Science Journal of Applied Mathematics and Statistics
2022; 10(2): 15-21
http://www.sciencepublishinggroup.com/j/sjams
doi: 10.11648/j.sjams.20221002.11
ISSN: 2376-9491 (Print); ISSN: 2376-9513 (Online)

The Normal, Chi-squared and Student's T Distributions in the Teaching of Biostatistics for Students of Medical Sciences

Rafael Mauro Avila Avila[1, *], Maria del Carmen Exposito Gallardo[2], Julio Cesar Pino Tarrago[3]

[1]Department of Mathematics, Faculty of Informatics and Mathematics, Holguin University, Holguin, Cuba

[2]Department of Informatics, University of Medical Sciences, Holguin, Cuba

[3]Civil Engineering, Manabi South State University, Manabí, Ecuador

Email address:

ravilaa62@gmail.com (R. M. A. Avila), mariaex@infomed.sld.cu (M. del C. E. Gallardo), jpinoecuador@gmail.com (J. C. P. Tarrago)
[*]Corresponding author

To cite this article:

Rafael Mauro Avila Avila, Maria del Carmen Exposito Gallardo, Julio Cesar Pino Tarrago. The Normal, Chi-squared and Student's T Distributions in the Teaching of Biostatistics for Students of Medical Sciences. *Science Journal of Applied Mathematics and Statistics*. Vol. 10, No. 2, 2022, pp. 15-21. doi: 10.11648/j.sjams.20221002.11

Received: December 21, 2021; **Accepted**: January 14, 2022; **Published**: April 8, 2022

Abstract: The normal, Chi-squared and Student's t distribution are three of the most important in the teaching learning process of Biostatistics for the specialty of Medicine. However, the absence of a deductive approach in the introduction to the topics, makes difficult to understand its origin. From the starting point of the systems of continue distributions of probability, it is proposed an alternative way to introduce the normal distribution in general and standard form, as well as the analytic expressions for the Chi-squared and Student's T distributions. Some epistemic gaps in the treatment of the thematic are identified by means of the analytic synthetic and documentary review methods. The essential objectives consist of the deduction the mathematical expression of such distribution as well as to value the possibility to introduce the basic elements of the used procedure, in the program of Biostatistics, emphasizing the notions of Differential Calculus developed in previous stage to the Integral Calculus. The essential conclusion is associated to the contribution of the suggested approach to the rigor and harmony of mathematical statistic knowledge in the discipline, although some concepts of Mathematical Analysis are necessary in order to facilitate the understanding of the practical applications in the career of Medicine and in other branches of the research in biomedical sciences.

Keywords: Biostatistics, Normal Distribution, Pearson's Distributions, Teaching Learning Process

1. Introduction

The normal distribution is perhaps the most important in classical statistics. This theoretical model named second Laplace's law, laplacian distribution, Gauss-Laplace distribution or simply Gauss distribution, requires a detailed knowledge because it is essential not only in the context of theory of measurement errors. Many random variables in medical and biological sciences satisfy this model, therefore the analysis of its behavior is a starting point in the context of Inferential Statistics [1].

The fundamental role of Gauss-Laplace distribution is based on the fact that under broad assumptions, as the number of summands increases, the distribution is asymptotically normal. This is linked with limit central theorem whose importance in this context can be summarized as follow: if the results of a random experiment is determined by a large number of random factors and the influence of each of them is so small that it can be neglected, then such an experiment is successfully approximated by a normal distribution, thus the mathematical expectation and variance are appropriate [2].

There are several approaches to the study of this distribution. Some classic books on probability theory deal with the topic in detail and rigorously [3, 4]. In other cases, the theme is introduced from the historic perspective in association with the works of Carl Friedrich Gauss (1777-1855) on theoretical foundations for the analysis of physical measurement errors. Some authors define the distribution in a more formal way in which there is no reference to its origin [5].

Freund presents the topic taking into consideration the behavior of a continuous distribution which approximates to the binomial in the case where the number of trials n is large and the probability of success is close to one-half (1/2) [6]. On the other hand Chistiakov introduces the concept of normal distribution from the definition of an absolutely continuous probabilistic space [7]. In that approach, any non-negative function, whose integral over the real axis, that is, from minus infinity (-∞) to plus infinity (+∞), is equal to 1, constitutes the distribution density of a certain random variable.

Torres introduces the main concepts related to the normal distribution in Biostatistics and Medicine by means of notions of Integral Calculus and geometric reasoning [1]. The density function is discussed in terms of distributions of theoretical frequencies using the idea of functional dependence with respect to a certain random variable x. However the mathematical expression of the aforementioned distribution is not deducted; instead of this, the expression is presented as function of x and the population parameters mean (μ) and standard deviation (σ). Other essential contents for the further understanding of statistics topics as Chi squared and Student's T distributions are discussed in essential features, although some epistemic gaps are detected.

The main objective of the paper is to deduct the density function of the normal distribution from one of the most studied systems of univariate continuous distributions, introduced by Karl Pearson (1857-1936). Other objectives consist of: to value the possibility of introducing this function using the notions of Differential Calculus, before explaining those associated with the definite integral, so that it may be useful in the teaching of Biostatistics and in the didactic work of the teachers; to discuss the organic relation of that functions with other distributions as Chi squared and Student's T as well as to formulate some recommendations in the teaching learning process of the normal distribution and its applications in the course of Biostatistics for students of Medical Sciences.

The research was carried out based in a methodology which took into consideration the qualitative paradigm represented by the document analysis [8, 9]. On this basis, the detailed revision included: study plans, different normatives of the Ministry of High Education (MES), analytical programs of several disciplines and matters for the specialty of Medicine, the wide recommended bibliography and documents prepared as part of the teaching material. The experiences accumulated by the main author during the teaching-learning process of the contents of Descriptive and Inferential Statistics at the Holguin Medical University, were a valuable source to observe the assimilation of the topics.

The analysis and synthesis made possible the exhaustive processing of the collected information while the historical-logical method was useful in identifying the main ideas about the genesis and evolution of the concepts, especially those related to the normal distribution. The deduction as an own method of Mathematics served as resource to obtain the mathematical explicit form of the normal probability density function using Pearson's distributions as a starting point. The specific type within these last distributions required the calculations of the first four moments and the corresponding parameters.

2. Normal Distribution and Integration: Difficulties for Its Understanding

The concept of integral is one of the most important in Mathematical Analysis. In particular, the definite integral of a function of one real variable and its interpretation, favor the understanding of ideas associated with the use of Statistics in biomedical sciences. However, the student of Medical Sciences at the undergraduate level exhibit a poor understanding of this concept. On the other hand, they have an incipient knowledge about other important concepts of the Higher Mathematics as function, limit, continuity, Differential Calculus, Integral Calculus in general as well as the relationship between the operations of derivation and integration of functions of one real variable.

The notion of definite integral, its geometrical interpretation as area under the curve and its direct relation with the concept of probability is not evident for the students. The teachers' efforts during the exposition of the topic are focused on convincing them about the following fact: with the indefinite increase of the number of observations the histogram and the frequency polygon approach to the ideal curve named theoretical model of probability distribution. Thus, the notion of limit appears in a subtle way, but it is not introduced; instead, it is preferred only to mention the term.

In addition to the aforementioned fact, the teachers reaffirm another: the referred approximation implies that the area of each rectangle is close to the theoretical frequency or probability that the value of an observation of the random variable is in that interval [1]. A direct consequence of this is the interpretation of area under the curve in an interval (a, b) as probability that the variable takes values between a and b. The conclusion that follows from all this is that the total area is nothing more than the probability that such a variable always adopts a value among the whole set of those observed, so that this probability is equal to unity.

The important concept of function f(x) is implicit in the above reasoning. This provides a good opportunity to discuss such a concept and not only present it as a mean of expression of theoretical distributions. The next step in the discussion of this topic is more formalized since the concepts of area, integral with its symbol and function are presented in the following terminology: the definite integral of function f(x), between the values x=a and x=b coincides with the area bounded by the graph of f(x), and the lines x=a and x=b and its symbolic representation is:

$$P(a < x < b) = \int_{x=a}^{x=b} f(x)dx \tag{1}$$

Then, the probability of the sure event is associated to the whole area under the curve, corresponding to the so called normalization condition given by the following expression:

$$\int_{-\infty}^{+\infty} f(x)dx = 1 \tag{2}$$

One more time the notion of limit appears when the lower and upper integration limits are represented by the symbols $-\infty$ and $+\infty$ respectively. In spite of the valuable and considerable effort to promote the understanding of such knowledges, a sort of rupture occurs when the expression for the calculation of probabilities is presented as follows:

$$P(a < x < b) = \int_a^b \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{1}{2}(\frac{x-\mu}{\sigma})^2} dx \tag{3}$$

The above formula is well known by mathematicians but not by the future physicians who might identify the presence of the density of normal distribution as the function f(x) in the integrand. As the technique to calculate such an integral in order to find the area under the curve requires a deeper knowledge of the subject and special skills, the emphasis is oriented to the parameters of the distribution μ and σ. However, the problem represented by the calculations, is replaced by the suggestion of using tables to determine the probabilities that a certain value of x is contained between the limits a and b. On the other hand, it is suggested the use of other properties of the integral, which are somewhat esoteric for future professionals of the Medicine in a first knowledge of the subject.

The mathematical form of the standard normal distribution is established by definition as in [10]. It is not so difficult for the students to substitute the values μ=0 and σ=1 and thus to obtain the following density function:

$$f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}, -\infty < x < +\infty \tag{4}$$

The general above approach of Torres and other authors, does not refer to any procedure to deduct the concrete mathematical form of the density function [1]. In this way, the formulas that represent the density distribution seem to be obtained by magic and overshadow the beauty and logic of the Mathematics. Therefore, this represents an epistemological gap that breaks the deductive coherence and the exposition of the teaching material and consequently, the minimum mathematical knowledge is offered to medical students in an incomplete manner.

One way of dealing with continuous probability distributions is the use of distribution systems. These satisfy differential equations in such a way that the concrete distributions are obtained deductively. In this way, it is possible to systematically apply methodologies for the generation of families of distributions and to obtain common characteristics for these families [11].

3. Deduction of the Normal Distribution Density Function

3.1. Normal and Pearson's Distributions

The family of distributions introduced by Pearson between 1894 and 1895 provides a flexible description of different distributions. These are commonly identified in biometric investigations. Pearson introduced new classes in 1916. The differential equation defining the Pearson distributions, written as a quotient between two polinoms of n and m degrees, looks as follows, as in [2]:

$$\frac{df(x)}{dx} = \frac{P_n(x)}{P_m(x)} f(x) \tag{5}$$

where:

$$P_{n=1}(x) = a_1 x + a_0 \tag{6}$$

$$P_{m=2}(x) = b_2 x^2 + 2b_1 x + b_0 \tag{7}$$

In the above expressions: a_0, a_1, b_0, b_1, b_2 are the parameters of the distribution which are completely determined with the first four central moments. The most general approach to moments is based on the concept of mathematical expectation E(x). If the random variable X is discrete and it takes values X_1, X_2,……, X_k with probability P_1, P_2,….., P_k respectively, the mathematical expectation is defined as the sum of the products of the values taken by the variable by their corresponding probabilities, that is:

$$E(x) = p_1x_1 + p_2x_2 + \cdots + p_nx_n = \sum_{k=1}^{n} p_k x_k \tag{8}$$

The equation (8) is modified for a continuous random variable x with probability density f(x); in this case, the sum is changed by the integral and the result is:

$$E(x) = \int_{-\infty}^{+\infty} xf(x)dx \tag{9}$$

The k-th moments of x and the k–th central moment of x-E(x) (if the mathematical expectation exists) can be expressed by means of integrals in the case of continuous random variables. As a result of the calculation of such integrals it follows that: the first moment (k=1) of x is precisely the mathematical expectation $M(x)^1 = E(x)$, while the second central moment (k=2) of x-E(x) is the variance: $M(x - M(x))^2 = D(x)$.

The standard deviation is given by $\sigma = \sqrt{D(x)}$. If the variance is zero $D(x) = 0$, the standard deviation vanishes and the probability that the random variable x takes the value equal to its mathematical expectation $(p(x = E(x)))$ is equal to 1. This fact represents one of its most important properties. The calculation of $M(x)$ and $D(x)$ for the specific case of normal distribution, confirms these results. In other words: the mathematical expectation for a random variable that is normally distributed is the mean or average μ $(M(x) = \mu)$, while its variance is σ^2 $(D(x) = \sigma^2)$. Such parameters refer to the population level.

If a random variable is described by a Pearson type distribution satisfying differential equation (5), the coefficients a_0, b_0, b_1, b_2, are related to the moments M_k and M^k of this distribution. In consequence, a_0 and b_0 are given by the following mathematical expression as in [2]:

$$a_0 = \frac{M_3(M_4 + 3M_2^2)}{P_{M_2M_3M_4}} \tag{10}$$

$$b_0 = -\frac{M_2(4M_2M_4 - 3M_3^2)}{P_{M_2M_3M_4}} \quad (11)$$

P_{M2M3M4} is a polynomial expression in terms of M_2, M_3 and M_4 given by:

$$P_{M_2M_3M_4} = 10M_4M_2 - 18M_2^3 - 12M_3^2 \quad (12)$$

If a_1=1, the equation (5) after the variable separation, becomes:

$$\frac{df}{f} = \frac{x+a_0}{b_2x^2+2b_1x+b_0}dx \quad (13)$$

Substituting b_1=0, b_2=0 in the above equation and integrating this, the solution has the form:

$$f(x) = ce^{\frac{(x+a_0)^2}{2b_0}} \quad (14)$$

C is a constant. Considering that M_1=0, $M_2=\sigma^2$, M_3=0, $M_4=3\sigma^4$ and substituting in the expressions (10) and (11), the results is: a_0=0 and $b_0=-\sigma^2$. Therefore, the formula (14), after the application of the normalization condition, becomes:

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{x^2}{2\sigma^2}} \quad -\infty < x < +\infty \quad (15)$$

If the standard deviation σ=1, the expression (15) coincides with the well-known probability density for the standard normal distribution given by formula (4), commonly used in books of Biostatistics.

Several continuous distributions satisfy the differential equation (13). It is known that knowledge of the first four moments makes it possible to estimate the four parameters that appear in it, so that a density is available, but the estimates in general are not efficient because they are performed by the method of moments [11].

The equation (13) can be found in the literature in a slightly different form [11, 12]. This last is obtained if the coefficients in equation (6) have the following values: a_0 =d, a_1 =-1. Apart from this, if b_1, and b_2 vanishes, the resulting differential equation is:

$$\frac{df}{dx} = (c_0 + c_1x)f \quad (16)$$

The integration of this equation gives:

$$f(x) = ce^{-\frac{c_1}{2}\left[-\left(x+\frac{c_0}{c_1}\right)^2+\frac{c_0^2}{c_1^2}\right]} \quad (17)$$

where c is the integration constant and c_1 and c_2 are given in terms of coefficients by the expressions:

$$c_0 = \frac{d}{b_0}, c_1 = -\frac{1}{b_0} \quad (18)$$

Formula (17) is similar to that deducted in [13]. The constant c can be determined by the appropriate choice of the initial condition or by the application of the normalization condition, which is more usual in Probability and Statistics. Finally the solution is:

$$f(x) = -\frac{c_1\sqrt{2\pi\left(-\frac{1}{c_1}\right)}}{2\pi}e^{c_1\frac{\left(x+\frac{c_0}{c_1}\right)^2}{2}} \quad (19)$$

The presence of the sign minus (-) inside the square root in the equation (19), should not be a cause for confusion because the term ($-1/c_1$) is positive. The equation can be transformed to a more usual form, taking into account the values given by expression (18). The substitution of these, transforms (19) in the known formula for the Gaussian probability density, corresponding to the normal distribution, with $d = \mu$ and $b_0 = \sigma^2$:

$$f(x) = \frac{1}{\sqrt{2\pi b_0}}e^{-\frac{(x-d)^2}{2b_0}} \quad (20)$$

3.2. On Chi-squared and Student's t-distributions

The notion of Chi-squared distribution or χ^2-distribution is given by Torres in order to solve problems in which discrepancies between observed and expected theoretical frequencies appear [1]. The general procedure to introduce the distribution considers two ways: a way by means of limit of the histogram or frequency polygon which is consistent with the introduction of the normal distribution; another way consist of algebraic treatment which is usual in other books as [10].

Despite the fact that this distribution play an important role and its use is widely discussed in books of Biostatistics, the mathematical expression for the probability density is not given usually. This could generate doubts and at the same time curiosity since the mathematical formula is not presented as in the case of the normal distribution. However, its graphic is shown in different cases in order to explain the behavior of the distribution shape with the degrees of freedom in association with the concept of skewness as the measure of the asymmetry of the probability distribution [1, 10, 14].

It is said that a random variable has a χ^2-distribution with n degrees of freedom, if its density has the following mathematical expression:

$$f(x) = \begin{cases} \frac{1}{2^{\frac{n}{2}}\Gamma\left(\frac{n}{2}\right)}x^{\frac{n}{2}-1}e^{-\frac{x}{2}} & x > 0 \\ 0 & x \le 0 \end{cases} \quad (21)$$

One of the difficulties to introduce this distribution with the help of the formula (21) consists of the presence of the so called Gamma function denoted by Γ. Due to the medical students don't have a solid mathematical foundation, this function might be a problem. However teachers could at least explain the essentials about this, taking as the starting point its definition in terms of Eulerian Integral of the second kind:

$$\Gamma(x) = \int_0^{+\infty} e^{-t}t^{x-1}dt \quad (22)$$

where: x>0.

The function may be defined for all complex numbers, with the exception of non-positive integers. Some properties that are useful in the development of the topic about statistical distributions are the following:

$$\Gamma(1) = 1;\ \Gamma\left(\frac{1}{2}\right) = \sqrt{\pi} = 2\int_0^{+\infty} e^{-x^2}\,dx \quad (23)$$

$$\Gamma(x+1) = x\Gamma(x) \quad (24)$$

$$\Gamma(n+1) = n\Gamma(n) = n!;\ n \in \mathbb{N} \quad (25)$$

The theoretical model of probability distribution named Student's T has many application in the context of Biostatistics. The probability density function is given by the formula:

$$f(x) = \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}(1+\frac{x^2}{n})^{-\frac{n+1}{2}} \quad -\infty < x < +\infty \quad (26)$$

Here again n represents the degrees of freedoms; Γ is the same Gamma function.

The distribution has the following important property: with the increase of the degrees of freedom ($n\rightarrow\infty$) it tends to the standard normal distribution. The proof on this requires the applications of the property of limits, the Stirling's approximation and the limit definition of the transcendent number e. Therefore it is possible to pose:

$$\text{if } n \to \infty, \frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)} \to \frac{1}{\sqrt{2\pi}} \quad (27)$$

$$\text{if } n \to \infty, (1+\frac{x^2}{n})^{-\frac{n+1}{2}} \to e^{-\frac{x^2}{2}} \quad (28)$$

In consequence the limit approximation from an analytic point of view, gives:

$$\frac{\Gamma(\frac{n+1}{2})}{\sqrt{n\pi}\Gamma\left(\frac{n}{2}\right)}(1+\frac{x^2}{n})^{-\frac{n+1}{2}} \to \frac{1}{\sqrt{2\pi}}e^{-\frac{x^2}{2}} \quad (29)$$

The last limit show the close relation between the two distributions as the degrees of freedoms become larger.

Other interesting relations between χ^2 and normal distribution as well as the Pearson distributions and Student's T and its implications in the teaching learning process in Biostatistics for students of Medicine, are part of a future paper.

4. Implications of the Above Approach in the Teaching Methodology

The Pearson distributions require the notion of differentiation since they are defined from a differential equation which must be satisfied by the probability density function or density f(x). It is worth remembering that in the course of Higher Mathematics, Differential Calculus precedes Integral Calculus, even in the case of introductory notions or the exposition of its essential concepts [15].

Biostatistics has been traditionally studied as part of the discipline Medical Informatics, in the second year of the Medicine career in Medical Sciences Holguin University [16], although the new Study Plan E considers it as part of the Research Methodology in Health Sciences. The difficulties related to the insufficient mathematical background of students are well known. However, the presentation of the topic referred to the distribution function by means of the exposed approach, offers the alternative of giving continuity to the classical logical order in the introduction of the associated concepts as well as showing its historical-logical conception.

The explanation of the procedure for obtaining the concrete form of the theoretical distribution requires notions related to the concept of first order ordinary differential equation. There are excellent books about the topics, for example [17, 18]. These methods are not included in programs as well as they do not constitute a priority in the basic elements of Mathematics. But the notions of integration are dealt with in Biostatistics, so the exposition of the elementary ideas about differential equations offers the possibility of a qualitative discussion about the existing link between the two mutually inverse operations of Calculus: derivation and integration.

The fundamentals of Differential Calculus open the way for the understanding of the basic conceptual apparatus of integration. Some authors have suggested the treatment of some contents without the use of derivatives, especially linked to optimization problems as well as to other themes related to Differential Calculus [19].

The alternative of introducing the topic about the density function for the normal distribution as a particular case of Pearson distributions, familiarizes the students with the concepts of derivative and differential equation. This could complement the ideas about the resolution of problems contemplating time-varying functions and other contents which have high impact on the formation of the physician [20].

The calculation of the coefficients of the distribution in the above approach, has been associated with the moments of the distributions given by mathematical expressions (10) and (11). The concept of momentum is discussed in the framework of Descriptive Statistics and is summarized with other topics in the basic bibliography as [1]. In this case the development of the subjects is carried out starting from the measurements that make it possible to know the form of the distribution of the data and thus to obtain a more complete information.

In the above scope, the moments of order R with respect to zero and with respect to the mean are denoted by m_R and M_R respectively. These averages of the powers of order R of the direct data or their deviations are such that for R=1, the moment of order 1 with respect to zero is simply m_1 or the usual arithmetic mean while M_1 is zero; however if R=2 then $M_2=S^2$ which is the variance of the data. Consequently, the program of Biostatistics prepares the student with minimal information about the meaning of moments of the distributions.

The coherent character of mathematical knowledge and not its reduction to a contextualized framework as well as the relationship between various disciplines of Mathematics is evidenced by the association of the moments to the family of Pearson distributions and to the differential equation that satisfies the unknown function. This allows students to develop a solid thinking style not only in this matter but also in other fields of their professional activities as research, clinical practice or during the deepening of the specialty studies.

In relation with the χ^2 and Student's T distributions, there is not a secret that the analytic approach in an introductory lesson is probably more difficult than the topic related to the normal distribution. However many facts which involve the three distributions are commented in a qualitative way. That is for example the case in which, the discussion of the similarities between Student's T and the standard normal is introduced from the geometric considerations.

In the above sense, the center of both models, its bell shape and its approximation when the degrees of freedoms are greater than 30, are explained in general features. The analytic approach by means of the mathematical expressions serves as complement and confirmation of the geometric and intuitive ideas.

A problem in the teaching of those topics is the concept of functions of Γ type in terms of an integral representation and not a formula as it is known by students who are not familiarized with Higher Mathematics. But the knowledge of that function might be very useful in the explanation of the behavior of the χ^2 probability density function with the degrees of freedom. For example for n=1 or n=2, $\Gamma(1/2) = \sqrt{\pi}$ and $\Gamma(1) = 1$, therefore the substitution of these values in equation (21), gives the following formulas for the functions:

$$n = 1; f(x) = \frac{1}{\sqrt{2\pi}} x^{-\frac{1}{2}} e^{-\frac{x}{2}}, x > 0 \tag{30}$$

$$n = 2; f(x) = \frac{1}{2} e^{-\frac{x}{2}}, x > 0 \tag{31}$$

These last functions are more familiar to the students so its graphics might be build, specially the corresponding to function (31) which has an pure exponential behavior.

Other interesting application of Gamma function is based on the properties given by the equations (23) and (24). The last (24) is linked in an unsuspected way with the concept of factorial of a natural number given by (25) and its generalization. The use of the factorial is typical for the calculation of combinatorial numbers, in the case of random variables which distribution function is binomial type. Some problems associated with the use of factorial are developed and suggested in the book [1] as an illustrative example about the calculation of the number of possible samples of a given size.

5. On the Teaching Learning Process of the Normal Distribution for Medical Students

Random variations in very large populations or samples usually correspond to certain types of distributions. It is advisable an exposition of the basic ideas associated to the deduction of their mathematical form, combining the quantitative and qualitative reasoning.

The graphic of the normal distribution is a curve sometimes called the normal error curve [21]. It is recommended to keep in mind that the terms error and normal have very special meanings, so that they should not be confused with the usual meanings attributed to these words. For this reason, when referring to the curve or the distribution, it is suggested to use the word Gaussian and discuss the precise meaning of these terms.

The definition of the normality ranges of random variables is one of most illustrative application of the normal distribution during the medical career. It should be borne in mind that the aim is to find certain values so that multiplied by the standard deviation and the result of such multiplication added and subtracted from the average, allow to obtain the referred ranges, if the population parameters and the proportion of patients in that population are known.

The answer to problematic situations associated with normality ranges can be obtained in a general way by means of an interesting mathematical problem whose solution requires the knowledge of integration methods that are not part of the content of the programs of the Biostatistics. These, moreover, include the calculation of integral which can not be reduced to elementary way. However, in certain simplified cases, a less laborious and arduous way can be suggested, which the teacher must previously select with great care.

It is recommended to put special attention to the application of statistical methods in the case of random variables that follow normal behavior. Cautions should be taken when the samples are small. The use of methods in this situation has to be rigorously justified in order to admit that such a sample has been randomly selected from a normally distributed population, otherwise the inferred results may not be valid.

In case deviations of a large sample from normal behavior are detected, it is advisable to carefully analyze such deviations before applying the distribution. The causes of those deviation may be diverse. It is suggested to provide an analysis of them so as to clarify to the students by means of examples as the following: in the case of isoniazid concentration in plasma, after administration of a standard dose in groups of people with a genetic polymorphism for the metabolizing enzyme, the cause could be associated with the fact that the sample may have been chosen from a population that is not homogeneous for the studied variable [21]. Another cause is that although the population is homogeneous, the frequency distribution of the values is different. The choice of a measurement scale that is not appropriate can lead to such deviations. The choice of a measurement scale that is not appropriate can also lead to such deviations.

Useful applications of the normal distribution in Inferential Statistics topics related to Medical Sciences can be found in complementary books for the career [22]. More comprehensive treatments of such a distribution from the quantitative, qualitative point of view and its various uses can be suggested to students, although they require wider consultation of the available literature [23-27].

6. Conclusions

The mathematical expression of the normal distribution density function was deduced from Pearson system of univariate continuous distributions. This alternative approach to introduce the referred topic, brings consistency, rigor and harmony to the mathematical statistical knowledge included in the program of Research Methodology in Health Sciences.

The method to introduce the mathematical form of such distribution, in spite of requires the familiarization with concepts of the Differential Calculus, ordinary differential equations and their methods of solution, makes possible the analysis of the notions of derivation, integration and their respective interpretations in unity and coherence.

The deduction of the mathematical form of the distribution or at least the explanation in general terms, of the procedure for obtaining the analytical expression of the probability density, contributes to develop a solid thinking style which is necessary in the research activity of future physicians.

7. Recommendations

Taking into consideration the developed approach, it is possible to analyze the implications of the introduction of analytical methods to present other statistical distributions, in the teaching learning process of the Research Methodology in Health Sciences. It would be interesting to explore the link between Pearson system and Student's T distribution as well as its impact in the development of a probabilistic and statistical thinking, in the mathematic education of the future physicians, their scientific culture and their social performance.

References

[1] Torres Delgado, J. A., Rubén Quesada, M., Bayarre Vea, H., Garriga Sarría, E. P., Pría Borras, M. C., & Gran Álvarez, M. et al. (2004). Informática Médica, tomo II Bioestadística. Centro de Cibernética Aplicada a la Medicina (CECAM). Editorial Ciencias Médicas: La Habana, 2004.

[2] Koroliuk, V. V. (1986). Manual de la teoría de las probabilidades y estadística matemática. Moscú: Mir.

[3] Gnedenko, B. V. (1997). The theory of probability. Overseas Publishers Association: Netherlands.

[4] Whittle, P. (1982). Probability. Moscow: Nauka (in Russian Languaje).

[5] Guerra Bustillo, C. W., Menéndez Acuña, E., Barrera Morera, R., & Egaña Morales, E. (1991). Estadística. La Habana: Pueblo y Educación.

[6] Freund, J. E., Perles, B. M. (2007). Modern Elementary Statistics (12th International Editions). Pearson Prentice Hall: New Jersey.

[7] Chistiakov, B. P. (1982). Course on Probability Theory. Moscow: Nauka (in Russian Languaje).

[8] Ander-Egg, E. (1995). Métodos y Técnicas de investigación social. Vol. III: Cómo organizar el trabajo de investigación. Lumen: Buenos Aires.

[9] Giacosa, N., Vergara, M. L., Zang, C., López, J., Galeano, R., Godoy, N. & et al. (2015). Libros de texto y Programas Analíticos de Física en carreras de Ingeniería de la UNaM. Revista de Enseñanza de la Física, 199-207, 27 (no extra).

[10] Van Belle, G.; Fisher, L. D.; Heagerty, P. J., & Lumley, T. (2004). Biostatistics. A Methodology for the Health Sciences. New Jersey: John Wiley & Sons.

[11] Herrerías Pleguezuelo, R., & Callejón Céspedes, J. (2017). Los sistemas de Pearson como generadores de distribuciones de probabilidad. Aplicaciones estadísticas y económicas. http://www. ugr. es/~callejon/lossistemas. pdf

[12] Weisstein, Eric W. (2017) Pearson System. MathWorld, A Wolfram. http://mathworld.wolfram.com/PearsonSystem.html.

[13] Ávila, Ávila, R. M, Pino Tarragó, J. M., Expósito Gallardo, M. C., & Domínguez Gálvez, D. L. (2020). La distribución normal en ciencias biomédicas: un enfoque a partir de las distribuciones de Pearson. Revista Sinapsis; Vol 1, No 16.

[14] Cherniick, M. R., & Riss, R. H. (2003). Introductory Biostatistics for the Health Sciences. Modern Applications Including Bootstrap. New Jersey: John Wiley & Sons.

[15] Valdés Castro, C., & Sánchez Fernández, C. (2011). Introducción al Análisis Matemático. La Habana: Félix Varela.

[16] Escalona Fernández, L. A. (2020). Alternativa didáctica para desarrollar el proceso de enseñanza-aprendizaje de la Bioestadística en la carrera de Medicina. Acta Latinoamericana de Matemática. Vol. 33, Número 1.

[17] Elgoltz, L. E. (2010). Ecuaciones diferenciales y cálculo variacional. Félix Varela: La Habana.

[18] Robinson, J. C. (2004). An introduction to Ordinary Differential Equation. University Press: Cambridge.

[19] Escalona Fernández, L. A., & Velázquez, J. R. (2012). Resolución de problemas de optimización sin el uso de límites y derivadas. Interpretaciones médicas". En: Flores, R. (ed.). Acta Latinoamericana de Matemática Educativa 25, 365-374.

[20] Escalona Fernández, L. A. (2013). Resolución de problemas matemáticos aplicados a la medicina y su impacto en la formación del médico general". Correo Científico Médico Holguín; volumen: vol. 17, No. 24.

[21] Bowman, W. C., & Raud, M. J. (1984). Farmacología. Bases bioquímicas y patológicas. Aplicaciones clínicas. Interamericana: México D. F.

[22] Oliva González, L., & O'Farril Mons, E. (1988). Bioestadística y computación: guía de estudio. La Habana: Pueblo y Educación.

[23] D'Agostino, R. B., Sullivan, L. M., & Beiser, A. S. (2006). Introductory applied biostatistics. Thomson, Brooks/Cole: Toronto.

[24] Bluman, A. G. (2009). Elementary Statistics: A Step by Step Approach. McGraw-Hill Companies, Inc.: New York.

[25] Lancaster, H. O. (2011). Quantitative Methods in Biological and Medical Sciences: A Historical Essay. Springer: New York.

[26] Anderson, S. J. (2012). Biostatistics. A Computing Approach. Boca Raton: CRC Press.

[27] Indrayan, A., & Kumar Malhotra, R. (2018). Medical Biostatistics. Boca Raton: CRC Press.