SciencePG
Science Publishing Group

Establishments of Empiric Intervening Formulas Methods, Procedures and Mathematical Elementary Algorithms: Medical Applications

Luis Alberto Escalona-Fernández

Department of Medical Information Technology, Faculty of Medical Sciences. Holguín University of Medical Sciences, Holguín, Cuba

Email address:
luisalbert@infomed.sld.cu

Abstract: *Background*: updating methods, procedures and mathematical elementary algorithms, constitute a tool of work for the reliability of the results of any study, and in medical applications. *Aims*: confirming the behavior of the experimental intervening data the modelation, through the linear regression equations, according to the methods, procedures and mathematical elementary algorithms, the ones that they are not necessary in knowledge and abilities of differential calculus. *Method*: they use the theoretic methods: Analysis synthesis, induction deduction and abstraction concretion. Process of understanding, explanation and interpretation. Methods, procedures and mathematical algorithms, as well as statistical methods and information-technology professional programs are applicable. *Results*: empiric formulas follow on from a mathematical model, which confirms the behavior of the experimental data by means of the simulation, a medical problem gets worked out, the correlation coefficient is checked for sampling, a confidence interval of the linear correlation coefficient estimates itself, helped in information-technology professional programs. *Conclusions*: they indicate methods, procedures and the mathematical elementary algorithms to construct a model, which confirm the behavior of the experimental data, and theoreticians for simulation, as from nature and the logical order of the study, helped for information-technology professional programs.

Keywords: Experimental Data, Mathematical Elementary Methods, Medical Applications

1. Introduction

It is important to understand, explain and interpret the experimental data, by means of a model to establish, which verifies the information, the nature and the logic of the data, according to the experimentation for the collection, processing, presentation and interpretation of the processed data, from the efficient use of mathematical and statistical tools designed for these purposes [1-3].

It is necessary to master the concepts and definitions, methods, procedures, and mathematical algorithms, which facilitate the handling of information for the interpretation of the scientific problem that is solved in a study. [1-3]

Use as examples those that involve parameters and statistics known to students, such as the linear correlation of random variables, and examples carefully selected for their simplicity and importance according to the results obtained from the available scientific literature [1-3]. The importance of these variables, and the relationships to be established, are assumed from the components: academic, labor and research. [1]

It is necessary to have professional problems, in which statistical inference techniques are used so that they value as a result making the decisions, [1-3] to be followed in diagnoses and therapies to be applied in individuals, families and communities, where they provide their services of Health.

In general, significant guidelines and regulations are specified and clarified for the optimal development of the training of health professionals in undergraduate and

graduate studies, from the didactic and scientific-methodological point of view.

In the teaching-learning process, patterns should not be set in the way of proposing and solving different situations by way of examples, exercises, and problems to avoid stereotypes, which are transmitted among the teachers of a discipline, for years of work, based on the results obtained.

Therefore, the following scientific problem is defined: How to establish a linear regression model for medical experimental data, based on the mastery of the contents of elementary mathematics, without resorting to the classical knowledge and skills of differential calculus?

The objective of this work is to confirm the behavior of the experimental data represented, according to elementary mathematical methods, procedures and algorithms, in which the knowledge and skills of differential calculus are not necessary.

2. Materials and Methods

Theoretical methods are used: analysis - synthesis, induction - deduction, abstraction - concretion. The method to construct curves of elementary functions, [4-6] comprehension, explanation and interpretation processes. Mathematical and statistical methods, procedures and algorithms are applied.

The empirical methods applied are mainly framed in participant and non-participant observation, in order to delve into the problem and obtain information on the main difficulties that affect the teaching-learning process, from the classes, as forms of teaching organization. Socialization workshops are held to assess the viability of the effectiveness of the proposed didactic alternative.

They select 10 professors from the discipline of Health Research Methodology of the Department of Medical Informatics of the Holguín Medical Sciences University are nominated, the intended sample according to their work experience in teaching, and interviews are carried out with open questions to identify: The beliefs, opinions, and motivations of health professionals about the usefulness of inferential statistics in professional practice. The following difficulties are revealed:

Variations in the way of analyzing and using the concepts studied are not always favored in the different teaching activities in undergraduate and graduate studies in order to stimulate creative thinking in medical problem situations that require it.

1. There is no consolidated work strategy to delve into alternatives that develop creative capacities of health professionals.
2. Little use of professional statistical packages to solve professional problems and ratify the results obtained.
3. The creation and elaboration of situations that break with the schemes of traditional teaching, based on the systematization and integration of undergraduate and postgraduate content, is insufficient.

A documentary review of the programs of the discipline Methodology of Health Research is carried out, in general the curricula of Medical Sciences, previous and current, the programs of the Master's Degree in Primary Health Care, scientific review of books, articles, and monographs referring to inferential statistics.

3. Results

The objectives proposed in different programs, the methodological guidelines, stimulate the creative thinking of teachers and students, thus stimulating the development of mental capacities, given the bilateral nature of the teaching-learning process.

It is important to create health situations to be faced by the General Practitioner in his professional work that show the need to interpret the association between random variables that describe the evolution of processes in individuals, families, and communities to consolidate the knowledge and skills related to the Concepts addressed in inferential statistics, from experimental data.

The analysis of the following problems is proposed, which are previously oriented, which guarantees the active and creative participation of health professionals. [7-9]

The empirical formulas to be established are obtained in the updated scientific literature on the application of the contents of Higher Mathematics; However, this discipline is not taught in the Medicine career, which is why it constitutes a challenge, the search for an alternative from the basic mathematical knowledge and skills to deduce these equations from the least squares method (MCM) [10-12].

Problem 1. Establish the empirical formulas for the points $(x_i; y_i)$ located approximately on the line r, defined by the equation $y_i = a + b \, x_i$, the ordinate and abscissa of the points $(x_i; y_i)$ with i=1, 2,…, N; They are obtained as results of experimental measurements that are examined as data, for the coefficients a and b of the equation of the line, which constitute unknown quantities to be determined.

The following illustrates how medical problems are solved using the MCM. [12, 13]

Problem 2. Given the results of the measurements of the magnitudes of doses in milligrams (mg) of a certain drug, and the reduction of the heartbeat per minute (beats/min).

a. Determine the regression equation.
b. Mean the regression equation according to the study.
c. What is the dose for a reduction of 13 beats/min?
d. Establish the rate of change of the reduction in heartbeat and the dose of the drug in patients.
e. Check the error of the experimental data and the regression equation [14, 15].

It is important to check the results obtained in the study sample at the population level, for which the following is proposed:

Problem 3. Verify the correlation of the results, from contrasting the following statement: The correlation coefficient is less than or equal to 0.70 at the population level with a significance level of 5%. Given the following measurements of the dose (mg) of a drug to reduce

(beats/min) of the heart in patients, based on a randomly selected sample, making a comparison [15-20].

It is not enough to confirm the results, it is revealing to specify between which values the linear correlation coefficient is found, for other samples of that population.

Problem 4. Estimate a 95% confidence interval for the correlation value in the population of the Reduction (beats/min), according to the dose (mg) of the drug, proposed in problem 3.

Table 1. *Results of the measurements of the measurements of the dose (X) and the beats for minutes (Y).*

Records	Dosis $(X)_{mg}$	Reduction $(Y)_{beats/min}$	Square (X^2)	Product $X*Y$
1	0.5	10	0.25	5
2	0.75	8	0.5625	6
3	1	12	1	12
4	1.25	12	1.5625	15
5	1.5	14	2.25	21
6	1.75	12	3.0625	21
7	2	16	4	32
8	2.25	18	5.0625	40.5
9	2.5	17	6.25	42.5
10	2.75	20	7.5625	55
11	3	18	9	54
12	3.25	20	10.5625	65
13	3.5	21	12.25	73.5
$\sum$	26	198	63.375	442.5

4. Discussion

Problem 1: The differences of the ordinate are given by the equation

$$y_i - a - bx_i = \varepsilon_i \tag{1}$$

To determine the coefficients a and b in such a way that the errors are in absolute value as close to zero (as small as possible); thus the sum of the squares of the errors is as small as possible, it is designated

$$U = \sum_{i=1}^{n} \varepsilon_i^2 \tag{2}$$

In other words, the sum of the errors in absolute value is minimal. When substituting, the following expression is displayed:

$$U(a,b) = \sum_{i=1}^{n}(y_i - a - bx_i)^2 \tag{3}$$

The difference of the function U, which depends on a and b, around the point $(a_0; b_0)$ follows as:

$$U(a_0 + (\pm\delta_1); b_0 + (\pm\delta_2)) - U(a_0; b_0) \tag{4}$$

values positive reals δ_1 and δ_2. The difference is developed ; [9-11]:

$$U(a_0 + (\pm\delta_1); b_0 + (\pm\delta_2)) - U(a_0; b_0) \tag{5}$$

thus the following addends are obtained:

$$2(na_0 + (\sum x_i)b_0 - (\sum y_i))(\pm\delta_1) + 2((\sum x_i^2)b_0 + (\sum x_i)a_0 - (\sum x_i y_i))(\pm\delta_2) + 2((\sum x_i)(\pm\delta_1)(\pm\delta_2) +$$

$$n(\pm\delta_1)^2 + (\sum x_i)(\pm\delta_2)^2 \tag{6}$$

reaches its minimum value, if the difference:

$$U(a_0 + (\pm\delta_1); b_0 + (\pm\delta_2)) - U(a_0; b_0) \geq 0 \tag{7}$$

the experimental data are chosen in such a way that

$$\sum_{i=1}^{n} x_i \geq 0 \tag{8}$$

the product coefficients and the squares of δ_1 and δ_2 are positive real values, which are eliminated to reduce the difference of the increment of the function at the point (a_0, b_0), by imposing the required conditions, cancel the linear coefficients of δ_1 and δ_2. The following system of linear equations is obtained (9):

$$\left.\begin{array}{l} na_0 + (\sum_{i=1}^{n} x_i)b_0 = \sum_{i=1}^{n} y_i \\ (\sum_{i=1}^{n} x_i)a_0 + (\sum_{i=1}^{n} x_i^2)b_0 = \sum_{i=1}^{n} x_i y_i \end{array}\right\} \tag{9}$$

The formulas for the coefficients a_0 and b_0 are determined. When

$$\frac{\sum_i^n x_i}{n} \neq \frac{\sum_i^n x_i^2}{\sum_i^n x_i} \tag{10}$$

then, the system of equations has a unique solution, which is possible because they are conditions that the researcher imposes on his experimental design, and is part of the scientific rigor, to guarantee that the information collected provides the equation of the mathematical model, which adjusts the results obtained.

Sometimes given the nature, and the behavior of the logical order of the experimental data, do not allow to impose these required conditions; In other words, there is not always a straight line that optimizes the results obtained by experimentation, that is, it does not have a square error close to zero.

The inequality

$$U(a_0 + (\pm\delta_1); b_0 + (\pm\delta_2)) - U(a_0; b_0) \geq 0 \tag{11}$$

holds for any positive real values δ_1 and δ_2. That is, the function U (a; b) that depends on the independent variables a and b, has a

$$U_{mínimo\ relativo}(a_0; b_0) \leq U(a_0 + (\pm\delta_1); b_0 + (\pm\delta_2)), \tag{12}$$

which is classified as an absolute extreme. Comparing to the proposal accomplished by Escalona-Fernández [6, 10].

To carry out calculations and checks, it is recommended to use computer programs, such as Liber Office Calc, which provides valuable help. The experimental data are organized in tabular form, as presented in problem 2, in accordance with the formulas to be used. A file must be previously created with the records of the information to be processed.

In the worksheet, the data option is activated, then options are displayed, when pointing in the last panel of the "Statistics" list, options are offered with a double click, access to "Regression", in this the columns with the registered data x_i, y_i; Thus, the values of the "Slope" and the

"Intersection" are obtained, that is, the numerical values of the coefficients of the linear regression line. Other options allow determining the sample correlation coefficient, T, F, Z tests; Chi Square, among others.

Problem 2: a. Record 6 in the table 1:

$$\sum_{i=1}^{n=13} x_i = 26; \sum_{i=1}^{13} x_i^2 = 63.375; \sum_{i=1}^{13} y_i = 198;$$
$$\sum_{i=1}^{13} x_i y_i = 442.5 \tag{13}$$

They substitute themselves in (9):

$$13*a_0 + 26*b_0 = 198,$$

$$26*a_0 + 63.375*b_0 = 442.5$$

The condition

$$\frac{\sum_i^n x_i}{n} = \frac{26}{13} = 2 \neq \frac{\sum_i^n x_i^2}{\sum_i^n x_i} = \frac{63.375}{26} \tag{14}$$

is checked to guarantee the existence and the uniqueness of the solution, it is obtained approximately that: $a_0=7.05$ and $b_0=4.09$; are the coefficients of the regression equation, which is defined as:

$$\text{Reduction }_{beats/min} = 7.05 + 4.09 * (\text{dose }_{mg}) \tag{15}$$

b. It is established that the reduction in heart beats in patients is 4.09 times the applied dose, and 7.05 beats per minute are added to the reduction due to other factors not considered in the study.

c. Then, $13_{beats/min} = 7.05 + 4.09 * (\text{dose }_{mg})$, then dose $_{mg} = (13 - 7.05) / (4.09)$, that is, the dose is equal to 1.45 mg of the drug for a reduction of 13 beats per minute.

d. Increasing the dose of the drug applied to patients by one milligram (1 mg) reduces approximately 4 (4.09) beats per minute.

e. The experimental data are evaluated in the (theoretical) regression equation, the calculated error is expressed, this is almost zero, therefore, the efficiency of the established mathematical model is ratified, the real and theoretical results are compared by simulation.

$$\sum_{i=1}^{13} \varepsilon_i = \sum_{i=1}^{13} (y_i - 7.05 - 4.09 x_i) \approx 0,00000000000001 \tag{16}$$

Making a comparison [11].

An effective behavior of the drug is manifested, from different doses, in different patients, according to the study measurements.

The ratification of the regression model is related to the verification of sufficient conditions, based on the necessary conditions imposed. On the contrary, when the error is not small, the regression model: Linear, logarithmic and potential does not fit the experimental results.

Problem 3: 1) Hypothesis formulation:

H_0: $\rho \leq 0.70$, the linear correlation at the population level is less than or equal to 0.70.

H_1: $\rho > 0.70$, the correlation is greater than 0.70.

2) The significance level is selected $\alpha=0.05$

3) The T-student distribution is used for the calculated

statistic: $T_{cal}=4.85$

4) $T_{0.95;\ 12\ GL}=2.18$. As a criterion statistic (the statistician) belongs to the critical domain (because $T_{0.95;\ 12GL}=2.18 < T_{cal}=4.85$). Therefore the null hypothesis is rejected.

5) It is convenient to assume the correlation coefficient at the population level is greater than 0.70 at the population level with a significance level of 5% and 12 degrees of freedom.

Problem 4: It is known that to determine the confidence interval for ρ:

$$[Z_{\frac{\alpha}{2}} < \frac{Z_r - Z_\rho}{\frac{1}{\sqrt{n-3}}} < Z_{1-\frac{\alpha}{2}}] = 1 - \alpha. \tag{17}$$

Where:

$$Z_r \text{ asymt} \sim N\left(\frac{1}{2}\ln\frac{1+\rho}{1-\rho}; \frac{1}{n-3}\right) \tag{18}$$

For $Z_{r=0.95}=1.83$

The confidence interval:

$$(Z_1 = Z_r - (Z_{\frac{\alpha}{2}})\frac{1}{\sqrt{13-3}}; \ Z_2 = Z_r + (Z_{1-\frac{\alpha}{2}})\frac{1}{\sqrt{13-3}}) \tag{19}$$

The values are determined ($Z_1 = 1.21; Z_2 = 2.45$)

$$(\rho_{inf} = \frac{e^{2Z_1}-1}{e^{2Z_1}+1}; \ \rho_{sup} = \frac{e^{2Z_2}-1}{e^{2Z_2}+1}) = (0.84; 0.99) \tag{20}$$

Making a comparison [15-20].

With a confidence level of 95% and 12 degrees of freedom, the true value of the correlation ρ in the population is between the values 0.84 (lower limit); and 0.99 (upper limit), that is: $0.84 < \rho < 0.99$.

In general, it is important to establish an association between random variables, whose conditions to be imposed meet the requirements demanded by the researchers, according to the design of the experiment, and the aspects to be checked, which are illustrated in the solutions to problems 1, 2, 3 and 4; These are basic considerations to be followed as an alternative to the proposed didactic strategy.

The group of teachers, from the bank of problems created to develop the topic referred to inferential statistics. It proposes a didactic strategy to analyze and use the studied concepts applied to professional problems, in order to stimulate creative thinking.

The alternatives to develop creative capacities in teachers and health professionals, as well as the problems to be elaborated must be based on health situations to be faced by the General Physician, previously oriented; the results of which are verified by professional computer programs.

5. Conclusions

The basic consideration of the didactic strategy that is proposed as an alternative to be followed in this work indicates:

1. Methods, procedures and elementary mathematical algorithms to build a linear regression model, which confirm the behavior of experimental and theoretical

data by simulation, according to the nature and logical order of the study.

2. The empirical formulas are established by the MCM, in which the knowledge and skills of differential calculus are not necessary, these procedures; and formulas are applied to medical problems, supported by professional computer programs, the sample correlation coefficient is verified, which is estimated by an interval at the population level.

It is recommended to carry out methodological and collective workshops with the professors of the discipline Methodology of Health Research to deepen the design of the didactic strategy that is proposed in order to stimulate creative thinking, and perfect the bank of problems, as well as updating methods, and statistical inference techniques.

References

[1]   MINSAP. Statistics. Master's Degree in Primary Health Care. Havana, 2006.

[2]   De-la-Horra J. Applied Statistics. Editorial Díaz de Santos S. A. Madrid, Spain [Internet]. 2016 [cited 2018 28 Mar]: [approx. 165 p.]. Available at: https://estadisticaunicaes.files.wordpress.com/2012/05/uned-estadc3adstica-aplicada-julic3a1n-de-la-horra.pdf.

[3]   Efimov A, Korakulin P, Pospelov P, Terésechenko A, Vokolov V, Zemskov Y, et al. Problems of Higher Mathematics. Volume III. Editorial Mir. Moscow [Internet]. 2016 [cited 2017 Apr 14]: [approx. 607 p.]. Available at: https://es.scribd.com/document/353534585/Problemas-de-Las-Matematicas-Superiores-III-Efimov-Demidovich.

[4]   Koroliuk VS. Manual of probability theory and mathematical statistics. Editorial Mir, Moscow [Internet]. 2016 [cited Apr 14, 2017]. Available at: https://www.casadellibro.com/libro-manual-de-la-teoria-de...y.../183104.

[5]   Escalona-Fernández LA, Velázquez JR. Solving optimization problems without the use of limits and derivatives. Medical interpretations. México, DF: Draft Latin American of Educational Mathematics; Volume 25 [Internet]. 2012 [cited 2019 May 3]. [approx. 7 p.]. Available at: http://www.clame.org.mx.

[6]   Numa M, Martín A, Diéguez R, Sánchez A. University statistical training aimed at solving professional problems. University Pedagogy. 2014; 19 (1): 30-47. Available at: http://www.cvi.mes.edu/peduniv/index.php/peduniv.

[7]   Infante YP, Bayés E. Proposal for a Biostatistics exercise booklet. MEDISAN [magazine on the Internet]. 2016 [cited 2018 Sep 3]; 20 (12): [approx. 7 p.]. Available at: http://www.medisan.sld.cu/index.php/san/article/view/167.

[8]   Zimmermann W, Cunningham S. Visualization in Teaching and Learning Mathematics (editors) [Internet]. 1991 [cited 2018 22 Feb]: [approx. 27 p.].(Mathematical Association of America, Notes, 19). Available at: http://www.sciepub.com/reference/106471.

[9]   Escalona-Fernández LA. Construction and interpretation of curves of elementary functions and the resolution of optimization problems to solve medical professional problems. Thesis as an option to the scientific degree of Doctor in Pedagogical Sciences, University of Holguín. Cuba. 2019 [accessed 2020 January 10]; [approx. 216p.]. Available at: http://tesis.sld.cu/index.php?P=Home.

[10]   Kudriávtsev VA, Demidóvich BP. Brief Course of Higher Mathematics. Editorial Mir, Moscow, 2017, [approx. 608p.].

[11]   Barnett R, Ziegler M, Byleen KL. Applied Calculus for business and economics, life sciences, and social sciences, 13th Edition. San Francisco, California, United States of America, Dellen Macmillan. 2015 [cited 2019 December 16]; [approx. 670p.]. Available at: https://www.pearson.com/us/higher-education/product/Barnett-Calculus-for-Business-Economics-Life-Sciences-and-Social-Sciences-13th-Edition/9780321869838.html.

[12]   Lipkus IM, Peters E. Understanding the role of numeracy in health: proposed theoretical framework and practical insights. Health EducBehav. 2009 [cited 2019 Feb 22]; 36 (6): 1065-1081. Available at: https://doi.org/10.1177%2F1090198109341533.

[13]   Rebollar A, Ferrer M. Teaching based on problems and exercises: a didactic conception to stimulate the learning management of the teacher and the student. Athens. 2014, [cited 2019 December 29]; (2014).2 (26). Available at: http://www.redalyc.org/html/4780/478047202003/.

[14]   Bayarre H, Couturejuzón L. Correlation and Simple Linear Regression. Havana: National School of Public Health. 2004. Available on CD.

[15]   Fliess J L. Statistical Methods for Rates and Proportions. Second Edition. New York. John Wiley and Sons; 1981

[16]   Armitage P, Berry G. Statistical Methods in Medical Research. 3rd ed. Oxford: Blackwell Scientific Publications, 1994.

[17]   Daniel WW. Bioestadística. Base para el análisis de las ciencias de la salud. 3ra edición. México. D. F: Limusa; 1997.

[18]   Norman GR, Streiner DL. Bioestadística. España: Hartcourt Brace; 1998.

[19]   Dawson-Saunders B, Trapp RG. Bioestadística. 2da Edición. México: El Manual Moderno, 1999.

[20]   Swinscow, TVD. Statistics at square one. 9th edition. BMJ publishing group; 1997.