



Двадцать ошибок статистического анализа в биомедицинских публикациях

Критические обзоры убедительно свидетельствуют, что примерно **в половине научных работ**, выполняемых с использованием статистических методов анализа результатов, этот анализ проводится с ошибками.

Исследования высокого методологического качества **заслуживают** соответствующего изложения материала.

Хорошее представление результатов является **важной частью** исследования, такой же, как сбор и анализ данных.

Первая попытка обсудить вопросы, связанные со статистической вероятностью, в медицинской литературе была предпринята **в 30-х годах XX века**.

С тех пор исследователи в разных областях медицины обнаружили **множество ошибок** при проведении статистического анализа даже в работах, результаты которых были опубликованы в наиболее авторитетных научных журналах.

Неверное отображение статистических данных представляет собой **давнюю и широко распространенную проблему**, чреватую серьезными последствиями.

Проблема недостаточно хорошо осознана, несмотря на то что большинство ошибок возникает при использовании простейших статистических методов.



С распространением подходов доказательной медицины возросло понимание актуальности проблем, связанных с низким качеством изложения статистического материала.

В основе доказательной медицины лежит использование опубликованных в медицинской литературе исследований, поэтому она зависит от методологического качества статей.

Несколько авторских коллективов предложили собственные руководства по представлению результатов различных исследований и изложению результатов статистического анализа.



Ошибка 1. Количественные данные представлены с излишней точностью

Большинство из нас легче воспринимают количественные данные, представленные одной или двумя цифрами, чем тремя и более - округление улучшает восприятие материала. Во многих случаях необязательно приводить максимально точные значения.

Если масса тела больного составляет 60 кг, то использование показателя 60,18 кг только затруднит восприятие, несмотря на то что формально он соответствует действительности. >

Пример: Количество участников (женщин) на момент окончания исследования примерно в 3 раза превышает таковое в начале.

Этот факт становится очевидным лишь после округления соответствующих показателей:

Число женщин возросло с 29 942 до 94 347

Число женщин возросло с 29 900 до 94 300

Число женщин возросло примерно с 30 000 до 94 000.



Ошибка 2. Непрерывные данные представлены в виде порядковых без объяснения причин и способа преобразования

Для облегчения статистического анализа непрерывные данные можно представить в виде двух и более порядковых категорий, например, рост в см как низкий, нормальный и высокий.

Такое упрощение уменьшает точность результатов и вариабельность данных.

Автор должен объяснить, почему он сделал это. Следует описать критерии выбора диапазона значений в рамках каждой из порядковых категорий, чтобы избежать возможности появления систематической ошибки.

В некоторых случаях преобразование непрерывных данных в порядковые **имеет целью подгонку** конечных результатов под желаемую схему

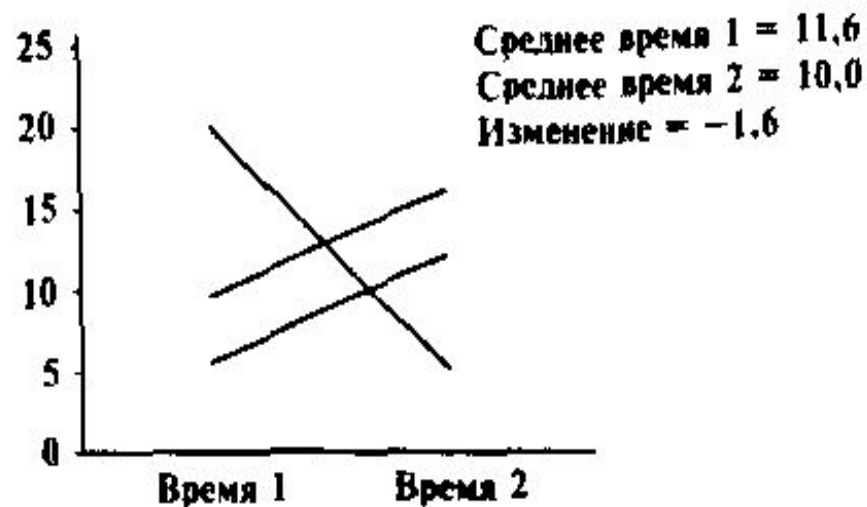
Ошибка 3. Представлены средние групповые значения для парных данных без сообщения размера изменений внутри групп

Данные, относящиеся к одному и тому же участнику исследования, называются парными. При последовательных измерениях признака величина как средних групповых, так и индивидуальных значений может различаться от измерения к измерению.

Если в статье представлены только групповые средние значения, читатель может не заметить изменения индивидуальных показателей.

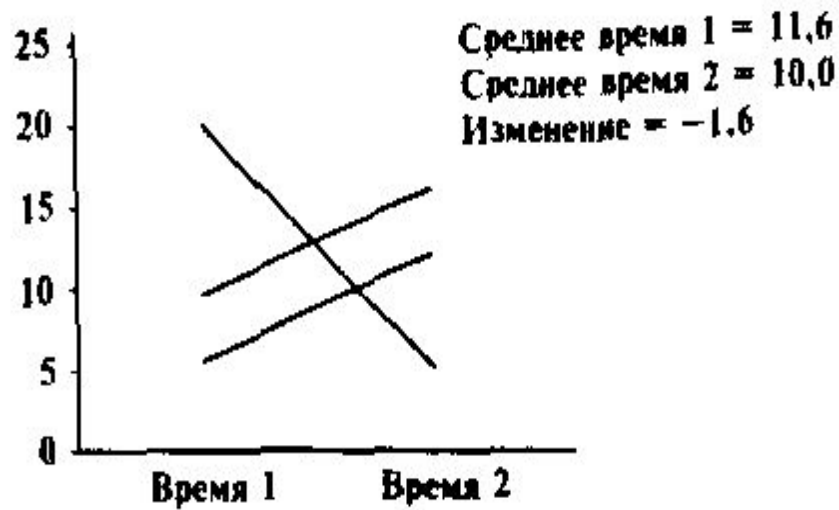
Пока не будут отображены индивидуальные значения, альтернативный вариант трактовки данных будет неочевиден.





Результаты, приведенные на рисунке, можно интерпретировать как среднее уменьшение показателя в группе от момента 1 к моменту 2, либо как увеличение показателя у 2 из 3 участников.

И то и другое соответствует истине, но если в статье представлен лишь один из этих выводов, читатель может сделать неверное заключение о результатах.



Парные данные должны быть представлены таким образом, чтобы были очевидными изменения как индивидуальных, так и групповых характеристик.

В данном примере результат можно интерпретировать как среднее уменьшение показателя на 1,6 либо как его увеличение у 2 из 3 участников



Ошибка 4. Неправильно используются статистические характеристики данных

При описании непрерывных данных наиболее часто используют понятия - средняя величина (M) и среднее квадратическое отклонение (СКО – сигма). Эти показатели применимы только при условии **нормального или Гауссова распределения** значений.

При нормальном распределении в 68% случаев результаты измерений лежат в пределах ± 1 СКО от среднего значения, в 95% случаев – в пределах ± 2 СКО, в 99% случаев – в пределах ± 3 СКО.

При **асимметричном** распределении эта закономерность отсутствует. В этом случае используют такие показатели, как медиана (50-й центиль, или точка, которая делит данные на две равные части) и межквартильный диапазон (обычно от 25-го до 75-го центиля).

Хотя для определения средней величины и СКО достаточно результатов двух последовательных измерений, эти показатели недостаточно хорошо описывают данные исследований с **небольшим числом участников (малые выборки)**.

Большинство биологических показателей не подчиняются нормальному распределению.

Исходя из этого, в медицинской литературе термины - медиана, диапазон и межквартильный диапазон должны встречаться чаще, чем средняя величина и СКО.



Ошибка 5. Стандартная ошибка средней величины используется для описательного анализа данных или в качестве показателя точности оценки

Средняя величина и СКО (сигма) описывают центральную тенденцию и вариабельность данных, подчиняющихся нормальному распределению.

Средняя величина и стандартная ошибка средней величины (СОС - m) являются точечной оценкой для характеристики популяции.

СОС всегда меньше, чем СКО, поэтому иногда представляют именно ее, чтобы результаты измерений выглядели более точными.


В пределах «средняя величина ± 1 СОС» лежит популяционная средняя с вероятностью 68% - доверительный интервал, (ДИ) 68%.

В медицинских исследованиях предпочтительно




Среднюю величину и СОС применяют для характеристики как выборки, так и популяции. Чтобы избежать путаницы, следует запомнить: **среднюю величину и СКО** предпочтительно использовать для **обобщенной характеристики данных**, подчиняющихся **нормальному распределению**, а **среднюю величину и 95% ДИ** – в качестве **точечной оценки и уровня ее точности**.

Например, если средняя масса тела у 100 мужчин составляет 72 кг, а СКО – 8 кг, то (при условии нормального распределения значений) примерно в $\frac{2}{3}$ (68%) случаев результат измерения будет лежать в диапазоне **от 64 до 80 кг**. Данный пример показывает правильное использование средней величины и СКО для характеристики распределения значений.



Для точечной оценки и определения уровня ее точности предпочтительно использовать среднюю величину и 95% ДИ (диапазон значений в пределах ± 2 СОС).

В рассмотренном выше примере правильной будет фраза: средняя масса тела составляет 72 кг при 95% ДИ от 70,4 до 73,6 кг.

Это означает, что при повторном измерении данного показателя в аналогичной (случайной) выборке мужчин в той же популяции примерно в 95% случаев средняя масса тела составит от **70,4** до **73,6** кг. 

Ошибка 6. Для описания различий между группами используется только величина p

Использование величины p (*уровень значимости*) для оценки статистической значимости часто **неоправдано**. Даже при условии корректного применения данный показатель имеет целый ряд ограничений.

В большинстве случаев вместо величины p , либо дополнительно к ней, следует указывать **абсолютное различие** в частоте изучавшегося события между группами (относительное или выраженное в процентах различие может быть истолковано неверно) **и его 95% ДИ**.



Ниже приводятся встречающиеся в статьях формулировки в порядке возрастания их методологического качества.

«Эффект от применения лекарственного средства оказался статистически значимым». Данная формулировка не позволяет определить ни величину эффекта, ни его клиническую, ни статистическую значимость. Читатель может заключить, что характеристика эффекта как «статистически значимого» в данной ситуации означает целесообразность использования препарата.



«Эффект от использования средства, заключающийся в снижении уровня диастолического артериального давления (АД), оказался статистически значимым ($p < 0,05$)». Отсутствует указание на величину эффекта, поэтому его клиническая значимость остается неясной.

Величина p может составлять 0,049, что близко к пороговой величине (0,05), и практически не отличается от 0,051, т. е. статистически незначимого уровня.

Наличие подобной условной черты (0,05) представляет собой одну из проблем при использовании величины p .



«Среднее диастолическое АД в основной группе уменьшилось со 110 до 92 мм рт. ст. ($p=0,02$)».

Такая формулировка встречается наиболее часто.

В ней отражены результаты измерений до и после вмешательства, однако не указано различие между ними.

Среднее уменьшение показателя на 18 мм рт. ст. статистически значимо, однако это лишь точечная оценка.

В отсутствие 95% ДИ нельзя определить, насколько она точна, и, следовательно, практически значима.



«Использование препарата привело к снижению уровня диастолического АД в среднем на 18 мм рт. ст. (со 110 до 82 мм рт. ст.) при 95% ДИ от 2 до 34 мм рт. ст. ($p=0,02$)».

Границы ДИ свидетельствуют, что при использовании данного препарата в 100 выборках, аналогичных изучавшейся, среднее снижение АД в 95 из них будет лежать в пределах от 2 до 34 мм рт. ст. Уменьшение на 2 мм рт. ст. клинически незначимо в отличие от снижения на 34 мм рт. ст.

Таким образом, хотя среднее уменьшение АД оказалось статистически значимым, эффект от использования препарата в других испытаниях может оказаться клинически незначимым, т. е. полученные в исследовании результаты не позволяют сделать окончательного вывода о целесообразности вмешательства.



Если оба показателя, определяющих как верхний, так и нижний пределы ДИ, клинически значимы, можно полагать, что вмешательство клинически эффективно.

Когда оба показателя клинически незначимы, вероятнее всего, вмешательство неэффективно.

Может оказаться, что клинически значим только один из показателей; в таком случае следует провести исследование с большим числом участников.



Ошибка 7. Отсутствует подтверждение того, что анализируемые данные соответствуют предположениям, лежащим в основе использованных статистических методов

При несоблюдении критериев использования того или иного метода полученный результат может оказаться неточным.

В тексте статьи должно присутствовать название использованного метода и подтверждение того, что он применим для анализа имеющихся данных.

Например: полученные результаты подчиняются нормальному распределению, что позволяет использовать t-тест.



Наиболее характерные ошибки:

- **использование параметрических методов** (основанных на предположении о нормальном распределении данных) **для анализа данных, не подчиняющихся нормальному распределению** - при сравнении двух групп нередко используют критерий Стьюдента, хотя более оправдано применение критерия Вилкоксона или другого непараметрического метода;
- **использование методов, предназначенных для независимых выборок, при анализе парных данных** - применяют критерий Стьюдента, а не парный t-тест.



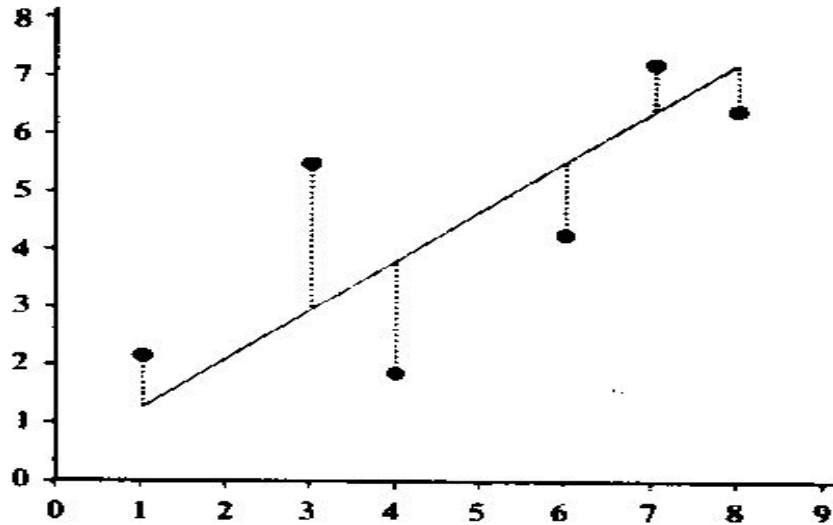
Ошибка 8. Использование линейной регрессии без подтверждения линейного характера связи

Подтвердить линейный характер зависимости можно с помощью изучения остатков, т. е. различий между наблюдаемыми и прогнозируемыми при помощи модели величинами.

Если при отображении в виде графика остатки представляют собой прямую линию, а их значения приближаются к нулю, то можно говорить о линейном характере зависимости.

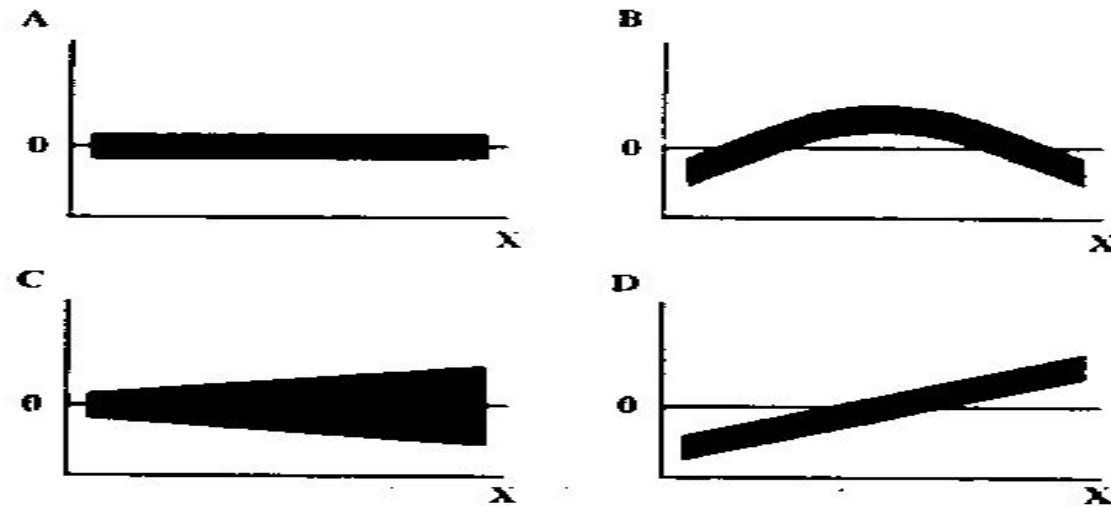
Изучение остатков необходимо, поскольку сам по себе график линейной регрессии не всегда позволяет верно оценить характер зависимости.





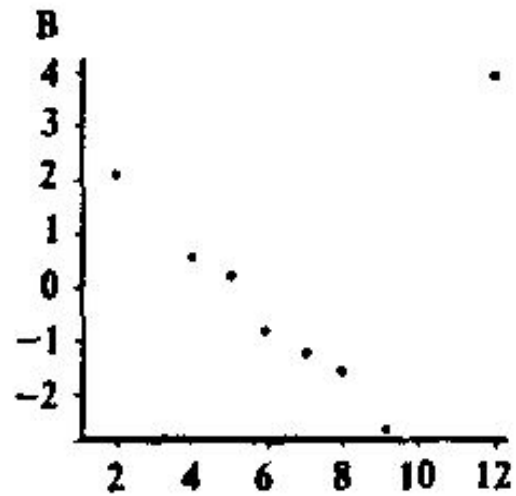
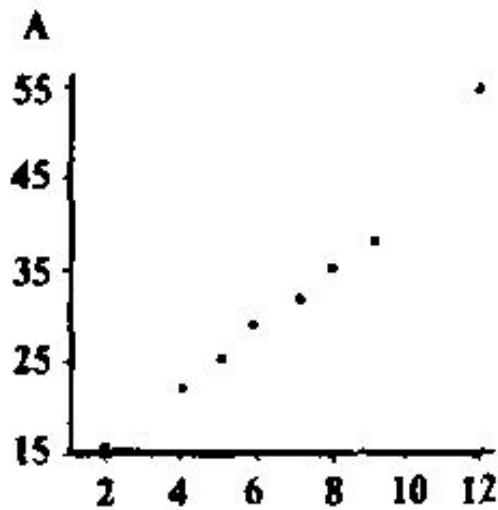
Остатки представляют собой различие между наблюдаемыми и прогнозируемыми при помощи регрессионной модели величинами.





Когда график остатков свидетельствует, что их величины приближаются к нулю на протяжении всего диапазона значений, зависимость имеет линейный характер (A).

Иной вид графического изображения остатков (B, C, D) свидетельствует о нелинейном характере зависимости, для описания которой линейная регрессия непригодна. ➤



Впечатление о линейном характере зависимости может оказаться обманчивым.

В данном примере зависимость выглядит линейной (А), однако в действительности это не так, о чем свидетельствует график остатков (В).




Ошибка 9. В анализ включены не все данные и не все участники

Пропуски в данных встречаются довольно часто и крайне отрицательно сказываются на общем впечатлении от статьи, поскольку у читателя может возникнуть предположение, что **автор недостаточно внимателен или попросту ленив.**



При обнаружении пропущенных данных возникают следующие вопросы:

- **Причина пропуска.** Включены ли в анализ минимальные и максимальные результаты? Пропущены ли данные из-за ошибки в лаборатории? Возможно, данные пропущены, поскольку они противоречат выводам автора?
- **Воспроизводимость** полученных результатов. Является ли указанный диапазон значений таковым в действительности? Так ли невелико число выбывших из исследования?
- **Методологическое качество исследования в целом.** Если итоговые данные не совпадают в исследовании, насколько автор был точен при описании других аспектов работы? 

Ошибка 10. Не указано, использовалась ли поправка на многократность проверки

гипотез

В большинстве статей приводится несколько величин p , что повышает вероятность возникновения ошибки 1 рода (альфа-ошибки), т. е. ошибочного заключения об эффективности вмешательства, когда в действительности полученный результат случаен.

Предположим, что исследование включает шесть групп. Сравнение групп между собою требует проведения 15 парных статистических тестов, результатом которых будет определение 15 величин p .

В отсутствие поправки вероятность возникновения ошибки I рода возрастает с 5 на 100 (обычный уровень вероятности альфа-ошибки составляет 0,05) до 55 на 100 (т. е. 0.55).

Проведение серии сравнений во многих случаях можно признать целесообразным, но подобный поисковый анализ должен быть соответствующим образом обоснован и описан.

Однако «перетряхивание» данных путем вычисления множества величин p с целью обнаружить какие-нибудь статистически значимые различия **служит признаком низкого методологического качества исследования.**



Ошибка 11. Ненужное сравнение исходных характеристик в рандомизированных клинических испытаниях

В рандомизированных клинических испытаниях (РКИ) каждый участник имеет равную вероятность оказаться как в группе вмешательства, так и в контрольной, поэтому любое различие в исходных характеристиках групп случайно.

Следовательно, наличие статистически значимых различий в исходных показателях не свидетельствует о систематической ошибке.

Приняв во внимание, что вероятность альфа-ошибки составляет 0,05, в 5 сравнениях из 100 различие в исходных характеристиках окажется статистически значимым просто в силу случая.



Ошибка 12. Не указаны критерии нормы и отклонения от нормы при оценке эффективности диагностических методов

Значимость результата при использовании любого диагностического метода зависит от того, какие критерии были выбраны для определения нормы и отклонения от неё.

В медицине существует шесть определений того, что представляет собой норма.

Диагностическая норма: диапазон значений, в пределах которого показатель свидетельствует об отсутствии заболевания. Такое определение представляется целесообразным, поскольку имеет клинический смысл.

Терапевтическая норма: диапазон значений, вне пределов которого показатель свидетельствует о целесообразности терапии. И это определение



Другие определения - менее информативны, однако, к сожалению, нередко используются авторами:

Эпидемиологическое определение нормы: диапазон значений, в пределах которого показатель свидетельствует об отсутствии риска развития заболевания.

Например, высокий уровень холестерина в сыворотке крови сам по себе не представляет интереса; однако факт, что при этом повышается риск развития заболеваний сердца, заставляет считать высокий уровень холестерина отклонением от нормы.

Статистическое определение нормы: нормальным считается показатель, определенный у здоровых лиц. Данное определение подразумевает, что полученные результаты подчиняются нормальному распределению, т.е. при графическом изображении кривая имеет вид колокола. Следует учитывать, что многие результаты не подчиняются нормальному распределению.

Перцентильное определение норм: нормальным считается показатель, лежащий в пределах диапазона. Например, любой показатель в пределах нижних 95% всех значений определяется как норма, а в пределах оставшихся верхних 5% – как отклонение от нормы. И в данном случае критерием служит частота показателя вне зависимости от клинической значимости.

Социальное определение нормы: нормальным следует называть показатель, который принято считать таковым. Например, желаемая масса тела или возраст, к которому ребенок должен научиться самостоятельно ходить.

Подобные критерии не всегда клинически значимы.



**Статистические сравнения исходного состояния групп в РКИ.
Различие в концентрации альбумина в крови случайно ($p=0,03$);
оно не указывает на систематическую ошибку.**

В данном случае использование величины p необязательно

Признак	Группа контроля (n=43)	Группа вмешательства (n=51)	Различие	p
Средний возраст, годы	85	84	1	0,88
Мужчины (n, %)	21 (49)	21 (51)	3%	0,99
Медиана концентрации альбумина в крови (г/л)	30,0	33,0	3,0 г/л	0,03
Сахарный диабет (n, %)	11 (26)	8(20)	6%	0,83

Ошибка 13. Отсутствует объяснение, каким образом неопределенные (сомнительные) результаты учтены при вычислении операционных характеристик диагностического теста (таких, как чувствительность и специфичность)

Далеко не всегда использование диагностического метода позволяет получить однозначно положительный или отрицательный результат. Врач может не согласиться с интерпретацией клинических признаков.

Результаты, которые нельзя признать ни положительными, ни отрицательными, влияют на практическую значимость метода, поэтому их наличие и относительная частота должны быть приведены в статье.

Существует **три варианта** неопределенных результатов:

Промежуточные результаты занимают промежуточное положение между отрицательными и положительными.

Неопределенные результаты такие, которые не позволяют сделать ни положительного, ни отрицательного заключения. Например, ответы, полученные при психологическом тестировании, из которых неясно, страдает ли обследуемый алкогольной зависимостью.

Не поддающиеся интерпретации результаты получены при использовании метода с несоблюдением существующих стандартов проведения исследования. Например, определение уровня глюкозы в крови после



В тексте статьи должно иметься объяснение того, каким образом подобные результаты были учтены при определении чувствительности и специфичности метода.

Операционные характеристики будут зависеть от того, как неопределенные результаты учитывались: как положительные, отрицательные, либо не включались в анализ.

В стандартной таблице сопряженности 2×2 , столбцы и строки для сомнительных результатов отсутствуют.

Даже при условии высокой чувствительности, но при наличии значительного процента сомнительных результатов, практическая значимость метода будет невелика.



Стандартная таблица для определения операционных характеристик диагностического теста*

Таблица не включает в себя сомнительные результаты, которые нередко и не всегда оправданно игнорируют.

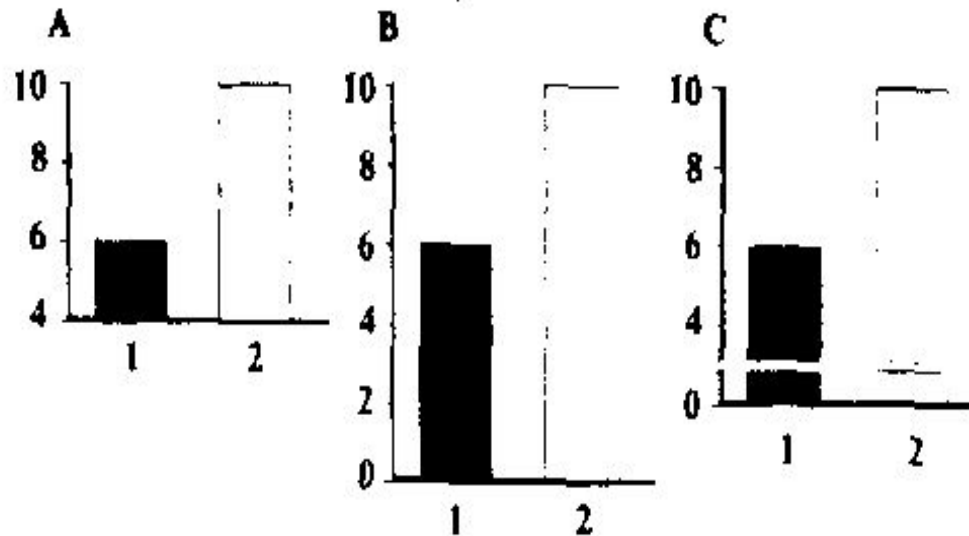
Результат	Заболевание		Всего
	имеется	отсутствует	
Положительный	a	b	a+b
Отрицательный	c	d	c+d
Всего	a+c	b+d	a+b+c+d



Ошибка 15. Несоответствие между внешним видом графика или диаграммы и данными, на которых они основаны

Очень важно, чтобы внешний вид графиков не искажал смысл данных, на которых они основаны. Одна из проблем возникает при необходимости построения графиков, начальной точкой которых служат ненулевые значения.





А. Гистограммы и графики с отсутствующими нулевыми значениями могут способствовать неверному восприятию материала.

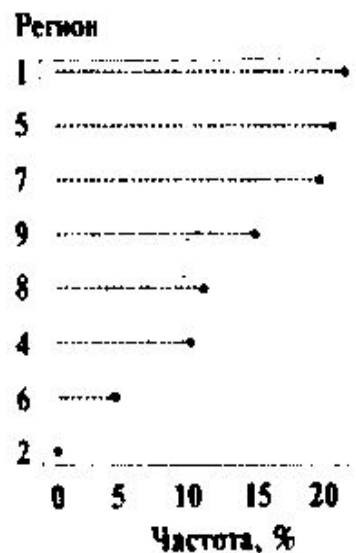
В. Гистограмма построена правильно: высота обоих столбцов соответствует действительности.

С. При отсутствии возможности построения гистограммы, включающей в себя нулевые значения, ось должна быть разорвана.

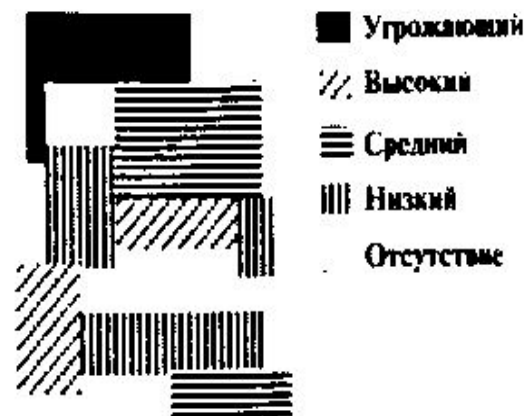


А. Число случаев, на регион В. Число случаев, на регион

Регион	Частота, %
1	23
2	0
3	17
4	11
5	22
6	5
7	21
8	12
9	16



С. Число случаев, на регион




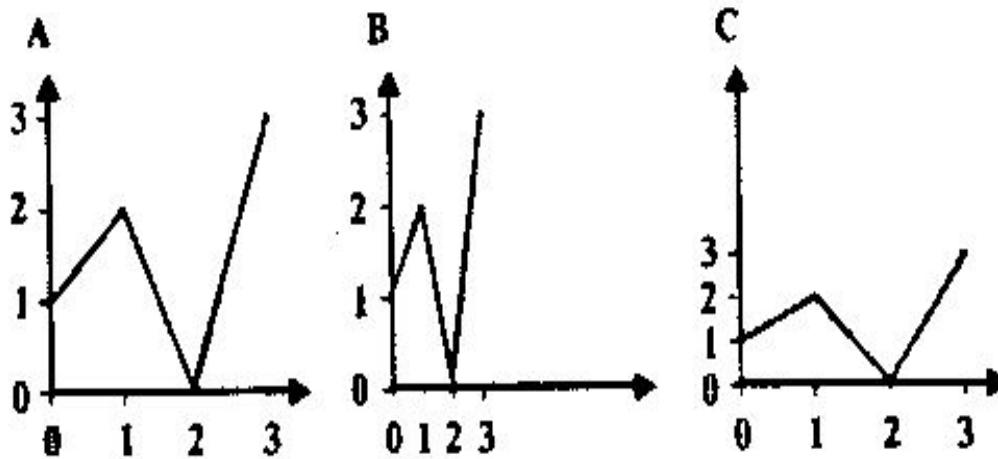
- А. Таблицы позволяют наиболее информативно отобразить точные количественные данные.
- В. Точечные (ленточные) диаграммы позволяют наиболее информативно отобразить общие закономерности и провести сравнение.
- С. Карты наиболее информативно отражают пространственные взаимоотношения.



Ошибка 14. Рисунки и таблицы используются лишь для «хранения» данных, а не с целью облегчить восприятие материала

В научных статьях, помимо собственно «хранения» информации, таблицы и рисунки должны служить для облегчения восприятия материала - таблицы и рисунки в статьях могут отличаться от тех, которые были созданы автором исключительно для регистрации данных и проведения анализа.

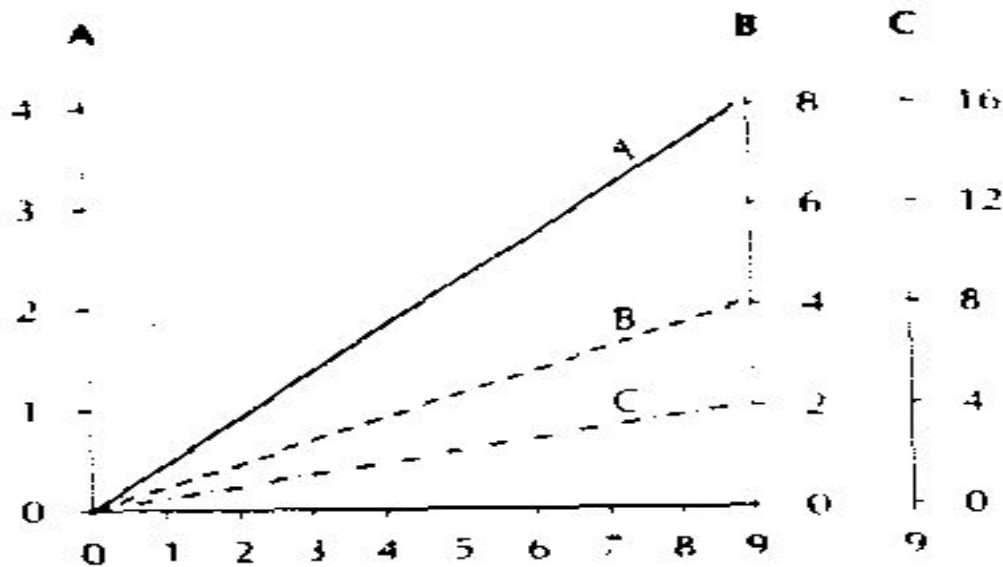
Легче всего сравнивать данные, расположенные рядом друг с другом, поэтому оптимальным следует считать именно такую структуру таблицы, подсказывающую читателю то или иное сравнение. 



Другая проблема заключается в «эластичности» графиков. Одна из осей может быть непропорционально сжата или растянута, что приводит к ошибочному восприятию данных. Неверно выбранный при построении графика масштаб может способствовать нарушению восприятия материала.

Уменьшение масштаба по оси X (в данном случае отражающей время, (B) приводит к тому, что изменения признака Y выглядят внезапными. Уменьшение масштаба по ординате приводит к впечатлению о постепенных изменениях Y.





Аналогичные трудности возникают при использовании двойных осей. Если шкала справа не связана математическим отношением с левой, изменение масштаба на одной из осей может приводить к изменению впечатления о связи признаков

При использовании графиков с несколькими осями, каждая из которых служит для отображения своего показателя, связь между последними может искажаться.

В данном примере подъем линии В представляется в 2 раза меньшим, чем линии А, а подъем линии С – в 4 раза меньшим.

Ошибка 16. Нечеткое определение понятия «объект исследования»

Термином «объект исследования» обозначают изучаемый предмет, событие или явление. Трудности возникают, если таким предметом служит не сам больной, а нечто иное.

Например, если в ходе исследования изучены исходы в отношении 50 глаз, то сколько больных в нем участвуют? И что означает 50%-я эффективность лечения?

Если исходом диагностического исследования является заключение специалиста, может быть необходимым исследование выборки специалистов, а не просто выборки результатов исследования.

В этом случае размером выборки является число специалистов, а не число полученных оценок.



Ошибка 17. Интерпретация статистически незначимых или полученных в исследованиях с малой статистической мощностью результатов как отрицательных, а не недостаточных

Статистическая мощность представляет собой вероятность выявления статистически значимого различия при условии, что оно действительно существует.


Статистически незначимые результаты, полученные в исследовании с малой статистической мощностью, **неверно считать отрицательными**; они недостаточны: «отсутствие гарантии не есть гарантия отсутствия».

К сожалению, многие исследования, в которых получены статистически незначимые результаты, характеризуются малой статистической мощностью. >

Ошибка 18. Непонимание различий между объяснительными (идеальная эффективность вмешательства) и прикладными (реальная эффективность вмешательства) исследованиями при планировании и интерпретации исследований

Задачей **фундаментальных исследований** служит объяснение патогенеза того или иного заболевания либо механизма действия лечебного вмешательства.


Они проводятся в «идеальных» или «лабораторных» условиях, позволяющих осуществлять тщательный контроль за отбором участников, процессом лечения и наблюдения.

Результаты таких исследований не всегда применимы в клинической практике, не поддающейся столь тщательному контролю. 

Задачей прагматических исследований (оценивающих реальную эффективность вмешательства) является помощь в принятии решений в клинике. Они проводятся в обычных условиях, в которых осуществляется работа врачей.

На конечный результат подобных исследований может влиять множество факторов, не поддающихся контролю, поэтому научная значимость полученных данных ограничена, однако практическая ценность велика.

Во многих случаях авторы пытаются объединить оба подхода, но, в конечном счете, ни один из них не реализуется в полном объеме.

Результаты исследования следует интерпретировать исходя **из природы вопроса**, для ответа на который оно предназначено. 

Различия между фундаментальным и прикладным исследованиями эффективности таблетированного препарата цинка при лечении простуды

Характеристика исследования	Фундаментальное	Прикладное
Критерии диагностики	Получение культуры Rhinovirus	Наличие 3 из 10 симптомов
Критерии оценки эффективности (исходы)	Количество отделяемого из носа (подсчет салфеток)	Уменьшение числа и длительности сохранения симптомов
Условия проведения	Стационар	Амбулаторно
Прием препарата	Под контролем исследователя	Под контролем самих больных
Структура	Маскированное, плацебо-контролируемое	Маскированное, плацебо-контролируемое
Задача	Действенность цинка как противовирусного препарата	Эффективность использования цинка при лечении простуды

Ошибка 19. Представление результатов не в клинически важных единицах

С клинической точки зрения наиболее информативно представление данных в виде связи между прилагаемыми усилиями и получаемым результатом, например, в виде числа пациентов, нуждающихся в лечении для получения одного положительного результата.

Такой способ представления данных позволяет сравнивать различные вмешательства с использованием единых критериев.



Во всех приведенных ниже примерах полученные результаты представлены клинически четко и грамотно, однако каждая из формулировок позволяет составить различное мнение об эффективности вмешательства.


Результаты, представленные в абсолютных показателях.

В Хельсинкском исследовании (мужчины, продолжительность наблюдения 5 лет) в группе плацебо инфаркт миокарда был отмечен у 84 (4,1%) из 2030 участников
в группе получавших гемфиброзил - 56 (2,7%) из 2051 (p<0,02);

Снижение абсолютного риска составило 1,4% (4,1%-2,7%=1,4%).




Результаты, представленные в относительных показателях. В Хельсинкском исследовании частота возникновения инфаркта миокарда в группах плацебо и гемфиброзила составила 4,1 и 2,7% соответственно. Абсолютное уменьшение риска на 1,4% соответствует снижению относительного риска развития инфаркта миокарда в группе вмешательства на 34% ($1,4\%/4,1\%=34\%$).

Результаты, представленные в виде связи между усилиями и результатом. В Хельсинкском исследовании, включавшем 4081 мужчину с гиперхолестеринемией, показано, что для предотвращения 1 случая инфаркта миокарда необходимо проводить лечение 71 участнику в течение 5 лет. 

Результаты, представленные в виде связи между усилиями и результатом (другой вариант).

В Хельсинкском исследовании, включавшем 4081 мужчину с гиперхолестеринемией, показано, что для предотвращения 1 случая инфаркта миокарда в течение 5 лет необходимо назначить около 200 000 доз гемфиброзила.

Результаты, представленные в виде отношений общей смертности.

В Хельсинкском исследовании от инфаркта миокарда в группах гемфиброзила и контроля умерли 6 и 10 участников соответственно. Снижение абсолютного риска составило 0,2%, снижение относительного риска – 40%. Для предотвращения 1 случая смерти от инфаркта миокарда в течение 1 года необходимо назначить лечение 2460 мужчинам. 

Ошибка 20. Смещение понятий статистической и клинической значимости

Даже несущественное различие, выявленное при сравнении больших групп, может оказаться статистически значимым, но не иметь при этом клинического значения.

Даже существенное различие, выявленное при сравнении небольших групп, может иметь клиническое значение, но не быть при этом статистически значимым.

Если в ходе исследования, включающего несколько больных в терминальном состоянии, хотя бы один из участников в какой-либо из групп выживет, такой результат будет клинически значимым, хотя статистически значимое различие в выживаемости между группами может отсутствовать.

Заключение

- ✓ **Главное решение** проблемы ошибок статистического анализа данных состоит **в изучении** методологии исследований и статистического анализа.
- ✓ Статистикам следует проявлять большую активность в вопросе обучения авторов, редакторов и читателей.
- ✓ Авторы должны привлекать статистиков на этапе **планирования исследования**, а не после его завершения.
- ✓ Редакторы должны систематически применять рекомендации по представлению статистических данных.
- ✓ Важно, чтобы в журналах **тщательно проверяли** статьи, в которых содержится статистический анализ.
- ✓ Читатели **должны обучаться** интерпретации статистических данных и требовать от авторов грамотного их представления

