

PARAMETRIC VERSUS NONPARAMETRIC TESTS IN BIOMEDICAL RESEARCH

Miodrag Stojanović^{1,2}, Marija Andjelković-Apostolović^{1,2},
Zoran Milošević^{1,2}, Aleksandra Ignjatović^{1,2}

Despite the wide use of statistics in biomedical research, simple ideas are sometimes misunderstood or misinterpreted by medical research workers, who have only limited knowledge of statistics. This article deals with basic biostatistical concepts and their application to enable postgraduate medical students and researchers to analyze and interpret their study data and to critically interpret published literature. The adequate choice of statistical tests has a strong influence on data interpretation. Understanding this choice is important for critical evaluation of biomedical research. The question often arises on whether to use parametric or nonparametric test. If we are planning a study and trying to determine how many patients/cases to include, a nonparametric test will require a slightly larger sample size to have the same power as the corresponding parametric test. In summary, nonparametric procedures are useful in many cases and necessary in individual, but they are not the perfect solution. Fortunately, the most frequently used parametric analyses have their non-parametric counterparts. This can be useful when the assumptions of a parametric test are violated and we can thus choose a nonparametric alternative instead.

Acta Medica Medianae 2018;57(2):75-80.

Key words: Biomedicine, parametric tests, nonparametric tests

¹Public Health Institute Niš, Center for Informatics and Biostatistics in Health Care, Niš, Serbia

²University of Niš, Faculty of Medicine, Niš, Serbia

Contact: Miodrag Stojanović
Public Health Institute Niš
Blvd. dr Zorana Djindjića 50, 18105 Niš, Serbia
E-mail: mstojanovic@izjz-nis.org.rs

Introduction

Few people among medical students and physicians understand the differences between parametric and nonparametric statistics and most do not realize how important it is to make the right choice (1, 2). Statistics is basically a way of thinking about data that are changeable. Despite the wide use of statistics in biomedical research, simple ideas are sometimes misunderstood or misinterpreted by medical research workers, who have only limited knowledge of statistics. This article deals with basic biostatistical concepts and their application to enable postgraduate medical students and researchers to analyze and interpret their study data and to critically interpret published literature. We will try to explain the differences between parametric and nonparametric statistics and why it is crucial to know which

type of test is appropriate to use and in what situations.

The adequate choice of statistical tests strongly influences data interpretation. Understanding this choice is important for the critical evaluation of biomedical research. The question often arises on whether to use one or another test.

Biostatistics: The concept

Statistics is just a methodology and without scientific application it has no purpose. Statistics may thus be defined as a discipline concerned with the analysis of numerical data derived from a group of statistical elements. These statistical elements may be human beings, animals, or other organisms. Biostatistics is a branch of statistics applied to biological or medical sciences. Biostatistics covers the applications and contributions not only from health, medicines and nutrition, but also from the fields such as epidemiology, biology and genetics (3). Biostatistics involves various stages, like setting the hypothesis, collection of data and application of statistical analysis. In order to draw valid conclusions, researchers should know about the data obtained during the research, its distribution, and its analysis.

The first step, before considering any statistical analyses, is data research. Statistical methods for analysis mainly depend on the type of data. Generally, data present the picture of variability and central tendency. Therefore, it is very important to under-

stand the types of data. There exist three types of data: nominal, ordinal, and interval data. Nominal or categorical data simply assigned "names" or categories are based on the presence or absence of certain attributes/characteristics without any ranking between the categories (4). For example, humans are categorized by gender as males or females; by marital status as married, not married, widowed and divorced. Ordinal data, also called ordered, are the type of data which are expressed as scores or ranks. There is a natural order among categories, and they can be ranked or arranged in an order (4). For example, burns may be classified into four ranks and another example is the APGAR score. Interval data (continuous data) are the third type, which are characterized by an equal and definite interval between two measurements (some of the examples are weight, hemoglobin, body mass index).

The next step is to choose an adequate test for the analyses based on the type of collected data and some key features of that data. Hence, looking at the data, we are looking at data distributions to estimate the center, shape and spread and describe how the validity of many statistical procedures relies on an assumption of approximate normality (5). There are several statistical tests that can be used to assess whether the data are derived from a normal distribution. The most popular are the Kolmogorov-Smirnov test and the Shapiro-Wilk test (6). These normality tests take into account both the skewness and kurtosis of the data, and, therefore, the application of normality tests is recommended. These tests compare the observed data to quantiles of the normal (or other specified) distribution. The null hypothesis for each test is H_0 : Data follow a normal distribution, versus H_1 : Data do not follow a normal distribution. If the test is statistically significant (e.g., $p <$

0.05), then data do not follow a normal distribution, and a nonparametric test should be used.

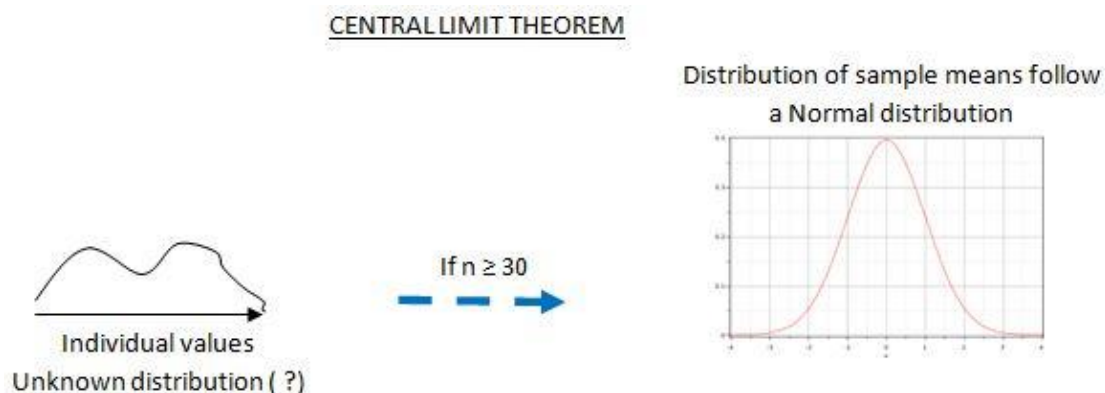
Therefore, we will try to explain the difference between parametric and nonparametric procedures. The principal difference for parametric versus nonparametric is:

- If measurement scale is nominal or ordinal, then we use nonparametric statistics;
- If we are using interval or ratio scales, we use parametric statistics.

Parametric tests

We can freely say that most people who use statistics are more familiar with parametric than nonparametric techniques. Parametric tests are based on the assumption that the data follow a normal or "bell-shaped" distribution. Parametric methods are often those for which we know that the population is approximately normal, or we can approximate using a normal distribution after we invoke the Central Limit Theorem. There are two parameters for a normal distribution: the mean and the standard deviation. Parametric tests are usually appropriate when examining either interval data or ratio data.

Altman states that „parametric methods require the observations within each group to have an approximately Normal distribution ... if the data do not satisfy these conditions ... a nonparametric method should be used" (7). According to the Central Limit Theorem (Graph 1), when the sample size is larger than 30, normality is not a main condition for a standard t (Student) or z hypothesis test: even though the individual values within a sample might follow an unknown, non-normal distribution, the sample means (as long as the sample sizes are at least 30) will follow a normal distribution.



Methods

If normality tests do not provide evidence for normal distribution, the data can be transformed to more normally distributed data. In some cases, the transformation of data will make it better to match

the assumptions. To transform the data, we perform a mathematical operation on each observation, and then use these transformed numbers in our statistical test. The most popular transformations are the log and square-root transformations (8). In situations when we cannot make the data more normally

distributed, we will select an equivalent nonparametric test. Commonly used parametric tests are described below.

Student t-Test:

The Student t-test is likely the most commonly applied parametric test. It was developed by a statistician William Sealy Gosset, who developed the „t-statistic“ and published it under the „Student“ pseudonym (9). A single sample t-test is used to determine whether the mean of a sample is different from a known average. A 2-sample t-test is used to establish that the means of two populations are equal. "Repeated measures" t-test is used to determine the differences between two responses measured on the same statistical units. One should know the mean, standard deviation, and number of samples to calculate the test statistic. In a data set with a large number of samples, the critical value for the Student t-test is 1.96 for an alpha of 0.05, and 2,58 for an alpha of 0.01, obtained from a t-test table.

The z-Test:

The z-test is very similar to the Student t-test. However, with the z-test, the variance of the standard population, rather than the standard deviation of the study groups, is used to obtain the z-test statistic. Using the z-chart, like the t-table, we can see what percentage of the standard population is outside the mean of the sample population. If, like the t-test, greater than 95% of the standard population is on one side of the mean, the p-value is less than 0.05 and statistical significance is achieved. The disadvantage of this test is that it should not be used if the sample size is less than 30.

The one-way analysis of variance (ANOVA):

The one-way analysis of variance (ANOVA) is used to determine whether there are any statistically significant differences between the means of three or more independent (unrelated) groups. The test statistic for ANOVA is called the F-ratio. As with the t- and z-statistics, the F-statistic is compared with a table to determine whether it is greater than the critical value. In interpreting the F-statistic, the degrees of freedom for both the numerator and the denominator are required. The degrees of freedom in the numerator are the number of groups minus 1, and the degrees of freedom in the denominator are the number of data points minus the number of group.

Further, to determine which specific groups differed from each other, we need to use a post hoc test (Bonferroni, Tukey, Duncan...), which represents a t-test modification. Two way ANOVA, also called two factors ANOVA, determines how a response is affected by two factors.

Pearson Correlation Coefficient:

The correlation coefficient (r) is a value that tells us how well two continuous variables from the same subject correlate to each other. An r value

may have values from -1 to +1: +1 means the data are completely positively correlated, an r of 0 means that the variables are completely random, and an r of -1 is completely negatively correlated. It is important to note that in biomedical research r could not be +1 or -1, because between the variables there is not any functional but statistical association. Further, the crucial thing to remember is that this is only an association and does not imply a cause-and-effect relationship.

Nonparametric tests

In biomedical sciences data often does not follow normal distribution (10) and the sample sizes are often small. Nonparametric tests are a satisfactory alternative to parametric tests for the data where there are skewness, extreme asymmetries and multimodality, especially in small samples. These tests are also called "distribution free tests" and represent statistical techniques for which we do not have to make any assumption of parameters for the population we are studying. According to Robson (11), non-parametric tests are usually appropriate when examining ordinal or nominal data when the assumptions of parametric test have not been achieved. A non-parametric statistical test is also a test whose model does not specify conditions about the parameters of the population from which the sample was taken. It does not require measurements as strong as that required for parametric tests. Non-parametric tests are generally appropriate when the data being examined is ordinal or nominal and is based on a small population sample or does not have a clear Gaussian function. In general, the measure of central tendency in nonparametric testing is median. Commonly used non-parametric tests are described below.

Pearson's chi-squared test

The Chi-square test is a non-parametric test of proportions. This test is not based on any assumption or distribution of any variable. This test, though different, follows a specific distribution known as Chi-square distribution, which is very useful in research. We use this test to determine whether there is a significant difference between the expected and observed frequencies in one or more categories. This test is used to investigate whether distributions of categorical variables differ from one another (10). The Chi-Square test of Independence is used to determine if there is a significant relationship between two nominal (categorical) variables. The frequency of one nominal variable is compared with different values of the second nominal variable. The data can be displayed in an $R \times C$ contingency table, where R is the row and C is the column. It has no alternative in parametric testing.

Mann-Whitney U test

This test is a nonparametric alternative for independent student t-test. It is used for continuous data, to compare the means of two independent or unrelated samples for significant differences. To com-

pute the U-test, data is ranked ordered and combined into a single dataset. This combination is used to determine if the rank ordering is random or clustered. If the data points of the sample are clustered, then there is evidence of a significant difference between the sample means. Conversely, randomly distributed rank ordered data would be the evidence that there is no significant difference between the means of the samples.

Wilcoxon signed-rank test

Wilcoxon signed-rank test is a nonparametric test that can be used to determine whether two dependent samples were selected from populations having the same distribution. It compares two related samples, matched samples, or repeated measurements on a single sample to assess whether their population mean ranks are different. It is used as an alternative to the paired Student's t-test, t-test for matched pairs, or t-test for dependent samples, when the population cannot be assumed to be normally distributed (12, 13).

Kruskal-Wallis test

Kruskal-Wallis test is a nonparametric test used for comparing two or more independent samples of equal or different sample sizes. It extends the Mann-Whitney U test when there are more than two groups. This test is the nonparametric equivalent of the ANOVA which can be used for both continuous and ordinal-level dependent variables. However, like most non-parametric tests, the Kruskal-Wallis Test is not as powerful as the ANOVA.

The Friedman test

The Friedman test is a non-parametric test for testing the difference between several related samples. This test is an alternative to Repeated measures analysis of variances which is used when the same parameter has been measured under different conditions on the same subjects.

Spearman rank correlation

Spearman rank correlation is a nonparametric alternative to the Pearson correlation coefficient. It assesses how well the relationship between two variables can be described using a monotonic function (10). This test measures the strength and direction of association between two ranked variables. Spearman rank correlation has less power than the Pearson correlation coefficient, and in situations when we can choose between the two, Pearson correlation coefficient is a better option.

Differences between parametric and non-parametric tests

The „power“ of a nonparametric test is lower

than that of its parametric counterpart. This means that to detect any given effect at a specified significance level, a larger sample size is required for non-parametric compared to parametric tests (13). They are generally less statistically powerful than the analogous parametric tests when the data are truly approximately normal. „Less powerful“ means that there is a smaller probability that the procedure will tell us that two variables are associated with each other when they in fact are really associated. Some people also debate if non-parametric tests are most appropriate when the sample sizes are small. However, when the data set is large, (e.g. $n > 30$), the Central Limit Theorem can be used, so it often makes little sense to employ nonparametric tests.

Another disadvantage associated with non-parametric tests is that their results are often more difficult to interpret than the results of parametric tests. Many nonparametric tests use data ranking values instead of using the actual data, hence the difference in mean ranks between two groups very often does not really contribute to our intuitive understanding of the data.

Non-parametric tests are appropriate for very small samples. However, if sample sizes as small as $N = 5$ are used, nonparametric tests have no alternatives. Non-parametric tests can treat samples made up of observations from several different populations, can treat data which are in ranks as well as data whose seemingly numerical scores have the strength in ranks. They are available to treat data which are classificatory, and are easier to learn and apply than parametric tests.

If we are planning a study and trying to determine how many patients/cases to include, a non-parametric test will require a slightly larger sample size to have the same power as the corresponding parametric test. In summary, nonparametric procedures are useful in many cases and necessary in individual, but they are not the perfect solution.

Fortunately, the most frequently used parametric analyses have their non-parametric counterparts. This can be useful when the assumptions of a parametric test are violated and therefore we can choose the nonparametric alternative. The examples are shown in Table 1.

Conclusion

The tests outlined here are commonly used in clinical studies. Understanding these tests will provide some framework for analyzing test results when critically reading journal articles. Inappropriate use of statistical tests will lead to incorrect conclusions. In general, we should try to avoid non-parametric tests whenever possible (because they are less powerful). In conclusion, the next time when you are having doubts about which test to employ, you should consult a statistician.

Table 1. Parametric tests and nonparametric counterparts

Statistical Tests		
Parametric test	Corresponding nonparametric test	Purpose of test
<i>t</i> test for independent samples	Mann-Whitney U test; Wilcoxon rank-sum test	Compares two independent samples
Paired <i>t</i> test	Wilcoxon matched pairs signed-rank test	Examines a set of differences
One way- ANOVA	Kruskal-Wallis analysis of variance by ranks	Compares three or more groups
Two way- ANOVA	Friedman Two way analysis of variance	Compares groups classified by two different factors
Pearson correlation coefficient	Spearman rank correlation coefficient	Assesses the linear association between two variables

References

- West CP, Ficalora RD. Clinician attitudes toward biostatistics. Mayo Clinic proceedings 2007; 82(8): 939-43. [[CrossRef](#)][[PubMed](#)]
- Windish DM, Huot SJ, Green ML. Medicine residents' understanding of the biostatistics and results in the medical literature. Jama 2007; 298(9): 1010-22. [[CrossRef](#)][[PubMed](#)]
- Dakhale GN, Hiware SK, Shinde AT, Mahatme MS. Basic biostatistics for post-graduate students. Indian Journal of Pharmacology 2012; 44(4): 435-42. [[CrossRef](#)][[PubMed](#)]
- Nerurkar RP. Basics of statistics for postgraduates. Indian journal of dermatology, venereology and leprology 2008; 74(6): 691-5. [[CrossRef](#)][[PubMed](#)]
- Neideen T, Brasel K. Understanding statistical tests. Journal of Surgical Education 2007; 64(2): 93-6. [[CrossRef](#)][[PubMed](#)]
- Ghasemi A, Zahediasl S. Normality Tests for Statistical Analysis: A Guide for Non-Statisticians. International Journal of Endocrinology and Metabolism 2012; 10(2): 486-9. [[CrossRef](#)][[PubMed](#)]
- Altman DG. Practical Statistics for Medical Research. London: Chapman and Hall; 1991. [[CrossRef](#)]
- Manikandan S. Data transformation. Journal of Pharmacology & Pharmacotherapeutics 2010; 1(2): 126-7. [[CrossRef](#)][[PubMed](#)]
- Raju TN. William Sealy Gosset and William A. Silverman: two "students" of science. Pediatrics. 2005; 116(3): 732-5. [[CrossRef](#)][[PubMed](#)]
- Gaddis ML, Gaddis GM. Introduction to biostatistics: Part 6, Correlation and regression. Annals of emergency medicine 1990; 19(12): 1462-8. [[CrossRef](#)][[PubMed](#)]
- Robson C. Experiment, design and statistics in physiology. 3rd ed. Pinguin books; 1994. [[CrossRef](#)]
- Wilcoxon F. Probability tables for individual comparisons by ranking methods. Biometrics 1947; 3(3): 119-22. [[CrossRef](#)][[PubMed](#)]
- Bridge PD, Sawilowsky SS. Increasing physicians' awareness of the impact of statistics on research outcomes: comparative power of the *t*-test and and Wilcoxon Rank-Sum test in small samples applied research. Journal of clinical epidemiology 1999; 52(3): 229-35. [[CrossRef](#)][[PubMed](#)]

Revijalni rad

UDC: 311:61
doi:10.5633/amm.2018.0212**PARAMETARSKI NASUPROT NEPARAMETARSKIM
TESTOVIMA U BIOMEDICINSKIM ISTRAŽIVANJIMA***Miodrag Stojanović^{1,2}, Marija Anđelković-Apostolović^{1,2},
Zoran Milošević^{1,2}, Aleksandra Ignjatović^{1,2}*

¹Institut za javno zdravlje Niš, Centar za informatiku i biostatistiku u zdravstvu, Niš, Srbija
²Univerzitet u Nišu, Medicinski fakultet, Niš, Srbija

Kontakt: Miodrag Stojanović
Institut za javno zdravlje
Bulevar dr Zorana Đinđića 50, 18105 Niš, Srbija
E-mail: mstojanovic@izjz-nis.org.rs

Uprkos širokoj upotrebi statistike u biomedicinskim istraživanjima, jednostavne ideje su ponekad pogrešno shvaćene ili pogrešno tumačene od strane medicinskih naučnih radnika, koji uglavnom imaju ograničeno znanje iz statistike. Ovaj članak se bavi osnovnim konceptima biostatistike i njene primene kako bi se studentima postdiplomskih studija medicine i istraživačima omogućilo da analiziraju i kritički tumače svoje podatke i dostupnu literaturu. Adekvatan izbor statističkih testova ima bitan uticaj na interpretaciju podataka. Razumevanje ovog izbora je bitno za kritičku procenu rezultata biomedicinskih istraživanja. Pitanje koje se često postavlja je da li koristiti parametarski ili neparametarski test. Ukoliko planiramo da sprovedemo određenu studiju i pokušavamo da utvrdimo koliko bolesnika/slučajeva je potrebno uključiti u nju, neparametarski test će zahtevati veću veličinu uzoraka da bi postigao istu snagu kao i odgovarajući parametarski test. Ukratko, neparametarskih testovi su korisni i neophodni u mnogim slučajevima, ali oni često nisu savršeno rešenje. Srećom, najčešće korišćene parametarske analize imaju svoje neparametarske ekvivalente. Ovo saznanje je korisno u slučaju kada raspred nije po tipu normalnosti, te stoga biramo neparametarsku alternativu.

Acta Medica Medianae 2018;57(2):75-80.

Ključne reči: *biostatistika, parametarski testovi, neparametarski testovi*

This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) Licence