

JUPYTER NOTEBOOK В ОПТИМИЗАЦИИ СТАТИСТИЧЕСКОЙ ОБРАБОТКИ ДАННЫХ ЭКСПЕРИМЕНТАЛЬНОЙ ПАТОЛОГИИ: ОЦЕНКА ПАРАМЕТРОВ ТРЕВОЖНОСТИ НА МОДЕЛИ БОКОВОГО АМИОТРОФИЧЕСКОГО СКЛЕРОЗА

Оробец М.А.¹, Кучеряну В.Г.², Голоборщев В.В.²

¹Первый Московский государственный медицинский университет имени И.М. Сеченова
Министерства здравоохранения Российской Федерации (Сеченовский университет),
Москва, Россия; margaret.orobets@gmail.com

²ФГБНУ «НИИ общей патологии и патофизиологии», Москва, Россия

Современные языки программирования стали неотъемлемым инструментом автоматизации обработки научных данных. В отличие от специализированных программ, они обеспечивают исследователям значительную методологическую гибкость: позволяют модифицировать алгоритмы в соответствии с требованиями конкретного исследования, интегрировать различные аналитические подходы и представлять результаты в настраиваемых форматах визуализации. Ярким примером такого инструментария служит Jupyter Notebook – интерактивная вычислительная среда на основе Python, R, Julia и др., сочетающая исполняемый код, текстовые аннотации и графические представления данных в едином документе.

Данный метод анализа экспериментальных данных, при условии компетентного применения общедоступных инструментов, обеспечивает комплексное решение задач статистической оценки результатов исследования, включая проверку нормальности распределения, расчёт значений p-value, определение мощности критериев, оптимизацию объёма выборок и генерацию репрезентативных графиков, в рамках единого процесса, адаптированного к специфике конкретного дизайна эксперимента. Сохранение алгоритма обработки данных и результатов уже проведённого (пилотного) исследования способствует обеспечению воспроизводимости научного эксперимента.

Так, для статистической оценки параметров тревожности у трансгенных животных, моделирующих боковой амиотрофический склероз, и контрольной группы с немодифицированным геномом в тесте «Приподнятый крестообразный лабиринт» нами был реализован алгоритм в среде Jupyter Notebook, обеспечивающий автоматизированный подбор статистических критериев и расчёт p-value для каждого исследуемого параметра. Анализ данных выявил тенденцию к увеличению времени пребывания генетически модифицированных мышей в преморбидной стадии (n=7) в закрытых рукавах лабиринта по сравнению с контролем (n=9), что может коррелировать с повышенной тревожностью экспериментальной группы (p = 0.056, Манна-Уитни U-test).

JUPYTER NOTEBOOK AS A TOOL FOR OPTIMIZING STATISTICAL DATA ANALYSIS IN EXPERIMENTAL PATHOLOGY: ASSESSMENT OF ANXIETY PARAMETERS IN THE AMYOTROPHIC LATERAL SCLEROSIS MODEL

Orobets Margarita A.¹, Kucheryanu Valerian G.², Goloborshcheva Valeria V.²

¹I.M. Sechenov First Moscow State Medical University of Ministry of Public Health of the Russian Federation (Sechenov University), Moscow, Russia; margaret.orobets@gmail.com.

² Institute of General Pathology and Pathophysiology, Moscow, Russia.

Modern programming languages have become an indispensable tool for automation of scientific data processing. Unlike specialized programs, it provides researchers with considerable methodological flexibility as it allows to modify algorithms according to the requirements of a particular study, to integrate various analytical approaches and to present results in customizable visualization formats. Jupyter Notebook is a striking example of such a toolkit. It is an interactive computing environment based on Python, R, Julia, etc. that combines executable code, text annotations, and graphical representations of data in a single document.

When implemented using properly utilized open-access tools, this analytical approach provides a comprehensive solution for the statistical evaluation of research outcomes. It integrates key tasks, including normality distribution assessment, p-value calculation, statistical power determination, sample size optimization, and generation of representative visualizations, within a unified workflow tailored to the experimental design. Preservation of the data processing algorithm and pilot study results facilitates reproducibility of the scientific experiment.

Thus, we implemented an algorithm in the Jupyter Notebook environment for the statistical evaluation of anxiety-related parameters in transgenic animal model of amyotrophic lateral sclerosis (ALS) and a control group with an unmodified genotype using the Elevated Plus Maze test. This algorithm enables automated selection of statistical criteria and calculation of p-values for each analyzed parameter. Data analysis revealed a tendency toward increased time spent in the closed arms of the maze by genetically modified mice in the premonitory stage (n=7) compared to controls (n=9), potentially correlating with elevated anxiety levels in the experimental group (p = 0.056, Mann-Whitney U-test).