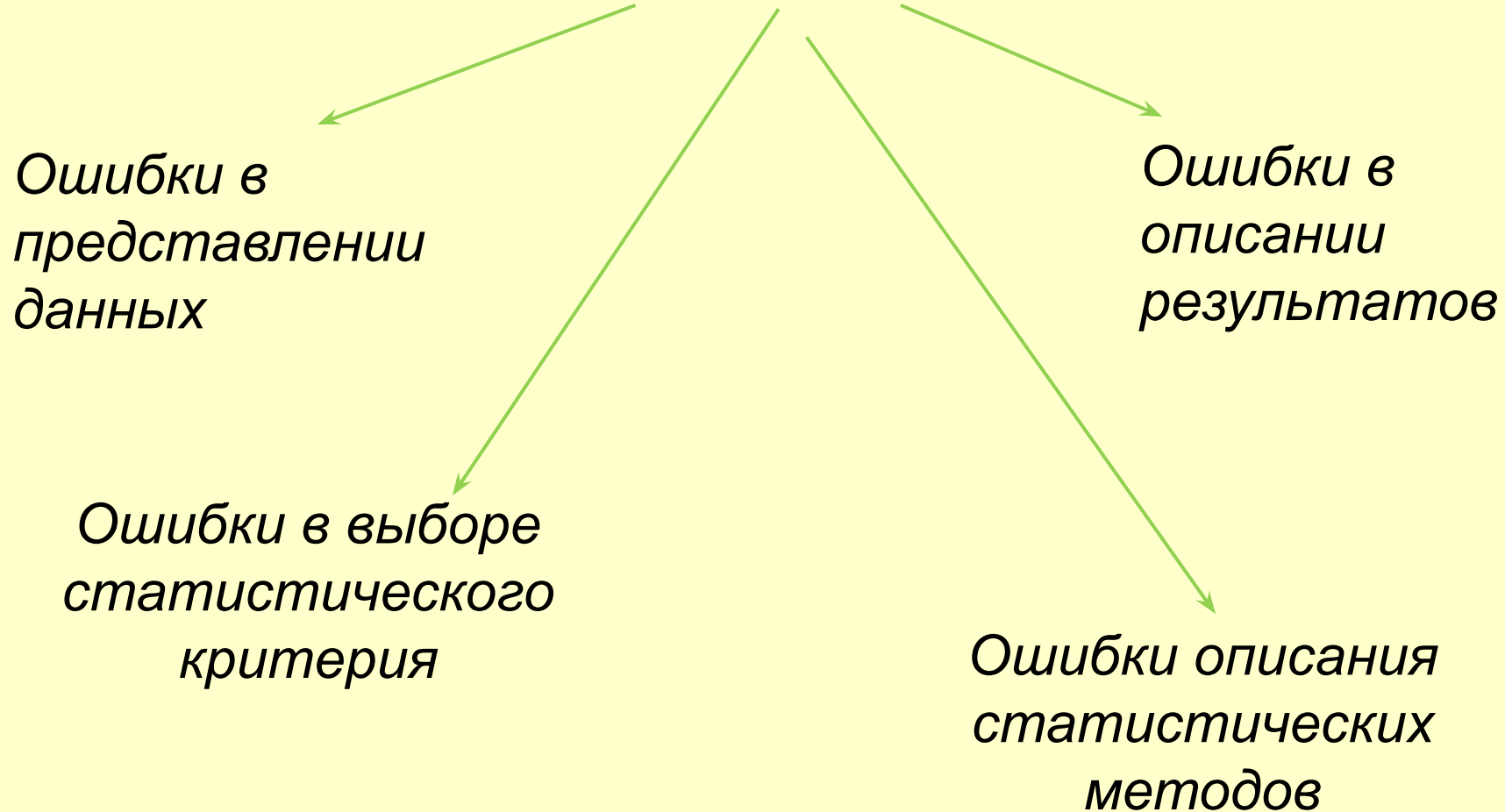


Распространенные статистические ошибки

Цель:

- ознакомиться с распространенными ошибками статистического анализа

ОШИБКИ СТАТИСТИЧЕСКОГО АНАЛИЗА ДАННЫХ



*Ошибки в
представлении
данных*

*Ошибки в выборе
статистического
критерия*

*Ошибки в
описании
результатов*

*Ошибки описания
статистических
методов*

Ошибки описания статистических методов

- Нет описания
- Шаблонные клише:
 - Результаты обработаны статистически.
 - Обработку данных проводили статистическими методами.
 - Результаты обработаны методами вариационной статистики (или: стандартными программами, общепринятыми методами и т.п.).
 - Результаты подвергнуты статистической обработке
 - Статистическая обработка материала произведена с использованием компьютера IBM PC, (Pentium II) по стандартным программам

Пример абсурдного описания

- Достоверность различия между средними арифметическими сравнениями вариационных рядов устанавливалась по степени вероятности положительной гипотезы в соответствии с описанием, приведенным в монографии Г.Ф. Лакина

Источник: http://www.biometrica.tomsk.ru/error_5.htm

Примеры из диссертаций

Источник: http://www.biometrika.tomsk.ru/error_5.htm

Описание из диссертации 1 (стр. 41)	Описание из диссертации 2 (стр. 53)	Описание из монографии {Зырянов, 1996} (стр. 48)	Описание из диссертации 3 (стр. 40)
Достоверность различия сравниваемых величин определяли с помощью показателя точности Р по таблице Стьюдента, где он располагается в зависимости от значений t и n.	Достоверность определяли с помощью абсолютного показателя точности «р», который находили по специальной таблице, где он расположен в зависимости от (t) Стьюдента и числа степеней свободы «n».	Достоверность <i>различий средних арифметических</i> определяли по абсолютному показателю точности процентных точек распределения Стьюдента в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).	Достоверность <i>различных</i> средних арифметических определяли по абсолютному показателю точности (<i>P</i>) по таблице процентных точек распределения Стьюдента в зависимости от коэффициента достоверности (t) и числа степеней свободы (n).

Ошибки представления данных

Ошибки представления данных – это ошибки, возникающие при передаче информации от источника к получателю. Они могут быть вызваны различными факторами, такими как человеческие ошибки, технические сбои или неправильная интерпретация данных.

Ошибки представления данных могут быть классифицированы по различным критериям, таким как масштаб, частота возникновения и влияние на процесс. Они могут быть системными или случайными, а также могут быть вызваны ошибками в сборе, обработке или анализе данных.

Ошибки представления данных могут привести к различным последствиям, таким как искажение информации, потеря данных или принятие неверных решений. Поэтому важно уделять внимание предотвращению и выявлению ошибок представления данных на всех этапах работы с информацией.

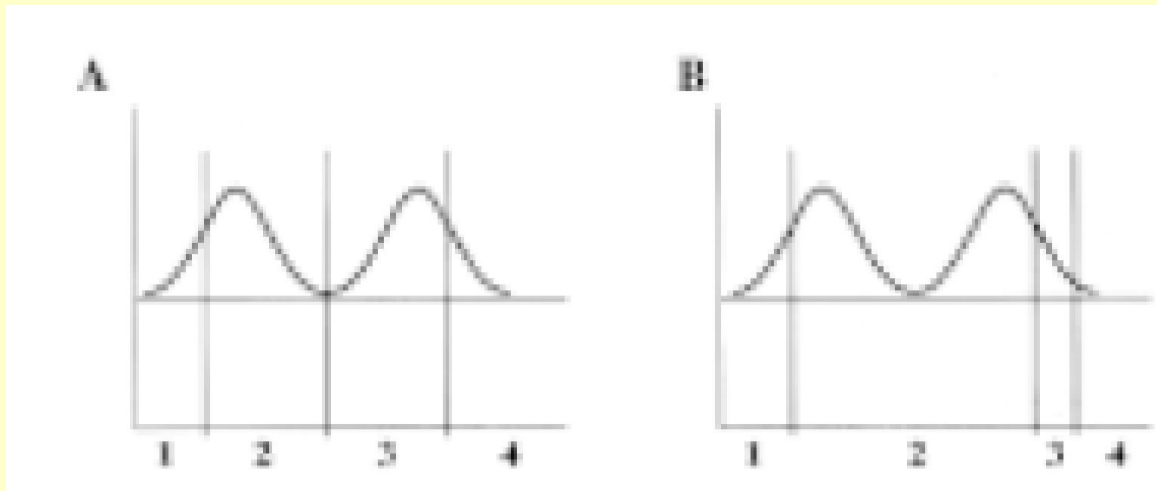
Для предотвращения ошибок представления данных необходимо использовать надежные методы сбора и обработки данных, а также проводить регулярные проверки и аудиты. Важно также обучать сотрудников, работающих с данными, правилам и стандартам, чтобы минимизировать человеческие ошибки.

В заключение, ошибки представления данных являются серьезной проблемой, которая может повлиять на качество и надежность информации. Поэтому важно уделять этому вопросу особое внимание и принимать меры для предотвращения и выявления ошибок на всех этапах работы с данными.

Спасибо за внимание! Если у вас есть вопросы или комментарии, пожалуйста, свяжитесь со мной. Мы будем рады обсудить эту тему более подробно.

ПОДМЕНА ТИПОВ ДАННЫХ

- Замена количественных данных качественными;
- Качественные данные анализируются как количественные.



для качественных данных нет понятия расстояния между значениями

ОКРУГЛЕНИЕ

Количественные данные представляются с излишней точностью

ПРАВИЛО: числовое значение результата измерений представляется так, чтобы оно оканчивалось десятичным знаком того же разряда, какой имеет погрешность этого результата.

Погрешности измерения сами определяются с некоторой погрешностью.

«Погрешность погрешности» обычно такова, что в окончательном результате погрешность приводят с одной-двумя значащими цифрами.

НЕПРАВИЛЬНОЕ ИСПОЛЬЗОВАНИЕ СТАТИСТИЧЕСКИХ ОЦЕНОК

Уровень глюкозы $8,2 \pm 7,5$ ммоль/л

Выраженность боли: $2,5 \pm 1,2$ балла
(1 – слабая, 2 – средняя, 3 – сильная)

Качественный номинальный признак – мода;

Ранговый (ординальный) признак – мода и медиана;

Количественный признак – мода, медиана, среднее.

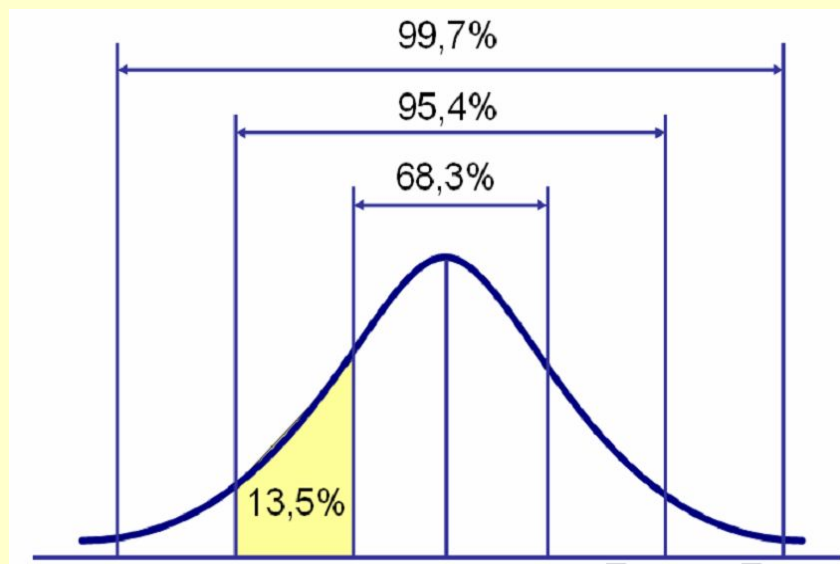
СТАНДАРТНАЯ ОШИБКА СРЕДНЕГО (SEM)

- Среднее – описывает центральную тенденцию;
- SD (s , σ) - вариабельность данных;
- SEM – показатель точности оценки среднего.

Пример: измеряем массу тела у $N=100$ мужчин, среднее $m=72$ кг, $SD=8$ кг, тогда $SEM=0,8$.

Вывод 1: примерно в 68% случаев результат измерений будет лежать в диапазоне (64; 80)кг.

Вывод 2: примерно в 68% случаев средняя масса тела составит (71,2; 72,8)кг.



Доверительные интервалы (ДИ, CI)

только единичные публикации представляют выборочные характеристики с указанием 95% ДИ, тогда как эпидемиологи заявляют, что «Определение доверительных интервалов стало обычным способом представления главных результатов клинических исследований, поскольку такой подход обладает многими преимуществами перед методом проверки гипотез (оценкой p)» (Флетчер Р. и соавт., 1998, с. 246).

Ошибки в выборе статистического критерия

АНАЛИЗИРУЕМЫЕ ДАННЫЕ НЕ СООТВЕТСТВУЮТ УСЛОВИЯМ КРИТЕРИЯ

- использование параметрических критериев для анализа данных, не подчиняющихся нормальному распределению;
- использование критериев для независимых выборок при анализе парных данных.
- использование t -критерия (критерия Манна-Уитни) для сравнения трех и более групп, а также для сравнения долей.

Ошибки в описании результатов

«Смутно пишут о том, о чем смутно
представляют»

М.В. Ломоносов



ПРИМЕР 1

«Статистическая обработка полученных результатов, оценка распределения показателей, определение границ нормального распределения проводилась с помощью пакета статистических программ Statistica (StatSoft. USA). Критический уровень значимости (p) при проверке статистических гипотез принимался за 0,05. Оценка соотношений генеральных дисперсий в большинстве сравниваемых выборок показала их неравенство. Также в большинстве выборок выявлено неправильное распределение. В связи с этим статистическая обработка данных проводилась непараметрическими методами. Учитывались средние значения (M), стандартные отклонения (SD), минимальные и максимальные значения в выборках ($Min-Max$). Для проверки статистической гипотезы разности средних значений использовался ANOVA, метод множественных сравнений, критерий наименьшей значимости».

1. Что такое «граница нормального распределения»? Зачем ее находили?
2. С помощью какого критерия проверялась гипотеза о виде распределения?
3. Что такое «неправильное распределение»?
4. Данные описаны с помощью среднего и стандартного отклонения.
5. ANOVA – параметрический критерий.

ПРИМЕР 2

7. «Статистическая обработка данных проводилась на персональном компьютере с вычислением средних значений выборок, дисперсии, коэффициента корреляции. Достоверность параметров оценивали при помощи t -критерия Стьюдента».

1. Гипотеза о виде распределения не проверялась.
2. Что такое «достоверность параметров»?
3. Гипотеза о равенстве дисперсий не проверяется.
4. Уровень значимости не указан.

Statistical analysis

Data analysis was conducted using SPSS 17.0 software. The study population was described using frequencies and percentages for categorical variables and means, standard deviations, medians, ranges and interquartile ranges (IQR) for continuous variables. Statistical comparisons were made using a chi-square test or Fisher exact test for categorical data, and a student's t-test or Mann–Whitney U test for continuous data. Dependence between continuous variables was determined with Pearson's correlation. P values less than 0.05 were considered significant.

<http://www.biomedcentral.com/1756-0500/5/363>

Statistical analysis

We calculated 95% confidence intervals. ... Multivariate analysis taking parity, single or multiple births, smoking, education, maternal age, country of birth, calendar period, and marital status into account was carried out on the material stratified on body mass index and restricted to women with a body mass index <30.

All statistical calculations were done using SAS software, version 9.3. The regressions were carried out in the proportional hazards regression (PHREG) procedure of SAS.

BMJ 2013;346:e8632 doi: 10.1136/bmj.e8632 (Published 15 January 2013)

ПРИМЕР 3

Статья "Влияние гиперлипидемии на чувствительность тимоцитов к апоптозу у мышей линии СВА и С57В1/С."

Киселева Е.П., Пузырева В.П., Огурцова Р.П., Ковалева И.Г.

Институт экспериментальной медицины РАМН, Санкт-Петербург.

Бюллетень экспериментальной биологии и медицины, вып. 8, 2000, стр. 200-202.

Цитаты из статьи	Наш комментарий
"Полученные данные обработаны статистически с использованием t критерия Стьюдента." Далее в тексте приведены выражения вида ($M \pm m$)" и результаты сравнения отдельных групп между собой. Для конкретных сравниваемых пар групп не сообщается объем выборок, однако в тексте статьи сказано, что объем выборок изменялся в интервале от 8 до 16.	В работе не сообщается о проверке условий необходимых и достаточных для использования t-критерия Стьюдента - нормальности распределения и равенства генеральных дисперсий (для всех признаков и во всех группах). Используя данные таблицы, проведем проверку гипотез о равенстве дисперсий для нескольких случайно выбранных пар. Поскольку для каждой конкретной группы сравнения в статье не указан объем выборки, то используем минимально возможное в данном случае значение, равное 8. Для пары $2,4 \pm 0,1$ и $6,0 \pm 0,3$ значение критерия Фишера $F = 9,719$ ($p=0,0048$). Для пары $2,3 \pm 0,1$ и $3,8 \pm 0,2$ значение критерия Фишера $F = 4$ ($p=0,044$). Для пары $1,6 \pm 0,1$ и $3,0 \pm 0,2$ значение критерия Фишера $F = 4$ ($p=0,044$). Для пары $17,6 \pm 0,1$ и $26,0 \pm 0,2$ значение критерия Фишера $F = 4$ ($p=0,044$). Для пары $17,2 \pm 0,1$ и $22,7 \pm 0,4$ значение критерия Фишера $F = 16$ ($p=0,0008$). Для пары $8,6 \pm 0,2$ и $13,1 \pm 0,4$ значение критерия Фишера $F = 4$ ($p=0,044$). Итак, поскольку достигнутый уровень значимости гораздо меньше 5%, то гипотеза о равенстве дисперсий для этих случаев отвергается! Вывод: если даже предположить, что во всех сравниваемых группах наблюдалось нормальное распределение, что само по себе весьма маловероятно, тем не менее, критерий Стьюдента не может быть использован в данных условиях вследствие неравенства генеральных дисперсий (см. проблему Беренса-Фишера). . Из чего следует, что выводы авторов не могут быть признаны корректно обоснованными методами статистики, а стало быть надежность их весьма сомнительна.

АЛГОРИТМ ВЫБОРА СТАТИСТИЧЕСКОГО КРИТЕРИЯ

ПРИЗНАК	ИССЛЕДОВАНИЕ				
	Две группы	Более двух групп	Группа до и после лечения	Одна группа несколько видов лечения	Связь признаков
Количественный (нормальное распределение*)	Критерий Стьюдента	ANOVA	Парный критерий Стьюдента	Дисперсионный анализ повторных измерений	Линейная регрессия, корреляция, или метод Блэнда-Алтмана
Качественный	Критерий χ^2 Z-критерий	Критерий χ^2	Критерий Мак-Нимара	Критерий Кокрена	Коэффициент сопряженности
Порядковый	Критерий Манна Уитни	Критерий Крускала Уоллиса	Критерий Уилкоксона	Критерий Фридмана	Коэффициент ранговой корреляции Спирмена

Рекомендации «Nature medicine» (Guide to authors, 2008, p. 5).

Надо указывать наименование стат. теста, количество (n) в каждом стат. анализе, обоснование выбора определенного теста (включая обсуждение нормальности распределения, если тест предназначен только для нормально распределенных данных), α -уровень для всех тестов, были тесты односторонними или двусторонними и фактические значения «p» для каждого теста (не просто « $p < 0,05$ »). Должно быть ясно указано, какой стат. тест использован для получения данного p-значения.

Рекомендации «Nature medicine» (Guide to authors, 2008, p. 5).

Выборочные данные должны быть представлены в виде описательных статистик: размеры выборки (n) для каждого набора данных; меры рассеяния, такие как стандартное отклонение или размах. Для малых выборок размах является более приемлемым, чем стандартное отклонение. На графиках должны быть ясно отображены границы ошибок (error bar — «усы») или доверительных интервалов. Авторы обязаны указывать, являются ли числа, следующие за знаком $\pm$, стандартными ошибками среднего (s.e.m.) или стандартным отклонением (s.d.).

Рекомендации «Nature medicine» (Guide to authors, 2008, p. 5).

Авторы обязаны обосновать применение конкретного критерия и объяснить, согласуются ли анализируемые данные с допущениями для теста.

Многие стат. критерии требуют, чтобы данные были распределены приблизительно нормально; используя эти тесты, авторы должны указать, как они проверили свои данные на нормальность. Если данные не соответствуют условиям применения теста, то должна использоваться непараметрическая альтернатива.

Рекомендации по подготовке графиков

Графики должны быть на белом фоне, следует избегать рамок, ненужных цветов, декоративных эффектов (таких как трехмерные графики), рисунков с высоким разрешением. Вертикальная ось гистограммы не должна быть усеченной, чтобы преувеличивать незначительные различия.

Рекомендации «Nature medicine» (Guide to authors, 2008, p. 5).

[http://www.nature.com/nm/authors/submit/Checklist_of_statistical_a
dequacy.doc](http://www.nature.com/nm/authors/submit/Checklist_of_statistical_a
dequacy.doc)».

Спасибо за внимание!

