The vision of biostatistics and biometrics on the analysis of biometric data: the Framingham risk score as an example

A visão da bioestatística e biometria na análise de dados biométricos: a pontuação de risco de Framingham como exemplo

Walmirton Bezerra D’Alessandro¹
Sávia Denise Silva Carlotto Herrera²
Fábio Pegoraro³
Aline Almeida Barbaresco D’Alessandro⁴
Maykon Jhuly Martins de Paiva⁵
Samara Tatielle Monteiro Gomes⁶
Sara Falcão de Sousa⁷
Mosaniel Falcão de França Junior⁸

¹ Doctor of Tropical Medicine, Post-Doctor of Health Care and Assessment by Universidade Federal de Goiás (UFG), Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: walmirton@unirg.edu.br Orcid: https://orcid.org/0000-0002-2897-9770
² Master in Health Sciences by Universidade Federal do Tocantins (UFT), Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: saviaherrera@unirg.edu.br Orcid: https://orcid.org/0000-0002-0695-9621
³ PhD in Production and Systems Engineering from Pontifícia Universidade Católica do Paraná (PUC-PR), Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: fabiopegoraro@unirg.edu.br Orcid: https://orcid.org/0000-0002-0715-0867
⁴ PhD in Tropical Medicine and Public Health, Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: aline.a.b.dalessandro@unirg.edu.br Orcid: https://orcid.org/0000-0003-0966-6098
⁵ Master in Science and Technology, Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: prof.maykon.paiva@gmail.com Orcid: https://orcid.org/0000-0002-6228-4550
⁶ PhD in Biology from Universidade Federal do Pará (UFPA), Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: samaratatielle@unirg.edu.br Orcid: https://orcid.org/0000-0001-7803-2860
⁷ PhD in Health Sciences from Universidade Federal de Goiás (UFG), Universidade de Gurupi (UNIRG), Av. Rio de Janeiro, 1585, St. Central, Gurupi - TO, CEP: 77403-090. E-mail: sarafalcao@unirg.edu.br Orcid: https://orcid.org/0000-0003-4172-0306
⁸ Master in Public and Collective Health, Faculdade de Ciências Sociais e Agrárias de Itapeva (FAIT), Pilão d’Água, Rod. Francisco Alves Negrão, Km 285, Itapeva - SP, CEP: 18412-000. E-mail: drmosanielfalcao@hotmail.com Orcid: https://orcid.org/0000-0002-7064-8705
Abstract
The fields of Biostatistics and Biometrics play a crucial role in unraveling the mysteries of human health and empowering healthcare professionals to make informed decisions. By mining health data for meaningful patterns, they enable medical practitioners to identify emerging trends and develop targeted, effective policies and practices. Meanwhile, the analysis of biological data unleashes a wealth of insights, enabling accurate diagnoses and unleashing innovative therapies that transform lives. Together, these fields are a powerful force for good, fueling breakthroughs that enable us to better understand and heal the human body (Clayton and Hills 2019).

Keywords: Biostatistics. Biometrics. Framingham Risk. Standard Error.

Introduction
The fields of Biostatistics and Biometrics play a crucial role in unraveling the mysteries of human health and empowering healthcare professionals to make informed decisions. By mining health data for meaningful patterns, they enable medical practitioners to identify emerging trends and develop targeted, effective policies and practices. Meanwhile, the analysis of biological data unleashes a wealth of insights, enabling accurate diagnoses and unleashing innovative therapies that transform lives. Together, these fields are a powerful force for good, fueling breakthroughs that enable us to better understand and heal the human body (Clayton and Hills 2019).
Biostatistics and Biometrics can assist in medical algorithms in various ways, providing tools for health data analysis and developing more accurate predictive models. Here are some ways in which these disciplines can contribute:

- Identification of risk factors: Biostatistics can be used to identify risk factors for diseases, allowing medical algorithms to be designed to identify individuals who are more likely to develop a particular disease. This can lead to earlier diagnoses and more effective treatments.
- Modeling health data: Biometrics can be used to model health data, helping to identify patterns and relationships that can be used to predict future diagnoses or outcomes. Medical algorithms can be designed to incorporate these models, allowing them to make more accurate predictions.
- Personalization of treatment: Biostatistics and Biometrics can be used to personalize treatment based on individual patient characteristics, such as age, sex, medical history, and genetics. Medical algorithms can be designed to take these factors into account, allowing treatment to be tailored to each patient's individual needs.
- Real-time data analysis: Biostatistics and Biometrics can be used to analyze health data in real-time, allowing medical algorithms to make real-time adjustments based on new information. This can improve the accuracy of diagnoses and treatment outcomes.

In summary, Biostatistics and Biometrics provide powerful tools for health data analysis and predictive modeling, allowing medical algorithms to be designed to provide more accurate diagnoses, personalize treatment, and make real-time adjustments based on new information.

Descriptive Statistics: Common measures of central tendency used in descriptive statistics include mean, median, and mode, which describe the typical or central value of a dataset. Measures of variability, such as range, variance, and standard deviation, describe the spread or dispersion of the data.

Measures of Central Tendency

The mean: It is represented mathematically by
The vision of biostatistics and biometrics on the analysis of biometric data: the Framingham risk score as an example

\[
\bar{x} = \frac{\sum x}{n}
\]

Where:

is the mean, \( \sum \) is the sigma symbolizing summation of sets of and is the total number of data set measured.

**The Median:** It is represented mathematically by Med (X). "For a data set with \( n \) elements, the formula for the median is:

- If \( n \) is odd: \( \text{Median} = \text{element at position } (n + 1) / 2 \)
- If \( n \) is even: \( \text{Median} = \text{arithmetic mean of the elements at positions } n / 2 \) and \( (n / 2) + 1 \).

**The mode:** It is represented mathematically by Mo. Where is the value that has the highest frequency in a distribution and is represented by Mo. In a set of numbers, it may or may not exist, and if it exists, it may or may not be unique, so we have modal, unimodal, bimodal, etc. sets.

**Parameters of Dispersion**

**Variance:** It is represented mathematically by

\[
\sigma^2 = \frac{\sum (x - \overline{X})^2}{n - 1} \quad \text{or} \quad s^2 = \frac{\sum x^2 - (\sum x)^2}{n - 1}
\]

The most important measure of variance, represented by \( s^2 \) for the sample or \( \sigma^2 \) for the population

Where:

\( \Sigma \) denotes the sum of the values; \( x \) represents the value of the \( i \)th observation; \( \overline{X} \) is the population mean; \( n \) is the sample size.

To calculate the variance, you need to follow these steps:

Revista Gestão e Secretariado (GeSec), São Paulo, SP, v. 14, n. 8, 2023, p. 12811-12818.
1) Calculate the mean \( \bar{X} \) of the data set.
2) Subtract the mean from each value in the data set to get the deviation from the mean 
\( (x - \bar{X}) \) for each observation.
3) Square each deviation \( (x - \bar{X})^2 \).
4) Add up all of the squared deviations \( \sum (x - \bar{X})^2 \).
5) Divide the sum by the sample size \( (n) \) to get the variance.

**The Standard Deviation:** It is represented mathematically by

\[
\sigma = \sqrt{\frac{\sum_{i=1}^{n} (x_i - \bar{x})^2}{n-1}}
\]

Standard deviation is another commonly used measure of dispersion, which is simply the square root of the variance. In other words, variance gives us an idea of how much the data is spread out, and the standard deviation gives us a more easily interpretable measure of this spread. Together, variance and standard deviation are powerful tools for understanding the distribution of a set of data.

**The Standard Error:** It is represented mathematically by

\[
S_x = \frac{S}{\sqrt{n}}
\]

Standard error or mean error: Different samples taken from the same population may have different means. The variation between this set of means is estimated through the standard error, which corresponds to the standard deviation of the means.

The standard error tells us how much we can expect the sample mean to vary from the population mean. It's an important tool for making inferences about a population based on a sample. The smaller the standard error, the more accurate the sample mean is likely to be as an estimate of the population mean.

In general, larger sample sizes tend to have smaller standard errors, because the larger the sample size, the more accurately the sample mean represents the population mean.

Measures of association and relative risk are essential for medical studies. They help
to identify the strength and direction of the relationship between two variables and to quantify the risk of an outcome based on exposure to a particular factor. These measures are commonly used in epidemiological studies to assess the association between risk factors and disease outcomes, which can inform the development of preventive measures and treatments. Understanding these measures is critical for medical professionals, as it enables them to make informed decisions about patient care and public health interventions (Goodwin and Ryu 2023).

The Formula for Calculating the Odds Ratio is

$$\text{OR} = \frac{(a/b)}{(c/d)}$$  \hspace{1cm} (1)

Where:

“a” is the number of individuals in the exposed group who experience the event, “b” is the number of individuals in the exposed group who do not experience the event, “c” is the number of individuals in the unexposed group who experience the event, and “d” is the number of individuals in the unexposed group who do not experience the event.

The Formula for Calculating the Relative Risk is

$$\text{RR} = \frac{(a/a+b)}{(c/c+d)}$$  \hspace{1cm} (2)

Where:

“a” and “b” are the same as in the odds ratio formula, and “c” and “d” are also the same as in the odds ratio formula. Formula (1) and (2) are widely used in epidemiological studies associated with statistical tests.

Indeed, the Framingham Risk Score is a prime example of the important role that biostatistics and biometric data play in medical research and clinical decision-making. The Framingham Risk Score is a statistical model used to estimate an individual's 10-year risk of developing cardiovascular disease, including heart attack and death from heart disease.

The score takes into account several risk factors, including age, sex, blood pressure, cholesterol levels, smoking status, and diabetes. By analyzing data from the Framingham Heart Study, which tracked the health of thousands of people over several decades, researchers were able to identify these risk factors and develop a mathematical model to predict an individual's risk of developing heart disease.

Revista Gestão e Secretariado (GeSec), São Paulo, SP, v. 14, n. 8, 2023, p. 12811-12818.
The Framingham Risk Score has become a widely used tool for physicians to help identify patients at high risk for heart disease and to guide decisions regarding preventive interventions, such as lifestyle modifications and medication therapy. However, it is important to note that the score is not perfect and may not accurately predict an individual's risk in all cases.

Overall, the Framingham Risk Score is just one example of how biostatistics and biometric data analysis can contribute to medical research and help improve clinical outcomes for patients.

Although the FHS provides valuable information for assessing the risk of cardiovascular diseases, it has some limitations that should be taken into account when using the Framingham Risk Score (FRS) and interpreting the results for individual patients. The FRS is only applicable to individuals who have not previously been diagnosed with heart disease, and it can only predict the development of coronary heart disease (CHD) and not other types of heart or vascular diseases. However, there are separate prediction models available for heart failure and stroke on the National Heart, Lung, and Blood Institute (NHLBI) website. It is important to note that the FRS should not be used as a substitute for medical evaluation, as indicated on the NHLBI website.

Artificial intelligence (AI) has the potential to revolutionize the healthcare industry by utilizing biometric and biostatistical data to predict the attitude for resolving health problems and developing new treatments. Analyzing large amounts of data allows AI algorithms to identify patterns, correlations, and associations that would be impossible to detect with traditional methods. In addition, AI can help improve the accuracy and efficiency of data analysis, thereby increasing the ability of healthcare professionals to make more precise diagnoses and informed decisions. According to Sendef and Robbins (2019), the use of AI in healthcare is constantly evolving, and new technological advancements are being developed to further enhance the application of these techniques in the medical field.

Below we have included a figure representing the idea of the interaction between biostatistics and biometrics in the identification of future diseases. It represents a forward-looking approach that can save lives through numbers, a vision (Figure 1).
The vision of biostatistics and biometrics on the analysis of biometric data: the Framingham risk score as an example

Figure 1. Interaction between statistics and biological data.
Figure created by power point and midjourney bot

References


Gabriela de Souza Ferreira1, Giuliana Von Berg de Oliveira1, Guilherme Artur Campos1*, Letícia Freire Rettore Mendes1, Lucas Augusto Moris Afonso1, Marcus Vinicius da Silva1, Tânia Maria Gonçalves Quintão Santana1 Cardiovascular risk by the Framingham score in cardiology services of a medium-sized city of Minas Gerais Rev Med Minas Gerais 2020;30 (Supl 4): S69-S76


Submetido em: 10.07.2023
Aceito em: 07.08.202