# МЕТОДЫ СБОРА, ГРУППИРОВКИ И ПРЕДСТАВЛЕНИЯ СТАТИСТИЧЕСКИХ ДАННЫХ. МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ В БИОМЕДИЦИНСКОЙ ИНЖЕНЕРИИ

**Каримов Амир Улугбекович, Нарзиев Камаридин Джамшидович**
**студенты лечебного факультета Самаркандского государственного медицинского университета**
**Вохидов Аликул Мелитошевич**
**Старший преподаватель, кандетат экономических наук**
**Самаркандского государственного медицинского университета**

## АННОТАЦИЯ

В данной научной статье подробно рассматриваются методологические основы сбора, первичного контроля и группировки статистических данных, построения вариационных рядов и представления результатов в табличной и графической формах в области биомедицинской инженерии. Освещены виды статистического наблюдения, принципы группировки количественных и качественных переменных, методика определения ширины интервала с использованием формулы Стерджеса, а также способы расчёта абсолютных и относительных частот. На практических и клинических примерах продемонстрированы автоматизированная обработка данных с использованием функций СЧЁТЕСЛИ и других инструментов в программной среде Microsoft Excel 2013, а также построение гистограмм и столбчатых диаграмм. В заключительной части статьи проанализированы типичные методологические ошибки, возникающие при представлении статистических данных, и рассмотрены способы их предотвращения.

**Ключевые слова:** статистические данные, статистическое наблюдение, вариационный ряд, вариант, частота, относительная частота, формула Стерджеса, интервальная группировка, статистическая таблица, гистограмма, Microsoft Excel, биомедицинская инженерия.

# METHODS OF DATA COLLECTION, GROUPING, AND PRESENTATION OF STATISTICAL DATA. METHODOLOGICAL FOUNDATIONS IN BIOMEDICAL ENGINEERING

**Amir Ulugbekovich Karimov, Kamaridin Jamshidovich Narziev**

**Students of the Faculty of Medicine, Samarkand State Medical University**
**Alikul Melitoshevich Vokhidov**
**Senior Lecturer, Candidate of Economic Sciences, Samarkand State Medical University**

**ABSTRACT**

This scientific article examines in detail the methodological foundations of collecting, initially checking, and grouping statistical data, constructing variational series, and presenting results in tabular and graphical forms in the field of biomedical engineering. It discusses the types of statistical observation, the principles of grouping quantitative and qualitative variables, the method for determining interval width using Sturges' formula, and methods for calculating absolute and relative frequencies. Practical and clinical examples demonstrate automated data processing using the COUNTIF function and other tools in Microsoft Excel 2013, as well as the construction of histograms and bar charts. The final section analyzes common methodological errors in the presentation of statistical data and considers approaches to preventing them.

**Keywords:** statistical data, statistical observation, variational series, variate, frequency, relative frequency, Sturges' formula, interval grouping, statistical table, histogram, Microsoft Excel, biomedical engineering.

**ВВЕДЕНИЕ**

Одним из начальных и определяющих этапов любого научно-практического и биомедицинского исследования является сбор объективных и достоверных статистических данных [1]. В области биомедицинской инженерии данные формируются с помощью современных медицинских сенсоров, диагностических аппаратов (ЭКГ, ЭЭГ, мониторов $SpO_2$), лабораторных исследований и клинических экспериментов [2].

Однако исходный массив необработанных данных, например значения артериального давления или частоты сердечных сокращений, полученные у разных пациентов, сам по себе не всегда является достаточным основанием для формирования обоснованных научных выводов. Каждый числовой показатель необходимо представить в информационной форме, содержащей

сведения о его единице измерения, содержательном значении и объекте, к которому он относится [3].

Цель данной статьи заключается в методологическом обосновании методов сбора и систематизации статистических данных в биомедицинской инженерии, построения вариационных рядов, интервальной группировки, а также научно корректного представления результатов в виде таблиц и гистограмм с использованием программной среды Microsoft Excel.

**АНАЛИЗ ЛИТЕРАТУРЫ**

В международной научной практике накоплен значительный опыт исследования методов сбора статистических данных и их визуализации. Как подчёркивает Э. Тафти [1] в своих классических работах по визуальному представлению количественной информации, графики и таблицы должны наглядно и без искажения отражать скрытые закономерности и структуру данных.

Вопросы группировки медицинских и биомедицинских данных, а также построения вариационных рядов подробно рассматриваются в исследованиях Б. Рознера [2] и С. Гланца [3]. По мнению этих авторов, высокая вариабельность биологических показателей требует особого внимания к выбору интервалов и принципам группировки, поскольку некорректное распределение данных по группам может скрыть характерные различия между исследуемыми совокупностями.

Формула определения количества групп, предложенная Г. А. Стерджесом [4], и её применение в инженерной и медицинской статистике получили дальнейшее обоснование в работах Д. К. Монтгомери [5]. Д. Г. Альтман [6] проанализировал основные методологические ошибки, встречающиеся в статистических таблицах и диаграммах, и показал, что неправильный выбор типа графического представления может привести к искажённой интерпретации научных результатов.

**МЕТОДОЛОГИЯ И ТЕОРЕТИЧЕСКИЕ ОСНОВЫ**

Статистическое наблюдение представляет собой процесс сбора информации об изучаемых явлениях и процессах в соответствии с заранее разработанными планом и программой [2].

В биомедицинской инженерии применяются три основные формы статистического наблюдения:

Непосредственное наблюдение. Исследователь непосредственно измеряет и регистрирует показатели с помощью медицинских сенсоров и датчиков.

Документальное наблюдение. Используются существующие медицинские карты, результаты лабораторных анализов и оцифрованные базы данных.

Наблюдение на основе анкетирования. Количественные, категориальные и субъективные сведения собираются у участников исследования посредством специально разработанных анкет.

Перед проведением статистического анализа необходимо осуществить логический и арифметический контроль исходных данных [3]. Например, при обработке набора значений температуры тела пациентов (36,6; 36,8; 36,5; 37,1) следует проверить корректность ввода каждого значения и исключить возможные ошибки регистрации.

В процессе логической проверки контролируются наличие пустых ячеек, повторяющихся записей, экстремальных значений, а также соответствие используемых единиц измерения. При обнаружении ошибок исправления должны осуществляться на основании первоисточника, а не путём произвольного изменения наблюдаемых значений.

Вариационный ряд, частота и относительная частота

Исходный неупорядоченный массив данных располагают в порядке возрастания или убывания значений, формируя вариационный ряд.

Варианта $x_i$ — отдельное числовое значение, встречающееся в вариационном ряду.

Частота $f_i$ — количество повторений соответствующей варианты в исследуемом массиве данных.

Относительная частота $w_i$ — отношение частоты определённого значения к общему количеству наблюдений $n$:

$$w_i = \frac{f_i}{n}$$

где:

$w_i$ — относительная частота;

$f_i$ — частота соответствующего значения;

$n$ — общее количество наблюдений.

Для выражения относительной частоты в процентах используется формула:

$$w_i(\%) = \frac{f_i}{n} \times 100\%$$

Пример. Значение 72 уд/мин зарегистрировано в двух из восьми наблюдений.

$$w = \frac{2}{8} = 0{,}25$$

В процентном выражении:

$$w(\%) = 0{,}25 \times 100\% = 25$$

Следовательно, в 25 % наблюдений частота сердечных сокращений составляла 72 уд/мин.

Интервальные вариационные ряды и формула Стерджеса

При наличии большого объёма количественных данных, особенно при (n>30), целесообразно объединять наблюдения в интервальные группы. Для определения количества групп (k) используется формула Г. А. Стерджеса [4]:

$$k = 1 + 3.322 \times log_{10}(n)$$

где $n$ — объём совокупности наблюдений.

Ширина интервала $h$ определяется как отношение разности максимального и минимального значений к количеству групп:

$$h = \frac{x_{max} - x_{min}}{k}$$

где $X_{max}$ — максимальное значение, $X_{min}$ — минимальное значение, $k$ — количество групп.

Полученное значение количества групп при необходимости округляют до целого числа. При формировании интервалов следует обеспечить однозначное отнесение каждого наблюдения к соответствующей группе.

**РЕЗУЛЬТАТЫ И ПРАКТИЧЕСКИЙ ПРИМЕР**

Пример интервальной группировки в биомедицинской инженерии

В ходе исследования с использованием разработанного датчика мониторинга сердечного ритма была измерена частота сердечных сокращений у 20 пациентов.

Исходные данные (уд/мин):

**68, 72, 75, 72, 80, 78, 75, 82, 76, 72, 80, 74, 78, 75, 82, 70, 76, 80, 78, 72.**

Упорядоченный вариационный ряд:

**68, 70, 72, 72, 72, 72, 74, 75, 75, 75, 76, 76, 78, 78, 78, 80, 80, 80, 82, 82.**

Результаты подсчёта частот представлены в таблице.

| **Варианта $x_i$, уд/мин** | **Частота $f_i$** | **Относительная частота $w_i$** | **Процент, %** |
|---|---|---|---|
| 68 | 1 | 0,05 | 5 |
| 70 | 1 | 0,05 | 5 |
| 72 | 4 | 0,20 | 20 |
| 74 | 1 | 0,05 | 5 |
| 75 | 3 | 0,15 | 15 |
| 76 | 2 | 0,10 | 10 |
| 78 | 3 | 0,15 | 15 |
| 80 | 3 | 0,15 | 15 |
| 82 | 2 | 0,10 | 10 |
| **Итого** | **20** | **1,00** | **100** |

Представленная таблица позволяет определить частоту встречаемости каждого значения, оценить распределение показателя в исследуемой выборке и подготовить данные для последующего графического представления.

Автоматизированная обработка данных в Microsoft Excel 2013

Для автоматического подсчёта частот в Microsoft Excel 2013 используется формула:

=СЧЁТЕСЛИ($B$2:$B$21;C2)

В данном случае диапазон $B$2:$B$21 содержит исходные значения частоты сердечных сокращений, а ячейка C2 — значение, частоту которого необходимо определить.

Относительная частота рассчитывается с помощью формулы:

=D2/СЧЁТ($B$2:$B$21)

После вычисления результат можно отобразить в процентном формате, используя соответствующую настройку формата ячейки.

Для визуального представления распределения данных строятся гистограммы и столбчатые диаграммы. В зависимости от типа диаграммы и конфигурации Excel 2013 для её построения используется меню «Вставка» → «Диаграммы» с выбором соответствующего типа графика.

Диаграмме рекомендуется присвоить название: «Распределение частоты сердечных сокращений пациентов».

При этом следует учитывать, что гистограмма применяется для представления распределения количественной переменной по интервалам, тогда как столбчатая диаграмма преимущественно используется для сравнения отдельных категорий или дискретных значений.

## ОБСУЖДЕНИЕ И ТИПИЧНЫЕ МЕТОДОЛОГИЧЕСКИЕ ОШИБКИ

При сборе, группировке и представлении статистических данных могут возникать следующие методологические ошибки [6]:

Группировка данных без предварительной проверки и упорядочивания исходных значений. Смешение абсолютных частот и относительных частот, выраженных в процентах.Отсутствие единиц измерения и информативных заголовков в статистических таблицах.

Отождествление гистограммы, предназначенной для представления распределения количественных данных, со столбчатой диаграммой, используемой для сравнения категорий.

Выбор чрезмерно большого или недостаточного количества групп и интервалов, приводящий к искажению структуры распределения данных.

Формирование поспешных клинических выводов исключительно на основании визуального представления результатов без проведения необходимой статистической проверки гипотез.

Предотвращение указанных ошибок предполагает систематический контроль качества исходных данных, обоснованный выбор методов группировки, корректное применение статистических инструментов и соблюдение требований к оформлению таблиц и графиков. В биомедицинских исследованиях графическая интерпретация должна дополнять, а не заменять полноценный статистический анализ.

**ЗАКЛЮЧЕНИЕ**

Сбор и проверка первичных данных в биомедицинской инженерии являются фундаментальной основой объективности и достоверности научного исследования.

Построение вариационных рядов и интервальная группировка с использованием формулы Стерджеса позволяют систематизировать большие массивы данных, делая их более компактными и доступными для анализа.

Корректное применение инструментов Microsoft Excel обеспечивает автоматизированный расчёт частот и относительных частот, сокращает время обработки данных и упрощает создание наглядных гистограмм и диаграмм.

При составлении статистических таблиц и графиков необходимо строго соблюдать методологические требования, правильно указывать единицы измерения и выбирать способы визуализации, соответствующие типу исследуемых данных.

Таким образом, грамотная организация сбора, группировки и представления статистических данных способствует повышению качества биомедицинских исследований и обоснованности выводов, получаемых при анализе физиологических и клинических показателей.

## СПИСОК ЛИТЕРАТУРЫ


1. Tufte, E. R. (2001). *The Visual Display of Quantitative Information*. Graphics Press.
2. Rosner, B. (2015). *Fundamentals of Biostatistics*. Cengage Learning.
3. Glantz, S. A. (2012). *Primer of Biostatistics*. McGraw-Hill Medical.
4. Sturges, H. A. (1926). The choice of a class interval. *Journal of the American Statistical Association*, 21(153), 65–66.
5. Montgomery, D. C., & Runger, G. C. (2018). *Applied Statistics and Probability for Engineers*. John Wiley & Sons.
6. Altman, D. G. (1991). *Practical Statistics for Medical Research*. CRC Press.
7. Mamatqulov B. Tibbiyot statistikasi (biostatistika) asoslari. — Toshkent: G‘afur G‘ulom nomidagi nashriyot-matbaa ijodiy uyi, 2005. — 140 b.