifth and fourth measurements. This means that low, regular and good grades can be distinguished, as well as a clear trend towards improving ratings registered for this dimension as implementation of ES progressed (See Table 1). Therefore, it is considered that ES positively influenced the general knowledge about research methodology.

Secondly, the comparison between five measurements made for dimension "Integration of knowledge on research levels" is presented. Figure 2 shows that there were significant differences between measurements ($p = 0.001$). Post-hoc test tells us that these differences allow construction of two subsets. Subset 1 includes first, third and second measurement, while subset 2 includes four measurements; third, second, fifth and fourth. In general terms we can also observe a trend towards improvement of ratings registered for this dimension (Table 2). ES positively influenced the integration of knowledge specific to research levels.

Thirdly, comparison made of the measurements of dimension "Application of knowledge in data analysis" is shown. In this case, no significant differences were observed between records that were made ($p = 0.157$). You can see a slight increase in grades obtained, and it is worth emphasizing that first measurement and third measurement show non-approving grades, while rest of measurements already show approval levels of academic performance in this dimension. From statistical analysis we can suggest that with size of the sample used, it was not possible to demonstrate that ES improved in application of knowledge in data analysis.



Source: Periodic measurements during course.

Figure 1: General knowledge on research methodology. Comparison between measurements made before and during implementation of educational strategy.

Mean ± Standar error

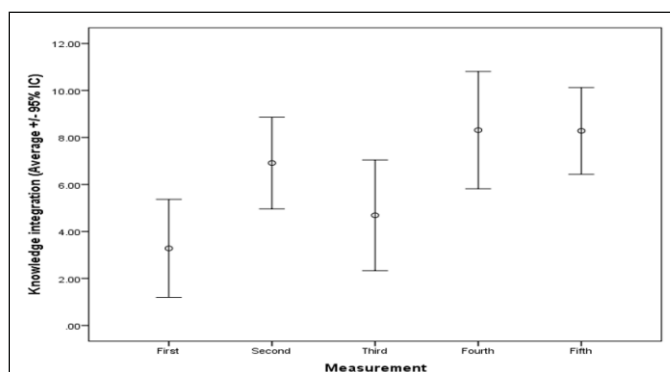
ANOVA: Value F = 9.972; $p = 0.000$

Table 1: Subsets obtained with application of Post-hoc test for General knowledge on research methodology dimension.

Measurements		N	Subset for alpha = 0.05		
			1	2	3
HSD Tukey	First	8	2.75		
	Third	8	4.75	4.75	
	Second	8		6.25	6.25
	Fifth	8		7.463	7.463
	Fourth	8			8.325
	Value p		0.281	0.07	0.248

Source: Periodic measurements during course.

Tukey test, Significance level ≤ 0.05

**Figure 2:** Integration of knowledge on research levels. Comparison between measurements made before and during implementation of educational strategy.

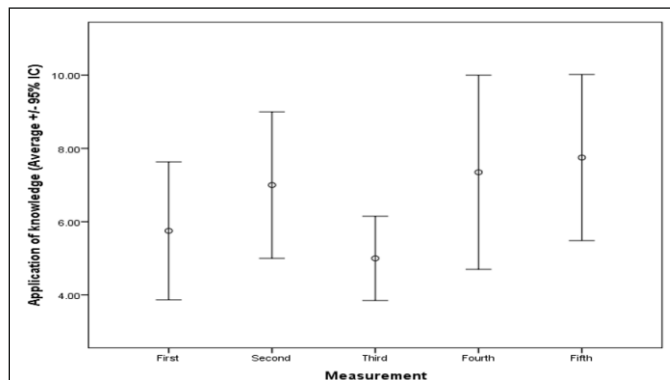
Mean $\pm$ Standar error

ANOVA: Value F = 6.013; p = 0.001

Table 2: Subsets obtained with application of Post-hoc test for Integration of knowledge on research levels dimension.

Measurements		N	Subset for alfa = 0.05		
			1	2	3
HSD Tukey ^a	First	8	3.281		
	Third	8	4.688	4.688	
	Second	8	6.912	6.912	
	Fifth	8		8.281	
	Fourth	8		8.313	
	Value p		0.058	0.059	

Source: Periodic measurements during course.



Source: Periodic measurements during course.

Figure 3: Application of knowledge in data analysis. Comparison between measurements made before and during implementation of educational strategy.

Mean $\pm$ Standar error

ANOVA: Value F = 1,770; p = 0.157

DISCUSSION:

The Educational Strategy applied had the propose of proving whether this strategy improved some aspects related to methodology and data analysis. There are many and very diverse investigative competences that must be developed in the students during their scientific training, however, in this particular case, our ES had a relevant importance during the experiment. The ES focused specifically on strengthening of skills for data analysis. All exercises included in the didactic support material (Alonso and Alonso, 2018) integrated an approach, a context, a set of data and specific questions that required the analysis of data. Our ES positively influenced, albeit partially, quality of data analysis. The obtained results indicate that only in dimensions "General knowledge about research methodology" and "Integration of knowledge on research levels", the ES influenced positively, but not in dimension "Application of knowledge in data analysis" "Knowledge included in dimensions "General knowledge about research methodology" and "Integration of knowledge on research levels", had the characteristic that they were taught in the theoretical part of the ES, reinforced through the support material, the exercises in laboratory and teacher's advice, which could explain positive impact observed in these dimensions.

For dimension "Application of knowledge in the analysis of data", knowledge that characterized it was contained in 32 exercises of didactic material, which should have been resolved during 96 hours of work in computer laboratory, in a collaborative way and with advice of teacher.

Ours results indicate that ES does not have a positive impact on this dimension, which may have been due to fact that students were required to have previous knowledge of statistical software management, as well as punctuality and attendance at sessions in computer laboratory. Another factor that perhaps prevented ES from influencing this dimension positively, could have been poor communication between teacher and student during development of exercises, because in most cases, students prefer to solve their doubts among themselves.

In addition to above, the factor that perhaps influenced more strongly, was that at the time of making measurements of three dimensions, collaborative work was no longer allowed, nor copying answers, nor communication via messages between students and problems to resolve increased their degree of difficulty as semester progressed. During measurements, only didactic support material was allowed to be consulted. The rote aspect may not have had to do with student learning, since algorithms used to run the hypothesis tests or statistical procedures were included in teaching material and could be consulted whenever desired.

Similar results have already been observed in other studies that indicate that when the complexity of the exercise increases and students must solve them individually (without copying or without the help of other classmates), the ability to solve a problem decreases (Alonso, 2015, Alonso, Alonso and Valadez, 2015, Alonso, Cuevas and Alonso, 2017, Pulido, 2009).

Inzunza and Jiménez (2013) have pointed out that even students in the area of mathematical sciences have problems understanding some concepts related to the application of hypothesis tests to data analysis. They suggest that in order to overcome the problem, the teacher must not only master statistical procedures, but must also assume a positive attitude to ensure that student learning is meaningful (Inzunza and Jiménez, 2013).

Medina (2014) observed deficient grades in both cognitive and procedural competences when investigating competences for research in Nursing students (Medina, 2014). On the other hand, Cuevas (2012) states that the contents taught in the Research Methodology module allow the student to successfully develop their skills to investigate, although it does not specify whether it is cognitive, procedural or attitudinal competences (Cuevas, Guillén, Martínez, Ortiz, and Rocha, 2012).

In this work, we assume the position that learning is a shared responsibility between teacher and student. Currently, in the so called knowledge society, data analysis has become a strategic resource that is involved in most of the disciplines that integrate human knowledge. Large amount of data that is generated daily in companies, health services, education, commerce, banking services and others, will require professionals who know technological tools for data analysis (Oppenheimer A, 2018).

For this reason, it is worrying that the dimension least influenced by ES has been precisely that of application. One possible explanation could be that measurement and evaluation of application of knowledge requires procedural skills, such as mastery of statistical software, the management of databases and, above all, reading of the statements inherent to each question. It is not only about reproducing algorithms mechanically, but it is in this dimension, in which students match many of their competences, such as, for example, software management, comprehension reading, statistical reasoning, interpretation of results and contextualization of everything the above. This could be achieved if students take responsibility for their own learning and make sense of it.

At the school, as Savater points out, knowledge learned must fundamentally be one that gives meaning to personal, social and professional life of the student (Savater, 1999). The ES uniformly applied to participants is possibly not compatible

with individual learning style. In addition, personal problems of each student may have influenced impact of ES on quality of data analysis.

CONCLUSION:

We concluded that the educational strategy could not improve the three dimensions in which we divided quality of data analysis performed by nursing students.

Educational Strategy showed benefits in dimensions "General knowledge on research methodology" and "Integration of knowledge on research levels", but there were no significant changes in dimension "Application of knowledge in data analysis". It is suggested that dialogue between teacher and student should be strengthened during application of strategy, and probably initial basic training on use of statistical software should help students to be better prepared to assume educational strategies such as it was proposed in this study.

REFERENCES:

- Alonso, T.J. (2015). Actitud hacia la investigación y logro escolar durante el aprendizaje de la Estadística de estudiantes de Enfermería. *Revista Cuidarte*, 4(8), pp.8–16.
- Alonso, T.J., Alonso, R.A., y Valadez, D.D. (2015). Actitud hacia la investigación científica de estudiantes de Enfermería. *Revista Cuidarte*, 4(7), pp. 24–35.
- Alonso, T.J., Alonso, R.A., Zamora, C.F., and Medina, T.C. (2017). Aprender haciendo: Modelos predictivos en la enseñanza y aprendizaje de Estadística. *Social Science Learning Education Journal*, 2(10), pp. 11–20.
- Alonso, T.J., Cuevas, G.L., y Alonso, R.A. (2017). Uso del ritual de la significancia estadística y su impacto sobre el aprendizaje de la misma. *Revista CuidArte*, 6(12), pp.16.
- Alonso, T.J., Valadez, D.D., Carrasco, Y.M., y Guzmán, G.A.L. (2016). Aprendizaje experiencial aplicado al proceso enseñanza-aprendizaje de Estadística Inferencial. Experiencia docente. Memoria de. VI Encuentro Internacional en la Enseñanza de la Probabilidad y la Estadística (EIEPE), Puebla, México: Facultad de Ciencias Físico Matemáticas de la Benemérita Universidad Autónoma de Puebla.
- Alonso, T. J., Alonso, R. A., Ordoñez, M. A. L., y Cruz, M. L.D. (2018). Methodological quality on academic research reportes from nursing students. *International Educational Applied Scientific Research Journal*, 3(5), pp. 3–6.
- Alonso, T. J., y Alonso, R. A. (2018). Análisis de datos para la investigación científica (En prensa). México: FES Iztacala, UNAM.
- Bellochio, Mabel. (2009). Educación basada en competencias y constructivismo. Un enfoque y un modelo para la formación pedagógica del siglo XXI. México: ANUIES.
- Coriat B. (1996). El taller y el robot. México: Siglo XXI editores.
- Cuevas, G.L., Guillén, C.D., Martínez, C.J.L., Ortiz, L.G.G., y Rocha, R.V.E. (2012). Competencias en Investigación desarrolladas por un grupo de estudiantes de Enfermería. *Revista Cuidarte*, 1(2), pp. 14–21.
- Díaz Barriga Arceo F. (2006). Enseñanza situada: Vínculo entre la escuela y la vida. México: McGraw-Hill.
- Flores OR. (2000). Hacia una pedagogía del conocimiento. México: McGraw-Hill.
- Inzunza, C.S., y Jiménez, R.J.V. (2013). Caracterización del razonamiento estadístico de estudiantes universitarios acerca de las pruebas de hipótesis. *Revista Latinoamericana de Investigación en Matemática Educativa*, 16(2), pp. 179–211. <https://doi.org/10.12802/relime.13.1622>
- Landero, H.R., y González, R.M.T. (2009). Estadística con SPSS y metodología de la investigación. México: Trillas.
- Lara, B.A., Alonso, T.J., Pineda, O.J., Ibarra, Ch.L., López, S.G, y Paredes, B.L. (2006). Programa del módulo de Metodología de la investigación en Enfermería. FES Iztacala, UNAM.
- López, Z.R. (2011). Ética de la profesión académica. Valores del profesorado en la sociedad del conocimiento. México: Universidad Autónoma de Sinaloa / Juan Pablos editor.
- López, Z.R y Félix, S.V. (2010). Ética y cultura en el profesorado universitario. En: Perfiles culturales de la formación universitaria. México: Juan Pablos Editor, UAS.
- Medina, A.G.I. (2014). Competencias en investigación en estudiantes de la licenciatura en Enfermería de la FES-I UNAM. Tesis de Licenciatura. UNAM, Facultad de Estudios Superiores Iztacala. México.
- Oppenheimer, A. (2018). Sálvese quien pueda. El futuro del trabajo en la era de la automatización. México: Penguin Random House. Grupo editorial.
- Prieto, V.L., y Herranz, T.I. (2010). ¿Qué significa estadísticamente significativo?. España: Ediciones Díaz de Santos.
- Pulido, J.E. (2009). Enseñanza de la Estadística a partir de la actitud del alumno. *Laurus*, 15(30), 42–70.
- Rifkin, J. (1996). El fin del trabajo. Nuevas tecnologías contra puestos de trabajo: El nacimiento de una nueva era. México: Paidós.
- Rivera, H.M, Arango, P.L, Torres, V.K, Salgado, B.R, García, G.M.F, y Caña, D.L. (2009). Competencias para la investigación. Desarrollo de habilidades y conceptos. Trillas / Universidad Simón Bolívar.
- Ruiz, C. (1998). El reto de la educación superior en la sociedad del conocimiento. México: ANUIES.
- Savater, F. (1999). El valor de educar. México: IEESA; SNTE.
- Supo, C.J. (2014). Seminario de investigación científica. Arequipa, Perú: Bioestadístico.com. Retrieved from Bioestadístico.com; 2014. Recuperado a partir de: <http://seminariosdeinvestigacion.com/sinopsis>
- Supo, J. (2017). Criterios para elegir un procedimiento estadístico. Arequipa, Perú. Retrieved from <https://www.youtube.com/watch?v=gPNUKsKz4mw>
- Tobón, T.S. (2005). Formación basada en competencias. Pensamiento complejo, diseño curricular y didáctica. Bogotá: ECOE ediciones.
- Universidad Nacional Autónoma de México. (2015). Reglamento General de Exámenes. UNAM. Retrieved from <http://www.ddu.unam.mx/index.php/reglamento-general-de-examenes>