

# Принципы описания статистики в биомедицинских публикациях

Леонов Василий Петрович,  
редактор сайта

БИОМЕТРИКА

E-MAIL: [leo.biostat@gmail.com](mailto:leo.biostat@gmail.com)

(Лекция в Якутске, 13.11.2009)

# Почему авторы используют статистику?

Обобщающие полёты ума в сфере патологических явлений совершенно необходимы, так как **запас детальных фактов в настоящее время очень велик** и получает характер, за недостатком обобщений, **тяготеющего ума балласта**, с которым едва может справиться наша память.

В.В.Пашутин.

Курс общей и экспериментальной патологии.

(СПб., 1885, т.1, стр. 581)

# Почему авторы используют статистику?

Как видим, еще более 100 лет назад ведущие российские ученые понимали, что возрастание массы необобщённых фактов приводит к тому, что **они постепенно превращаются в балласт.**

В наше время актуальность этого тезиса стала еще сильнее.

# Почему авторы используют статистику?

Статистика и является инструментом такого обобщения. Часть авторов, которые разделяют этот тезис, **пытаются осознанно применять статистические методы**, чтобы на основе результатов анализа выборочных наблюдений сделать заключение относительно всей исследуемой популяции.

Другие используют статистический анализ потому, что **так делает большинство их коллег**, либо же считая, что **наличие статистических выводов поднимает престижность публикации**, вызывает большее доверие к ней.

# Почему авторы используют статистику?

Третьи - **удовлетворяя требованиям одного из соавторов**, как правило, научного или номинального руководителя - основного автора. Обычно такой руководитель, ещё на заре своей научной карьеры, удачно применил тот или иной метод анализа, и с тех пор старается воспроизводить его и в последующих работах.

## Почему авторы используют статистику?

Большинство авторов публикаций искренне надеются, что применение статистических методов должно дать им, и читателям их работ, некоторое новое знание. Однако можно «... как блестящие идеи, так и научные нелепости одинаковым образом облечь во впечатляющий мундир формул и теорем... Наряду с математизацией знаний происходит и математизация глупостей; язык математики, как ни странно, оказывается пригодным для выполнении любой из этих задач».

В.В.Налимов. Вероятностная модель языка. М.: Наука, 1979. стр.176.

## Почему плохи имеющиеся рекомендации по описанию статистики?

Причин этого несколько, но главная из них в том, что пишутся они не профессионалами в области анализа данных, а редакционными работниками, имеющими самое общее представление о специфике этой области деятельности. Поэтому в результате эти рекомендации содержат много общих фраз и пожеланий, нежели конкретных примеров и описаний, подобных тем, которые приводятся в государственных и отраслевых стандартах.

## Почему плохи имеющиеся рекомендации по описанию статистики?

В последние годы ряд журналов принимает за основу таких рекомендаций **требования так называемой Ванкуверской группы - Международного комитета редакторов медицинских журналов - ICMJE**. Если сравнить тексты трёх последовательных требований ICMJE начиная с 1978 г. по настоящее время, то можно заметить, что раздел «Статистика» в этих требованиях практически не изменился. В зависимости от того, в каком из российских журналов печатался перевод этих рекомендаций, контекст содержания этого раздела оставался в целом одним и тем же, изменялись же только некоторые нюансы, обусловленные предпочтениями переводчиков.



## Почему плохи имеющиеся рекомендации по описанию статистики?

Рассмотрим две версии одного и того же фрагмента, заимствованные из публикаций разных российских журналов. «Описывайте статистические методы настолько детально, чтобы грамотный читатель, имеющий доступ к исходным данным, мог проверить полученные Вами результаты.» В другом журнале (Кардиология, вып. 6, 1994, стр. 158) этот фрагмент выглядел так: «Статистика (статистические методы) - описывайте статистические методы настолько детально, чтобы квалифицированный читатель, имеющий доступ к оригинальным данным, смог проверить полученные Вами результаты.

## Почему плохи имеющиеся рекомендации по описанию статистики?

Что такое «**квалифицированный**» или «**грамотный**» читатель? Тот, который знает азбуку и умеет читать? Или же тот, который имеет диплом о высшем образовании, или диплом кандидата или доктора наук? Но это ведь тоже не гарантия того, что такой читатель поймет ваше описание. Как видим, такая рекомендация больше порождает вопросов и уводит в сторону от сути дела.

# Предъявите Ваши данные!

Идем далее: «...имеющий доступ к исходным данным, мог проверить полученные Вами результаты». Прочитав этот фрагмент, задайтесь вопросом: **ЧЬИ** «исходные данные» подразумевали авторы этих рекомендаций? «Исходные данные» авторов статьи, или же «исходные данные» читателя?

# Предъявите Ваши данные!

Предположим речь идет об «исходных данных» авторов статьи. Многие ли из авторов согласятся поделиться собственными «исходными данными»? Вы еще строите далеко идущие планы по «эксплуатации» этих данных, и вдруг кто-то их просит (требует?) у Вас... Тем более, если читатель сообщит, что они ему нужны для того, чтобы «проверить полученные Вами результаты». Как видим и здесь ценность подобной рекомендации весьма сомнительна...

# Предъявите Ваши данные!

Помимо чисто морально-этической стороны этой проблемы возникает и другие аспекты, в частности правовые. Поскольку уважаемый автор публикации, как правило, добывал-создавал эти самые «исходные данные» в процессе работы, то его работодатель вправе запретить такую передачу.

# Предъявите Ваши данные!

Другая причина заключается в том, что автор публикации может быть не полностью уверен в корректности выводов, полученных (и уже опубликованных!). Боязнь того, что более квалифицированный исследователь, проведя корректно анализ авторских данных, докажет ошибочность, или же некорректность продекларированных автором выводов, также вызовет отрицательную реакцию автора на такую просьбу читателя.

# Предъявите Ваши данные!

Может ли читатель «проверить полученные ... Результаты» имея собственные «исходные данные»?  
Тоже маловероятно.... Во-первых, **такие данные не будут идентичны** тем, которые имеет автор статьи. Во-вторых, **он может не располагать тем вычислительным ресурсом**, которым располагал автор статьи.

# Предъявите Ваши данные!

Сам **контекст рекомендации**  
"чтобы ... читатель ... мог  
проверить полученные Вами  
результаты" изначально  
**ориентирует читателя на**  
**недоверие авторским выводам**



# Предъявите Ваши данные!

**"На Западе давным-давно перестали верить в идеальный портрет ученого-подвижника, живущего интересами науки, для которого истина превыше всего. Реальность современного мира - ученый, которому не чуждо все человеческое, включая и пороки и достоинства. Он подвержен и честолюбию, и стремлению к материальному достатку, и прочим стимулам, способным склонить его к обману.**

# Предъявите Ваши данные!

Действительно, если итогом долгих размышлений и экспериментов являются отрицательные результаты; если журналы не принимают публикации с такими результатами, а карьера зависит от этого, искушение подправить данные может быть трудно преодолимым. Так, **из 32 авторов, опубликовавших статьи в одном из научных журналов США, 21 под разными предлогами отказались представить редакции исходные материалы**" [Сент-Джеймс-Роберт Ай. Заслуживают ли доверия ученые? //Этический кодекс ученого. Реферативный сборник. М.: ИНИОН, 1980. - 227с.]

# Предъявите Ваши данные!

В ведущих журналах США и Англии 15 лет назад появились требования о предоставлении исходных данных. Так, согласно требованиям журнала JAMA - "Журнал Американской Медицинской Ассоциации" ( Vol. 290 No. 1, July 2, 2003 ) (18 отдельных международных изданий, распространяется более чем в 150 странах на 12 языках), автор публикуемой статьи принимает на себя следующее обязательство:  
**"Я удостоверяю, что по требованию редактора я предоставляю данные или буду содействовать в полной мере предоставлению тех данных, на которых базируется данная статья ДЛЯ ИХ ПРОВЕРКИ редактором или его представителями".**

# Предъявите Ваши данные!

Такой уровень требований вполне объясним, поскольку ежегодно журнал получает порядка 4000 рукописей, однако публикуются из них порядка 10%. Пожалуй, данная формулировка требований журнала JAMA наиболее адекватна и разумна, поскольку здесь речь идет не о каком-то абстрактном читателе, желающем произвести проверку авторских результатов, а о редакторе журнала, или его представителе. Иными словами, сомнительность или неточность описания должна устраняться на самом первом этапе - этапе работы с рукописью, которая еще не стала статьей, т.е. не опубликована. Очевидно, что для этого в составе журнальных редакционных коллегий должны быть специалисты, способные такую работу выполнять достаточно качественно.

# "А ВОЗ И НЫНЕ ТАМ?"

Следуя примеру зарубежных журналов, в частности таких как JAMA и BMJ, **аналогичные пункты требований стали вводить и отечественные журналы**. К примеру, ещё в 1995 г. журнал "Кардиология" прямо говорил о необходимости статистической цензуры, что за рубежом уже давно стало нормой. Так, **в вып. 6 за 1995 г. (стр. 93)** в "Правилах подготовки рукописей ...", редакция написала следующее: **"Все статьи, принятые к публикации, при необходимости будут проверяться на адекватность применения статистических методов и интерпретации данных"**.

# "А воз и ныне там?"

В том, что такая необходимость действительно существует, читатели могут убедиться прочитав нашу статью **"Применение методов статистики в кардиологии (по материалам журнала "Кардиология" за 1993 — 1995 гг.)"** опубликованную в журнале Кардиология (№1 1998). Однако наличие такой необходимости вовсе не означает, что она автоматически будет реализована силами редакции. Прочитав вышеупомянутую статью, можно увидеть многочисленные примеры того, как эта самая необходимость либо не была осознана редакцией журнал, либо осознана, но последующая проверка не была произведена.

# "А ВОЗ И НЫНЕ ТАМ?"

В 2001 г. были разработаны и опубликованы в "**Сибирском медицинском журнале**" новые редакционные требования. В них достаточно подробно описывались и требования к описанию и использованию статистики. В том же 2001 г. автор этих строк начал статистическое рецензирование в этом журнале, а с 2003 г. в журнале начала публиковаться "**Колонка биостатистика**". В этой рубрике автор данной статьи анализировал опубликованные материалы с точки зрения возможности применения в описываемых исследованиях иных, не использованных авторами статистических методов.

# "А ВОЗ И НЫНЕ ТАМ?"

Однако через некоторое время эти усилия встретили сопротивление со стороны руководителей журнала. Последние годы журнал никем не финансировался и существовал за счёт оплаты авторами публикуемых статей. Понятно, что в этом случае **«Кто платит, тот и заказывает музыку»...**



# "А воз и ныне там?"

Летом 2006 г. на заседании редколлегии я обратил внимание главного редактора журнала, академика РАМН, председателя Томского научного центра Сибирского отделения РАМН, директора НИИ кардиологии Р.С. Карпова на то, что **уже несколько лет в журнале отсутствует практика представления письменных рецензий на поступающие рукописи**. При этом мною было также отмечено, что большинство публикуемых в настоящее время статей противоречат действующим редакционным требованиям в части использования и описания статистических методов. Ответ был таков: **«Решение о публикации статей будет принимать врач, а не статист»**.

# "А воз и ныне там?"

Когда академик РАМН путает статиста со статистиком, то иного отношения к статистике и не стоит ожидать. В своё время другой академик, Т. Лысенко, уже говорил нечто подобное: «... нас, биологов, и не интересуют математические выкладки, подтверждающие практически бесполезные статистические формулы менделистов»

# "А воз и ныне там?"

Продолжим критический анализ имеющихся рекомендаций. "По возможности, **подвергайте полученные данные количественной оценке и представляйте их с соответствующими показателями ошибок измерения и неопределенности (такими как доверительные интервалы).**" Как видим, здесь решение о необходимости количественной оценки остается за автором. Однако и в этом случае непонятно, о каких конкретно "**количественных оценках**" и "**соответствующих показателях**" идет речь, поскольку таких оценок и показателей может быть достаточно много.

# "А воз и ныне там?"

Что же, перечислять все возможные оценки? Но ведь на каждый случай всё нельзя предвидеть. Более разумно дать некоторый рекомендуемый перечень таких оценок, а также сформулировать принципы построения такого описания. Кроме того, если все рукописи, содержащие статистические оценки, будут рецензироваться специалистом по анализу данных, то все оставшиеся проблемы будут разрешаться в процессе общения автора и рецензента.

# "Доверяй, но проверяй", или "Не доверяй, и потому проверяй"

Изначальная редакционная **ориентация читателя на недоверие** к опубликованным в статьях выводам, можно трактовать как **признание того, что публикуемые статьи могут содержать сомнительные выводы**. Другими словами, это можно рассматривать как самопризнание участия в "игре в научный фокус": ".. в надежность результатов исследования изначально никто не верит, свидетельством его научности становится сам факт публикации, а важнейшим условием публикации - упоминание о статистических "гитиках". (Цитата из статьи президента издательства Медиа-Сфера С.Е. Бащинского "**СТАТИСТИКА УМЕЕТ МНОГО ГИТИК**", Международный журнал медицинской практики, 1998; N4, с.13-15. ")

## "Доверяй, но проверяй", или "Не доверяй, и потому проверяй"

Продолжим анализ требований. "Укажите, какие компьютерные программы, доступные для широкого пользователя, применялись в Вашей работе". Эту рекомендацию надо понимать так, что если вы использовали, по вашему мнению, не доступные для широкого пользователя программы, то упоминать о них не следует? Мало того, помимо того, что вы, как автор, считаете себя специалистом в своей предметной области, вы должны быть ещё и специалистом по маркетингу статистических программ. Иначе как вы узнаете, доступна или нет использованная вами программа "широкому" или "узкому" пользователю?

# "Доверяй, но проверяй", или "Не доверяй, и потому проверяй"

Итак, какую позицию должен занять читатель, знакомясь с той или иной статьей: **"Доверяй, но проверяй"**, или же **"Не доверяй, и потому проверяй"**? Очевидно, что выбор будет определяться в основном тем, насколько полно, доказательно и понятно авторы публикации смогут донести до читателя детали использованных методов анализа. При этом важно не скатиться в другую крайность, когда забывается примат объекта и цели исследования и подчиненная им роль статистического анализа. В этом случае подробное описание методов анализа становится самодовлеющей целью, за которой теряется и сама цель исследования.

# Принципы описания статистики

Нередко один и тот же метод можно использовать сразу в трёх "ипостасях", т.е. используя его и для описания, и для анализа, и для предсказания. К примеру, логистической регрессией мы **ОПИСЫВАЕМ** параметры генеральной совокупности, но одновременно мы и **АНАЛИЗИРУЕМ** взаимосвязи в этой совокупности, результат же логистической регрессии мы применяем для **ПРЕДСКАЗАНИЯ**.



# Принципы описания статистики

Конструируя описание использованных методов биостатистики в своей публикации, необходимо помнить, **КАКУЮ ЦЕЛЬ** вы преследуете, давая будущему читателю это описание. Итак, **сформулируем первый принцип описания статистики в публикациях: описание должно иметь цель.** Например, приводя результаты **АНАЛИЗА**, мы можем использовать эти результаты для целей **ОПИСАНИЯ**.

# Принципы описания статистики

В опубликованных ранее требованиях говорится: "Дайте определение статистическим терминам, сокращениям и большинству символов". Трудно не согласиться с таким требованием, и следовать ему, конечно же, необходимо. Однако давайте задумаемся, почему редакторы сочли необходимым включить эту фразу в свои рекомендации? Разве это не является само собой разумеющимся требованием применительно к любой научной терминологии? Это говорит об отсутствии в журналах устоявшихся, единых требований к используемым в статьях статистическим терминам, сокращениям и символам.

# Принципы описания статистики

Пожалуй наиболее часто об этом требовании забывают те, кто пишут в статье **выражения вида  $5,6 \pm 1,4$** , при этом нигде не уточняя, какие именно величины соединены знаком  $\pm$ . Вот и ломает голову потом бедный читатель, вспоминая нехорошими словами авторов: что такое 1,4 - или это **стандартное отклонение**, или же это **ошибка среднего**. Поскольку относительно 5,6 вариантов меньше, видимо это среднее.

# Принципы описания статистики

**Второй принцип описания  
статистики в публикациях:**

**описание должно  
соответствовать контексту.**

Согласно этому же принципу необходимо тщательно согласовывать между собой всю терминологию присущую именно той модели, о которой идет речь.

# Принципы описания статистики

Задачи АНАЛИЗА чаще всего относятся к статистическому анализу причинности, установлению наличия или отсутствия причинно-следственных связей. Поскольку такие связи имеют вероятностную природу, оценки параметров таких связей всегда имеют **различные ошибки**. Непосредственную связь с величинами таких ошибок имеет **уровень статистической значимости**.

# Принципы описания статистики

Задачи **предсказания** невозможны без предшествующего этапа **АНАЛИЗА**, поскольку именно он дает практический инструментарий для такого предсказания. Отсюда следует **третий принцип** ("принцип трех ножек табуретки"): **описания разных этапов должны взаимно дополнять друг друга.**

# Принципы описания статистики

Если вы уверены в том, что все процедуры статистического анализа проведены корректно и полученные выводы адекватны действительности и важны, то **смело описывайте проведенные статистические исследования настолько подробно, насколько хватит вашей эрудиции.** Цель такой подробности единственная - убедить будущего читателя в том, что вашим результатам можно доверять, что **их не стоит ПРОВЕРЯТЬ, а нужно ИДТИ ДАЛЬШЕ,** т.е. либо использовать эти результаты в своей конкретной практике, либо же продолжить эти исследования, опираясь, отталкиваясь от ваших результатов.

# Принципы описания статистики

**Четвёртый принцип: степень детализации и объём описания статистики в публикации должны быть адекватны той роли, которую использованные методы сыграли в получении обсуждаемых в работе выводов.**



# Принципы описания статистики

**Пятый принцип: помните, что читателей во много раз больше, чем авторов публикаций.** Из этого следует, что рано или поздно найдётся опытный читатель, который обнаружит ваши ошибки или неточности описания.

## Описание использованных методов статистики (специальная часть)

В отдельной таблице приведены рекомендации по описанию более 50 методов и критериев, наиболее часто используемых в статистическом анализе биомедицинских данных (см. **ниже**).

№ п/п	Метод проверки статистических гипотез или статистический критерий	Содержание рекомендуемого описания
1	Проверка гипотез о законе распределения	Наименование закона распределения, на соответствие которому производится проверка. Название статистического критерия, с помощью которого производится проверка гипотез, полученная величина данного критерия и отвечающее ему значение достигнутого уровня значимости.
2	Проверка равенства двух законов распределения вероятностей	Сформулировать причины проверки данной гипотезы, указать статистический критерий для проверки, привести его значение и достигнутый уровень значимости. Сделать вывод о том, какая из гипотез принимается и какой из этого следует вывод применительно к целям исследования.
3	Проверка нормальности распределения вероятностей	Сформулировать причины проверки данной гипотезы, указать статистический критерий для проверки, привести его значение и достигнутый уровень значимости. Сделать вывод о том, какая из гипотез принимается, и какой из этого следует вывод, применительно к целям исследования. Желательно привести график распределения на «вероятностной бумаге».
4	Критерий Колмогорова-Смирнова	Указать, какой именно из семейства критериев использовался в конкретном случае и цель его применения (сформулировать проверяемые гипотезы). Привести значение критерия и достигнутый уровень значимости. По результатам проверки сформулировать вывод относительно выдвигаемых гипотез.
5	Уровень значимости "р = ...."	Обязательное указание, к какому конкретному статистическому критерию относится данное значение уровня значимости. Не употреблять выражения вида «р < ...» или «р > ...».
6	Оценка дескриптивных статистик	Объем выборки (подгрупп), среднее, стандартное отклонение, ошибка среднего. При сравнении вариабельности двух и более признаков - коэффициент вариации. Обязательно сообщить, использовались ли методы оценки аномальных наблюдений (выбросов) и если – да, то какие. Сообщить, применялись ли методы робастного оценивания (Пуанкаре, Винзора, Хубера и т.д.)
7	Проверка гипотез о равенстве дисперсий с помощью F критерия Фишера	Предварительно провести проверку имеющегося ограничения на использование F критерия Фишера. Вычислить значение критерия и достигнутого уровня значимости. Сделать вывод о том, какая из конкурирующих гипотез принимается, дать интерпретацию этого результата.
8	Проверка гипотез о равенстве нескольких дисперсий с помощью критериев Кохрэна, Бартлетта и т.д.	Сформулировать проверяемую гипотезу и указать используемый для этого статистический критерий. Вычислить значение критерия и достигнутого уровня значимости. Сделать вывод о том, какая из конкурирующих гипотез принимается, дать интерпретацию этого результата. При использовании критерия Бартлетта обязательно привести результаты проверки нормальности во всех сравниваемых группах.
9	F-критерий Фишера	Сообщить, для проверки каких именно статистических гипотез использовался данный критерий, степени свободы для него и достигнутый уровень значимости.
10	Сравнение двух выборок с помощью критерия знаков	Дать описание природы количественного и группирующего признаков. Сообщить объемы наблюдений в сравниваемых группах. Сформулировать гипотезу, которая проверяется с помощью данного критерия. Привести вычисленное значение z-критерия и величину достигнутого уровня значимости. Результат проверки гипотезы интерпретировать.

11	Проверка гипотез о равенстве средних критерием Стьюдента	Привести объемы выборок, результаты проверки нормальности распределения (оно должно быть нормальным) и равенства генеральных дисперсий (они должны быть равны), значение t-критерия Стьюдента и значение достигнутого уровня значимости « $p = \dots$ ».
12	t-критерий Стьюдента при проверке иных гипотез	При проверке иных статистических гипотез (например, значимости коэффициентов корреляции или коэффициентов регрессии и т.п.) в пакетах программ могут не выводиться значения самого t-критерия Стьюдента, а только отвечающий ему уровень значимости. Поэтому можно ограничиться только уровнем значимости « $p = \dots$ ».
13	Критерий Манна-Уитни	Привести величину критерия и достигнутый уровень значимости. Привести значения средних величин сравниваемых групп.
14	Сравнение двух групп с помощью критерий серий Вальда-Вольфовица	Сформулировать гипотезу, проверяемую с помощью данного критерия. Привести значение критерия и достигнутый уровень значимости для него. Обязательно указать объем выборки, по которой проводилась проверка гипотез.
15	Сравнение двух выборок с помощью критерия Ван дер Вардена.	Сформулировать гипотезу, проверяемую с помощью данного критерия. Привести значение критерия и достигнутый уровень значимости для него. Обязательно указать объем выборки, по которой проводилась проверка гипотез.
16	U-критерий Вилкоксона, X-критерий Ван-дер-Вардена	Привести величину критерия и достигнутый уровень значимости. Привести значения средних величин сравниваемых групп.
17	Последовательный анализ Вальда	Привести аргументы в пользу выбранного метода анализа. Сформулировать проверяемые гипотезы. Привести значения статистических критериев и объемов наблюдения на момент окончания анализа. Указать, какая из конкурирующих гипотез была принята.
18	Дисперсионный анализ Краскела-Валлиса	Дать определение сравниваемых групп, указать количественную переменную, пояснить мотив выбора непараметрического ДА. Привести значение H-статистики Краскела-Валлиса, достигнутый уровень значимости. В случае отклонения нулевой гипотезы желательно провести множественные сравнения, результаты которых обсудить.
19	Дисперсионный анализ	Сообщить, был ли данный анализ параметрическим (по Фишеру), или же непараметрическим. В первом случае сообщить результаты проверки нормальности для всех (NB!) сравниваемых между собой групп, а также результаты сравнения генеральных дисперсий для этих групп (дисперсии должны быть равны). Сообщить, проверялась ли модель с фиксированными эффектами (модель 1-го типа), случайными факторами (модель 2-го типа), или же это была смешанная модель. Привести значение критерия Фишера, степени свободы и достигнутый уровень значимости. Для однофакторного дисперсионного анализа желательно сообщить и значение коэффициента детерминации. Для однофакторной модели с числом уровней более 2, после отклонения нулевой гипотезы провести множественные сравнения, используя линейные (или нелинейные) контрасты; результаты обсудить. Для многофакторного анализа привести и обсудить как значимые, так и не значимые эффекты взаимодействия.
20	Множественные контрасты в дисперсионном анализе.	Сообщить, какой вид контрастов (линейные, нелинейные) использовались и их название (Шеффе, Дункана, Бонферрони и т.д.). Для обсуждаемых результатов контрастов привести полученные оценки (значения критерия и достигнутого уровня значимости). Желательно привести средние значения для сравниваемых групп.
21	Ковариационный анализ	Указать количественный признак, группирующий признак и ковариаты. Указать, являются ли ковариаты переменными или фиксированными. В случае влияния ковариат на межгрупповой фактор привести скорректированные средние.

22	Проверка гипотез о векторах с помощью лямбда-критерия Уилкса	Сформулировать проверяемые гипотезы, описать количественные признаки и группирующий признак. Сообщить результаты проверки гипотез нормальности в сравниваемых группах, а также результат проверки гипотезы о равенстве ковариационных матриц. Привести вычисленное значение лямбда-статистики, либо функций от этой статистики с величиной достигнутого уровня значимости. Дать интерпретацию принятой гипотезы.
23	Дискриминантный анализ	Дать описание дискриминируемых групп, их число и набор количественных переменных. Привести результаты проверки предположений для сравниваемых групп. Сообщить алгоритм оценки дискриминантных функций. Привести коэффициенты дискриминантных функций и канонических осей, обсудить их структуру, а также привести графики рассеяния в канонических осях. Привести таблицу классификации с использованием дискриминантных функций. При необходимости обсудить причины неправильной переклассификации отдельных наблюдений.
24	Оценка парных коэффициентов корреляции	Указать какой коэффициент корреляции оценивается (Пирсона, Спирмэна, Кендэла и т.д.). Для корреляции Пирсона обязательно сказать о результатах проверки нормальности для обоих признаков, привести значение коэффициента корреляции и значение достигнутого уровня значимости для него.
25	Оценка частных коэффициентов корреляции	Пояснить необходимость вычисления парциальных коэффициентов корреляции. Записать пару признаков, для которых вычисляется данный коэффициент и перечень элиминируемых признаков. После вычисления парциального коэффициента провести проверку его значимости, сравнить его величину с парным коэффициентом, дать интерпретацию имеющегося различия.
26	Проверка гипотез о равенстве коэффициентов корреляции	Укажите для какой пары признаков будет проводиться проверка гипотезы и равенстве коэффициентов корреляции, и сформулируйте саму гипотезу. В частности, сообщите вид сравниваемых коэффициентов корреляции, для какого количества коэффициентов проверяется гипотеза, и что представляют собой субпопуляции, в которых проведена оценка этих коэффициентов.
27	Коэффициент непараметрической корреляции Гамма	Обосновать выбор данной статистики. Указать пару признаков, для которых производится оценка корреляции. Привести значение статистики Гамма и величину достигнутого уровня значимости. Интерпретировать полученный результат.
28	Коэффициент непараметрической корреляции Тау-Кендалла.	Обосновать выбор данной статистики. Указать пару признаков, для которых производится оценка корреляции. Привести значение статистики Тау-Кендалла и величину достигнутого уровня значимости. Интерпретировать полученный результат.
29	W - коэффициент конкордации Кендалла	Сформулировать проверяемую гипотезу, дав описание анализируемых объектов и признаков. Привести вычисленное значение W-коэффициента, значение достигнутого уровня значимости и интерпретацию принятой гипотезы.
30	Ранговая корреляция Спирмена	Обосновать выбор данной статистики. Указать пару признаков, для которых производится оценка корреляции. Привести значение статистики Спирмена и величину достигнутого уровня значимости. Интерпретировать полученный результат.
31	Анализ таблиц сопряженности	Сообщить смысл отдельных градаций (уровней) анализируемых признаков. Уточнить, как вычислялся критерий Пирсона (классический метод, максимального правдоподобия, с поправкой Йэйтса и т.д.), привести значение критерия, число степеней свободы, достигнутый уровень значимости, а также наиболее адекватный показатель интенсивности связи признаков. Желательно обсудить вклады отдельных клеток таблицы в статистику Пирсона.
32	Проверка гипотез для нескольких частотных таблиц с помощью Q-критерия Кохрена	Сформулировать проверяемую гипотезу, дав описание каждой из частотных таблиц. Привести вычисленное значение Q-статистики и достигнутый уровень значимости. Интерпретировать принятую статистическую гипотезу.

33	Методы теории планирования экспериментов	Привести цель планирования экспериментов, аргументы в пользу выбранного плана, его матрицу. Привести таблицы с оценками параметров полученных зависимостей, результаты проверки предположений а также критерии согласия полученной модели и реальных наблюдений. В случае проведения процедуры оптимизации описать алгоритм оптимизации, привести графики в канонических осях.
34	Множественный регрессионный анализ	Сообщить о том, какой именно анализ использовался (линейный, нелинейный, методом наименьших квадратов либо какой-то иной). Сообщить об используемом алгоритме оценки коэффициентов регрессии (принудительное включение предикторов, пошаговый отбор, наличие/отсутствие свободного члена, метод всех регрессий, максимального значения коэффициента детерминации и т.д.). Привести результаты проверки нормальности остатков, некоррелированности и гомоскедастичности. Привести значения размерных и безразмерных коэффициентов регрессии и результаты проверки их значимости. Обсудить соотношения безразмерных коэффициентов регрессии. Привести результаты проверки адекватности всего уравнения в целом (дисперсионный анализ). Привести значения множественного коэффициента корреляции и коэффициента детерминации.
35	Анализ канонических корреляций	Сформулировать гипотезу, которую предполагается проверить данным методом. Перечислить состав каждого множества признаков. Привести оценки параметров выбранного канонического уравнения, а также значение коэффициента канонической корреляции и результат проверки его значимости. Привести график распределения объектов в осях канонических переменных. Дать интерпретацию имеющейся связи между двумя множествами.
36	Нелинейный регрессионный анализ	Сообщить о том, какой именно алгоритм оценки коэффициентов регрессии использовался. Привести результаты проверки адекватности всего уравнения в целом (дисперсионный анализ).
37	Оценка аллометрических уравнений	На вербальном уровне описать модель связи между исследуемыми признаками. Записать в явном виде искомое аллометрическое уравнение. Указать метод оценки параметров уравнения, в частности, используются ли линеаризация, или же оценка производится иными методами. Если последнее, то указать какими. Привести значения оценок параметров и проверить их значимость. Привести график фактических и расчетных значений зависимой переменной. Обсудить полученное уравнение.
38	Нелинейные преобразования переменных	Сформулировать цель применения нелинейных преобразований. Явно указать используемой преобразование. Если есть особые точки в преобразовании, уточнить, как поступали с наблюдениями в этом случае. Прокомментировать, была ли достигнута поставленная цель с помощью этого преобразования. Если использовались параметрические преобразования типа Бокса-Кокса или Бокса-Тидвелла, указать метод оценки искомых параметров преобразований, и результаты его использования.
39	Анализ таблиц дожития и оценка уравнений кривых выживаемости	Указать выбранную модель выживаемости - модель пропорционального риска Кокса, экспоненциальная регрессия, нормальная и логнормальная регрессия, стратифицированный анализ, метод множительных оценок Каплана-Мейера и т.д. Привести таблицы оценок параметров и результатов проверки значимости полученных уравнений, а также графики с функциями выживаемости. В случае сравнения выживаемости двух групп привести используемый критерий (Гехана – Вилкоксона, F-критерий Кокса, критерий Кокса-Мантеля, логранговый критерий, критерий Вилкоксона-Пето и т.д.)
40	Оценка кривых "Доза-Эффект"	Привести выражение нелинейного уравнения регрессии, для которого производится оценка зависимости "Доза-Эффект". Сообщить, в каких единицах обозначается доза и эффект. Указать метод оценки параметров уравнения (метод наименьших квадратов, функция потерь, метод взвешенных наименьших квадратов, метод максимума правдоподобия, максимум правдоподобия и логит/пробит модели и т.д.). Привести характеристики пригодности модели, объясненную долю дисперсии, критерий согласия, график наблюдаемых и предсказанных значений и т.д.
41	Критерий Пирсона хи-квадрат	Сообщить, как вычислялся критерий Пирсона (классический метод, максимального правдоподобия, с поправкой Йейтса и т.д.), привести значения критерия, степеней свободы, достигнутого уровня значимости. Для таблиц сопряженности привести наиболее адекватный показатель интенсивности связи признаков.
42	Анализ главных компонент	Сообщить на основе какой именно матрицы (ковариационной или корреляционной) выполнялся данный анализ. Указать каким методом производился отбор используемых главных компонент из всех возможных компонент. Обсудить структуру отобранных компонент и дать их интерпретацию.
43	Факторный	Сообщить об алгоритме выделения факторов из корреляционной матрицы (главные компоненты,

	анализ	использование общностей, метода максимального правдоподобия, центроидный или метод главных осей) а также об алгоритме вращения осей. Отметить, являются ли факторы после вращения ортогональными или использовался алгоритм косоугольного вращения. Привести аргументацию выделения необходимого количества факторов, их структуру, дать интерпретацию генеральных, общих и характерных факторов. Привести данные об информативности выделенных факторов. Желательно привести графику распределения собственных значений и распределение нагрузок признаков в осях факторов, а также результаты проверки значимости собственных значений.
44	Анализ соответствий	Сообщить, является ли данный анализ парным или множественным. Перечислить признаки, используемые в данном анализе, в случае множественного анализа указать группирующий признак. Привести значение статистики Пирсона и достигнутый уровень значимости для нее. Привести таблицу координат строк и столбцов в новых осях. По результатам анализа привести график распределения анализируемых признаков в осях новых координат, выделив при этом градации группирующего признака, обсудив взаимное расположение точек на данном графике. Привести показатели качества отображения, относительную инерцию и величину косинус-квадрат.
45	Кластерный анализ	Сообщить с какой целью использовался кластерный анализ, в частности, сформулировать некоторые гипотезы, которые предполагалось проверить с помощью этого вида анализа. Уточнить, что являлось объектом кластеризации (наблюдения или признаки), указать используемую метрику и алгоритм кластеризации (иерархический КА, метод k-средних, метод поиска сгущений и т.д.). Обязательно указать используемый функционал качества кластеризации. Привести результаты кластеризации, которые могут иметь разную природу в зависимости от алгоритма кластеризации. Весьма желательно привести графические результаты кластеризации. Обсудить соответствие результатов кластерного анализа и сформулированных выше гипотез. Указать пути дальнейшего использования результатов кластерного анализа.
46	Многомерное шкалирование	Сформулировать гипотезу, для проверки которой был использован метод многомерного шкалирования. Укажите используемый метод шкалирования (метрический или неметрический). Перечислите используемые в процедуре шкалирования переменные, указать искомую размерность отображения объектов, привести аргументацию выбора количества осей, показатели качества отображения (стресс, диаграмму Шепарда). Дайте интерпретацию полученных результатов, сопровождая ее графиком распределения объектов в осях новых шкал.
47	Анализ временных рядов	Указать, был ли весь ряд эквидистантным, а также какой из алгоритмов анализа использовался. Если производилось сглаживание ряда, указать алгоритм сглаживания. Идентифицировать модель временного ряда. Если производилась оценка тренда – привести результаты такой оценки; то же самое для сезонности. При оценке функции автокорреляции привести ее график и оценки значимости коэффициентов. При использовании модели АРСС (Бокса и Дженкинса) привести оценки параметров модели а также указать алгоритм оценки (квазиньютоновский максимизации правдоподобия, приближенный метод максимального правдоподобия МакЛеода и Сейлза, приближенный метод максимального правдоподобия с итерациями назад, точный метод максимального правдоподобия по Меларду и т.д. При использовании одномерного анализа Фурье, кросс-спектрального анализа либо быстрого преобразования Фурье привести их результаты в виде оценок параметров и графиков (периодограммы, спектральные плотности и т.д.).
48	Оценка отношения шансов	Описать, каким образом организовывались перекрестные исследования, какой смысл имели отдельные уровни признаков, между которыми изучалась связь. Привести оценку отношения шансов, и дать его словесную интерпретацию. Привести стандартную ошибку оценки отношения шансов, которая дает представление о его точности. Желательно также привести величину относительного риска, а также рассмотреть возможность использования логарифма отношения шансов и логистической модели. Используя соответствующий статистический критерий провести проверку значимости отношения шансов.
49	Проверка гипотез о равенстве относительных частот	Пояснить целесообразность проверки сформулированной гипотезы, сообщить выборочные значения относительных частот, записать проверяемую гипотезу, указав число пропорций. Пояснить какой метод использовался для проверки этой гипотезы. В зависимости от количества и вида признаков, идентифицирующих разные пропорции, эти методы могут отличаться.
50	Построение доверительного интервала для относительной частоты	Сообщить, для какого показателя получена оценка относительной частоты и для чего необходимо построение доверительного интервала. Указать величину доверительной вероятности. Учитывая, что существуют разные методы оценки доверительного интервала для относительной частоты, указать метод построения доверительного интервала.

51	Логистическая регрессия с биномиальной или мультиномиальной переменной отклика	Указать название и градации зависимой переменной, а также число предикторов, предлагаемых для включения в уравнение. Если предикторов немного, перечислить их, если же достаточно много (порядка нескольких десятков), описать их в виде отдельных групп признаков. Сообщить число наблюдений по каждой градации зависимого признака, метод оценки параметров уравнения, коэффициенты логистического уравнения и отношения шансов, критерии согласия фактических и предсказанных состояний объектов по градациям зависимой переменной. Желательно для наиболее интересных результатов записать в явной форме уравнение логистической регрессии.
52	Лог-линейный анализ	Сформулировать гипотезу, которую предполагается проверить с помощью лог-линейного анализа. Указать зависимую переменную и предикторы, а также алгоритм построения зависимости (принудительное включение признаков, автоматический поиск оптимального подмножества, использование эффектов взаимодействия и т.п.). Указать, имелись ли в наличии структурные нули, и если имелись, то для каких именно комбинаций признаков. Привести для конечного варианта значение статистики Пирсона и указать, как она вычислялась (классический метод или метод максимального правдоподобия). Обсудить имеющиеся маргинальные и частные связи. Привести график наблюдаемых и расчетных частот. При необходимости привести значение критерия Мантеля-Ханзеля.



# Работа с профессионалами – гарантия качества!

Центр БИОСТАТИСТИКА проводит выездные семинары по биостатистике и выполняет заказы по статистическому анализу данных для аспирантов, докторантов и соискателей. За последние 20 лет мы выполнили анализ данных более чем для 2000 человек. С вопросами по организации семинаров и анализу данных обращаться по адресу [leo.biostat@gmail.com](mailto:leo.biostat@gmail.com)