Main Issues of Statistical Analysis in Medical Research

Voxidov Alikul Melitoshevich Samarkand State Medical University, Samarkand

Voxidov Dilshod Alikulovich Samarkand State Medical University, Samarkand

ABSTRACT

This work is devoted to the main issues of statistical analysis of clinical trials. Statistical processing is a complex multi-stage process, the level of scientific organization of which decisively determines the quality of the accumulated statistical data, the results of their processing and comprehension.

Keywords: statistical processing, clinical trials

The goal: The article highlights the main issues of statistical analysis in clinical trials. Statistical processing is a complex multi-stage process, the level of scientific organization of which decisively determines the quality of the accumulated statistical data, the results of their processing and comprehension.

Modern medicine cannot be imagined without the use of accurate and reliable methods of analysis and forecasting. It is known that the theory of probability and mathematical statistics arose in the middle of the 17th century as a result of the development of society and commodity-money relations.

Statistics, like any other science, arose from the practical needs of people. Even in the ancient world, the need to collect taxes, perform military service and other social activities led to the need to take into account the population, its location, occupation.

Results: Like any science, statistics has its own subject of study and its own specific methods. The subject of statistics is the quantitative aspect of mass phenomena in any field, inextricably linked with their qualitative content. At the same time, the quantitative expression of the patterns of development of these phenomena is carried out taking into account the specific conditions of place and time [1,2,6,7].

The specific methods by which statistics studies mass phenomena form the statistical methodology (or method of statistics).

In any statistical study, three successive stages can be distinguished:

1. Statistical observation, that is, the collection of primary statistical material.
2. Grouping and summary of observation results.
3. Calculation of the specific characteristics of the phenomenon under study and analysis of the obtained summary and calculation materials, formulation of conclusions and proposals [1,2].

At these stages of statistical research, specific methods are used, which together form the content of the methodology of statistical science.

A common feature of the information that makes up statistics is that they always do not refer to one single (individual) phenomenon, but cover a whole group of such phenomena (elements) with related characteristics.

Statistical processing is a complex multi-stage process, the level of scientific organization of which decisively affects the quality of the accumulated statistical data, the results of their processing and comprehension.

Currently, a large number of scientific studies are being carried out in various areas of medical activity, but statistical processing does not always meet modern requirements, which raises serious doubts about the results provided. This situation causes some bewilderment, since there is a sufficient amount of literature on the peculiarities of statistical processing of medical data3, as well as a wide range of software (the package of applied statistical programs Biostat, Statistica 6.0 from Install Shield Software Corporation (USA); software package SPSS (PASW Statictics) for Windows, etc.).

The application of statistical methods is not a formal procedure, it is a creative activity. And, like any creative activity, planning research and interpreting the results obtained require deep knowledge in the field of mathematics2.

The responsibility is also great - often it is the statistical evaluation of the results of clinical trials that determines the decision in favor of this method of treatment. In practice, statistical analysis is either formally performed by physicians themselves, or, after data collection, their analysis is entrusted to mathematicians who are not related to medical applications. In addition, often insufficient attention is paid to research planning.

As a result, already after the end of the information collection stage, when trying to apply statistical tests, it may turn out that the collected data is not enough to form a statistically significant conclusion about the advantages of one compared treatment method over another. And here, even the most complex mathematical methods for analyzing the results obtained will not save the situation if they were obtained in the course of an illiterately planned study [2,6].

Pearson correlation coefficient

And yet, why can't regression analysis be used to describe the tightness of the relationship? The residual standard deviation could be used as a measure of the closeness of the relationship. However, if you swap the dependent and independent variables, then the residual standard deviation, like other indicators of the regression analysis, will be different. We choose a hypothetical example: in ancient times, there were two types of Homo Sapience on earth: the first group was tall, and the second group was pygmies (short).

According to the sample we know of from 12 ancient earthlings, two regression lines were built. In one case, the weight is the dependent variable, in the second it is the independent variable. The regression lines are noticeably different. It turns out that the relationship of height with weight is one, and weight with height is another. The asymmetry of regression analysis is what prevents it from being directly used to characterize the strength of a relationship. The correlation coefficient, although its idea stems from regression analysis, is free from this shortcoming. Here is the formula[3,6,7].

$$r = \frac{\sum(X - \bar{X})(Y - \bar{Y})}{\sqrt{\sum(X - \bar{X})^2 \sum(Y - \bar{Y})^2}}$$

where $\bar{X}$ and $\bar{Y}$ are the mean values of the variables X and Y.

The expression for r is "symmetrical" - by swapping X and Y, we get the same value. The correlation coefficient takes values from -1 to +1. The closer the relationship, the greater the absolute value of the correlation coefficient. The sign shows the direction of the connection. At $r > 0$, they speak of a direct correlation (with an increase in one variable, the other also increases), with $r < 0$, they speak of an inverse one (with an increase in one variable, the other decreases). the strongest direct correlation: $r = +1$. the strongest inverse correlation: $r =– 1$. the correlation is direct, also quite strong: r =

0.8. there is no relationship between the features: $r = 0$.

Let's take an example from 12 ancient earthlings, which we have already considered from the point of view of regression analysis. Let's calculate the correlation coefficient. The initial data and intermediate results of calculations are given in the table below.

Sample size n = 12, average height $\bar{X}$=$\sum\frac{X}{n}=\frac{1402}{12}=116{,}8$ and weight $\bar{Y}$=$\sum\frac{Y}{n}=\frac{222}{12}=18{,}5$

We find: $\sum(X-\bar{X})(Y-\bar{Y})=143{,}0$; $\sum(X-\bar{X})^2=241{,}7$; $\sum(Y-\bar{Y})^2=103{,}0$;

Let's substitute the obtained values into the formula for the correlation coefficient:

$$r=\frac{143{,}0}{\sqrt{241{,}7\times 103{,}0}}=0{,}91$$

The value of r is close to 1, which indicates a close relationship between height and weight. To get a better idea of which correlation coefficient should be considered large and which should be considered insignificant, take a look at the correlation coefficients

Table. The steps taken to calculate the correlation coefficient for group 2:

№	X	Y	$X-\bar{X}$	$Y-\bar{Y}$	$(X-\bar{X})(Y-\bar{Y})$	$(X-\bar{X})^2$	$(Y-\bar{Y})^2$
1	120	21	3,2	2,5	7,9	10,0	6,3
2	121	22	4,2	3,5	14,6	17,4	12,3
3	118	19	1,2	0,5	0,6	1,4	0,3
4	116	17	-0,8	-1,5	1,2	0,7	2,3
5	119	20	2,2	1,5	3,3	4,7	2,3
6	117	18	0,2	-0,5	-0,1	0,0	0,3
7	115	16	-1,8	-2,5	4,6	3,4	6,3
8	108	13	-8,8	-5,5	48,6	78,0	30,3
9	113	18	-3,8	-0,5	1,9	14,7	0,3
10	122	19	5,2	0,5	2,6	26,7	0,3
11	123	24	6,2	5,5	33,9	38,0	30,3
12	110	15	-6,8	-3,5	23,9	46,7	12,3
	116,8	18,5	0,0	0,0	143,0	241,7	103,0

The table shows the main calculated values for calculating the correlation coefficient in our example, which we analyzed above.

Relationship between regression and correlation

We initially used examples of correlation coefficients to build regression lines. Indeed, there is a close relationship between the correlation coefficient and the regression analysis parameters, which we will now demonstrate. Different ways of presenting the correlation coefficient, which we will get in this case, will allow us to better understand the meaning of this indicator.

Recall that the regression equation is constructed in such a way as to minimize the sum of squared deviations from the regression line.

Table. Examples of correlations

Example	Coefficient correlations r	Volume sample n
Height and weight of ancient earthlings (table)	0,91	12

Conclusion: Thus, the processing and presentation of data obtained in the course of the Pearson correlation coefficient in clinical trials includes several main stages: determining the nature of the analyzed feature (quantitative, qualitative); dependent or independent groups; determination of the type of distribution (normal or non-normal). The adequacy of the choice of statistical methods of analysis is decisive in the correct interpretation of the data obtained.

List of used literature:

1. Статистический портал StatSoft. http://www.statsoft.ru/home /portal/default. asp
2. Реброва О. Ю. Описание процедуры и результатов статистического анализа медицинских данных в научных публикациях. http://www.mediasphera. ru/mjmp/2000/4/r4-00-21.htm
3. Гланц С. Медико-биологическая статистика. М., 1999: стр. 250-255
4. Ефимова М. Р., Петрова Е. В., Румянцева В. Н. Общая теория статистики. М., 2005: 413 с.
5. Ребров О. Ю. Статистический анализ медицинских данных. Применение пакета прикладных программ STATISTICA. М., 2005: 305 с.
6. Voxidov A. M. et al. TIBBIY-BIOLOGIK TADQIQOTLARDA STATISTIK TAHLIL JARAYONLARI //Academic research in educational sciences. – 2022. – Т. 3. – №. 3. – С. 287-293.
7. Voxidov A. M., Malikov M. R., Voxidov D. A. TIBBIYOTDA DIFFERENSIAL TENGLAMALARNI FARMATSIYA SANOATIDA QO'LANISHI //Academic research in educational sciences. – 2021. – Т. 2. – №. 12. – С. 1096-1102.
8. Абдурахимов Р. М., Вохидов А. Клинико-лабораторные особенности течения синдрома эндогенной интоксикации у детей раннего возраста с пневмонией //Вестник Авиценны. – 2020. – Т. 22. – №. 3. – С. 403-408.
9. Вохидова Д. А. и др. PROBLEMS OF BIOLOGY AND MEDICINE.
10. Ne'matov N. I. TIBBIY VEB SAYTLAR YARATISH SAMARADORLIGI.
11. Malikov, M. R., Rustamov, A. A., & Ne'matov, N. I. (2020). STRATEGIES FOR DEVELOPMENT OF MEDICAL INFORMATION SYSTEMS. Theoretical & Applied Science, (9), 388-392.
12. Berdiyevna, A. S., & Olimjonovna, T. F. (2022). INNOVATIVE APPROACHES IN THE EDUCATION SYSTEM TO INCREASE YOUTH PARTICIPATION. Web of Scientist: International Scientific Research Journal, 3(3), 674-677.
13. Esirgapovich, K. A. (2022). THE EASIEST RECOMMENDATIONS FOR CREATING A WEBSITE. Galaxy International Interdisciplinary Research Journal, 10(2), 758-761.
14. Toxirova, F. O., Malikov, M. R., Abdullayeva, S. B., Ne'matov, N. I., & Rustamov, A. A. (2021). Reflective Approach In Organization Of Pedagogical Processes. European Journal of Molecular & Clinical Medicine, 7(03), 2020.
15. Ne'matov, N., & Rustamov, T. (2022). SANATORIYLAR ISHINI AVTOMATLASHTIRISH: BRON XIZMATI VA UNING STRUKTURASI. Eurasian Journal of Academic Research, 2(11), 763-766.