sciendo

STUDIES IN LOGIC, GRAMMAR AND RHETORIC 68 (81) 2023

DOI: 10.2478/slgr-2023-0006

This work is licensed under a Creative Commons Attribution BY 4.0 License (http://creativecommons.org/licenses/by/4.0)

Robert Milewski
Department of Biostatistics and Medical Informatics
Medical University of Białystok, Poland
e-mail: robert.milewski@umb.edu.pl
ORCID: 0000-0001-6241-8283

Magdalena Roszak
Department of Computer Science and Statistics
Poznan Medical University, Poland
e-mail: mmr@ump.edu.pl
ORCID: 0000-0001-6495-6771

THE ISSUE OF SELECTION OF APPROPRIATE METHODOLOGY AND METHODS OF GRAPHICAL REPRESENTATION OF DATA IN BIOMEDICAL RESEARCH

Abstract. The development of medicine is based on reliable medical research. This is a process that must be planned in detail and performed in accordance with the accepted study protocol, as any negligence – even a small one – or deviation from the protocol may result in distorted research results and – in consequence – to false conclusions. One of the key stages of research is the selection of the appropriate methodology, particularly in terms of tests that verify the posed hypotheses. Each statistical method has a set of assumptions that have to be met in order for the method to be used. This paper explains why the appropriate selection of methods, verification of the required assumptions, and correct graphical representation of the obtained results are so important from the practical perspective.

Keywords: statistical analysis, biostatistics, biomedical research, normal distribution, assumptions of statistical tests, graphical representation.

Medicine, in a broad sense of the term (which includes pharmaceutical sciences and health sciences), is one of those fields of knowledge where statistics undeniably plays a key role. Statistics makes it possible to plan an experiment in terms of its set aim, choose the appropriate sample size, collect data that contain the relevant information, and last but not least, analyse the collected data and make inferences that implement the research aim and are articulated in the form of the conclusion of the study. The aforementioned stages appear in the majority of original papers published in medical

journals. Only the introduction, which introduces the reader to the subject matter of the study; and the discussion of the obtained results with references to the published literature, also containing the an interpretation of the results, are not listed above. It might seem that statistics is a universal tool that can be used in all scientific disciplines. In the case of medicine, however, its unique character resulted in the creation of specialized methods to be used in medical research, which have been combined under a common name, i.e. "biostatistics". The term is understood as the adaptation of statistical methods for the needs of research studies in the area of broadly understood medicine (from biology, through pharmaceutical sciences or health sciences, to physical culture or veterinary medicine) [1].

This paper presents issues connected with the character of data collected in medical research; methods of their description and representation; the importance of selection of the appropriate research methodology and verification of assumptions that are required by the performed tests or the built models.

Normal distribution – based on the Gaussian curve – plays a key role in statistics [2]. Indeed, many methods have been created that require that the assumptions connected with the normality of distribution of the analysed variables should be met [3]. Normal distribution is associated with nature itself as its essence consists in the action of numerous minute random factors, which when – combined in sufficiently large numbers – result in a shape similar to the Gaussian curve. In medicine, normal distribution is mainly associated with physiology. However, as the essence of medicine lies in studying pathologies, which usually deform distributions of the analysed variables, they are in effect considerably removed from the ideal normal distribution. Such variables are usually asymmetric, due to the pathological increase in the value of the analysed factor, or its decrease. It is obvious then that asymmetric distribution cannot be close to normal distribution, which makes it impossible to use many methods correctly, i.e. those methods that require that the assumption of normality should be met (e.g. Student's t-test, analysis of variance, or Pearson's correlation) [4–5].

The shape of distribution of the analysed variables is also connected with the issue of appropriate indicators used for their description [6]. The two main types of such indicators are average measures (e.g. arithmetic, geometric, and harmonic means, the median, or the mode) and measures of dispersion (standard deviation, standard error, variance, quartiles, etc.). Measures such as the arithmetic mean and standard deviation have become extremely popular in medical papers. Nevertheless, in many cases

their use is incorrect – these measures work well in representations of distributions that are at least symmetric, while the occurrence of asymmetry, particularly outliers, may result in the presented measures losing their intended role [7].

The arithmetic mean is very sensitive to asymmetry and outliers, especially in cases of small sample sizes. This can be clearly illustrated through the following theoretic example:

- in a 1000-element sample consisting of 999 values "1" and one value "100", the arithmetic mean is 1.099
- in a 100-element sample consisting of 99 values "1" and one value "100", the arithmetic mean is 1.99
- in a 10-element sample consisting of 9 values "1" and one value "100", the arithmetic mean is 10.9

Mathematically speaking, the above calculations are perfectly correct, but what would be the purpose of presenting such measures? In the third presented case, i.e. the 10-element sample, the arithmetic mean is 10.9. It would be difficult to call the value the average measure as it has nothing to do with the nine "1" elements, not to mention the "100". Of course, the above example is purely theoretical, but its simplicity makes it possible to clearly illustrate the impact of outliers on the arithmetic mean. The average measure that is much better suited to such situations is the median, whose intuition consists in presenting the "middle" value. The median is resistant to asymmetry and – more importantly – outliers. In all the three analysed cases, the median values are exactly "1", which is consistent with the intuition of the average value. Authors of scientific papers who decide to use the arithmetic mean to represent the average measure usually argue that other authors in literature overwhelmingly use this particular indicator as well, therefore by referring to other papers, they adopt the same representation. It should be remembered, though, that the median is more universal; in situations when the distribution of a variable is symmetric, the median equals the arithmetic mean, whereas the greater the asymmetry, the less effective is the arithmetic mean in representing the average measure (even though it is correct mathematically) [8]. Hence, when one wishes to refer to literature – where the mean is much more prevalent than the median – the latter can be used without problems, by formulating a reference, e.g. "the average value in the analysed paper was Median=..., whereas in his/her paper, author X presented the mean value, i.e. Mean=...".

Standard deviation (SD) illustrates the dispersion of elements in a sample. Similarly to the arithmetic mean, it is sensitive to outliers, but above all, it requires symmetry of distribution. In literature, average measures

and dispersions are usually represented as Mean±SD, where it is understood that the SD is set on both sides of the mean value. However, in cases of considerable asymmetry, dispersion on one side is much greater, while on the other it is much smaller than the presented SD. The measure of dispersion that is resistant to outliers while also allowing to vary the range on both sides of the average value are quartiles (Q_1 and Q_3). They can be thought of as medians of both halves of the analysed sample, divided by the value of the main median. Thus, a more universal method of representation of the average measure and dispersion, resistant to asymmetry and outliers, is the representation of Median (Q_1; Q_3) instead of Mean±SD [9].

Apart from representing the appropriate average measures and dispersion, a key element of each research paper is the selection of the correct algorithms for the analysis of the collected data. Again, parametric methods dominate in the literature, which are used for the assessment of the values of parameters selected for a given distribution of the population from which a sample is drawn. Such parameters are, e.g., the mean, variance, or standard deviation [4].

Even one of the most popular tests, i.e. Student's t-test, is a parametric method used for the comparison of mean values between two groups [10]. For normal distribution, when the arithmetic mean and the standard deviation are known, the full distribution of a variable can be plotted, therefore, when two distributions are to be compared, the problem can be simplified to a comparison between two means, exactly as is the case with Student's t-test. This approach, however, requires that assumptions concerning the distribution of the analysed variable being close to normal distribution should be met in both the compared groups, as should variance homogeneity. Hence, if the distributions of the analysed variable differ significantly from normal distribution, then comparing the mean values makes no sense; worse still, it may lead to deeply false conclusions. This is clearly shown by the following theoretical example:

- two samples with sizes of N=10 each consist of the following elements:
  $Sample_1$: 44, 95, 48, 41, 46, 42, 45, 49, 47, 43
  $Sample_2$: 58, 55, 52, 59, 51, 54, 57, 53, 56, 5
- do the values of the above variable differ significantly between the samples?

Glancing at the listed values, it can easily be noticed that the first sample mainly consists of values in the 5^{th} ten, whereas the second sample mainly contains values in the 6^{th} ten. Thus, it can be assumed that the second sample contains higher values that the first sample. However, after

calculating the mean values and standard deviation, it turns out that these are identical for both samples:

$\text{Mean}_1 = \text{Mean}_2 = 50$

$\text{SD}_1 = \text{SD}_2 = 16.02$

Therefore, performing the parametric Student's t-test, which *de facto* compares two means, results in no significant differences ($p \approx 1.00$), similarly to the graphical representation of distributions by means of mean and SD (Fig. 1).

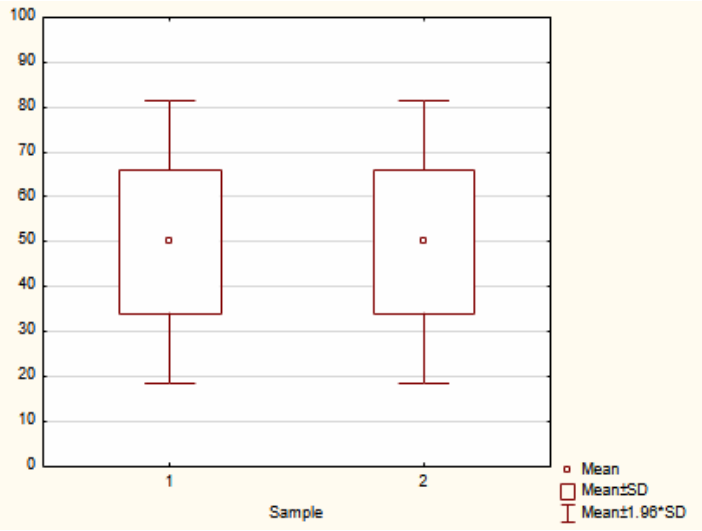


Figure 1. Comparison of the values of a variable between two samples using mean and SD

The situation is altogether different when the Mann-Whitney test, which is the nonparametric counterpart of Student's t-test, is used [11]. In this case, the differences are statistically significant ($p = 0.02$), even despite the small sizes of both samples. The graphical representation of the distributions, which uses the median and quartiles, gives the impression of illustrating a situation that is totally different from the one above (Fig. 2).

The simple example presented above is, again, purely theoretical. Yet, it clearly shows the huge importance of selecting the appropriate statistical method, testing the assumptions, and creating the correct graphical representation of the obtained results.

An example well-known in literature that illustrates the importance of graphical representation of data in relation to their statistical analysis is the so-called Anscombe's Quartet, a data set created in 1973 by the British statistician Francis Anscombe (Fig. 3) [12].

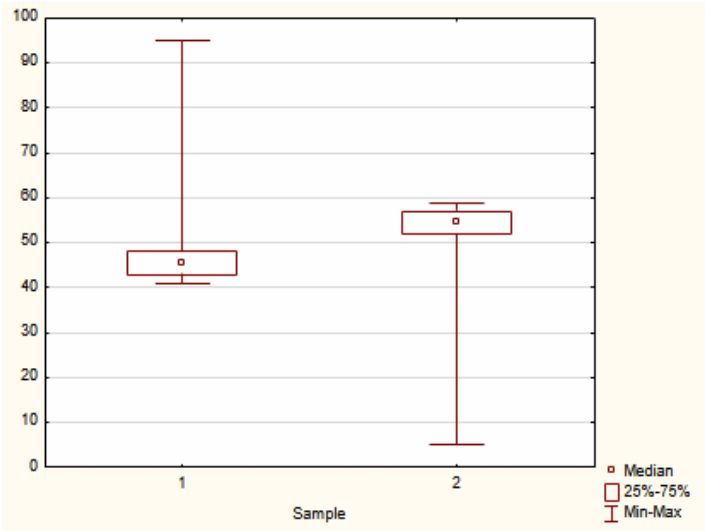


Figure 2. Comparison of the values of a variable between two samples using the median and quartiles

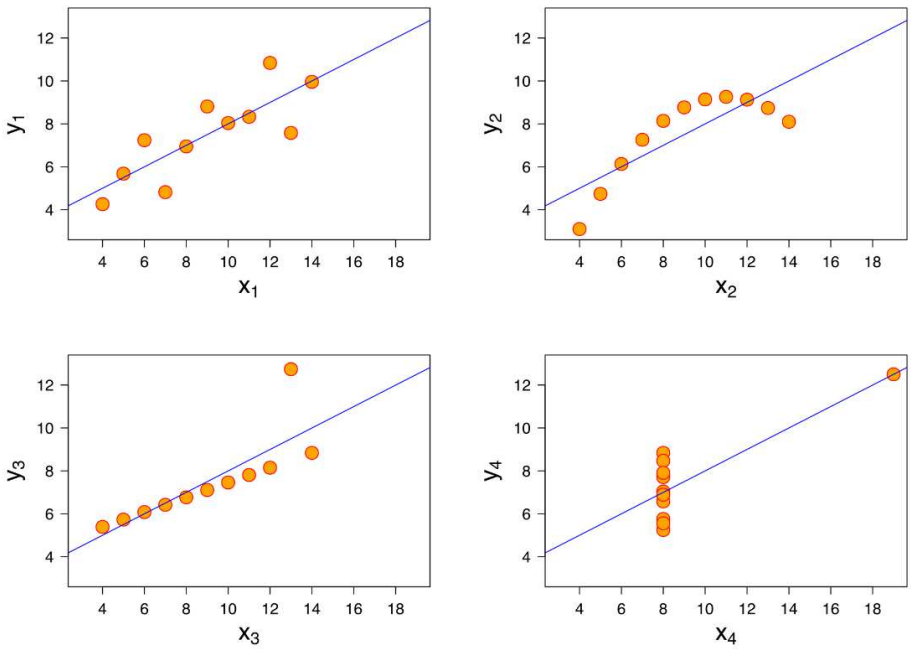


Figure 3. Anscombe's Quartet

Anscombe's Quartet is a set of four pairs of variables with identical statistical characteristics such as the arithmetic mean, variance, correlation coefficient, or the linear regression equation, with each of the pair looking totally different when represented graphically. The 1st set is data of the type that is often encountered in medicine, i.e. the existence of a relatively strong positive relationship between both variables is indicated. The 2nd set is a perfect curvilinear relationship based on the shape of the quadratic function. The 3rd set is a perfect linear relationship with one outlier; the 4th set is a set of points with the same X-axis value, also with one outlier [13–14].

Both examples presented above make it clear that, first of all, graphical representation of data is extremely important both in the course of biomedical research and at the stage of data presentation in scientific papers. Secondly, the selection of methods of presentation, similarly to the selection of the appropriate statistical tests (after testing the required assumptions) are absolutely essential in order to obtain true results, whereas lack of reliability in this area may dramatically change the results, thus leading to false conclusions drawn from the performed study [15–16].

Moreover, statistical models that are very extremely popular in biomedical papers (e.g. linear or logistic regression models, Poisson regression model, or Cox Proportional-Hazards model, which is the most popular model used in survival analysis) require specific assumptions to be met [17]. Examples of errors resulting from lack of verification of the correctness of the built statistical models would have been much more complex and less understandable that the simple examples presented above, which is why they have been omitted from this paper. It should be remembered, however, that the risk of obtaining false conclusions in the case of non-verified models is also very high, similarly to the case of lack of verification of the necessary assumptions when simple statistical tests are used.

Unfortunately, practice shows that a large proportion of medical literature consists of papers that contain methodological errors, connected either with method selection, testing of the required assumptions, appropriate graphical representation of results, or verification of the built models [18]. A large number of the published biomedical papers still contain the aforementioned errors, which results in them becoming a tool for the dissemination of – to a greater or lesser extent – false contents. It is also concerning that only few journals pay attention to the problem and verify in practice the reliability and correctness of analyses the papers they were sent had been based on. Involving a statistical editor in the publishing process also would not solve the problem as in order to verify the correctness of the performed analyses they would have to be granted access to raw data and

repeat the process of their analysis, paying attention to the stages of verification described above. Only when the procedure of verification of the correctness of statistical analyses becomes prevalent will it solve the issue of publishing reliable content in biomedical journals to a considerable degree. However, this process is too slow and many years will probably have passed before the majority of respectable biomedical journals adopt such procedures. Hence, what can be done "here and now" is taking care of the high level and wide scope of biostatistical education among students of biomedical courses – it is them who – in a few years or more – will be conducting key research, driving the global development of biomedicine. Whether they choose correctly between the proverbial Student's t-test and Mann-Whitney test in their research may in the end have an indirect impact on human life, and thus on our future and the future of the whole world.

REFERENCES

[1] Szczech L. A., Coladonato J. A. and Owen W. F. Key Concepts in Biostatistics: Using Statistics to Answer the Question "Is There a Difference?". Seminars in Dialysis. 15 (5), 2002, pp. 347–351.
[2] Lyon A. Why are Normal Distributions Normal? The British Journal for the Philosophy of Science, 65, pp. 621–649, 2014.
[3] Thode H. C. Testing For Normality. New York: CRC Press; 2002.
[4] Geisser S. Modes of Parametric Statistical Inference, John Wiley & Sons, 2006.
[5] Cox D. R. Principles of Statistical Inference, Cambridge University Press, 2006.
[6] Alexander R. G., Schmidt J., Zelinsky G. Z. Are summary statistics enough? Evidence for the importance of shape in guiding visual search. Visual Cognition. 22 (3–4), 2014, pp. 595–609.
[7] Gómez-Déniz E., Calderín-Ojeda E. and Gómez H. W. Symmetric and Asymmetric Distributions: Theoretical Developments and Applications III. Symmetry, 14, 2022, 2143.
[8] von Hippel P. T. Mean, Median, and Skew: Correcting a Textbook Rule. Journal of Statistics Education, 13, 2, 2005.
[9] Skovlund E. Bootstrapping – pulling oneself up by one's hair? Tidsskr Nor Laegeforen. 139 (17), 2019.
[10] Kalpić D., Hlupić N. and Lovrić M. Student's t-Tests. In: Lovric, M. (eds) International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg, 2011.
[11] Mann H. B., Whitney D. R. On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other. Annals of Mathematical Statistics, 18 (1), 1947, pp. 50–60.
[12] Anscombe, F. J. Graphs in Statistical Analysis. The American Statistician, 27, (1), 1973, pp. 17–21.
[13] Tufte, E. R. The Visual Display of Quantitative Information. Cheshire, Conn.: Graphics Press, 2001.

[14] Chatterjee S. and Firat A. Generating Data with Identical Statistics but Dissimilar Graphics: A Follow up to the Anscombe Dataset. The American Statistician, vol. 61, no. 3, 2007, pp. 248–54.
[15] Milewska A. J., Milewski R., Górska U. and Jankowska D. Statistical methods in Polish medical publicatons. Studies in Logic, Grammar and Rhetoric 21 (34), 2010, 81–90.
[16] Glantz S. A. Biostatistics: how to detect, correct and prevent errors in the medical literature. Circulation, 1980; 61, 1.
[17] Jarque C. M. and Bera A. K. Efficient tests for normality, homoscedasticity and serial independence of regression residuals. Economics letters, 6, 1980, pp. 255–259.
[18] Altman D. G. The scandal of poor medical research. BMJ 1994; 308: 283–284.