

ВИЗУАЛИЗАЦИЯ ДАННЫХ С ПОМОЩЬЮ КОМБИНИРОВАННЫХ ДИАГРАММ В СРЕДЕ ПАКЕТА R ПРИ ИССЛЕДОВАНИИ МИКРОБИОМА ЧЕЛОВЕКА

А.С. Кройдер¹, М.В. Комарова¹, А.А. Трупакова², Д.Р. Рустянова²

1 «Самарский национальный исследовательский университет
имени академика С.П. Королёва», г. Самара

2 Самарский государственный медицинский университет, г. Самара

Визуализация данных — неотъемлемая часть научных исследований. Часто бывает необходимо показать не только средние или медианные значения количественных признаков, но и графически отразить их форму распределения. Диаграмма вида «усатый ящик», или `boxplot` представляет общую характеристику распределения значений количественной переменной. При её построении используют устойчивые оценки центральной тенденции и разброса: медиана, квартили, минимум и максимум. Скрипичные диаграммы, также называемые диаграммы типа виолончель, или `violinplot` позволяют оценить плотность распределения данных [1]. Они актуальны при анализе признаков с распределением, резко отличным от нормального: с выбросами или с двумя выраженными модами. Именно с такими случаями сталкиваются исследователи микробного пейзажа человека, когда в силу экологических законов происходит доминирование одних видов микроорганизмов и супрессия других.

Удобным средством визуализации таких данных служит построение комбинации диаграмм «`boxplot&violinplot`» среде R. R — свободно распространяемое программное обеспечение (<http://cran.r-project.org>) с большим количеством модулей, встроенных функций и гибкими настройками.

Для иллюстрации построения комбинированных графиков «`boxplot&violinplot`» использованы данные о количественном содержании лактобактерий (десятичные логарифмы геном-эквивалентов) в вагинальной микрофлоре у 220 девочек в возрасте от 2 до 18 лет. Графики построены с помощью модуля `ggplot2` пакета R 4.0. Применен код, показанный на рисунке 1. После символа `#` приведены комментарии. Использован массив данных `mydata`, лактобактерии закодированы в переменной `bac001`, стадии полового созревания — в переменной `stadia`. Полученный график представлен на рисунке 2.

```
tiff(filename="lacto.tif",width=1275, height=695, res=300, compression = "lzw")
# Задание имени и параметров выходного графического файла

ggplot(mydata, aes(x= bac001, y= stadia, fill = factor (stadia))) + # Выбор данных
geom_violin (trim= TRUE) + # Добавление вида графика violinplot
geom_boxplot (position=position_dodge(.9), width = 0.06, show.legend =FALSE)+ #
#Добавление графика boxplot и настройка его параметров
theme_bw () + theme(text = element_text(size=10)) + # Настройка размера текста
guides(fill= FALSE)+ # Отключение указателя
scale_fill_manual (values=c ("#FFFFFF", "#FFFFFF", "#FFFFFF", "#FFFFFF",
"#FFFFFF")) + # Выбор цветов графика
xlim (-1, 9) + # Выбор диапазона значений по оси x
xlab("Лактобактерии, ГЭ/обр") + # Название оси x
ylab ("Стадии полового созревания") # Название оси y

dev.off() # Завершение сохранения рисунка
```

Рисунок 1 – Код для создания комбинированных графиков в модуле ggplot2 пакета R

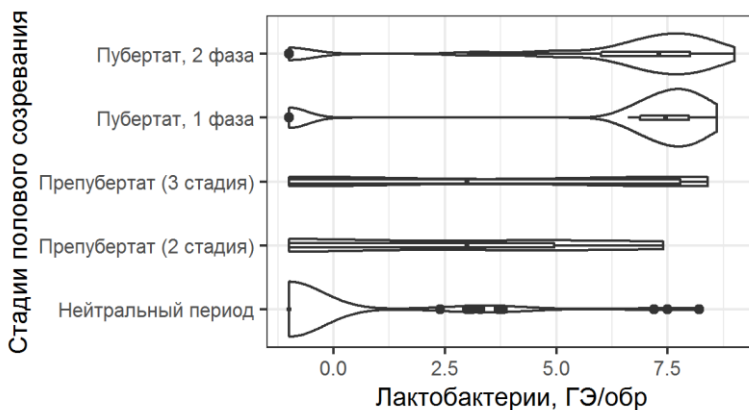


Рисунок 2 – Комбинированная диаграмма «boxplot&violinplot» для количества вагинальных лактобактерий у девочек на различных стадиях полового созревания

Построенные комбинированные гистограммы наглядно отражают биологические закономерности: преобладающее отсутствие лактобактерий в нейтральном периоде (скошенное влево распределение), всевозможное их содержание в препубертате (равномерное или U-образное распределение) и выраженное доминирование в пубертате (скошенное право распределение).

Список использованных источников

1. Мاستицкий, С.Э. Статистический анализ и визуализация данных с помощью R, [Текст] / С. Э. Мастицкий, В. К. Шитиков ; М32. – М.: ДМК Пресс, 2015 – 496 с.: цв. ил. ; Библиогр.: с.89-91. – ISBN 978-5-97060-301-7.

Кройдер Анастасия Сергеевна, студент кафедры ЛБС гр. 6369-120304D E-mail kroidernastya@yandex.ru

Комарова Марина Валериевна, к.б.н., доцент кафедры ЛБС, E-mail marinakom@yandex.ru

Трупакова Анна Андреевна врач акушер-гинеколог, соискатель кафедры акушерства и гинекологии Института педиатрии СамГМУ, E-mail anantru@mail.ru

Рустянова Дарья Рафиковна врач акушер-гинеколог, ассистент кафедры акушерства и гинекологии Института педиатрии СамГМУ, E-mail d.g.rustyanova@samsmu.ru

УДК 535.3; 51-73; 616-006

АНАЛИЗ РАМАНОВСКИХ СПЕКТРОВ КОЖИ МЕТОДОМ РАЗРЕШЕНИЯ МНОГОМЕРНЫХ КРИВЫХ

И.А. Матвеева

«Самарский национальный исследовательский университет
имени академика С.П. Королева», г. Самара

Ключевые слова: новообразования, метод разрешения многомерных кривых, рамановский спектр, рамановская спектроскопия.

В последние годы для исследования различных новообразований кожи все чаще применяют рамановскую спектроскопию [1]. Во время развития новообразования происходят биохимические изменения, которые приводят к изменению измеренных рамановских спектров [2].

Однако, несмотря на развитие технических средств для регистрации рамановского рассеяния света, анализ зарегистрированных спектров все еще остается сложной задачей, так как спектры содержат чрезвычайно большой объем информации обо всех химических веществах, которые могут быть в коже.

Среди существующих методов анализа рамановских спектров можно отметить метод главных компонент (РСА) и нейронные сети. Однако, эти методы имеют существенное ограничение в применении к реальным данным – нефизичность, то есть невозможность физической интерпретации результатов. Актуальным является поиск новых методов анализа рамановских спектров, которые смогли бы преодолеть этот недостаток.

В этом исследовании предлагается использовать метод разрешения многомерных кривых (MCR) для анализа рамановских спектров кожи. В последнее время этот метод нашел широкое применение в биологии и медицине для анализа спектральных данных [3].

Суть метода MCR состоит в разделении оптимальным образом матрицы рамановских спектров D на две меньшие матрицы C и S^T , где C – матрица концентраций компонентов образца кожи, S – матрица спектров этих компонентов. Уравнение, иллюстрирующее этот подход, можно записать в следующем виде:

$$D = C \cdot S^T + E \quad , \quad \text{где } E \text{ – матрица ошибок [3].}$$