

ОГЛАВЛЕНИЕ

ОТ АВТОРОВ	6
Глава I. ТЕРМИНОЛОГИЯ.....	8
1.1. Основные термины математической статистики.....	8
1.2. Смысл понятия «случайная величина»	11
1.3. Понятие о качественных и количественных величинах.....	11
1.4. Термины описательной статистики (Descriptive statistics).....	12
1.5. Другие термины математической статистики	16
Глава II. РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНЫХ ДАННЫХ	20
2.1. Виды распределения данных.....	20
2.2. Тестирование выборки на нормальность распределения	24
(Distribution fitting).....	
Глава III. ПРОВЕРКА ГИПОТЕЗ МЕТОДАМИ МАТЕМАТИЧЕСКОЙ	
СТАТИСТИКИ	31
3.1. F-критерий Фишера (сравнение двух выборочных дисперсий; F-test).....	31
3.2. Проверка гипотезы о равенстве двух средних при помощи t-критерия Стьюдента для независимых выборок	32
3.3. Одновыборочный тест (One sample t-test).....	35
3.4. Сравнение двух средних значений связанных выборок при помощи t-критерия Стьюдента (разностный метод; Paired t-test).....	36
3.5. Множественные сравнения.....	37
3.6. Сравнение нескольких групп с контрольной	38
3.7. Однофакторный дисперсионный анализ (ANOVA — analysis of variance; One-way analysis of variance).....	39
3.8. Непараметрические критерии (Nonparametric tests).....	40
Глава IV. АНАЛИЗ КАЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ.....	43
4.1. Анализ относительных значений.....	43
4.2. Таблицы сопряженности.....	45
4.3. Критерий χ^2 в анализе таблиц сопряженности.....	47
4.4. Точный критерий Фишера	51
4.5. Критерий Мак-Нимара	52

Глава V. АНАЛИЗ ВЗАИМОСВЯЗЕЙ	54
5.1. Функциональная, корреляционная и стохастическая зависимость.....	54
5.2. Коэффициент линейной корреляции (Correlation analysis)	55
5.3. Коэффициенты ранговой корреляции	56
5.4. Регрессионный анализ.....	57
Глава VI. АНАЛИЗ ВЫЖИВАЕМОСТИ (Кляшторный В.Г.).....	60
6.1. Графическое представление данных об изучаемом событии.....	61
6.2. Методы анализа данных типа «время до наступления события»	62
6.3. Построение таблиц времен жизни.....	63
6.4. Метод Каплана — Мейера.....	64
6.5. Сравнение кривых выживаемости	67
6.6. Регрессия Кокса	68
6.7. Примеры использования методов анализа выживаемости	70
Глава VII. АНАЛИЗ КАЧЕСТВА ДИАГНОСТИЧЕСКИХ МЕТОДОВ.....	74
7.1. Воспроизводимость метода исследования.....	74
7.2. Чувствительность диагностического метода (Sensitivity)	75
7.3. Специфичность диагностического метода (Specificity).....	76
7.4. Точность диагностической процедуры (Accuracy)	76
7.5. Применение таблицы сопряженности формата 2 × 2 для сравнения диагностической эффективности двух методов исследования	76
7.6. Отношение правдоподобия (Likelihood ratio).....	78
Глава VIII. ВИЗУАЛИЗАЦИЯ ДАННЫХ.....	79
8.1. Основные правила иллюстративного представления результатов научного исследования.....	79
8.2. Систематизация иллюстраций, применяемых для представления результатов научного исследования	80
8.3. Диаграммы, демонстрирующие особенности распределения выборочных данных	81
8.4. Диаграммы, предназначенные для иллюстрации базовых показателей описательной статистики	83
8.5. Диаграммы, иллюстрирующие изменение показателя (диаграммы динамики).....	85
8.6. Диаграммы, предназначенные для сравнения показателей	88
8.7. Диаграммы, демонстрирующие доли (структурные диаграммы)	93
8.8. Схематические изображения	98
8.9. Пустые иллюстрации.....	101
Глава IX. АЛГОРИТМЫ СТАТИСТИЧЕСКОГО АНАЛИЗА.....	103
СПИСОК ЛИТЕРАТУРЫ.....	107

ПРИЛОЖЕНИЯ

Приложение 1.	Критические значения критерия χ^2 при разных числах степеней свободы.....	109
Приложение 2.	Критические значения двустороннего t -критерия Стьюдента при разных числах степеней свободы	113
Приложение 3.	Критические значения коэффициента асимметрии, используемого для проверки гипотезы о нормальности распределения	118
Приложение 4.	Критические значения коэффициента эксцесса, используемого для проверки гипотезы о нормальности распределения	119
Приложение 5.	Критические значения критерия Шапиро — Уилка W при разных уровнях значимости.....	120
Приложение 6.	Критические значения критерия знаков (Z), соответствующие разным уровням значимости и объему выборки (n).....	121
Приложение 7.	Критические значения W -критерия Вилкоксона, применяемого для сравнения выборок с попарно связанными вариантами.....	122
Приложение 8.	Критические значения двустороннего F -критерия Фишера на уровне значимости $\alpha = 0,05$	123
Приложение 9.	Правила отбраковки полученных результатов	126
Приложение 10.	Правила округления результатов исследования	128
Приложение 11.	Обзор возможностей компьютерных программ для статистического анализа	129
Приложение 12.	Перечень англоязычных терминов, аббревиатур, символов и условных сокращений, используемых в англоязычной научной литературе и компьютерных программах для статистического анализа	131

Глава II. РАСПРЕДЕЛЕНИЕ ВЫБОРОЧНЫХ ДАННЫХ

2.1. Виды распределения данных

Очень часто исследователи медицинских явлений проверяют гипотезы о статистической значимости различий между эмпирическими выборками при помощи критерия Стьюдента, а выраженность связи между явлениями оценивают линейной корреляцией, однако эти пространственные методы следует использовать лишь при нормальном распределении. Поэтому изучение вида распределения данных должно стать неотъемлемой частью исследования, в котором применены параметрические критерии. Из этого следует, что, желая воспользоваться t -критерием Стьюдента, необходимо быть уверенным в нормальности распределения сравниваемых выборок.

Распределение Гаусса (*Gaussian distribution*)

Этот тип распределения данных можно обнаружить, если на исследуемый признак одновременно оказали и/или оказывают влияние большое число малых независимых друг от друга факторов. Данное утверждение подтверждается так называемой центральной предельной теоремой, имеющей важнейшее значение для теории статистического анализа. Этот тип распределения данных нередко встречается в биологии и медицине, поэтому его назвали нормальным. Нормальное распределение детально изучил в 1809 г. профессор математики университета в Геттингене Карл Фридрих Гаусс (*Johann Karl Friedrich Gauss*, 1777–1855). Немалый вклад в изучение этого распределения сделал Лаплас (*Pierre-Simon Laplace*, 1749–1827), поэтому его нередко называют распределением Гаусса — Лапласа. График плотности нормального распределения имеет вид колокола (рис. 1). Пик соответствует показателям центра распределения. Одной из важнейших характеристик нормального распределения является постепенное симметричное уменьшение частот по мере удаления от центра распределения (рис. 2). Другая важная характеристика нормального распределения — совпадение средней арифметической, моды и медианы. Показатели коэффициента асимметрии и эксцесса при нормальном распределении равны нулю.

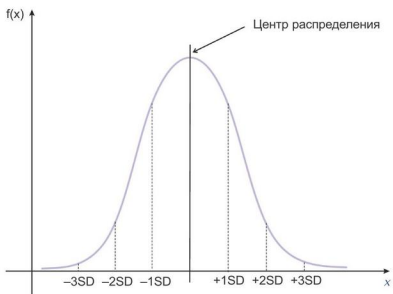


РИС. 1. Плотность нормального распределения

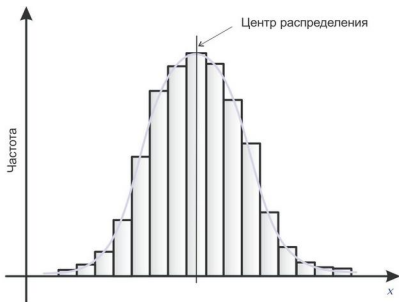


РИС. 2. Зависимость частоты признака от интервалов значений признака при нормальном распределении данных

Глава III. ПРОВЕРКА ГИПОТЕЗ МЕТОДАМИ МАТЕМАТИЧЕСКОЙ СТАТИСТИКИ

Рассматриваемые в научных исследованиях гипотезы разнообразны, исследователь формулирует их самостоятельно, основываясь лишь на полученных знаниях и приобретенном опыте, поэтому математическая проверка гипотез является, по сути, основной задачей статистического анализа. В главе I уже были даны понятия о нулевой и об альтернативной гипотезах. Принятие или отклонение нулевой гипотезы на основе статистических вычислений является доказательной базой любого научного исследования в биологии и медицине.

Чаще всего исследователи медицинских и биологических явлений проверяют гипотезы о статистической значимости различий между эмпирическими выборками. Используя различные теоретические подходы для реализации этой цели, сравнивают эмпирически полученные выборки с некоторыми теоретическими распределениями. Гипотеза отклоняется или принимается на определенном уровне значимости. Последний отражает выраженность недоверия к гипотезе об отсутствии различий между сравниваемыми выборками. При небольших значениях уровня значимости (в медицинских и биологических исследованиях меньше 0,05) следует сделать вывод о неслучайном различии и отклонить нулевую гипотезу.

3.1. *F*-критерий Фишера (сравнение двух выборочных дисперсий; *F*-test)

Для проверки гипотезы о равенстве дисперсий двух выборок используют *F*-критерий Фишера. Он способен корректно оценить дисперсии лишь в том случае, если обе выборки независимы и имеют нормальное распределение. Следует отметить, что в компьютерных программах для статистического анализа этот критерий отсутствует, поскольку на сегодняшний день он имеет лишь историческую ценность.

Для вычисления выборочных дисперсий необходимо использовать раздел *Descriptive Statistics* любой программы для статистического анализа. Также можно применять формулу вычисления дисперсии, пред-

Глава IV. АНАЛИЗ КАЧЕСТВЕННЫХ ПОКАЗАТЕЛЕЙ

4.1. Анализ относительных значений

Исследователю законов биологии и медицины довольно часто приходится работать с относительными значениями. Такие значения еще называют показателями структуры, соотношения, частоты или доли. При анализе подобных величин качественный признак или есть, или нет, поэтому такой анализ нередко называют альтернативным.

Для представления результатов исследования в относительных значениях используют проценты, промилле, продецимилле. Формулы для их расчета весьма просты:

$$\text{Проценты (\%)} = \frac{n}{N} \times 100,$$

где n — число наблюдений с изучаемым признаком; N — общее число наблюдений.

Бесспорным и огромным преимуществом показателя доли, выраженного в процентах, является его простота и великолепная наглядность, и именно поэтому вычисление процентов бывает удобно и полезно в различных ситуациях. Однако этот простой и наглядный относительный показатель имеет одно существенное ограничение. Его нельзя применять при числе наблюдений менее 20. При общем числе наблюдений от 20 до 49 проценты следует вычислять лишь в случае крайней необходимости и всегда с оговоркой, что вычисленный показатель доли неточен, поскольку число наблюдений невелико. При числе наблюдений 50–100 точность этого показателя доли также занижена, но вычисление процентов в целых числах все-таки возможно без оговорок. Наиболее эффективно данный показатель демонстрирует долю лишь при числе наблюдений более 100. В этом случае разрешено использовать один знак после запятой.

$$\text{Промилле (‰)} = \frac{n}{N} \times 1000;$$

$$\text{Продецимилле (‱)} = \frac{n}{N} \times 10\,000.$$