АНАЛИЗ ДАННЫХ В МЕДИКО-БИОЛОГИЧЕСКИХ ЭКСПЕРИМЕНТАЛЬНЫХ ИССЛЕДОВАНИЯХ

Абрамова А.Ю.

ФГБНУ «НИИ нормальной физиологии им. П.К. Анохина», Москва, Россия
nansy71@mail.ru

В настоящее время значимость корректного использования статистических методов анализа в современных исследованиях является одной из самых актуальных проблем медико-биологической науки. Высокое качество статистического анализа является обязательным условием востребованности научных результатов и транспарентности научных достижений отдельных исследователей и исследовательских коллективов в международном научном сообществе [2].

Основная цель применения различных статистических методов – сведение к минимуму случайных ошибок в научном исследовании. При этом могут решаться следующие задачи: выдвижение и определение гипотез; планирование объемов выборок; статистическая проверка выдвигаемых гипотез. Следует подчеркнуть, что любой статистический метод будет хорош настолько, насколько качественными являются входные данные (англ. “garbage in - garbage out” или “хлам на входе - хлам на выходе”).

Статистический анализ полученных данных проходит несколько этапов. Первым из них является проведение мероприятий по описательной статистике, то есть расчет количественных статистических показателей. Наиболее часто используемыми из них являются среднее значение и медиана. Здесь же рассматривается вопрос о соответствии распределения нормальному распределению вероятности. В зависимости от статистически выявленного распределения, последующие этапы анализа можно условно разделить на две большие «ветки»: анализ данных с применением параметрических и непараметрических статистических критериев. Помимо указанного выбора существуют принципиальные различия между анализом данных двух групп сравнения, повторными измерениями (динамикой) и множественным сравнением [1].

На последнем этапе статистического анализа при множественных проверках статистических гипотез для исключения ошибки второго рода (ложная гипотеза) используются несколько поправок. Наиболее часто применяемые – это поправка Бонферрони (*Bonferroni correction*) [4] и FDR-контроль (*False Discovery Rate control*) [3]. Применение той или иной поправки (или критерия) зависит от ряда факторов, в том числе, от количества сравнений и мощности самого критерия.

В заключении следует отметить, что широкий диапазон специализированных программ и подходов в биомедицинской статистике, с одной стороны, дает возможность исследователю представить статистически значимые результаты, с другой – при наличии множества методов часто возникает вероятность ошибки в выборе алгоритма, необходимого для анализа данных.

Литература.

1. Гланц С. Медико-биологическая статистика / Пер. с англ. - М., Практика, 1998. 459 с.

2. Гржибовский А.М., Иванов С.В., Горбатова М.А. Сравнение количественных данных двух парных выборок с использованием программного обеспечения Statistica и SPSS: параметрические и непараметрические критерии // Наука и здравоохранение. 2016. № 3. С. 5-25.
3. Benjamini Y., Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. // Journal of the Royal Statistical Society. Series B (Methodological). 1995. Vol. 57. N 1. P. 289-300.
4. Shi Q., Pavey E.S., Carter R.E. Bonferroni-based correction factor for multiple, correlated endpoints // Pharm. Stat. 2012. Vol. 11. N 4. P. 9-300.

DATA ANALYSIS IN BIOMEDICAL EXPERIMENTAL RESEARCH

Abramova A. Yu.

**P.K. Anokhin Research Institute of Normal Physiology, Moscow, Russia
nansy71@mail.ru**

Currently, the importance of the correct use of statistical methods of analysis in modern research is one of the most pressing problems of biomedical science. High quality statistical analysis is a prerequisite for the demand for scientific results and the transparency of scientific achievements of individual researchers and research teams in the international scientific community [3].

The main purpose of applying various statistical methods is to minimize random errors in scientific research. In this case, the following tasks can be solved: the nomination and definition of hypotheses; sample size planning; statistical verification of hypotheses. It should be emphasized that any statistical method will be as good as the input data are good ("garbage in - garbage out").

Statistical data analysis goes through several stages. The first of these is the descriptive statistics, that is, the calculation of quantitative statistical values. The most commonly used of these are the *mean* and *median*. It also addresses the issue of whether the distribution is consistent with the normal probability distribution (*Kolmogorov-Smirnov test* and/or *Shapiro-Wilk's test*). Depending on the statistically determined distribution, the subsequent stages of the analysis can be divided into two large "branches": data analysis using parametric and non-parametric statistical criteria. In addition to this choice, there are fundamental differences between the data analysis of two comparison groups, repeated measurements (dynamics) and multiple comparisons [2].

At the last stage of statistical analysis with multiple tests of statistical hypotheses, several corrections are used to eliminate the error of the second kind (false hypothesis). The most commonly used ones are *Bonferroni correction* [4] and *FDR control* (False Discovery Rate control) [1]. The application of an amendment (or criterion) depends on

a number of factors, including the number of comparisons and the power of the criterion itself.

In conclusion, it should be noted that a wide range of specialized programs and approaches in biomedical statistics, on the one hand, allows the researcher to present statistically significant results, on the other hand, in the presence of a variety of methods, there is often a probability of error in choosing the algorithm necessary for data analysis.

Literature.

1. Benjamini Y., Hochberg Y. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. // Journal of the Royal Statistical Society. Series B (Methodological). 1995. Vol. 57. N 1. P. 289-300.
2. Glants S. Mediko-biologicheskaya statistika / Per. s angl. - M., Praktika, 1998. 459 p.
3. Grzhibovskiy A.M., Ivanov S.V., Gorbatoва M.A. Sravnenie kolichestvennykh dannyykh dvukh parnykh vyborok s ispol'zovaniem programmogo obespecheniya Statistica i SPSS: parametricheskie i neparametricheskie kriterii // Nauka i zdavookhranenie. 2016. № 3. P. 5-25.
4. Shi Q., Pavey E.S., Carter R.E. Bonferroni-based correction factor for multiple, correlated endpoints // Pharm. Stat. 2012. Vol. 11. N 4. P. 9-300.