

Введение в статистический контроль качества клинических лабораторных анализов

Москва, igbalakh@yandex.ru тел (495)432-33-20, февраль.2009

Поскольку результат лабораторного анализа не может быть абсолютно точным, речь идет о том, как определить размер погрешности и решить, допустима ли она. Ответ зависит от имеющейся информации. Точность оценивается по размеру медицинского доверительного интервала, в который с заданной вероятностью (обычно 0.95) укладывается различие между истинной величиной и результатом, полученным в лаборатории. Если контрольные анализы не проводились, но имеется достаточный массив данных пациентов, минимальное значение среднего квадрата этого различия, оценивается методом дисперсионного анализа. Если выполнено два или более контрольных анализов материала известного состава, но параметры распределения погрешностей неизвестны, вероятность ошибки оценивается по распределению Стьюдента. В случае когда известна дисперсия внутрисерийной погрешности и выполнен хотя бы один контрольный анализ вероятность погрешности оценивается по теореме Байеса, Наконец, если имеется наиболее полная информация о свойствах аналитической системы - известны дисперсии внутрисерийной и межсерийной погрешностей и выполнены контрольные анализы, наиболее точный ответ дает метод совместного распределения контрольных и опытных проб.

В работе также раскрываются основные понятия и термины, используемые при статистическом контроле точности лабораторных анализов, освещаются возможные причины ошибок и особенности использования методов теории вероятностей в клинической лаборатории, кратко формулируются основные трудности и проблемы. Описываются принципы создания компьютерных программ для оценки точности.

1. Введение	2
1.1. Что такое контроль качества и зачем он нужен	3
1.2. Виды контроля качества	3
1.3. Особенности измерений в клинической лабораторной диагностике	4
1.4. Аналит и мезюрэнт	5
1.5. Точность (неопределенность) и правильность результата измерения	6
1.6. Доверительный интервал и медицинский интервал	7
1.7. Калибровки	9
2. Некоторые положения теории вероятностей	10
2.1. Случайные величины	10
2.2. Оценка вероятности по частоте	11
2.3. Нормальное распределение и распределение хи-квадрат	12
2.4. Погрешности	13
3. Контроль качества по данным анализов пациентов	14
3.1. Гистограммы	14
3.2. Критерий лямбда (Колмогорова-Смирнова)	16
3.3. Дисперсионный анализ.	17
3.3.1 Оценка дисперсии среднего	17
3.3.2 Автокорреляционная функция.	18
4. Контроль качества по контрольным пробам	19
4.1. Методы, основанные на оценке качества контрольных проб	20
4.1.1. Метод, основанный на распределении хи-квадрат	20
4.1.2. Метод, основанный на распределении Релея	21
4.2. Методы, основанные на оценке параметров распределения.	22
4.2.1 Распределение Стьюдента	22
4.2.1.1 Два контрольных анализа.	23
4.2.1.2 Больше чем два контрольных анализа.	25
4.2.2. Оценка точности результатов анализов на основе теоремы Байеса	27
4.2.3. Оценка вероятности попадания в медицинский интервал по совместному распределению нескольких контрольных анализов	31
4.2.3.1. Оценка межсерийной и внутрисерийной дисперсий погрешностей	32
4.2.3.2. Вероятность попадания результата анализа в медицинский интервал когда параметры распределения погрешностей известны	33
4.2.3.4. Оценка вероятности попадания серии результатов анализов в заданный медицинский доверительный интервал	36
5. Достоверность подсчета лейкоцитарной формулы	36
6. Заключение и практические рекомендации	41

Введение

Под влиянием успехов технического прогресса все большее значение приобретают результаты так называемых парамедицинских исследований, которые выполняются не непосредственно лечащим врачом, а его коллегой, специалистом узкого профиля. Вопрос о том, насколько можно полагаться на эти данные всегда актуален, особенно когда речь идет о результатах лабораторного исследования биологического материала. Ведь современная лаборатория способна в буквальном смысле слова завалить врача ворохом цифр, которые не всегда просто интерпретировать, тем более важно уметь оценить их достоверность. Вопрос отнюдь не простой, тем более, что требования к качеству анализов зависят от характера заболевания и проводимого лечения. Так при диспансерном обследовании достаточно знать находятся ли результаты заведомо в пределах «нормы», т.е. быть уверенным, что пациент не попадает в группу риска, а при контроле специфической терапии нужна большая точность. Чувствительный, простой и дешевый метод определения активности аминотрансфераз по Райтману и Френкелю явно не пригоден для клиники, но при эпидемиологических обследованиях детских коллективов он оправдывает себя.

1.1. Что такое контроль качества и зачем он нужен

Контролем качества называется сумма мероприятий, проводимых клинической диагностической лабораторией с целью контролировать качество и своевременность выдаваемых результатов. Он может проводиться как в стенах лаборатории ее сотрудниками, тогда это называется внутренним контролем, так и с участием других лабораторий и организаций, тогда он называется внешним. Это комплексная работа.

В последние годы этому уделяется много внимания – проводятся международные конгрессы, выходят приказы Министерства здравоохранения и социального развития России, издаются книги. Однако, делается явно недостаточно и лаборатория, которые хочет выдавать правильные результаты, (и, соответственно клиницисты, которые хотят получать их), должна полагаться в первую очередь на свои силы, ясно понимая, что официальная система контроля качества позволяет исключать лишь самые грубые ошибки. Для этого есть несколько причин – это и быстрый технический прогресс (особенно в компьютерной технологии), за которым никакая бюрократическая машина не может угнаться, и невозможность применять ко всем лабораториям одинаковые требования, и просто недостаточная научная разработка вопроса и подготовленность кадров. Дело в том, что контроль качества клинических лабораторных исследований, как и медицинская статистика и биометрика, основывается на теории вероятностей, но использует совсем другие главы и методы этой сложной математической дисциплины. Механический перенос традиций и методов одного раздела на другой чреват ошибками.

В идеале, передавая результат анализа лечащему врачу, лаборатория должна и трактовать его. Это сложная задача, которая, что греха таить, не всегда по силам лаборатории. Ведь, прежде всего надо ответить на два вопроса – насколько достоверен результат анализа и превышает ли он возможные физиологические колебания, т.е. клинически информативен. Однако, современная, оснащенная компьютером клиничко-диагностическая лаборатория, почти всегда может и должна выполнять более простую задачу – информировать врача о том, насколько аналитически достоверен данный результат. Иными словами лаборатория обязана уметь оценивать неопределенность выдаваемых ей результатов и информировать об этом лечащего врача. А оценка собственных возможностей это и есть внутренний контроль качества.

У проблемы качества лабораторных анализов есть и формально юридический аспект: в случае врачебной ошибки, или даже подозрения на такую ошибку, надо уметь доказать достоверность данных, на основе которых было принято решение.

Когда все идеально – оборудование новое, реактивы свежие, персонал опытный – контроль качества всего лишь перестраховка, но такое бывает не всегда. Зачастую приходится разрешать себе те или иные «вольности», и только правильно организованный внутренний контроль дает ответ, насколько эта «вольность» допустима. Внешнему контролю, особенно международному, обязательно должен предшествовать внутренний – какой смысл тратить деньги и силы, если нет уверенности, что заключение будет положительным?

1.2. Виды контроля качества

Практически все эффективные способы контроля качества требуют компьютера, соединенного «в линию» с анализатором. Простые приемы – рисование контрольных графиков, правила Вестгарда, косум и др., которые можно выполнить вручную, не дают достоверной картины, их выводы носят интуитивный характер и могут рассматриваться только как ориентировочные. Это никак не означает, что они бесполезны, они весьма полезны, так как дисциплинируют работников, приучают их оценивать свою работу. Просто возможности этих методов очень ограничены.

Различают оперативный контроль, который позволяет оценить погрешность выполненных сегодня анализов, до передачи их в клинику, и ретроспективный, когда речь идет

о состоянии дел за прошедший период. Методы работы и в том и другом случае очень близки, все дело в количестве обрабатываемого материала, но задачи они решают разные. Оперативный контроль отекает плохо выполненные серии анализов, ретроспективный же не только фиксирует качество работы – визитную карточку лаборатории, но и позволяет отобрать наиболее эффективные методики, сравнить разную аппаратуру, наборы реактивов и даже отдельных исполнителей.

Внешний контроль качества состоит в том, что из единого центра рассылаются контрольные материалы в проверяемые лаборатории, затем центр сравнивает результаты. Обычно вычисляют среднее всех лабораторий, которое принимают за истинное значение. Работа конкретной лаборатории оценивается в зависимости от того, насколько полученный в ней результат отличается от общего среднего. Чем больше это различие, тем ниже ранг качества работы. Такая система имеет, по крайней мере, два недостатка – чтобы занять высокий ранг надо не только самому хорошо работать, но и чтобы другие работали хуже. Как говорится, в стране слепых и кривой король! Другой недостаток существеннее – поскольку все биохимические методики включают калибровки, очень многое зависит от того, насколько совместимы разные калибраторы. Если их происхождение не оговорено, то получается, что выигрывают те, которые используют наиболее распространенный калибратор – именно он и определяет среднюю величину. Получается, что соревнуются не лаборатории, а фирмы торгующие реактивами. При правильной организации работы центр должен был бы рассылать не только контрольные, но калибровочные материалы.

1.3. Особенности измерений в клинической лабораторной диагностике

Наука об измерениях – метрология и раздел теории вероятностей, касающийся биологических объектов – биометрика, существуют давно, их рекомендации хорошо разработаны, но не всегда применимы в клинической лабораторной диагностике. Дело в том, что и в физике, и в технике, и в биометрии стараются либо повторять одно и то же измерение по несколько раз, либо обследовать группы однородных объектов – например, длину костей или вес зерен. Если бы врачи-лаборанты имели возможность повторять каждый анализ по несколько раз, можно было бы воспользоваться хорошо разработанными методами биометрии и метрологии, тогда задача очень упростилась бы. К сожалению, это невозможно. При анализе биологического материала он чаще всего разрушается, поэтому повторные измерения невозможны, можно только анализировать другую пробу из того же материала, и думать, что она не отличается от предыдущей. Но повторные анализы настолько дороги, что, за редким исключением, невыполнимы. Поэтому статистическая обработка результатов измерений в классическом варианте невозможна, приходится использовать совсем другие подходы.

Экспериментатор обычно статистически обрабатывает данные, чтобы доказать различие между группами, например, показать эффективность лечения. При этом существует что-то вроде «презумпции невиновности» – в сомнительных случаях, когда не удается доказать что различие существует, считается, что его нет, хотя на самом деле, оно, возможно, и есть. Оценивая качество работы врач-лаборант решает обратную задачу – он должен доказать, что различия нет, что сегодня аналитическая система работает также, как работала вчера и позавчера, данные сопоставимы. Поэтому лабораторная диагностика требует «презумпции виновности» – надо доказать, что результаты правильные, если таких доказательств нет, значит результаты работы неудовлетворительные. Это означает, что вся «серая зона» – когда вероятность что ответ правильный примерно равна вероятности, что он неправильный должна отбрасываться. Методы математической обработки данных, рекомендованные приказом Минздрава России, позволяют отбрасывать лишь заведомо ошибочные ответы, сомнительные же попадают в категорию достоверных. Вот какую цену приходится платить за механическое перенесение правил из соседней дисциплины!

Помимо всего прочего, клиника отличается от эксперимента и биометрии тем, что имеет дело не с группами, а индивидами. Данные лабораторных клинко-диагностических анализов можно, а иногда и нужно, усреднять, но при этом не надо выбрасывать ребенка вместе с водой и забывать, что мы имеем дело не с группой, а с конкретным пациентом.

Вопрос что важнее – отклонить правильный результат или пропустить неправильный, возникает потому, что оценка вероятностная, ошибка всегда возможна. Уменьшая вероятность ложного отклонения результата, мы тем самым увеличиваем опасность пропустить ошибку и, наоборот, устанавливая более строгий фильтр, отсекающий возможную ошибку, мы увеличиваем вероятность ложного отклонения. В данном случае ответ ясен – если мы сомневаемся в правильности результата и отвергаем его – мы мало рискуем, так как все данные и сам материал остаются в лаборатории. Они могут быть подвергнуты более обстоятельной проверке, и, если опасения не подтвердятся, переданы в клинику с некоторым запозданием. Обратная ошибка значительно хуже – дальнейшая судьба, вышедшего из лаборатории недостоверного результата ей неподконтрольна – никогда нельзя знать, как он будет использован лечащими врачами и к каким последствиям приведет. Поэтому критерии качества должны быть такими, чтобы из стен лаборатории не выходили не только недостоверные, но даже сомнительные результаты.

1.4. Аналит и мезюрант

Международные стандарты ISO /FDIS 15189 и ISO/FIDS 15195 вводят понятие мезюранта – т.е. измеряемой величины. Согласно сложившейся традиции, объект выполняемого клинической лабораторией анализа называется аналитом, им могут быть как химические вещества, так и клетки. Однако трудно говорить об анализе, если измеряется, например, диаметр эритроцита, тем самым аналит это частный случай мезюранта.

Практически все лабораторные показатели, будь то химические вещества или клетки, это не индивидуальные объекты, точное количество (концентрация) которых реально существует, а целые семейства, иногда довольно обширные. Их компоненты обладают неодинаковыми физиологическими и химическими характеристиками, поэтому объединение их в один показатель часто вынужденная и искусственная мера. Хотя Международный стандарт выделяет так называемые СИ-аналиты, куда попадают те вещества, химическая структура которых хорошо известна, поэтому концентрации могут быть выражены в единицах СИ, это не значит, что в организме все их молекулы одинаковы. Как писал французский философ Бергсон, занимаясь классификацией мы невольно становимся последователями Платона, который считал, что каждый реальный предмет всего лишь отражение некоторой вечной и нетленной идеи. От себя добавим, что поскольку такой идеи нет, мы ее каждый раз придумываем.

Что следует считать гемоглобином (что такое платоновская идея гемоглобина) – только ли полноценные молекулы, переносящие кислород, или также карбокси- и метгемоглобин, нитропроизводные, свободные молекулы гема и глобина и т. д. Насколько оправдано такое объединение и где его границы? А что есть глюкоза – только ли это сумма фуранозной и пиранозной форм, или также лабильные шиффовы основания и частично обратимые продукты гликирования белков? То же относится к ферментам сыворотки крови, которые не есть отдельные химические вещества с четко детерминированными свойствами, а группы, состоящие из нескольких изозимов. Любой из этих вопросов заслуживает серьезного обсуждения в каждом конкретном случае, и от того, какой будет ответ, зависят и границы нормы и допустимая погрешность измерения. Так результаты редуктометрических методов определения глюкозы зависит от того, сколько у данного пациента в крови восстанавливающих веществ, имитирующих глюкозу. То же касается и клеток – нет «естественной» границы между палочкой и зрелым сегментом, между ретикулоцитом и эритроцитом, разные школы могут (и фактически проводят) ее не совсем одинаково. Единый подход ко всем ситуациям отсутствует и вряд он будет выработан в ближайшее вре-

мя. Поэтому получение «правильных», полностью сопоставимых результатов лабораторных исследований это концептуальная, а не чисто техническая проблема.

Чтобы выйти из положения используют так называемую «установленную величину», которая в английской литературе называется «целевой величиной» (target value). Это концентрация или количество, которое содержится в определенном препарате и используется для калибровки. Она указывает, что должно получиться, если усреднить результаты многих повторных анализов, выполненных данным методом в данном материале в идеальных условиях. Тем самым истинное значение заменяется установленным. Информация о том, как эта величина получена, входит в понятие прослеживаемости, так как часто ее устанавливают путем сравнения со стандартом более высокого уровня. Такая постановка вопроса, вообще говоря, не может считаться идеальной, было бы лучше, если бы ее вычисляли по определенным правилам исходя, например, из спектральных характеристик материала или его атакующести специфическим ферментом, но в настоящее время такой возможности нет.

Аналитическая работа клинико-диагностических лабораторий опирается на два типа препаратов – калибровочные и контрольные. Когда-то считалось, что для калибровки используют химически чистые вещества, а для контроля надо брать препараты, имитирующие биологические субстраты. Это давно уже не так и на первом месте стоят технические характеристики – стабильность, возможность использовать один препарат для нескольких методов и т.д. Производители, как правило, не раскрывают их состав, и принципиальной разницы между калибраторами и контрольными материалами нет. Однако используют их по-разному: калибратор для получения калибровки, а контрольный материал для ее контроля. Очень важно, что используются два разных препарата, это позволяет повысить надежность работы, хотя, когда проверка не может подтвердить калибровку, возникает сложная ситуация. В этом случае непросто разобраться кто прав, лучше всего заменить оба препарата.

Иногда в разговорах, да и в литературе, вместо названий «калибраторы» и «контроли» говорят о «стандартах» или «эталопах». Это нежелательно, так как в метрологии эти слова употребляются иначе.

1.5. Точность (неопределенность) и правильность результата измерения

Такие, в житейском смысле очевидные понятия как «норма» и «точность» разными людьми могут пониматься по-разному, дать им однозначные, непротиворечивые определения очень не просто, терминологии разных документов не совпадают. Наилучшие, пригодные для практического употребления, термины дают Международные Стандарты ISO, хотя и они не лишены недостатков. Прежде всего, надо понимать, что точность и правильность это разные вещи. Правильность это качество, которое характеризует способность аналитического метода определить именно данный аналит, она не зависит от работы практической лаборатории, а только от возможностей заложенных в самом принципе определения. Так, например, результаты редуктометрического метода определения глюкозы зависят от присутствия в исследуемом материале других восстанавливающих веществ (глутатион, аскорбиновая кислота), самое тщательное выполнение аналитических процедур не может устранить этот дефект. При анализе электролитов ионоселективными электродами определяется активность ионов, но для сопоставимости с фотометрией пламени и атомной абсорбциометрией, результат выдается как концентрация. Чтобы ответ был правильным, надо калибровать электрод раствором с той же концентрацией белка и рН, как в анализируемой пробе, что практически невозможно, поэтому используют «стандартную» калибровку, которая предполагает, что белка 75 г/л, а рН 7,35; таким образом, сам метод калибровки таит в себе источник погрешности.

Согласно Международному стандарту точность есть качественная характеристика отдельного измерения, которая не может иметь количественное выражение, а может толь-

ко быть «достаточной» или «недостаточной». Количественной мерой, того, что в жизни мы называем точностью, служит неопределенность измерения, которая оценивается через разброс данных посредством математических понятий среднего квадратичного (по западной терминологии стандартного) отклонения и дисперсии. Таким образом точность и неопределенность это практически синонимы и нет необходимости проводить между ними различие. Хотя в стандарте упоминается понятие доверительного интервала, вопрос этот явно не доработан и требует уточнения.

1.6. Доверительный интервал и медицинский интервал

Физики и инженеры обычно делают несколько повторных измерений по данным, которых вычисляют среднюю (\bar{X}), и среднюю квадратичную (S). Как доказывают в курсах теории вероятностей, \bar{X} являются наилучшей оценкой истинного значения измеряемой величины, которое называется математическим ожиданием и обозначается M . Доверительным интервалом обычно называют интервал вокруг \bar{X} в котором с заданной доверительной вероятностью находится M . К сожалению, такой подход не пригоден для клинической лабораторной диагностики. Ниже этот вопрос будет рассмотрен более подробно, и введено понятие медицинского интервала. Те, кто не интересуется деталями, могут пропустить этот раздел, приняв на веру, что медицинский доверительный интервал отличается от традиционного доверительного интервала тем, что первый откладывается вокруг истинного значения измеряемой величины, а второй вокруг ее оценки, кроме того традиционный интервал не всегда симметричен, а медицинский – всегда. Удобно чтобы этот интервал совпадал с тем, что Фразер назвал «уровнем принятия решения» – минимальным изменением параметра, которое имеет клиническое значение.

В дальнейшем мы будем исходить из того, что R – результат любого лабораторного анализа можно представить себе в виде суммы двух величин M – «истинной» концентрации аналита и погрешности анализа δ , которая является случайной величиной.

$$R = M + \delta$$

Это относится в равной мере к пробам пациентов, калибровочным и контрольным анализам. В результате калибровки получается функция $M = \varphi(G)$ устанавливающая соответствие между G сигналом прибора и M концентрацией мезюранта (измеряемого вещества) в калибровочном материале (калибраторе). M это уже упоминавшаяся установленная величина (в англоязычной литературе целевая- target value). Когда речь идет о контрольном материале ее называют также аттестованным значением данного контрольного материала. Если калибровка выполнена хорошо – а именно этот случай мы рассматриваем, средняя величина (математическое ожидание) δ равно нулю, т.е. погрешности равновероятны в обе стороны. В этом случае M является средней величиной (математическим ожиданием) всех возможных результатов анализа данного материала выполненных данным методом. Тем самым «истинное» значение лабораторного показателя биологического материала зависит от метода, это средняя величина многих выполненных в одних и тех же условиях, в том числе в разные дни, повторных анализов. Слова «в данных условиях» надо понимать широко, имея в виду всю сумму условий в которых проводилась калибровка.

Случайная величина δ , точнее ее характеристики, служит наилучшим показателем качества работы. То, что она характеризует, на бытовом уровне обычно называют точностью анализа, международный стандарт использует термин «неопределенность». Ее лучше всего характеризует σ_δ^2 дисперсия случайной величины δ .

Дисперсия, это абстрактная математическая характеристика, в повседневной работе удобнее пользоваться более наглядным доверительным интервалом – участком вокруг M куда с заданной доверительной вероятностью попадает результат анализа R . Тем самым используются два параметра – размер доверительного интервала и вероятность попадания в него. Можно сказать, что ширина доверительного интервала характеризует точность, а

доверительная вероятность надежность измерения. В медицинских и биологических исследованиях почти всегда доверительную вероятность устанавливают на уровне 0,95, лишь изредка 0,99, поэтому обычно неопределенность характеризуется размером 95% доверительного интервала. Часто (без достаточных оснований), принимают, что этот интервал симметричен, тогда удобно указать его полуширину и отделять ее от результата знаком \pm . Например $12,3 \pm 0,5$ означает, что с вероятностью 0,95 истинная величина находится между 11,8 и 12,8.

Нетрудно видеть, что приведенное выше, приспособленное к нуждам лабораторной диагностики, определение доверительного интервала, отличается от общепринятого. Чтобы избежать путаницы, мы будем называть его медицинским доверительным интервалом или просто медицинским интервалом. Дело в том, что согласно установившейся традиции, понятие доверительного интервала относят к результатам измерения одной физической величины, в нашем случае к одному конкретному анализу. Этот интервал строится вокруг оценки измеряемой величины, в качестве которой служит усредненный результат серии измерений, он с заданной вероятностью покрывает математическое ожидание. В другой серии измерений (или при исследовании материала другого пациента) получается другой, может быть очень близкий к первому, но отличный от него, доверительный интервал. Сколько серий измерений, столько и доверительных интервалов. Мы же хотим иметь один интервал, характеризующий все измерения (анализы всех пациентов).

Другое различие сложнее и очень существенно. Доверительную вероятность β определяют как вероятность того, что разность между математическим ожиданием M и результатом измерения X по абсолютной величине меньше ε :

$$\beta = P\{|M - X| < \varepsilon\}$$

В этом случае говорят, что отрезок $X - \varepsilon$ $X + \varepsilon$ покрывает математическое ожидание («истинную величину») с вероятностью β . Зная закон распределения вероятности погрешности, можно указать такую погрешность измерения ε , которая настолько велика, что представляется очень маловероятной. Это позволяет утверждать, что с точки зрения техники измерений, вероятность того, что M находится за пределами отрезка $X - \varepsilon$ $X + \varepsilon$, мала, но еще не означает, что эта малая вероятность не реализуется. Могут существовать внешние обстоятельства, влияющие на это. Например, результат анализа может быть с аналитической точки зрения правдоподобным, но находиться за пределами тех величин, которые возможны у данного пациента, например, у ходячего больного. Поэтому доверительный интервал, как он обычно понимается и определен выше, означает только, что экспериментальные данные не противоречат тому, что математическое ожидание находится на отрезке $X - \varepsilon$ $X + \varepsilon$. Исследователь, измеряющий физический параметр, строит доверительный интервал вокруг среднего результата серии измерений X , а лаборант, оценивающий качество работы, строит медицинский доверительный интервал вокруг аттестованного (т.е. установленного) значения M контрольного материала. Медицинский доверительный интервал без всяких оговорок обозначает отрезок, на который с заданной вероятностью попадает результат измерения.

Приведенное выше определение доверительного интервала, широко распространено, так как удобно при решении физических задач, но не единственное возможное. Другой возможный подход, который В.И.Романовский называет классическим, опирается на теорему Байеса:

$$\beta = P\{X - \varepsilon < M < X + \varepsilon\} = \frac{\int_{X-\varepsilon}^{X+\varepsilon} \phi(M) f(X-M) dM}{\int_{-\infty}^{+\infty} \phi(M) f(X-M) dM}$$

Здесь $\phi(M)$ плотность вероятности того, что математическое ожидание равно M , а $f(X-M)$ плотность условной вероятности того, что результат измерения X , если математи-

ческое ожидание M . Сказанное в равной степени применимо и к результатам единичных измерений (X) и к средним величинам серии измерений (\bar{X}), различие только в величинах дисперсии.

При измерении физических параметров – например, длины волны света или температуры замерзания раствора, речь идет о величинах по определению неслучайных. В этом случае классический (по В.И.Романовскому) подход трудно использовать, так, так как смысл $\phi(M)$ непонятен. Однако, при выполнении лабораторных анализов, объект измерения случайная величина (у каждого пациента своя, заранее неизвестная, концентрация аналита). В этом случае $\phi(M)$ это плотность вероятности результата анализа в данной группе пациентов, ее можно оценить по частоте. Поэтому классический подход для вычисления доверительного интервала имеет определенные преимущества. В частности, хотя $f(X - M) = f(\varepsilon)$ симметрична, так как речь идет о погрешности измерения, но произведение $\phi(M)f(X - M)$ не симметрично относительно X (рис.1). Так как обычно при выполнении лабораторных анализов

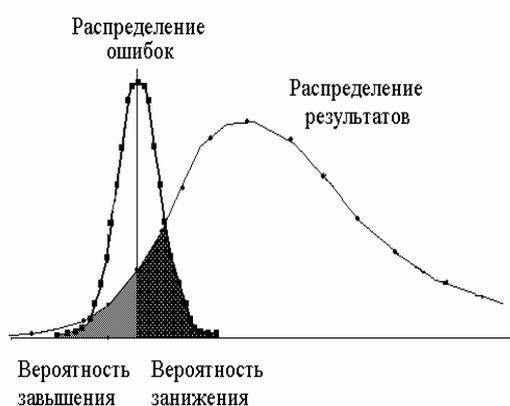


Рис. 1.1 Асимметрия вероятности ошибки

нормальные величины встречаются чаще патологических, указанная несимметричность означает, что пациент с нормальным значением исследуемого параметра имеет больше шансов получить ошибочный патологический результат, чем пациент с патологией по ошибке получить нормальные данные. Такая перестраховка, в целом устраивает лабораторную медицину.

1.7. Калибровки

Большинство современных химических лабораторных методов основано на цветных реакциях и фотометрировании, т.е. измерении оптической плотности. Обычно проводится калибровка, которая устанавливает соответствие между интенсивностью окраски и концентрацией аналита. Для этого используют изготовленный промышленностью специальный калибровочный материал или калибратор. Давно прошло то время, когда каждая лаборатория выполняла эту работу самостоятельно, сейчас это практически невозможно. Дело в том, что интенсивность окраски зависит не только от концентрации самого аналита, но и от других присутствующих в биологическом материале компонент, сам аналит, как уже отмечалось, это не одно вещество, а сумма нескольких соединений. Поэтому калибратор готовят на основе так называемого «матрикса», который имитирует состав сыворотки и имеет очень сложный состав. Изготовитель калибратора часто сам калибрует его по эталону более высокого уровня, поэтому очень важно, чтобы калибратор был «прослежен» т.е. было известно, на какие эталоны он опирается. Хороший калибратор единственный способ уменьшить систематическую погрешность, но этого недостаточно. После того как калибровка выполнена, ее надо систематически проверять, используя аттестованный контрольный материал – некоторый суррогат биологического материала, установленная концентрация аналита в котором известна. Очень важно чтобы калибратор и контрольный материал были сопоставимы, если же этого нет, и результаты контроля постоянно говорят о неудовлетворительной работе, причину которой найти не удается, надо заменить оба материала. Это, конечно, самый крайний случай.

Методы определения ферментов не могут калиброваться подобно прочим анализам, здесь понятие калибровки имеет другой смысл. Ведь определяется не сам фермент, а его активность, а активность это не вещество, а функция. Мерой активности фермента служит

количество продукта реакции, образовавшееся за единицу времени (или разрушенного субстрата), в лаборатории определяется не сам фермент, а этот продукт. Раньше по нему и калибровали. Теперь же, обычно используют синтетические субстраты, которые подобраны так, что всегда образуются одно те же вещество (чаще всего паранитрофенол или НАД), спектральные характеристики которого известны. Необходимость калибровки отпадает, расчет активности делается по константам, записанным производителем набора реактивов в инструкции. Это с одной стороны позволяет избежать ошибок при калибровке, с другой стороны чревато опасностью, поскольку небольшие изменения условий реакции (например, кислотности) заметно меняют спектральные характеристики продуктов реакции, что сказывается на величине константы. Кроме того, константа рассчитывается для света определенной длины волны и степени монохроматичности, изготовитель реактивов может не знать, какой прибор будет использован в лаборатории. По этим причинам в наборах для формально одних и тех же методов могут указываться разные константы, вызванные этим ошибки обнаружить и исправить очень трудно. Калибровка по материалу с известной ферментативной активностью (и, соответственно, проверка по нему) снимает часть проблем, но создает новые. Как уже отмечалось, ферменты сыворотки крови это смесь нескольких изоформ, соотношение которых у разных людей разное, и в ряде случаев зависит и от природы заболевания. В контрольном материале ферменты совсем другой природы, их реакция на изменение условий анализа иная, в ряде случаев может быть противоположной. Поэтому использование калибровочных и контрольных материалов для определения ферментов может создавать ложное ощущение благополучия там, где его на самом деле нет.

2. Некоторые положения теории вероятностей

2.1. Случайные величины

Статистический контроль качества основывается на теории вероятностей, поэтому стоит вкратце перечислить некоторые ее, хорошо известные положения, имеющие непосредственное отношение к существу вопроса.

Величины бывают случайные и неслучайные. Если брошен камень, его траектория строго определена начальными условиями – направлением и скоростью полета, поэтому место падения можно вычислить по законам механики, это величина неслучайная. Если же брошена игральная кость, невозможно точно предсказать на какую грань она упадет, это дело случая, поэтому число выпавших очков это случайная величина. Результат лабораторного анализа тоже случайная величина, потому что заранее неизвестно какой именно биологический материал будет анализироваться сегодня. По традиции случайную величину обозначают прописной буквой, например x , а результат конкретного анализа называют экземпляром случайной величины x и обозначают заглавной буквой обычно с индексом – например, X_1 , X_2 и т.д.

Если случайная величина с равной вероятностью может принять любое значение, говорят, что она распределена равномерно, но чаще приходится сталкиваться с такими случайными величинами, которые какие-то значения принимают чаще, чем другие, тогда говорят о распределении. Функция, которая показывает вероятность того, что случайная величина X меньше неслучайной величины Y называется функцией распределения X . При увеличении Y всегда также растет и функция распределения, которая, однако, никогда не бывает больше единицы. Скорость, с которой она растет, называется плотностью распределения, колоколообразную кривую плотности нормального распределения (кривую Гаусса) все когда-либо видели. По определению вся площадь под кривой плотности распределения равна единице, вероятность попадания экземпляра случайной величины на какой-либо участок под ней равна доли, которую занимает площадь участка.

Если какое-то измерение повторить несколько раз, получается серия результатов измерений или другими словами серия экземпляров случайной величины. Среднее бесконечного числа повторных измерений (что, конечно же, невозможно!), называется математическим ожиданием, это и есть истинное значение измеряемой величины. Чем больше измерений в серии, тем ближе ее среднее к математическому ожиданию – идеалу, к которому можно стремиться, но который нельзя достичь! Обычно математическое ожидание обозначают буквой M , результаты отдельных измерений X_1, X_2, \dots, X_n , а среднее серии, которое также называется выборочной оценкой среднего или просто оценкой – \bar{X} .

Разброс экземпляров случайной величины (например, результатов повторных анализов) характеризуется дисперсией – средним квадратом отклонения, которая обозначается σ^2 . На практике в ряде случаев удобнее пользоваться средним квадратичным, которое равно квадратному корню из дисперсии и обозначается σ . Дисперсия это средний квадрат всех возможных экземпляров случайной величины – т.е. в нашем случае бесконечного числа повторных анализов, такая же абстракция, как и математическое ожидание. На практике оно тоже оценивается по среднему квадратичному отклонению выборки и обозначается S . Таким образом, мы всегда имеем дело как с идеальными значениями параметров распределения M и σ , так и с их, полученными в эксперименте, выборочными оценками \bar{X} и S .

Вероятность – это абстрактное понятие, которое касается возможности появления какого-либо события (например, того, что ошибка больше некоторой величины), обычно она заранее неизвестна. Ее можно оценить только, когда накоплен определенный опыт и частота появления события известна. Это называется оценкой вероятности по частоте.

2.2. Оценка вероятности по частоте

Вероятность события недоступна непосредственному наблюдению, мы видим только его частоту. Например, если в крови пациента 25% всех лейкоцитов это лимфоциты, то вероятность P при подсчете лейкоцитарной формулы встретить лимфоцит составляет 0,25. Однако, при просмотре 100 клеток мы не обязательно встретим точно $P^*=25$ лимфоцитов, их может быть и $P^*=24$ и $P^*=26$ и даже $P^*=30$, не зависимо от опыта или прилежания работника, а только от случая. Как же узнать, сколько их могло бы быть, если бы можно было просмотреть все клетки? Прямо на этот вопрос ответить нельзя, зная частоту можно только указать диапазон в котором с заданной вероятностью находится ответ.

Каждая клетка либо лимфоцит, либо не лимфоцит, подобно тому, как брошенная монета падает либо орлом, либо решеткой. Такое распределение называется биномиальным. Среднее квадратичное отклонение вычисляется по формуле:

$$S = \sqrt{\frac{P(1-P)}{n}}$$

Здесь P – вероятность того, что в поле зрения именно клетка данного вида (т.е. истинная их доля), $1-P$ – вероятность что это другая клетка, n – число посчитанных клеток. Зная эти параметры, можно оценить каких результатов следует ожидать при подсчете формулы. На практике обычно приходится решать обратную задачу – при просмотре мазка получились такие-то результаты (известна частота события), что можно сказать о вероятности? Более подробные расчеты показывают, что если событий больше 10% и меньше 90%, границы доверительного интервала математического ожидания (т.е. истинного значения) достаточно точно можно найти по формулам:

$$P_n = P^* - t \sqrt{\frac{P^*(1-P^*)}{n}} \quad P_g = P^* + t \sqrt{\frac{P^*(1-P^*)}{n}}$$

Здесь P_n и P_g нижняя и верхняя границы доверительного интервала, P^* частота (найденная доля клеток), n общее число событий (число просмотренных клеток), t – коэффициент, который зависит от доверительной вероятности:

Доверительная вероятность	t
0,90	1,643
0,95	1,960
0,99	2,576

Так, если при подсчете лейкоцитарной формулы посчитано 100 клеточных элементов, и из них P^* оказались лимфоцитами, то с 95% вероятностью их истинная доля находится между:

$$P^* - 1,96 \sqrt{\frac{P^*(P^* - 1)}{n}} \text{ и } P^* + 1,96 \sqrt{\frac{P^*(P^* - 1)}{n}}$$

То же самое относится и к любым другим событиям или явлениям, например результатам опытов или числу заболеваний данной болезнью – P^* наблюдаемая доля (частота), n общее число наблюдений или опытов.

2.3. Нормальное распределение и распределение хи-квадрат

Теория нормального распределения бала заложена еще Гауссом в 18 веке, затем развита и разработана Лапласом, однако условия его возникновения раскрыл только Ляпунов, который доказал центральную предельную теорему. Суть ее заключается в том, что если какая-то величина есть результат сложения многих независимых случайных величин, она распределена нормально, не зависимо от того, по какому закону распределены составляющие ее слагаемые. В качестве примера можно привести время ожидания поезда метро. Каждый пассажир приходит независимо от других и имеет равную вероятность прождать любой отрезок времени между поездами. Такое распределение называется равномерным. Легко убедиться на примерах, что среднее время ожидания двумя случайными пассажирами уже не распределено равномерно, а имеет пик в середине интервала. Если же наблюдать за 6 или 7 пассажирами, то распределение их среднего времени ожидания практически не отличается от нормального.

Аналогичная картина складывается и при выполнении лабораторных анализов – если грубые погрешности исключены, а результаты не совпадают в силу мелких случайных причин: легкой липемичности, хлопьев белка или инородных частиц, пузырьков воздуха, загрязнения реактивов, мерцаний источника света, контаминации предыдущей пробой и т.д. , распределение результатов повторных исследований одного и того же материала должно подчиняться нормальному закону. Если этого нет, вероятнее всего, имеется какая-то доминирующая погрешность, которая может быть устранена.

Плотность нормального распределения задается функцией:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{x-M}{\sigma} \right)^2}$$

Вероятность, что случайная величина больше X_1 и меньше X_2 равна соответствующему участку площади под кривой, т.е. интегралу. Он называется функцией Лапласа, обозначается греческой буквой Φ , не может быть выражен в элементарных функциях, но опубликован в специальных таблицах. Обычно, когда говорят о нормальном распределении, имеют в виду одномерное нормальное распределение, это значит, что речь идет об одной случайной величине. График плотности ее распределения можно нарисовать на плоскости. Если же случайных величин две, например, выполнено два анализа одного и того же контрольного материала, распределение результатов описывается двумерным нормальным распределением, вид которого представлен на рис. 2.

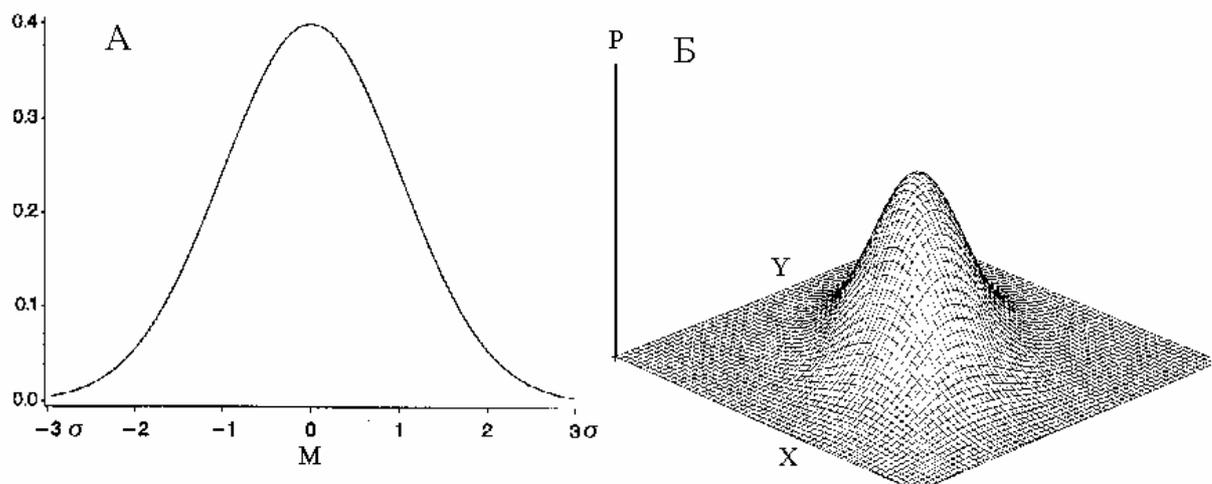


Рис.2 Плотности вероятностей одномерного (А) и двумерного (Б) нормальных распределений.

Это уже пространственная фигура, она показывает плотность вероятности всех возможных комбинаций обеих величин, в этом случае говорят об их совместном распределении. Аналогично, если случайных величин три речь идет о трехмерном нормальном распределении, в общем случае о многомерном распределении. Естественно встает вопрос как посчитать вероятности таких комплексных событий. Оказывается, что если все случайные величины распределены по нормальному закону с одинаковыми параметрами, т.е. являются разными экземплярами одной и той же нормально распределенной случайной величины, вероятность каждой их комбинации определяется суммой квадратов. Если известны среднее значение M и средняя квадратичная σ , которые должны получиться при анализе данного контрольного материала, то статистики говорят, что математическое ожидание генеральной совокупности всех правильно выполненных анализов M , а дисперсия σ^2 . Допустим, выполнено три анализа, результаты которых X_1 , X_2 и X_3 несколько отличаются от M . Чтобы определить, случайное это отличие или неслучайное, надо узнать совместную вероятность того, что X_1 , X_2 и X_3 взяты именно из генеральной совокупности всех правильно выполненных анализов. Если она больше чем 0,05 (или при более строгом подходе больше чем 0,01), считается что различие случайно (не значимо), и результаты контроля хорошие. В противном случае считается, что полученные результаты не являются случайной выборкой из совокупности всех правильно выполненных анализов. О вероятности судят по сумме квадратов отклонений, которая по традиции обозначается греческой буквой хи во второй степени χ^2 (читается хи-квадрат) и вычисляется по формуле:

$$\chi^2 = \frac{(X_1 - M)^2}{\sigma^2} + \frac{(X_2 - M)^2}{\sigma^2} + \frac{(X_3 - M)^2}{\sigma^2} \dots$$

Вычислив χ^2 по таблицам распределения хи-квадрат, находят искомую вероятность.

2.4. Погрешности

Никакое измерение не может быть абсолютно точным, всякое таит в себе возможность ошибки или погрешности, это одно и то же. Аналитик всегда имеет определенное мнение о вероятности ошибки, даже если он его четко не формулирует – кто бы стал выполнять анализ, если бы считал, что результат может быть любым, ошибка не предсказуема!

Разрабатывая научные методы статистического контроля качества, надо четко формулировать законы распределения погрешностей измерения. Весь опыт работы аналити-

ческих лабораторий говорит о том, что чаще всего результаты повторный анализ распределены по нормальному закону, а также что существует два типа погрешностей – внутри серии (быстрые) и между сериями (медленные). Погрешность внутри серии при каждом повторном измерении своя, она характеризуется дисперсией σ_r^2 и средней 0, погрешность между сериями характеризуется дисперсией σ_d^2 средней 0, внутри каждой серии она одинаковая. Реальная погрешность анализа есть сумма этих двух погрешностей.

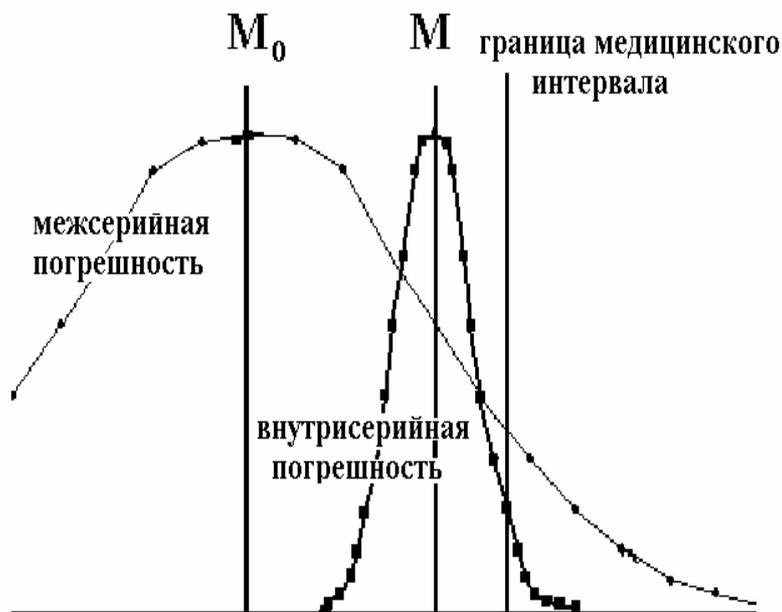


Рис.3 Межсерийная и внутрисерийная погрешности.

Конечно, никто не может утверждать, что погрешности медицинских лабораторных анализов всегда распределены по нормальному закону, тут возможны любые ситуации, поэтому контроль работы лаборатории должен начинаться с того, чтобы проверить, действительно ли результаты повторных анализов одного и того же материала распределены по нормальному закону. К сожалению, это сделать трудно – когда материала мало, заподозрить, что распределение ненормальное можно только, если оно явно асимметрично.

3. Контроль качества по данным анализов пациентов

Использование контрольных образцов для проверки качества работы всегда дорого – надо покупать сами образцы, требуются затраты на реактивы и работу, поэтому объем проверок неизбежно ограничен и возникает вопрос о статистической достоверности. Контроль по данным пациентов практически ничего не стоит – анализы ведь все равно выполняются, надо только иметь достаточно материала, компьютерную программу и умение. В большой лаборатории, которая ежедневно обслуживает сотню пациентов, контроль по их данным статистически очень достоверен, но не дает абсолютных величин и сопоставление результатов разных лабораторий (контингенты которых могут различаться) затруднено. Поэтому надо использовать оба варианта, разумно комбинируя их. Есть еще и промежуточный подход – так называемый метод «расщепленных» проб – когда несколько образцов проанализированного биологического материала, повторно исследуют на следующий день, это очень хороший способ оценить воспроизводимость результатов в разные дни.

Ниже описаны три способа использования данных пациентов для контроля качества.

3.1. Гистограммы

Просто и наглядно можно судить о качестве работы по данным пациентов строя гистограммы. В лаборатории, где ежемесячно выполняется несколько тысяч однотипных ис-

Иногда ошибки делят на случайные и систематические, Систематические называют также сдвиг и биос, понимая под этим различие между средней многократно повторенного анализа и истинным значением. Такое разделение неоправданно, так как систематическая ошибка непостоянна: она может увеличиваться или уменьшаться в зависимости от обстоятельств, поэтому тоже является случайной величиной. Правильнее говорить не о систематической, а о межсерийной ошибке.

следований, и данные находятся в компьютере, это легко сделать. Чтобы построить гистограмму, все результаты разбивают на разряды, желательно одинаковой ширины, подсчитывают число случаев в каждом из них и рисуют столбики. Число случаев в каждом разряде это частота событий, руководствуясь описанными выше общими правилами по ней можно вычислить доверительный интервал вероятности и отложить его на графике. Это облегчает сравнение данных. Такая гистограмма приведена на рис.4.

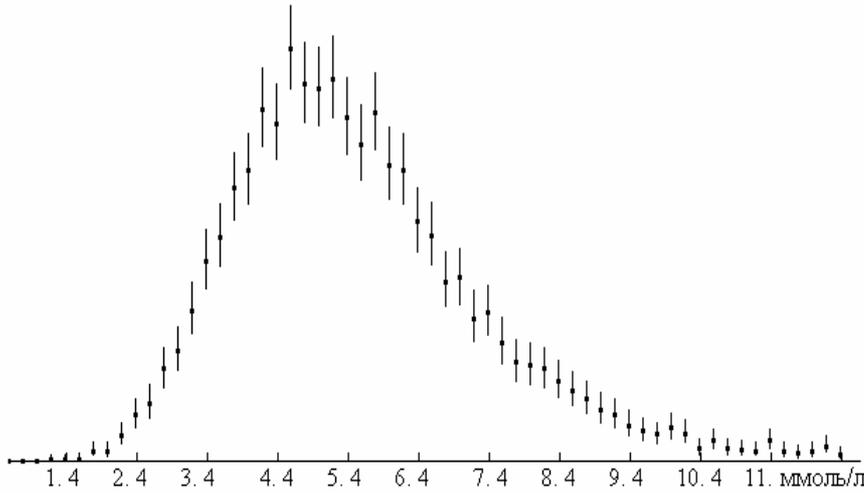


Рис.4 Гистограмма концентраций мочевины в крови, отложен 95% доверительный интервал, 6689 анализов, шаг 0,2 ммоль/л.

Самое сложное это удачно выбрать ширину разряда – если она широка – теряется информативность, если узка – в каждом разряде мало случаев и различия могут быть случайными. Чтобы обойти эту трудность можно заменить гистограмму функцией распределения (рис. 5), когда на графике откладывается не доля анализов, которые попали в данный диапазон

значений, а доля результатов, которые меньше данной величины. В отличие от гистограммы, функция распределения практически непрерывна, она получается более плавной,

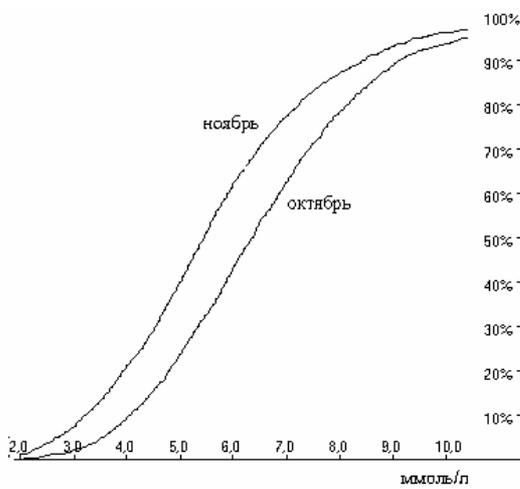


Рис. 5 Функция распределения концентрации мочевины крови. Те же данные на рис.7 и 8.

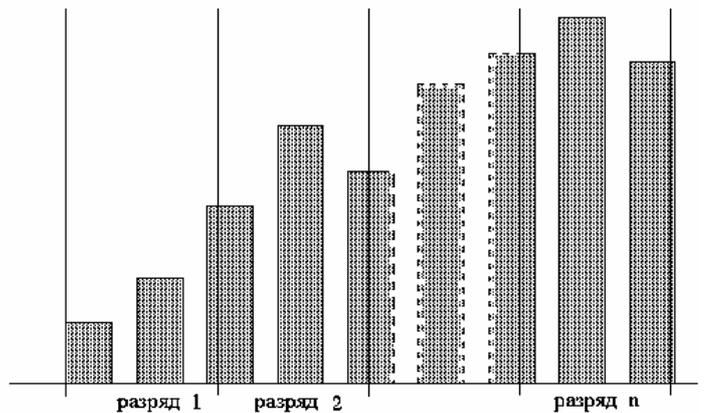


Рис. 6 Схема сглаженного разделения данных на разряды для построения гистограмм.

но различия не так бросаются в глаза.

Чтобы сделать гистограмму стандартной и «сглаженной» без потери информативности можно использовать следующий прием. Диапазоны разрядов выбираются так, чтобы в первые 15 разрядов попала половина результатов. Для этого находят такой результат анализа t , чтобы n_1 – число результатов меньше t , было по возможности равно n_3 – числу результатов, которые больше t . Понятно, что точно они редко совпадают, так как каждый результат – это дискретное число, которое встречается много раз. В том случае, когда результат анализа t встречается n_2 раз, мы делим n_2 на две части αn_2 и $(1-\alpha)n_2$ так, чтобы выполнилось равенство:

$$n_1 + \alpha n_2 = n_3 + (1-\alpha)n_2$$

Величину $m+\alpha$ считаем медианой множества, а ширину разрядов устанавливаем $(m+\alpha)/15,5$. Таким образом в разряд R попадают результаты, которые больше $(R-1)(m+\alpha)/15,5$ и меньше $R(m+\alpha)/15,5$. Те результаты, которые граница режет «по живому» делятся на две части, пропорционально участкам, которые отделяет граница.

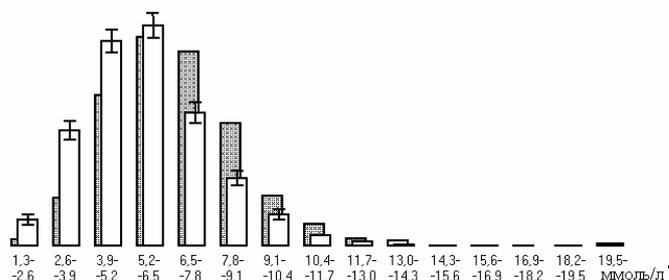


Рис. 7 Гистограммы мочевины крови в октябре (■ 3320 случаев) и ноябре (□ 4308 случаев). Отложен 95% доверительный интервал.

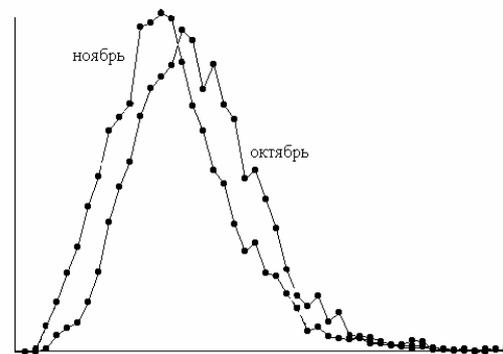


Рис. 8 Стандартные гистограммы мочевины крови в октябре (3320 случаев) и ноябре (4308 случаев).

На рисунках 7 и 8 приведены разные формы гистограмм тех же двух массивов, которые на рис. 5 представлены в виде функции распределения. Все три рисунка немного по-разному показывают одно и то же – в октябре все результаты были примерно на 1 ммоль/л выше, чем в ноябре. Наверяд ли это можно объяснить изменением контингента обследуемых, более вероятно, что прибор перекалибровали и новая калибровка отличается от старой. Гистограмма концентраций некоторых аналитов симметрична и по форме похожа на колоколообразную кривую нормального распределения, в других случаях она асимметрична с длинным хвостом в сторону больших величин. Это свойство самого аналита, которое определяется физиологией его обмена и не связано с качеством аналитической работы. Однако, если на гистограмме появляются пики или провалы, это плохой признак, вероятнее всего были существенные нарушения процесса анализа. Когда на одном графике отложены гистограммы за 2-3 месяца, сразу видно насколько стабильны результаты. Конечно, кривая распределения может сдвинуться и в результате изменения контингента больных, но это бывает редко и несложно проконтролировать. Если же гистограмма достоверно не меняется, это уже важный признак хорошей работы.

3.2. Критерий лямбда (Колмогорова-Смирнова)

Характер распределения результатов количественно оценивают по критерию Колмогорова-Смирнова. Этот критерий позволяет сравнить две совокупности данных и количественно оценить вероятность того, что различия между их распределениями носят не случайный характер. Принцип критерия следующий – для каждого R – возможного значения исследуемого показателя, подсчитывается частота – доля числа случаев, когда результат оказался меньше R . В первой группе частота обозначается $fR1$, во второй $fR2$. Выбирается та пара значений $fR1$ и $fR2$ различие между которыми максимально по абсолютной величине $f_{\max} = \max |fR1 - fR2|$. Критерий Колмогорова-Смирнова (его называют также критерием λ) вычисляется по формуле:

$$\lambda = f_{\max} \sqrt{\frac{n_1 n_2}{n_1 + n_2}}$$

Здесь n_1 – общее число анализов в первой совокупности, а n_2 во второй. Если величина $\lambda > 1.36$ то вероятность того, что различие между распределениями случайно 0.05 или меньше, если $\lambda > 1.63$ эта вероятность 0.01 или меньше. Соответственно, если $\lambda < 1.36$ мож-

но быть уверенным, что характер распределения достоверно не изменился, это говорит о том, что работа выполнена хорошо. Если же он больше 1.36, а тем более >1.63 надо искать причину – она может быть вызвана изменением контингента больных, но вероятнее всего аналитическими погрешностями. Из формулы видно, что чем больше сравниваемые совокупности, тем легче получить большое значение λ и тем самым доказать различие между ними. Когда данных мало, небольшая величина критерия, еще не означает, что различия нет, просто у нас нет материала, чтобы доказать, что оно есть. Опыт показывает, что критерий Колмогорова-Смирнова хорошо работает, когда в наименьшей из сравниваемых совокупностей порядка 100 результатов анализов. В практической работе выбирается какой-то достаточно длинный период, например месяц, который принимается за эталон, с ним сравниваются данные каждого рабочего дня. Можно сравнивать текущий рабочий день и с объединенными данными предыдущей недели.

По замыслу использование критерия Колмогорова близко к описанному в литературе методу «средних величин нормальных результатов» (Average of Normals, AON), поскольку и в том и в другом случае оценивается вероятность получить данную совокупность результатов. Но, используя критерий Колмогорова, мы непосредственно судим о вероятности события по его частоте, а в методе АОН косвенно – по средней величине, которая очень зависит от нескольких крайних значений. Для критерия Колмогорова-Смирнова это не имеет значения, так как учитываются не сами величины, а число случаев, когда результат превышает определенный уровень, а на сколько он его превышает не имеет значения.

Критерий лямбда считается мало чувствительным, но в нашем случае это достоинство – если уж он показал различие, значит, есть о чем задуматься.

3.3. Дисперсионный анализ.

Как уже отмечалось, основной «враг хорошей работы» – это невоспроизводимость в разные дни, которая вызвана тем, что величина так называемой систематической ошибки, т.е. по существу межсерийной погрешности, меняется. Чаще всего это происходит из-за различий в качестве реактивов или изменения температурного режима. Если лаборатория знает пределы этих колебаний, она тем самым знает, с какой точностью работает. В лаборатории, где выполняется много анализов, этот параметр можно оценить методами теории вероятностей, конкретно дисперсионного анализа, используя его в двух вариантах – оценки дисперсии среднего и автокорреляционной функции.

3.3.1 Оценка дисперсии среднего

Известно, что оценка средней, или как ее еще называют, выборочная средняя – \bar{X} сама является случайной величиной, которая имеет среднее квадратичное отклонение $m = S / \sqrt{n}$, где n число случаев, в выборке, а S среднее квадратичное отклонение всего массива данных. Если мы располагаем большим (порядка тысячи и больше) результатов анализов, мы можем посчитать m двумя разными способами – во-первых, по приведенной формуле, а кроме того непосредственно разбив всю последовательность результатов на группы, в каждой из которых n анализов, выбрав n так, чтобы оно примерно соответствовало числу анализов, данного вида, за один день. Для каждой группы подсчитывают среднее и среднее квадратичное отклонение средних, которое обозначают m^* . Опыт показывает, что, как правило, m^* значительно больше, чем m , это говорит о том, что систематическая погрешность все время колеблется, причем эти колебания нельзя объяснить случайностью – вся случайность находится в m . Для проверки разбивают тот же материал на группы другим способом – в первую относят первый анализ, $n+1, 2n+1, 3n+1 \dots$, во вторую анализы $2, n+2, 2n+2, 3n+2 \dots$ и т.д. Для этих групп делают те же расчеты, в этом случае обычно оказывается, что фактическая ошибка среднего совпадает с вычисленной

по формуле. Дальнейшие рассуждения просты – достоверность различия между m^* и m проверяется известным методом хи-квадрат (по критерию Фишера), и если оно достоверно, то вычисляется средняя квадратичная ежедневной систематической погрешности по формуле:

$$E_{\text{день}} = \sqrt{m^{*2} - m^2}$$

$E_{\text{день}}$ характеризует ту точность, которую фактически удалось получить в данной лаборатории в данный период. В силу центральной предельной теоремы теории вероятностей $E_{\text{день}}$ распределена по нормальному закону, поэтому с 95% вероятностью систематическая погрешность определения не превышает $1.96E_{\text{день}}$. Практически это означает, что различия между результатами анализов достоверны только, если они превышают $2E_{\text{день}}$, в противном случае они могут зависеть не от динамики заболевания, а от неточности анализов. Теоретически $E_{\text{день}}$ равна средней квадратичной межсерийной погрешности σ_d , вычисленной по воспроизводимости в разные дни, но надо иметь в виду, что обе эти величины случайные, поэтому точно совпадают только их математические ожидания.

3.3.2 Автокорреляционная функция.

Информацию о точности работы аналитической системы можно получить также исследуя автокорреляционную функцию. В самом деле, пусть δ означает межсерийную погрешность измерения, общую для всех анализов данной серии, независимо от состава исследуемого материала, ε индивидуальное внутрисерийное (случайное) отклонение, а \bar{X} среднюю величину пула данных, состоящего из нескольких серий измерений. Тогда каждый результат анализа

$$X = \bar{X} + \delta + \varepsilon$$

Заметим, что ε значительно больше, чем δ , так как отражает не только и не столько методическую ошибку анализа, сколько индивидуальный состав исследуемой пробы, а межсерийная погрешность δ это так называемая систематическая ошибка измерения, которую называют также сдвигом или биосом. Результаты двух последовательно выполненных анализов с номерами $2n-1$ и $2n$

$$X_{2n-1} = \bar{X} + \delta_{2n} + \varepsilon_{2n-1} \quad \text{и} \quad X_{2n} = \bar{X} + \delta_{2n} + \varepsilon_{2n}$$

Введем новую переменную Y , которая есть произведение отклонений от средней величины результатов двух последовательно выполненных анализов:

$$Y_n = (X_{2n-1} - \bar{X})(X_{2n} - \bar{X}) = (\delta_{2n} + \varepsilon_{2n-1})(\delta_{2n} + \varepsilon_{2n}) = \delta_{2n}^2 + \delta_{2n}(\varepsilon_{2n-1} + \varepsilon_{2n}) + \varepsilon_{2n-1}\varepsilon_{2n}$$

По условиям задачи ε_{2n-1} и ε_{2n} независимые случайные величины, которые с равной вероятностью могут быть и положительными и отрицательными, поэтому математическое ожидание суммы

$$M\left(\frac{\sum Y}{n}\right) = M\left(\frac{Y_1 + Y_2 + \dots + Y_n}{n}\right) =$$

$$M\left(\frac{(X_1 - \bar{X})(X_2 - \bar{X}) + (X_3 - \bar{X})(X_4 - \bar{X}) + \dots + (X_{2n-1} - \bar{X})(X_{2n} - \bar{X})}{n}\right) =$$

$$M\left(\frac{\delta_1^2 + \delta_2^2 + \dots + \delta_n^2}{n}\right) = M\left(\frac{\sum \delta^2}{n}\right) = \sigma_d^2 \quad \text{это дисперсия межсерийной погрешности, величина,}$$

которая лучше всего характеризует точность измерения.

Однако, выполнив ограниченное число анализов, мы получаем не математическое ожидание, а всего лишь его выборочную оценку $\bar{Y} = \frac{\sum \delta^2}{n}$, поэтому еще надо понять насколько \bar{Y} отличается от σ_d^2 . Ответ дает оценка дисперсии средней величины \bar{Y} , которую в англоязычной литературе называют стандартным отклонением средней и обычно обозначают m^2 , она равна дисперсии, деленной на число наблюдений в выборке:

$m^2 = \frac{\sigma^2}{n}$. Дисперсия также неизвестна, как и математическое ожидание, но, учитывая большое число наблюдений, можно заменить ее выборочной оценкой, считая выборкой всю сумму наблюдений. Вычисляют по общепринятой формуле

$$S_y^2 = \frac{\sum Y^2 - \frac{(\sum Y)^2}{n}}{n-1} \quad \text{и} \quad m_y = \sqrt{\frac{S_y^2}{n}} \bar{Y}$$

Теперь можно узнать доверительный интервал для σ_d^2 вокруг \bar{Y} . Согласно общепринятым представлениям, математическое ожидание Y , т.е. средняя квадратичная межсерийной погрешности, с вероятностью 95% находится в диапазоне $\bar{Y} \pm 1,96m_y$. Практически это означает, что число наблюдений должно быть таким, чтобы $m_y \ll \bar{Y}$.

4. Контроль качества по контрольным пробам

Контроль качества лабораторных анализов по данным пациентов дешев, информативен и легко выполним (при наличии компьютера), но его возможности ограничены сравнением качества работы одной и той же лаборатории в разные периоды времени. Чтобы оценить абсолютную погрешность измерения, надо либо принять, как это предложил М.И.Прищепа, что средняя величина всех анализов данного вида, выполненных в разных лабораториях, постоянна, либо анализировать контрольные образцы. Идея М.И.Прищепы очень перспективна, но может быть реализована только при определенных обстоятельствах, сейчас она не рассматривается.

Идеология контроля качества по контрольным образцам внешне очень проста – оценивается точность (неопределенность) результатов анализов проб, для которых установленные значения известны, полученная оценка без всяких поправок переносится на пробы пациентов. Однако, за внешней простотой скрывается много проблем – какое различие допустимо, как обсчитывать и интерпретировать результаты контролей, какие показатели использовать. Точный ответ требует глубокого научного анализа и громоздкой математической обработки результатов контрольных анализов, что, однако, вполне по силам оснащенной компьютером лаборатории.

Важно понимать, что методов обработки и интерпретации результатов контрольных анализов очень много, некоторые из них описаны ниже, но это не исчерпывает все возможности. Вопрос о том, каким или какими реально пользоваться должна решать сама лаборатория. Все зависит от имеющейся информации о природе аналитических ошибок, точнее, какие усилия можно предпринять, чтобы эту информацию получить. Ведь ошибку можно найти только если она появляется закономерно. Если же она может быть какой угодно и когда угодно, аналитическая работа вообще невозможна.

Все изложенные ниже способы и методы обработки и интерпретации контролей основаны на допущении, что результаты повторных анализов одного и того же материала (неважно пациента или контрольного) нормально распределены вокруг истинного (или установленного) значения. Если параметры этого распределения неизвестны используется статистика Стьюдента, если известна внутрисерийная дисперсия, а межсерийная неизвестна подходит Байесова статистика, если известны и внутрисерийная и межсерийная дис-

персии можно использовать наиболее точный метод совместного распределения контрольных и опытных проб.

Возможна и другая классификация – в одних случаях мы просто говорим «да» или «нет» выполняются требования или нет, грубо говоря, оцениваем качество контрольных проб. В других случаях исследуются параметры распределения и дается развернутый ответ в виде вероятности того, что погрешность не превышает заданную, или же ширины интервала в котором находится 95% ошибок.

4.1. Методы, основанные на оценке качества контрольных проб

4.1.1. Метод, основанный на распределении хи-квадрат

Самый простой путь основан на допущении, что результаты правильно выполненных контрольных анализов образуют множество, распределенное по нормальному закону с математическим ожиданием μ и дисперсией σ^2 . При этом μ совпадает с установленной величиной контрольного материала, а дисперсия устанавливается каким-либо нормативным документом, например, инструкцией изготовителя или внутрилабораторными правилами качественной работы. Она задает величину медицинского доверительного интервала, т.е. желаемый уровень точности. Понятно, что фактическая дисперсия зависит в первую очередь от используемого метода и возможностей аппаратуры и реактивов, но также и от качества контрольного материала. Ведь он поступает в лабораторию в виде серии контейнеров, для каждого повторного анализа вскрывается новый контейнер, и его содержимое не должно отличаться от предыдущих. Если это не так, например, потому что скорость разрушения материала при хранении разная, дисперсия возрастает. В этом случае вопрос о том, какое отклонения результата контрольного анализа от ожидаемого допустимо, сводится к вопросу о вероятности того, что полученные результаты контролей есть случайная выборка из генеральной совокупности результатов всех правильно выполненных анализов.

В такой постановке это классическая задача теории вероятностей, решение которой хорошо известно. Строго доказать, что выборка сделана именно из данной совокупности невозможно, можно только оценить вероятность, что она сделана из нее. Если такая вероятность велика, это мало что доказывает, потому что есть много совокупностей, из которых данная выборка может быть сделана с высокой вероятностью. Зато обратное положение куда более четкое – зная закон распределения, можно уверенно сказать, что выборка не отсюда, потому что вероятность этого мала. Таким образом, мы можем отсеять заведомо отрицательные результаты, но не можем гарантировать положительный.

Насколько это важно? Следует ли добиваться, чтобы контрольные анализы были абсолютно идеальными, или же можно допустить какие-либо отклонения? Ответ дают более сложные методы вычислений, о которых речь пойдет ниже. Распределение же хи-квадрат и близкое ему распределение Релея, позволяя отвергнуть заведомые ошибки, в то же время дают повод надеяться, что все в порядке.

Если выполнен только один контрольный анализ, то по закону нормального распределения, можно определить вероятность того, что результат относится именно к данному множеству. Однако, если выполнено два или больше контрольных анализов, то суммарная вероятность события оценивается не по распределению Гаусса, а по распределению χ^2 (хи- квадрат):

$$\chi^2 = \sum \left(\frac{X_n - \mu_n}{\sigma_n} \right)^2$$

Здесь X_n – это результат контрольного анализа. Понятно, что если используется несколько разных контрольных материалов, то в каждом случае вычитается свое установленное значение. Сложнее с дисперсией, трудно понять, почему для одних контрольных материалов она должна быть меньше или больше, чем для других. Принципиальных возражений против того, чтобы дисперсии различались нет, но интерпретация затруднительна.

В специальных таблицах приводятся значения χ^2 в зависимости от вероятности p и числа степеней свободы. Надо подчеркнуть, что в формуле μ и σ – математические ожидания, которые нам задаются как параметры контрольного материала, поэтому число степеней свободы равно числу наблюдений (контрольных анализов). По таблице находим вероятность того, что контрольные анализы действительно являются случайной выборкой из генеральной совокупности всех правильных ответов. Если эта вероятность меньше 0,05, а тем более меньше 0,01, делается вывод, что результаты неудовлетворительные, контрольные анализы не могут подтвердить что анализы выполняются качественно.

Надо ясно представлять себе, что критерий хи-квадрат, как и критерий Колмогорова-Смирнова, не дает непосредственной информации о погрешности измерения. Располагая этими критериями, мы ничего не можем сказать о том, куда попадет результат следующего, еще не выполненного анализа. Поэтому эти подходы косвенные. Они опираются примерно на такое рассуждение: «Если предыдущий (контрольный) анализ оказался удачным, то, наверное, и следующий (анализ пациенту) тоже будет удачным». Но никаких количественных оценок такое рассуждение не содержит. Мы ничего не говорим о том, какая аналитическая погрешность (а какая-то погрешность неизбежна), с какой вероятностью приведет к тому, что результат контрольного анализа нас не устроит. Мы узнаем только вероятность того, что сегодняшние результаты контрольных анализов являются случайной выборкой из множества всех идеально выполненных анализов данного материала. Это ограниченная, но точная и в научном отношении безупречная информация.

4.1.2. Метод, основанный на распределении Релея

В клиничко-диагностических лабораториях часто в одной и той же аналитической серии выполняют два контрольных анализа – один с концентрацией аналита, характерной для здорового человека, другой для больного, на лабораторном жаргоне их называют – «нормальный» и «патологический» контроли. Используя аппарат, разработанный Релеем (распределение Релея) результаты можно представлены графически в виде эллипса рассеяния. По существу это частный случай распределения χ^2 , с той разницей, что речь идет не о двух экземплярах одной и той же нормально распределенной случайной величины (когда математические ожидания и дисперсии одинаковы), а о двух разных величинах с разными математическими ожиданиями и дисперсиями. На оси абсцисс откладывается отклонение одного результата от заданного, а на оси ординат другого. Релей показал, что если обе величины x и y распределены по нормальному закону, то вероятность попадания в эллипс полуоси которого $x=k\sigma_x$ и $y=k\sigma_y$ определяется выражением:

$$P(x, y) = 1 - e^{-\frac{k^2}{2}}$$

В нашем случае дисперсии обоих контрольных анализов одинаковы $\sigma_x = \sigma_y$, поэтому эллипс превращается в круг: $P(x, y) = 1 - e^{-\frac{1}{2}\left(\frac{r}{\sigma}\right)^2}$. Попадание точки, координаты которой равны результатам контрольных анализов, в круг радиусом $2,445$ заданного среднего квадратичного отклонения говорит о том, что с вероятностью $0,95$ данная комбинация контрольных анализов относится к совокупности идеально выполненных исследований. Если же она попала в круг радиусом $3,035\sigma$, то вероятность составляет $0,99$. В практическом отношении важнее обратные величины: если аналитическая система работает иде-

ально, то вероятность попасть за пределы внутреннего круга составляет 5%, а за пределы внешнего 1%. Это очень наглядно – сразу видно попала точка в круг или нет, если она за пределами внутреннего круга результат сомнителен, если за пределами внешнего откровенно плохой.

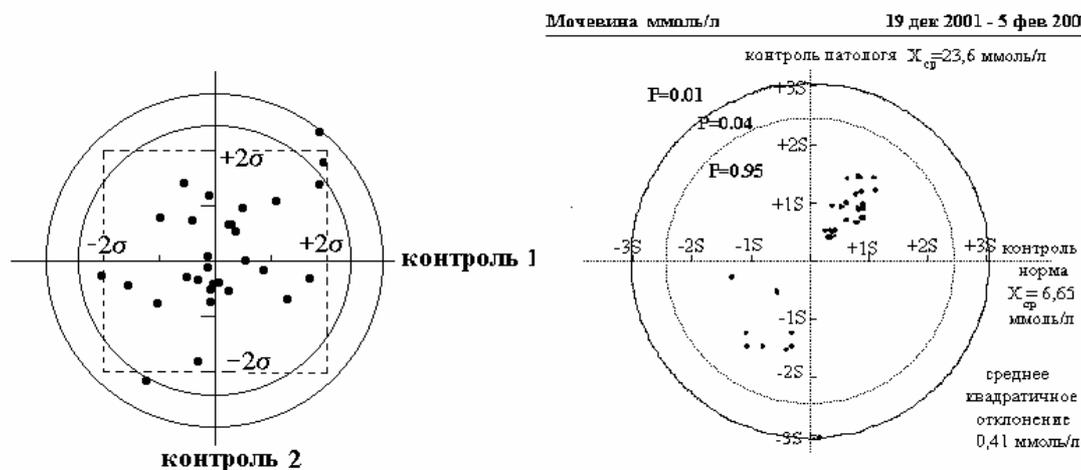


Рис. 9 График распределения Релея. Радиус внутренней окружности $2,45\sigma$, внешний $3,03\sigma$.

Рис.10 Образец циркулярного графика, который выдает программа.

На рис. 9 представлен такой круговой график для случая, когда результаты не очень хорошие, но приемлемые. В центре пунктирной линией намечен квадрат со сторонами 2σ . Если бы оценке делалась на основе одномерного Гауссова распределения (как это делается, например, в контрольных правилах Вестгарда), хорошим результатом следовало бы считать попадание точки в этот квадрат. Его контур не совпадает с кругом, поэтому ошибка в оценке может быть как в ту, так и в другую сторону.

На графике распределения Релея визуально хорошо прослеживается корреляция между результатами двух контрольных исследований разных материалов. Если она бросается в глаза, это важное свидетельство того, что межсерийная погрешность действительно меняется день ото дня и намного превышает внутрисерийную. Чем равномерней и плотней расположены точки вокруг центра, тем выше качество анализов.

4.2. Методы, основанные на оценке параметров распределения.

Рассмотренные выше методы оценки неопределенности результатов лабораторных анализов не позволяют количественно оценить вероятность попадания результата анализа пациенту в медицинский интервал вокруг истинного значения, возможность получить прямой ответ зависит от того, что известно о свойствах аналитической системы.

Как уже отмечалось, всякий аналитик имеет какое-то, пусть не формализованное, мнение о возможных погрешностях, ниже будут рассмотрены два варианта:

1) известно только, что погрешности распределены по нормальному закону, но параметры этого распределения неизвестны, в этом случае надо опираться на распределение Стьюдента,

2) известны также дисперсии внутрисерийной (σ_r^2) и межсерийной погрешностей (σ_d^2), вероятность попадания может быть оценена по совместному распределению контрольных и опытных проб.

4.2.1 Распределение Стьюдента

В биометрике часто используют так называемое распределение Стьюдента, чтобы найти доверительный интервал для математического ожидания в том случае, когда известны только оценки параметров нормального распределения. На самом деле термин

распределение в данном случае не очень удачный, скорее речь идет о преобразовании, которое сводится к замене случайной величины в неслучайных координатах на неслучайную величину, координата которой является случайным числом. Такое преобразование значительно упрощает поиск интервала, в котором с заданной вероятностью находится нормально распределенная случайная величина, если известны только оценки параметров ее распределения. Хотя, изложенное ниже формально не является частным случаем распределения Стьюдента, но метод преобразования тот же самый, поэтому не необходимости отказываться от привычного термина.

4.2.1.1 Два контрольных анализа.

Очень часто в каждой аналитической серии ставят по два контрольных анализа одного или разных материалов известного состава, и при этом получают две случайные величины A и B – разности между фактическими результатами и должными (аттестованными) значениями. Если есть основания считать, что эти погрешности распределены нормально, то, используя статистику Стьюдента, можно оценить вероятную ошибку при анализе пробы пациента.

Погрешности A и B , можно рассматривать как суммы $A = \mu + \sigma\eta_A$ и $B = \mu + \sigma\eta_B$, где μ и σ математическое ожидание (среднее) и среднее квадратичное отклонение, η_A и η_B стандартные формы нормально распределенных величин, когда среднее 0, а средняя квадратичная 1. Ошибка анализа пробы пациента X также может быть представлена в форме суммы $X = \mu + \sigma\eta_X$. В данном случае предполагается, что все три анализа выполнены одновременно, поэтому средние (μ) и средние квадратичные (σ) одни и те же. Задача заключается в том, чтобы определить диапазон в котором находится X – возможная погрешность анализа пробы пациента. Для этого делим разность между результатом анализа пациенту и полусуммой контролей на разность контролей:

$$\frac{X - \frac{A+B}{2}}{A-B} = \frac{\mu + \sigma\eta_X - \frac{\mu + \sigma\eta_A + \mu + \sigma\eta_B}{2}}{\mu + \sigma\eta_A - \mu - \sigma\eta_B} = \frac{\eta_X - \frac{\eta_A + \eta_B}{2}}{\eta_A - \eta_B} = \frac{r}{q} = t$$

Получилась безразмерная случайная величина t , которая не содержит конкретных параметров данной серии анализов, а зависит только от комбинации трех случайных чисел η_A , η_B и η_X

В то же время $X = t(A-B) + \frac{A+B}{2} = f_x(t)$ ошибка анализа пациенту, тесно связана с

t . Чем чаще t принимает то или иное значение, тем чаще возникает соответствующая по величине ошибка. Это связано с тем, что X монотонная и однозначная функция t , поэтому их функции распределения совпадают. Если t меньше некоторого значения t_k , то и $X = f_x(t)$ меньше соответствующего $X_k = f_x(t_k)$. Узнав закон распределения t , мы одновременно имеем и закон распределения X . Закон распределения t можно получить следующим образом.

Величина $t = \frac{r}{q}$ это частное от деления

$$r = \eta_X - \frac{\eta_A + \eta_B}{2} \text{ на } q = \eta_A - \eta_B.$$

Вероятность плотности ее распределения можно представить себе как плоскость где на одной оси отложен q , а на другой r . Все точки, которые соответствуют одному и тому же значению t_k , лежат на прямой T , проходящей через начало координат (см. рис.). Точки, соответствующие $t < t_k$ лежат правее и ниже линии T . Плотность вероятности функции

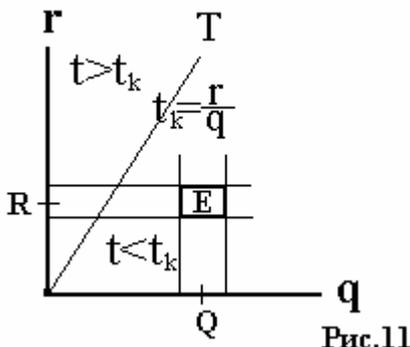


Рис.11

$t = \frac{r}{q}$ в точке QR есть произведение плотностей вероятностей q на r , поэтому вероятность попадания точки в элемент вероятности E

$$F_e(Q,R) = f_q(Q)f_r(R)\Delta q\Delta r$$

Знаменатель q это разность двух случайных величин с дисперсиями 1, поэтому их разность имеет дисперсию 2, среднее значение 0 и плотность вероятности

$$f_q(q) = \frac{1}{2\sqrt{\pi}} e^{-\frac{q^2}{4}}$$

Числитель r это сумма случайной величины η_X с дисперсией 1 и случайной величины $(\eta_A + \eta_B)/2$ с дисперсией 0,5, суммарная дисперсия 1,5, среднее 0 и плотность вероятности

$$f_r(r) = \frac{1}{\sqrt{3\pi}} e^{-\frac{r^2}{3}}$$

После подстановки получаем вероятность попадания в элемент E

$$F_e(Q,R) = \frac{1}{2\pi\sqrt{3}} e^{-\frac{q^2}{4} - \frac{r^2}{3}} \Delta q\Delta r$$

Если плоскость разбить на большое число квадратов со сторонами Δq и Δr , то суммируя вероятности попадания в каждый из них, можно посчитать вероятность того, что t меньше заданной величины t_k , и, одновременно, $X < X_k = t_k(A - B) + (A + B)/2$. Вычислив один раз t_k для обычно используемых вероятностей 0,99; 0,95 и 0,90, всегда можно рассчитать границы интервала в котором с заданной вероятностью находится ошибка измерения пробы пациента.

Практически всю плоскость удобно разбить на 10000 квадратов (100x100) с шагом 0,1 и провести вычисления с помощью компьютера. Критерием правильности вычисления служит сумма для всего большого квадрата, которая должна быть 0,25, так как рассматривается только четверть всего пространства. По условиям задачи и разность $(A-B)$ и сумма $(A+B)$ могут быть как положительными, так и отрицательными, а выше был рассмотрен только тот вариант, когда оба положительные.

Практически удобно всегда пользоваться $t=10$, которая соответствует вероятности 0,945. Получается очень простое правило:

94,5% ошибок лежит в диапазоне	
от $-10 A - B + (A + B)/2$	до $+10 A - B + (A + B)/2$

здесь $|A-B|$ абсолютная величина разности погрешностей контрольных анализов.

Как видно, доверительный интервал при двух контрольных анализах весьма протяженный. Это связано с тем, что мы имеем мало информации о природе ошибок и вынуждены растянуть доверительный интервал до такого размера, чтобы уверенно в него уложиться. Если же выполнять больше контрольных анализов, объем информации возрастет и можно точнее локализовать погрешности, поэтому величина t для той же доверительной информации будет меньше. Но это требует несколько иных вычислений о которых в следующем разделе.

Надо заметить, что описанный выше подход, когда точность работы оценивается по размеру интервала, в который ошибки попадают со стандартной вероятностью, удобен тем, что вычисления очень простые. Противоположный же путь – когда задан интервал и требуется определить вероятность попадания в него ошибки, также имеет право на существование, но требует сложных вычислений, поэтому может быть выполнен только специальной компьютерной программой. Он хорош тем, что позволяет сравнивать разные методы оценки точности, которые различаются по объему информации о природе ошибок.

Принцип вычисления сводятся к тому, что сначала находят величины t_n и t_k , соответствующие X_n и X_k - началу и концу доверительного интервала, а затем «по квадратикам» подсчитывают вероятность.

4.2.1.2 Больше чем два контрольных анализа.

Принцип тот же, что и для двух контролей, но вычисления сложнее. Как и в случае двух контрольных анализов, выражаем погрешности контролей $X_c = \mu + \sigma\eta$ в виде суммы, n общее число контролей. Соответственно для пробы пациента $X_p = \mu + \sigma\eta_p$.

В этом случае выборочные среднее и среднее квадратичное погрешностей контролей:

$$\bar{X}_c = \mu + \sigma \frac{\sum \eta}{n} \qquad S = \sigma \sqrt{\frac{\sum \eta^2 - \frac{(\sum \eta)^2}{n}}{n-1}}$$

Посмотрим как далеко от выборочного среднего \bar{X}_c находится ошибка анализа пациенту X_p . В качестве масштаба возьмем выборочную оценку среднего квадратичного S . Делим $X_p - \bar{X}_c$ на S и подставляем формулы выборочных среднего и среднего квадратичного :

$$\begin{aligned} \frac{X_p - \bar{X}_c}{S} &= \sqrt{n-1} \frac{\eta_p - \frac{\sum \eta_i}{n}}{\sqrt{\sum \eta_i^2 - \frac{(\sum \eta_i)^2}{n}}} = \sqrt{\frac{n-1}{n}} \frac{n\eta_p - \sum \eta_i}{\sqrt{n \sum \eta_i^2 - (\sum \eta_i)^2}} = \\ &= \sqrt{\frac{n-1}{n}} \frac{n\eta_p - \sum_1^n \eta_i}{\sqrt{\sum_{i \neq j}^n (\eta_i - \eta_j)^2}} = \sqrt{\frac{n-1}{n}} \frac{n\eta_p - \sum_1^n \eta_i}{\sqrt{\sum_{j=1}^m \omega_j^2}} = t \sqrt{\frac{n-1}{n}} \end{aligned}$$

здесь $m = n^2 - n = n(n-1)$, а ω слу-

чайные нормально распределенные числа с дисперсией 2 и средней 0.

$$t = \frac{n\eta_p - \sum_1^n \eta_i}{\sqrt{\sum_{j=1}^{j=m} \omega_j^2}} = \frac{r}{q}$$

безразмерная случайная величина, которая принимает разные

случайные значения в зависимости от комбинации $m+1$ случайных чисел. Так как $X_p = \bar{X}_c \pm tS \sqrt{\frac{n-1}{n}} = f(t)$ непрерывная неубывающая функция, $t < t_k$ с такой же вероятностью, как и $X = f(t) < X_k = f(t_k)$. Найдя закон распределения t , тем самым находим закон для X_p .

Плотность распределения t можно представить как двумерную функцию, откладываем на оси абсцисс $q = \sqrt{\sum_m \omega^2}$, а на оси ординат $r = n\eta_p - \sum_n \eta$.

r это случайная нормально распределенная величина со средней $\mu=0$ и дисперсией $\sigma^2 = n^2 + n = n(n+1)$, плотность ее распределения $f_r(r) = \frac{1}{\sqrt{2\pi n(n+1)}} e^{-\frac{r^2}{n(n+1)}}$.

q также случайная величина плотность распределения которой можно найти используя распределение χ^2 . Для этого вводим новую переменную $v = \frac{\sum \omega^2}{D} = \frac{q^2}{2}$ где $\sum \omega^2$ сумма квадратов случайных нормально распределенных чисел с математическим ожиданием 0 и дисперсией $D=2$. Известно, что если $F(v)$ функция распределения v , то ее производная $f_v(v) = d \frac{F(v)}{dv}$ называется плотность распределения и описывается законом χ^2 :

$$f_v(v) = \frac{v^{\frac{m}{2}-1} e^{-\frac{v}{2}}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)}$$

где m число степеней свободы, а Γ гамма функция.

После замены переменных, учитывая, что $\frac{dv}{dq} = q$ получаем плотность распределения q :

$$f_q(q) = f_v(v) \frac{dv}{dq} = \frac{q^{m-1} e^{-\frac{q^2}{4}}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)}$$

Также, как в случае двух контрольных анализов, разбиваем плоскость на элементы вероятности - квадратики со сторонами $\Delta q \Delta r$. Все точки, которые соответствуют одному и тому же значению t_k , лежат на прямой T , проходящей через начало координат (см. рис.). Точки, соответствующие $t < t_k$ лежат правее и ниже линии T . Плотность вероятности функции $t = \frac{r}{q}$ в точке QR есть произведение плотностей вероятностей q на r , поэтому вероятность попадания точки в элемент вероятности E

$$f_q(q)f_r(r)\Delta q\Delta r = \frac{q^{m-1}}{2^{\frac{m}{2}} \Gamma\left(\frac{m}{2}\right)} e^{-\frac{q^2}{4}} \frac{1}{\sqrt{2\pi(n+1)}} e^{-\frac{r^2}{2n(n+1)}} \Delta q\Delta r$$

Разбив плоскость на достаточное число квадратиков – в данном случае 200x100, и суммируя вероятности попадания в каждый из них, узнаем суммарную вероятность $P\{t < t_k\}$ того, что t меньше выбранного нами значения t_k (разумеется, $0 < t_k < 1$). Методом проб и ошибок подбираем такие t_k , чтобы $t_k \sqrt{(n-1)/n}$ соответствовало общепринятым стандартным вероятностям 0,99; 0,95 и 0,90. Эти величины приведены в таблице, сравнительно с аналогичными вероятностями для двух контрольных проб и нормальным распределением.

Величины $t_k \sqrt{(n-1)/n}$ для разных уровней доверительной вероятности и числа контрольных анализов, сравнительно с данными для двух анализов и нормального распределения

Доверительная вероятность	Число контрольных анализов				Нормальное распределение
	2	3	4	5	
0,99	>48	8,00	4,18	2,90	2,58
0,95	10,90	4,43	2,86	2,18	1,96
0,90	5,48	3,28	1,84	1,84	1,65

Поскольку по условиям задачи вероятность X_p ошибки при анализе пробы пациента непрерывная и неубывающая функция t , если $t < t_k$, то и соответствующие им вероятности

ошибок находятся в тех же отношениях $X(t) < X_k(t_k)$. Границы интервала в который с заданной вероятностью попадает погрешность анализа, рассчитываются по формуле:

Нижняя граница	$\bar{X}_c - S t_k \sqrt{(n-1)/n}$
Верхняя граница	$\bar{X}_c + S t_k \sqrt{(n-1)/n}$
\bar{X}_c - средняя различий между результатами анализов контролей и их аттестованными значениями, S - средняя квадратичная этих различий, $t_k \sqrt{n(n+1)}$ - коэффициент из таблицы.	

Как видно, при увеличении числа контрольных анализов t быстро уменьшается, соответственно резко сокращается доверительный интервал. Это не удивительно, так как чем больше мы знаем о природе погрешности, тем точнее мы можем ее локализовать. Отсюда следует практическое правило – если при выполнении двух контрольных анализов получился небольшой доверительный интервал, который целиком покрывается принятым в денной лаборатории для данной методики допустимым медицинским интервалом, можно быть уверенным, что работа выполнена качественно. Ведь доверительный интервал при двух контролях устанавливается с большим запасом. Если же обратная ситуация – доверительный интервал при двух контролях недопустимо велик, это еще не значит, что работа выполнена плохо надо увеличить число контролей и вычислить среднее и среднее квадратичное для всей их совокупности. Даже если получатся величины близкие к прежним, все равно доверительный интервал может оказаться приемлемым, так как коэффициент сократится.

4.2.2. Оценка точности результатов анализов на основе теоремы Байеса

Оценка качества работы по результатам анализов контрольных материалов известного состава по существу базируется на теореме Байеса, хотя обычно ее аппарат не используется в полной мере. Теорема Байеса или иначе теорема гипотез, рассматривает событие, которое уже произошло, поэтому достоверно, и анализирует вероятности обстоятельств, которые ему способствовали. Лучше всего разъяснить это на следующем примере. Имеются два, внешне одинаковые ящика, в первом белые шары, во втором черные. Из одного из них взят шар, из какого ящика он вынут? Если шар черный, очевидно, что вынут из второго ящика. Задача усложняется, если в первом ящике полвина белых, половина черных, а во втором только черные, опять вынут черный шар. Нетрудно сообразить, что с вероятностью $1/3$ он вынут из первого ящика, а с вероятностью $2/3$ из второго, конечно, если играли честно, и выбор был случаен.

Что-то подобное происходит и при оценке качества по результатам контролей – наша первая гипотеза (H_1), что аналитическая система работает хорошо, результаты могут быть приняты. Альтернативная гипотеза (H_2) состоит в том, что результаты аналитической серии не могут быть приняты. Чтобы решить какая из них верная, выполняется контрольный анализ, результат которого должен попасть в определенный диапазон, т.е. образно выражаясь быть белым. Поскольку всегда имеется некоторый разброс данных, существует вероятность, что даже при хорошо работающей аналитической системе не удастся уложиться в заданный диапазон, и в «ящик результатов» хорошей аналитической системы попадет черный шар. Возможна и обратная ситуация, когда в «ящике результатов» плохой аналитической системы случайно оказываются правильные данные. Спрашивается, как надо установить границы диапазона ожидаемых результатов, чтобы получить оптимальную информацию о качестве работы аналитической системы?

Теорема Байеса рассматривает условные вероятности событий, которые при записи разделяются вертикальной чертой – $P(H_1 | R)$ означает вероятность с которой произошло событие H_1 при условии, что имело место событие R . В данном случае H_1 означает, что первая гипотеза верна, аналитическая система работает хорошо, R – результат анализа уложился в заданный диапазон. По традиции, если событие R не произошло, результат не попал в заданный диапазон, это обозначается \bar{R} . Рассмотрим сначала событие, обозначенное R – результат контрольного анализа находится в заданном диапазоне – какова вероятность, что в этом случае верна первая гипотеза? Возможны только два варианта – либо система работает хорошо, верна гипотеза H_1 , условная вероятность события R обозначается $P(R | H_1)$, либо система работает плохо, хороший результат получился случайно, его вероятность обозначается $P(R | H_2)$. Нас интересует доля тех случаев, когда попадание контроля в заданный интервал служит признаком правильности первой гипотезы, поэтому

$$P(H_1 | R) = \frac{P(R | H_1)}{P(R | H_1) + P(R | H_2)}$$

Условные вероятности $P(R | H_1); P(R | H_2); P(\bar{R} | H_1); P(\bar{R} | H_2)$ суть характеристики аналитической системы, они зависят от методики измерения и природы возможных погрешностей. Это всего лишь вероятности, которые может быть реализуются, а может быть и нет. Проведя контрольный анализ и выяснив попал ли его результат в заданный диапазон, мы оцениваем вероятность того, что именно в этот раз результат правильный. Можно ли этот вывод распространить и на другие анализы той же серии зависит от того, что мы знаем о природе погрешностей. Если, как это часто бывает:

- 1) межсерийная погрешность (иначе называемая сдвигом или биосом) постоянна для всей аналитической серии и является основным источником погрешности измерения,
- 2) средняя квадратичная внутрисерийной погрешности известна, сделанный вывод можно распространить на все одновременно выполненные анализы. Задача заключается в том, чтобы оптимально выбрать заданный интервал R и, опираясь на известные медицин-

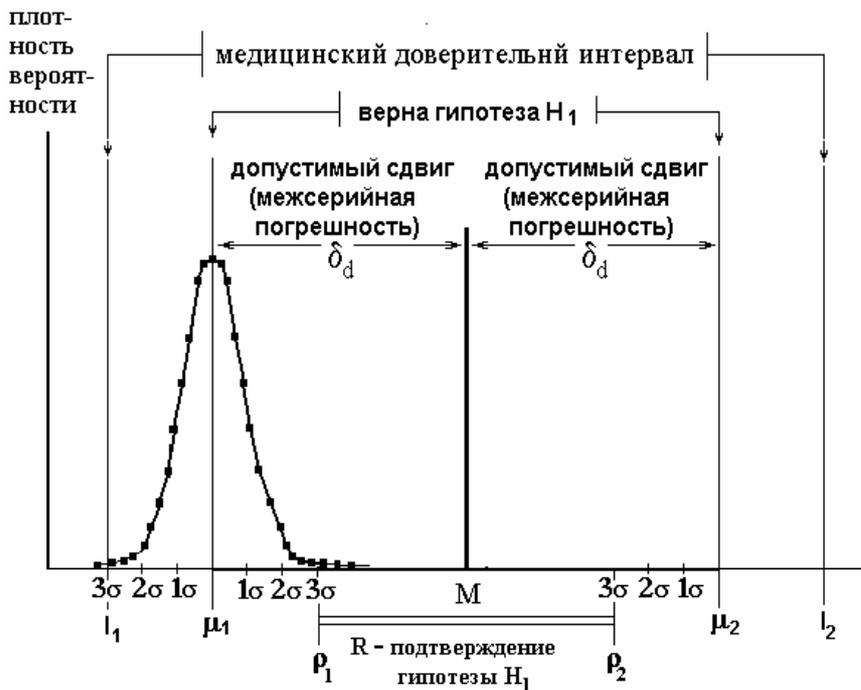


Рис. 12 Схема интервалов и гипотез

ский интервал и внутрисерийную погрешность, найти условные вероятности $P(R | H_1)$ и $P(R | H_2)$. Если ее удастся решить, то по результату одного контрольного анализа можно оценить достоверность всей серии. Ответ, разумеется, вероятностный, при этом не следует забывать, что чем выше достоверность положительной оценки, тем выше и вероятность ложного отклонения, правильно выполненного анализа.

Рассмотрим условия необходимые для того, чтобы по результатам одного контрольного анализа уверенно принять или отвергнуть всю аналитическую серию. Такой случай изображен на рис.12 где от-

резок $I_1 I_2$ это медицинский доверительный интервал, а δ_d максимальная величина сдвига (межсерийной погрешности) которую мы считаем допустимой. Когда гипотеза H_1 верна то сдвиг находится между точками μ_1 и μ_2 , мы будем говорить, что он в интервале H_1 . Размеры отрезков $I_1 \mu_1$ и $\mu_2 I_2$ выбраны так, чтобы в каждом умещалось по три средних квадратичных отклонения внутрисерийной погрешности (σ_r). Из рисунка видно, что, в этом случае результат анализа почти всегда попадает в медицинский интервал. Ведь даже в самом неблагоприятном случае, когда сдвиг максимален, за пределы медицинского интервала попадут только те результаты, которые отклоняются от средней более чем на $3\sigma_r$ – т.е. меньше 0,3%. Из того же рисунка видно, что если расстояние между границей интервала R и μ_1 (или, соответственно, μ_2) больше $3\sigma_r$, то результат «плохой» аналитической серии – когда гипотеза H_1 неверна, имеет очень мало шансов попасть в заданный интервал R . В данном примере попадание контроля в R означает что гипотеза H_1 верна с вероятностью больше чем 0,99, результаты серии могут быть приняты. Это высокая достоверность за которую приходится платить тем, что заданный интервал R мал – он на целых $12\sigma_r$ короче медицинского интервала! На практике такие высокие требования могут означать неразумно высокую долю ложных отклонений.

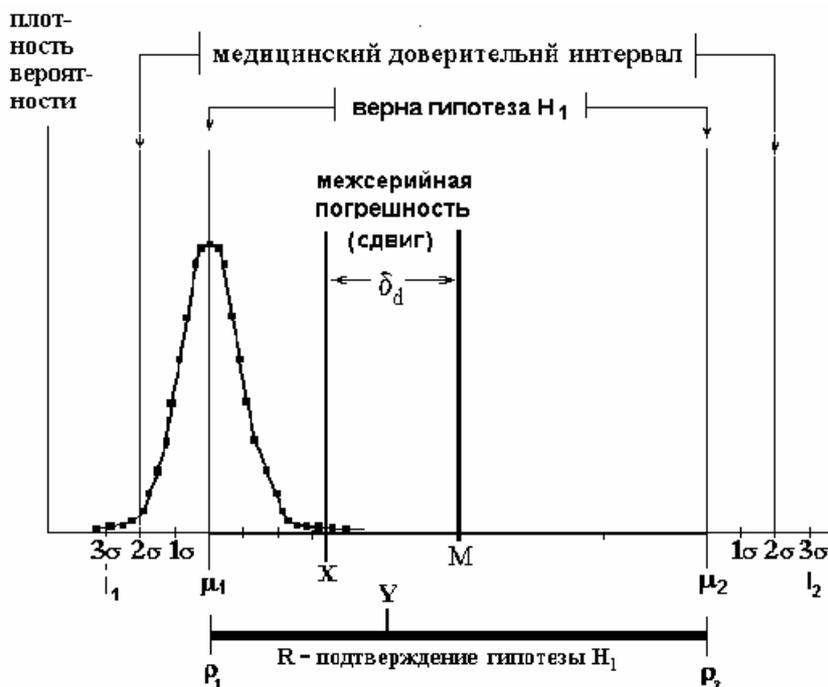


Рис. 13 Оценка точности по теореме Байеса

Более практичный и более общий случай приведен на рис. 13. Здесь требования мягче – гипотеза H_1 верна когда вероятность попадания результата в медицинский интервал не менее 0,95. Это означает, что расстояние между допустимой величиной сдвига (μ_1) и границей медицинского интервала сокращается до $1,96\sigma$, округлено 2σ . На этом рисунке X это математическое ожидание результатов контрольных анализов данной серии (т.е. установленная величина минус сдвиг), когда гипотеза H_1 верна $\mu_1 < X < \mu_2$ т.е., X находится в интервале H_1 . Заданный интервал R выбран,

таким, чтобы он совпадал с тем участком, где гипотеза H_1 верна, поэтому $\rho_1 = \mu_1$, а $\rho_2 = \mu_2$. Результат контрольного анализа обозначен Y , оценивается вероятность того, что гипотеза H_1 верна если $\rho_1 < Y < \rho_2$ т.е. результат контрольного анализа попал в R .

Случайная величина – межсерийная погрешность (иначе сдвиг) вообще говоря, распределена неравномерно, какие-то значения она принимает чаще, другие реже. Если эта закономерность известна, она может быть принята во внимание, в этом случае точность оценивается по совместному распределению результатов нескольких контрольных анализов (раздел 4.2.3). В настоящем разделе речь идет об оценке точности в условиях дефицита информации, когда ничего не известно о величине межсерийной погрешности, поэтому, все возможности равновероятны. Разумеется, в этом случае ответ получается менее точным, чем если бы мы знали закон распределения сдвига.

Для упрощения вычислений возьмем за начало координат установленную величину – точку M , тогда математическое ожидание данной серии равно сдвигу. Результат кон-

трольного анализа Y зависит от сдвига (X), поэтому речь идет о плотности условной вероятности

$$f(Y | X) = \frac{1}{\sigma_r \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2}$$

Нас интересуют вероятности следующих событий

- 1) Гипотеза H_1 верна (X находится в интервале H_1), а результат контрольного анализа (Y) в интервале R и
- 2) Гипотеза H_1 неверна (X вне пределов H_1), но результат контрольного анализа (Y) тем не менее попал в R .

Вероятность первого события получаем интегрируя плотность совместной вероятности по X в диапазоне от μ_1 до μ_2 и по Y от ρ_1 до ρ_2 .

$$p(R | H_1) = \frac{1}{\sigma_r \sqrt{2\pi}} \int_{X=\mu_1}^{X=\mu_2} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY$$

Интегрируя X от $-\infty$ до μ_1 получаем вероятность второго события

$$p(R | H_2) = \frac{1}{\sigma_r \sqrt{2\pi}} \int_{X=-\infty}^{X=\mu_1} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY$$

Попадание контрольного анализа в R свидетельствует о том, что гипотеза H_1 верна с вероятностью

$$P(H_1 | R) = \frac{\int_{X=\mu_1}^{X=\mu_2} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY}{\int_{X=\mu_1}^{X=\mu_2} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY + 2 \int_{X=-\infty}^{X=\mu_1} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY}$$

Напомним, что сама гипотеза H_1 предполагает, что с вероятностью 0,95 результат анализа отличается от своего истинного значения не больше, чем на полуширину медицинского интервала. Таким образом вероятность тут присутствует дважды.

Так как приведенный выше интеграл невозможно выразить через простые функции его заменяют суммированием, что не представляет сложности для современной вычислительной техники. Покажем, как это удобно сделать, заменив переменные. По условиям задачи нам известны две величины S – средняя квадратичной внутрисерийной погрешности и медицинский интервал $I_1 I_2$. Интервал H_1 мы выбрали так, чтобы обеспечить вероятность 0,95, а интервал R можно установить любым. Возьмем в качестве единицы измерения одну десятую средней квадратичной $s=S/10$, медицинский интервал в этом случае имеет длину $i = \frac{|I_1 I_2|}{s}$, интервалы R и H_1 короче медицинского интервала округлено на 4 средних квадратичных отклонения, поэтому имеют длину $(i-40)s$. За начало координат берем установленную величину M поэтому координата $\mu_1 = -(i-20)s$, а $\mu_2 = (i-20)s$. В новых координатах

интеграл $\int_{X=\mu_1}^{X=\mu_2} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY$ заменяется суммой $0,1 \sum_{k=-i+20}^{k=i-20} \left(0,1 \sum_{l=-i+20}^{l=i-20} e^{-\frac{1}{2} \left(\frac{k-l}{10} \right)^2} \right)$, а интеграл $\int_{X=-\infty}^{X=\mu_1} \int_{Y=\rho_1}^{Y=\rho_2} e^{-\frac{1}{2} \left(\frac{X-Y}{\sigma_r} \right)^2} dXdY$ суммой $0,1 \sum_{k=-i-20}^{k=-i+19} \left(0,1 \sum_{l=-i+20}^{l=i-20} e^{-\frac{1}{2} \left(\frac{k-l}{10} \right)^2} \right)$

Окончательно вероятность того, что при попадании результата контрольного анализа в заданный интервал R гипотеза H_1 верна и вся аналитическая серия может быть принята дает формула

$$P(H_1 | R) = \frac{\sum_{k=-i+20}^{k=i-20} \left(\sum_{l=-i+20}^{l=i-20} e^{-\frac{1}{2} \left(\frac{k-l}{10} \right)^2} \right)}{\sum_{k=-i+20}^{k=i-20} \left(\sum_{l=-i+20}^{l=i-20} e^{-\frac{1}{2} \left(\frac{k-l}{10} \right)^2} \right) + 2 \sum_{k=-i+19}^{k=-i+20} \left(\sum_{l=-i+20}^{l=i-20} e^{-\frac{1}{2} \left(\frac{k-l}{10} \right)^2} \right)}$$

Всякий вероятностный ответ таит в себе две возможных ошибки – во первых мы можем ошибочно принять плохие результаты, во вторых отклонить хорошие. Вероятность и той и другой ошибки зависит от того, какой размер контрольного интервала R задается.

Вероятности обосновано принять аналитическую серию или ложно отклонить ее в зависимости от размеров медицинского интервала (MI), контрольного интервала (R) и среднего квадратичного внутрисерийной погрешности (S)

Медицинский интервал/средне-квадратичное отклонение MI/S	Контрольный интервал равен медицинскому R=MI		Контрольный интервал меньше медицинского на 2 средних кв.откл. R=MI-2S		Контрольный интервал равен половине медицинского R=MI/2	
	Вероятность, что		Вероятность, что		Вероятность, что	
	ответ верный	отклонение ложное	ответ верный	отклонение ложное	ответ верный	отклонение ложное
5,0	0,603	0,001	0,869	0,017	0,923	0,041
6,0	0,669	0,001	0,902	0,006	0,973	0,025
7,0	0,716	<0,001	0,921	0,002	0,992	0,016
8,0	0,751	<0,001	0,934	0,001	0,998	0,011
9,0	0,779	<0,001	0,943	0,001	>0,999	0,008
10,0	0,801	<0,001	0,950	<0,001	>0,999	0,005

4.2.3. Оценка вероятности попадания в медицинский интервал по совместно-му распределению нескольких контрольных анализов

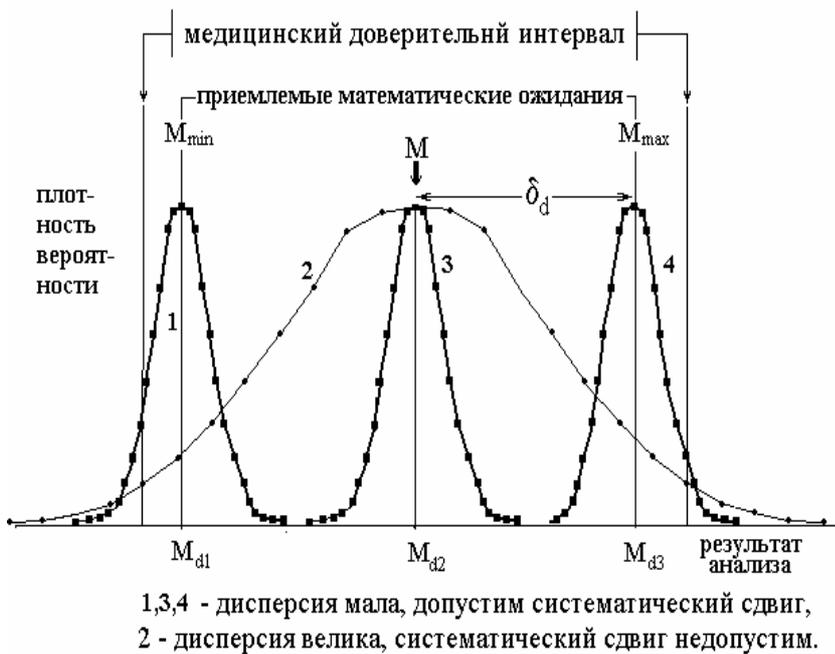
Обычно в лаборатории каждый день делается по 2-3 контрольных анализа, накопленные данные позволяет оценить средние квадратичные межсерийной (σ_d) и внутрисерийной (σ_r) погрешностей, что существенно увеличивает надежность статистической проверки качества работы. Изложенное ниже, с нашей точки зрения, должно быть основным методом внутреннего контроля качества в больших лабораториях.

Задача формулируется так: сделано несколько повторных анализов аттестованного контрольного материала (или материалов) результаты которых $R_1..R_n$ известны, дисперсии σ_d^2 и σ_r^2 также известны. Требуется найти вероятности того что:

- единичный анализ попадет в заданный интервал $|-I, +I|$,
- в серии из n анализов $(1-\alpha)n$ попадут в тот же интервал $|-I, +I|$.

Тем самым мы переходим от оценки неопределенности отдельного результата анализа к оценке качества работы лаборатории, потому что лечащего врача в первую очередь интересует достоверность результата анализа данному пациенту, но лаборатория должна смотреть шире – сколько из выполненных сегодня (или в этой серии) анализов могут оказаться недостоверными.

Доля анализов, которая укладывается в заданный медицинский интервал, зависит как от межсерийной (d_k) так и внутрисерийной (d_r) погрешностей. Если внутрисерийная



1,3,4 - дисперсия мала, допустим систематический сдвиг,
 2 - дисперсия велика, систематический сдвиг недопустим.
 рис. 14 Возможные варианты распределения
 результатов повторных анализов.

погрешность мала, можно мириться с увеличением межсерийной, вероятность попадания результата в медицинский интервал при этом останется высокой (рис. 14), Так как вероятность внутрисерийной ошибки во все дни одинакова, реальное качество анализов данной серии зависит от того, какая сегодня меж-серийная погрешность. Если бы можно было сделать много конт-рольных анализов, ее можно было бы вычислить и внести поправки в расчет, но по 2-3 точкам это делать

рискованно – различие может быть случайным и поправка превратится в источник ошибок. Поэтому правильнее взвешенно оценивать вероятность ошибки – т.е. обращать внимание не только на разброс и сдвиг результатов контролей, но и на то насколько они правдоподобны.

Такой подход требует достаточно громоздкой математической обработки, зато самый достоверный, поэтому больше всего подходит для повседневной работы. Громоздкость статистической обработки никого не должна смущать, так как в любом случае компьютер выполняет ее за считанные секунды.

4.2.3.1. Оценка межсерийной и внутрисерийной дисперсий погрешностей

Обозначим установленную величину контрольного материала μ , результат первого из двух выполненных в этот день параллельных анализов X_k , второго Y_k , межсерийную (т.е. систематическую) погрешность d_k , а внутрисерийные r_{kx} и r_{ky} . Межсерийная погрешность выполненных одновременно анализов одинакова

$$X_k = \mu + d_k + r_{kx}; Y_k = \mu + d_k + r_{ky}$$

где k это номер дня (или аналитической серии).

Ковариация по определению это:

$$\begin{aligned} K_{xy} &= \int \int_{-\infty}^{+\infty} (X_k - \bar{X})(Y_k - \bar{Y})f(X_k, Y_k)dXdY = \int \int_{-\infty}^{+\infty} (d_k + r_{kx})(d_k + r_{ky})f(X_k, Y_k)dXdY = \\ &= \int \int (d_k^2 f(X, Y) + d_k r_{kx} f(X, Y) + d_k r_{ky} f(X, Y) + r_{kx} r_{ky} f(X, Y))dXdY = \\ &= \int \int d_k^2 f(X_k, Y_k)dXdY = \sigma_d^2 \end{aligned}$$

Так как математические ожидания членов $d_k r_{kx}$, $d_k r_{ky}$ и $r_{kx} r_{ky}$ равны нулю, межсерийная дисперсия равна ковариации.

При вычитании дисперсии складываются, поэтому дисперсия разности двух одновременно выполненных анализов $X_k - Y_k = r_{kx} - r_{ky}$ равна удвоенной внутрисерийной дисперсии $\sigma_{X-Y}^2 = 2\sigma_r^2$. Применимость данной модели к конкретным результатам опытов, а заодно и правильность вычислений, легко проверить – общая дисперсия должна быть равна сумме дисперсий сдвига и разброса.

Результатов анализов двух контрольных материалов достаточно, чтобы оценить обе дисперсии, однако желательно ежедневно анализировать по три контрольных материала и, комбинируя попарно результаты, располагать тремя парами оценок, которые, в идеальном случае, должны совпадать. Если один из контрольных материалов некачественен, например, недостаточно гомогенен, совпадения может не быть. В этом случае наибольшего доверия заслуживает наименьшая дисперсия. Для количественной оценки можно использовать критерии Бартлетта, Кокрана и Фишера.

4.2.3.2. Вероятность попадания результата анализа в медицинский интервал когда параметры распределения погрешностей известны

Принимается, что результаты анализов распределены по нормальному закону, калибровка выполнена идеально, поэтому математическое ожидание равно установленной величине, аттестованные значения калибровочных материалов, также как межсерийная и внутрисерийная дисперсии известны. Вычисляется условная вероятность попадания X - результата анализа пациента в медицинский интервал при условии, что различия между фактическими и аттестованными значения (т.е. погрешности) нескольких контрольных анализов $R_1..R_n$ известны. Для этого вычисляется совместная плотность вероятности погрешностей всех $n+1$ анализов, а затем интегрируется по X в пределах медицинского интервала.

Межсерийную погрешность (сдвиг) обозначим M , погрешности результатов контрольных и опытного анализов R_n и X . Соответственно плотности вероятностей

$$P_M(M) = \frac{1}{\sigma_d \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{M}{\sigma_d} \right)^2}$$

$$P_R(R_n) = \frac{1}{\sigma_r \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{R_n - M}{\sigma_r} \right)^2}$$

$$P_X(X) = \frac{1}{\sigma_r \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{X - M}{\sigma_r} \right)^2}$$

Если события произошли одновременно, плотность их совместной вероятности есть произведение плотностей вероятностей всех событий:

$$P_{RMX}(M, R_1..R_n, X) = \frac{1}{\sigma_d \sigma_r^n (2\pi)^{\frac{n+2}{2}}} e^{-\frac{1}{2} \left(\frac{M}{\sigma_d} \right)^2 - \frac{1}{2} \left(\frac{R_1 - M}{\sigma_r} \right)^2 - \dots - \frac{1}{2} \left(\frac{R_n - M}{\sigma_r} \right)^2 - \frac{1}{2} \left(\frac{X - M}{\sigma_r} \right)^2}$$

преобразуем показатель степени

$$\begin{aligned} & -\frac{1}{2} \left(\frac{M}{\sigma_d} \right)^2 - \frac{1}{2} \left(\frac{R_1 - M}{\sigma_r} \right)^2 - \dots - \frac{1}{2} \left(\frac{R_n - M}{\sigma_r} \right)^2 - \frac{1}{2} \left(\frac{X - M}{\sigma_r} \right)^2 = \\ & = -\frac{M^2 \sigma_r^2 + \sum R^2 \sigma_d^2 - 2(\sum R) M \sigma_d^2 + n M^2 \sigma_d^2 + X^2 \sigma_d^2 - 2 X M \sigma_d^2 + M^2 \sigma_d^2}{2 \sigma_d^2 \sigma_r^2} = \\ & = -\frac{M^2 ((n+1) \sigma_d^2 + \sigma_r^2) - 2 M (\sum R + X) \sigma_d^2 + (\sum R^2 + X^2) \sigma_d^2}{2 \sigma_d^2 \sigma_r^2} = \end{aligned}$$

$$\begin{aligned}
&= -\frac{(n+1)\sigma_d^2 + \sigma_r^2}{2\sigma_d^2\sigma_r^2} \left(M^2 - 2M(\Sigma R + X) \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} + (\Sigma R^2 + X^2) \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} \right) + \\
&\quad - \frac{(n+1)\sigma_d^2 + \sigma_r^2}{2\sigma_d^2\sigma_r^2} \left(\frac{\sigma_d^4}{((n+1)\sigma_d^2 + \sigma_r^2)^2} (\Sigma R + X)^2 - \frac{\sigma_d^4}{((n+1)\sigma_d^2 + \sigma_r^2)^2} (\Sigma R + X)^2 \right) = \\
&= -\frac{(n+1)\sigma_d^2 + \sigma_r^2}{2\sigma_d^2\sigma_r^2} \left(M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X) \right)^2 - \\
&\quad - \frac{(n+1)\sigma_d^2 + \sigma_r^2}{2\sigma_d^2\sigma_r^2} \left(\frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R^2 + X^2) - \frac{\sigma_d^4}{((n+1)\sigma_d^2 + \sigma_r^2)^2} (\Sigma R + X)^2 \right) = \\
&= -\left(\frac{M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)}{\frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2}} \right)^2 - \frac{\Sigma R^2 + X^2}{2\sigma_r^2} + \frac{\sigma_d^2 (\Sigma R + X)^2}{2\sigma_r^2 ((n+1)\sigma_d^2 + \sigma_r^2)} = \\
&= -\left(\frac{M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)}{\frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2}} \right)^2 - \left(\frac{(\Sigma R^2 + X^2) - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)^2}{2\sigma_r^2} \right)
\end{aligned}$$

Таким образом функция $P_{RMX}(M, R, X)$ является произведением трех сомножителей

$$P_{RMX}(M, R, X) = \frac{1}{\sigma_d\sigma_r^2(2\pi)^{\frac{n+2}{2}}} e^{-\left(\frac{M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)}{\frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2}} \right)^2} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)^2}{2\sigma_r^2}}$$

Интегрирование по М по всему диапазону значений позволяет упростить эту формулу. Вводим переменную t

$$t = \frac{M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)}{2 \frac{\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2}} \quad \frac{dt}{dM} = \frac{(n+1)\sigma_d^2 + \sigma_r^2}{2\sigma_d^2\sigma_r^2}$$

$$\begin{aligned}
\text{теперь } \int_{-\infty}^{+\infty} P_{RMX}(M, R, X) dM &= \int_{-\infty}^{+\infty} \frac{1}{\sigma_d\sigma_r^2(2\pi)^{\frac{n+2}{2}}} e^{-\left(\frac{M - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)}{\frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2}} \right)^2} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)^2}{2\sigma_r^2}} dM = \\
&= \int_{-\infty}^{+\infty} \frac{1}{\sigma_d\sigma_r^2(2\pi)^{\frac{n+2}{2}}} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)^2}{2\sigma_r^2}} e^{-t^2} \frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2} dt = \\
&= \frac{2\sigma_d^2\sigma_r^2}{(n+1)\sigma_d^2 + \sigma_r^2} \frac{1}{\sigma_d\sigma_r^2(\sqrt{2\pi})^3} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2} (\Sigma R + X)^2}{2\sigma_r^2}} \int_{-\infty}^{+\infty} e^{-t^2} dt =
\end{aligned}$$

$$= \frac{2\sigma_d \sqrt{\pi}}{((n+1)\sigma_d^2 + \sigma_r^2)(2\pi)^{\frac{n+2}{2}}} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2}(\Sigma R + X)^2}{2\sigma_r^2}} = f(R, X)$$

поскольку $\int_{-\infty}^{+\infty} e^{-t^2} dt = \sqrt{\pi}$ (интеграл Эйлера-Пуассона).

Функция $f(R, X)$ это вероятность совместного распределения погрешностей контрольных и опытного анализов, но погрешности контролей $R_1..R_n$ известны, а опыта X нет. Придавая разные значения X , получаем распределение плотности условной вероятности по-

грешности опыта. Интеграл $\int_{-\infty}^{+\infty} f(R, X)dX$ дает полную величину условной вероятности

всех возможных погрешностей опыта при данных контролях, т.е. фактически вероятность данного расклада контролей. Нас интересует вероятность попадания результата опыта в заданный медицинский доверительный интервал $-I+I$, при условии, что контроли $R_1..R_n$

$$\int_{-I}^{+I} f(X | R)dX = \frac{\int_{-I}^{+I} f(X, R)dX}{\int_{-\infty}^{+\infty} f(X, R)dX} = \frac{\frac{2\sigma_d}{\sqrt{\pi}((n+1)\sigma_d^2 + \sigma_r^2)(2\pi)^{\frac{n+2}{2}}} \int_{-I}^{+I} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2}(\Sigma R + X)^2}{2\sigma_r^2}} dX}{\frac{2\sigma_d}{\sqrt{\pi}((n+1)\sigma_d^2 + \sigma_r^2)(2\pi)^{\frac{n+2}{2}}} \int_{-\infty}^{+\infty} e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2}(\Sigma R + X)^2}{2\sigma_r^2}} dX}$$

Коэффициенты перед знаком интеграла одинаковые в числителе и знаменателе, они сокращаются.

Таким образом, если погрешности n контролей известны, то вероятность того, что погрешность анализа, выполненного пациенту, уложится в установленный медицинский интервал есть отношение интеграла

$$\int e^{-\frac{\Sigma R^2 + X^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2}(\Sigma R + X)^2}{2\sigma_r^2}} dX$$

в пределах медицинского интервала к тому же интегралу в бесконечных пределах. Так как их нельзя выразить через элементарные функции, интегрирование заменяется суммированием

$$\sum_{i=-40}^{i=40} e^{-\frac{\Sigma R^2 + (i\Delta X)^2 - \frac{\sigma_d^2}{(n+1)\sigma_d^2 + \sigma_r^2}(\Sigma R + i\Delta X)^2}{2\sigma_r^2}}$$

Шаг ΔX изменения переменной X выбирается так, чтобы 10,5 шагов соответствовали I полуширине доверительного медицинского интервала. Программа отдельно суммирует результаты 21 шага – по 10 в обе стороны и нулевой, и относит их к общей сумме всех 81 членов.

4.2.3.4. Оценка вероятности попадания серии результатов анализов в заданный медицинский доверительный интервал

Пусть вероятность непопадания единичного результата в заданный медицинский интервал α , а попадания $(1-\alpha)$. Вероятность того, что в серии из n анализов результаты m анализов не попадет в него, а $n-m$ попадут, дается формулой Бернулли

$$P_n(m) = \frac{n(n-1)\dots(n-m+1)}{m!} \alpha^m (1-\alpha)^{n-m}$$

Сумма вероятностей всех случаев, когда число непопаданий $k=0,1,2 \dots ,m$

$$P_\beta = \sum_{k=0}^{k=m} \frac{n(n-1)\dots(n-m+1)}{m!} \alpha^n (1-\alpha)^{n-m}$$

и есть вероятность того, что в установленный медицинский интервал попадет не менее чем m из серии выполненных сегодня n анализов.

5. Достоверность подсчета лейкоцитарной формулы

Подсчет лейкоцитарной формулы, предложенный немецким гематологом Шиллингом в начале XX века, одно из самых распространенных лабораторных исследований. Все по опыту знают, что результаты повторной микроскопии одного и того же мазка почти всегда несколько различаются, обычно это объясняют неравномерностью распределения элементов, особенностями трактовки морфологии разными школами или даже небрежностью, не задумываясь о случайной природе возможных различий. С точки зрения теории вероятностей, случайные различия при подсчете лейкоцитарной формулы крови неизбежны, их размер и вероятность предсказуемы, это проистекает из самой сущности используемого метода. Поэтому врач-микроскопист должен представлять себе природу и возможный размер случайной погрешности, которая не зависит ни от квалификации и тщательности работника, ни от условий работы. О ней нельзя забывать, ни когда решается вопрос о том, находится ли полученный результат в границах нормы, ни когда оценивается достоверность изменения состава крови в динамике, ни когда проверяется качество работы по контрольному мазку. Последнее особенно важно, так как речь идет о профессиональной компетентности и ее оценка не должна зависеть от случайности.

Рассмотрим самый простой случай, когда есть только два типа клеток – один их них назовем сегментами другой несегментами. Допустим, что тех и других одинаковое количество. Если мы хотим судить о составе мазка по двум наугад выбранным элементам, то возможны четыре варианта 1) оба элемента оказались несегментами, 2) первый рассмотренный элемент сегмент, второй несегмент, 3) первый несегмент, второй сегмент и 4) оба элемента сегменты. Все четыре результата равновероятны поэтому если эксперимент повторять много раз, в половине случаев мы получим правильное представление о соотношении форменных элементов – что количество сегментов и несегментов одинаково, а вторая половина опытов даст заведомо неправильные результаты, при чем в четверти всех случаев мы будем думать, что сегментов совсем нет, а в другой четверти, что в мазке присутствуют только сегменты. Очевидно, что такой «халтурный» способ неприемлем, чтобы получить более точные данные надо увеличить число просмотренных клеток.

Если же мы посмотрим 4, случайно попавшие в поле зрения элемента, то возможны 16 равновероятных вариантов результата. Из них в 6 случаях число сегментов и несегментов одинаково, поэтому правильный результат получается в 6 из 16 опытов, т.е. в 37,5% случаев, в половине случаев ошибка будет «умеренной» – окажется, что сегментов 25% или 75%, а в 12,5% «грубой» – окажется, что сегментов вообще нет, или есть только одни сегменты.

Рассуждая подобным образом можно посчитать вероятность правильного ответа при просмотре 100 или любого другого числа наугад выбранных элементов. Математически

задача очень близка к той, которая возникает при игре в орлянку – когда подбрасывают монету и смотрят, какой стороной она упадет. Бернулли еще в 17 веке вывел формулу, которая показывает вероятность того, что при заданном числе бросаний монета столько-то раз упадет орлом, а столько-то решкой. Она так и называется формулой Бернулли и приводится во многих учебниках математики, и показывает вероятность P_n того, что в процессе n испытаний событие произойдет m раз. Применительно к нашей задаче, это означает, что если истинная доля сегментов составляет p , то из n просмотренных клеток m окажутся сегментами с вероятностью

$$P_n = \frac{n!}{m!(n-m)!} p^m (1-p)^{n-m}$$

Здесь $n! = 1 * 2 * 3 * \dots * n$ означает произведение всех целых чисел от 1 до n , то же самое для $m!$ и $(n-m)!$

В этой формуле учитываются вероятности двух взаимоисключающих событий – в нашем случае, что клетка сегмент или несегмент, при игре в орлянку – что монета упадет орлом или решкой. Такая конструкция называется биномом (двучленом) и, соответственно, распределение возможных результатов называется биномиальным. Приведенный выше пример возможных результатов подсчета двух и четырех клеточных элементов в мазке, это частных случай биномиального распределения и вычисления могли бы быть сделаны по приведенной формуле. В реальной жизни, когда просмотрено много элементов (на математическом языке говорят, что было много событий) – например, просмотрено 100 клеток в мазке, вычисления по формуле Бернулли становятся чрезвычайно громоздкими и практически невыполнимы. Однако, существует другой, более простой математический закон, который дает похожие результаты и может быть реально использован для расчетов. Это называется аппроксимацией (приближением) одной формулы другой. В данном случае известно, что если объем достаточно велик (считают 30 или больше клеток) и вероятности альтернативных результатов больше 0,04 и меньше 0,96 (т.е. что сегментов больше 4% и меньше 96%), биномиальный закон хорошо аппроксимируется законом нормального распределения (Гаусса). В этом случае средняя величина, полученная во многих опытах (она называется математическим ожиданием и обозначается буквой M) равна числу просмотренных в каждом подсчете элементов (n) умноженному на истинную долю элементов данного вида (P):

$$M = P * n$$

а среднее квадратичное отклонение

$$\sigma = \sqrt{nP(1-P)}$$

Эти формулы дают абсолютные количества клеток данного вида. Если же нас интересуют процентные соотношения, то надо разделить на n – число просмотренных клеточных элементов и умножить на 100:

$$M\% = 100 * P / n ;$$

$$\sigma\% = \frac{\sqrt{P(1-P)}}{\sqrt{n}} * 100$$

Так, если просмотрено 100 клеток в мазке где «истинная» доля сегментов составляет 0,5, то среднее квадратичное отклонение результата подсчета выраженного в процентах составляет

$$\sigma\% = \frac{\sqrt{0,5 * (1 - 0,5)}}{\sqrt{100}} * 100 = \frac{0,5}{10} * 100 = 5\%$$

Это означает, что даже когда просматривают 100 клеток, только в редких случаях ответ точно равен 50, чаще он несколько выше или ниже. При нормальном распределении 95% всех результатов укладываются

ся в интервал от $M-1,96\sigma$ до $M+1,96\sigma$, округляя $1,96\sigma$ это в нашем случае 10 элементов, поэтому в 95% случаев результат подсчета мазка будет от 40 до 60 элементов, а в 5% выйдет за эти пределы. Такова неизбежная погрешность, которую можно уменьшить, только увеличив число просмотренных клеточных элементов, так как случайный разброс результатов обратно пропорционален корню квадратному из числа просмотренных элементов. Поэтому увеличение их числа в два раза – с 100 до 200 клеток, уменьшает случайную погрешность только в $\sqrt{2}=1,41$ раза.

Границы доверительных интервалов для доверительной вероятности 0,95
при подсчете в мазке 100 и 200 клеток

Найдено клеток	100 клеток		200 клеток	
	Нижняя граница	Верхняя граница	Нижняя граница	Верхняя граница
5%	2,2	11,2	2,7	9,0
10%	5,5	17,4	6,6	14,9
15%	9,3	23,3	10,7	20,6
20%	13,3	28,9	15,0	26,1
25%	17,5	34,3	19,5	31,4
30%	21,9	39,8	24,1	36,7
35%	26,4	44,7	28,7	41,8
40%	30,9	49,8	33,5	46,9
45%	35,6	54,8	38,3	51,9
50%	40,4	59,6	41,3	56,9
55%	45,2	64,4	48,1	61,7
60%	50,2	69,1	53,1	66,5
65%	55,3	73,6	58,2	71,3
70%	60,4	78,1	63,3	75,9
75%	65,7	82,5	68,6	80,5
80%	71,1	86,7	73,9	85,0
85%	76,7	90,7	79,4	89,3

Приведенный пример важен главным образом для трактовки результатов внешнего контроля качества – когда квалификация врача-лаборанта проверяется по тому, насколько посчитанная им лейкоцитарная формула совпадает с тем, что было получено раньше. Надо

понимать, что точное совпадение возможно лишь случайно. Если посчитано 100 клеток, расхождение в 4-5 элементов вероятнее всего вызвано не ошибкой работника, а самой вероятностной природой метода. Чтобы улучшить результаты надо увеличить объем и считать по крайней мере 400 элементов, в этом случае средняя квадратичная погрешность в центре диапазона уменьшится с 5% до 2,5%, соответственно упадет и вероятность расхождения. В повседневной клинической работе нет ни возможности, ни необходимости просматривать больше, чем 100 форменных элементов. Поэтому встают два вопроса – насколько достоверно мы можем отличить норму от патологии, а при динамическом наблюдении какие изменения считать достоверными.

Для оценки точности подсчета лейкоцитарной формулы пригоден раздел теории вероятностей, который называется оценкой вероятности по частоте. Суждение выносится на основе того нормального распределения, которое лучше всего приближается к данному биномиальному. Вероятностью называется истинная доля клеточных элементов, т.е. та, которая получилось бы при тщательном просмотре огромного количества клеток. Частотой же называется их доля в конкретной выборке, т.е. среди тех 100 элементов, которые реально были просмотрены в данном мазке. Оценить точность – это значит сказать, в каком проценте случаев различие не превышает заранее установленной величины, например, 5 клеток. Эта величина называется доверительным интервалом для заданной довери-

тельной вероятности, чем он шире, тем вероятнее там находится истинная величина. Доверительный интервал может быть симметричным, в этом случае вероятность погрешности в обе стороны одинакова, или несимметричным, когда погрешность в одну сторону вероятнее, чем в другую. Теория вопроса изложена во многих учебниках, в том числе в рекомендованном для студентов высших учебных заведений курсе теории вероятностей Е.А.Вентцель. Там же приводятся формулы для расчетов, но этот аппарат несколько громоздок, поэтому мы ограничимся только результатом – таблицей границ доверительных интервалов для 95% доверительной вероятности при подсчете 100 или 200 клеток. Так, если при просмотре 100 клеток в мазке найдено 20 моноцитов, это значит, что их истинное число с 95% вероятностью находится между 13,3% и 28,9% клеток. На первый взгляд кажется странным, что ответ дробный, но это означает только, что если бы мы много раз в идеальных условиях пересчитали мазки, взятые одновременно у того же самого пациента, просмотрев каждый раз по 1000 клеток, то в 95 случаях из 100 результат был бы между 133 и 289. Конечно, применительно к лейкоцитарной формуле это проверить невозможно, но в других задачах, например при игре в орлянку, эти правила много раз с успехом проверялись. Как видно, сам метод подсчета лейкоцитарной формулы не очень точен.

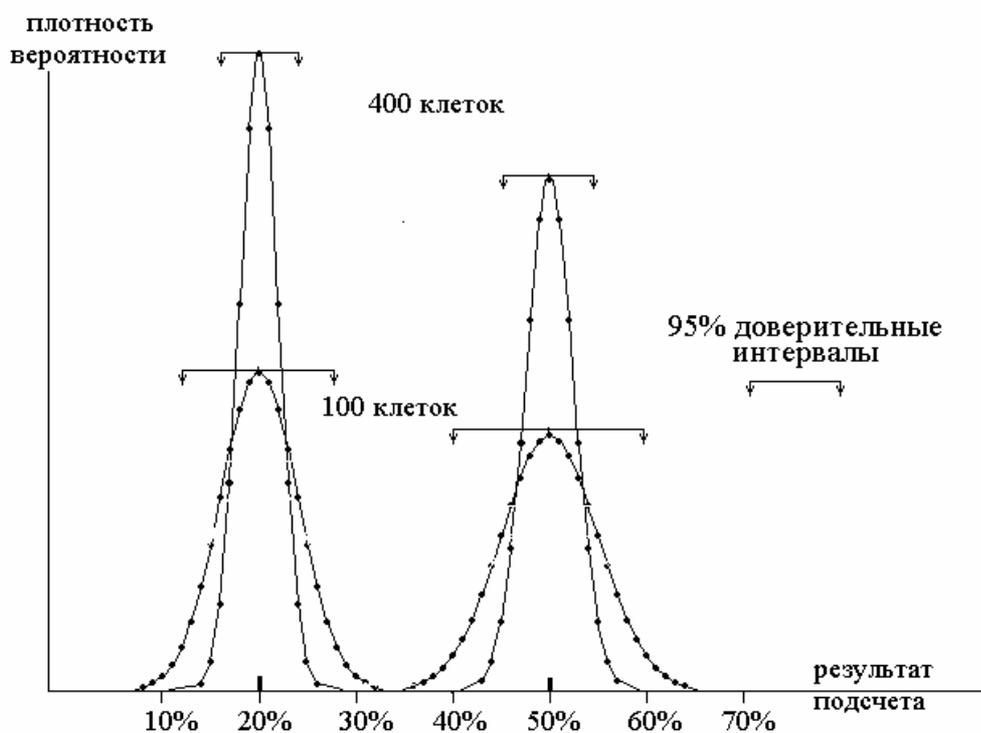


Рис. 15 Распределение результатов подсчета клеточных элементов в мазке для случаев, когда просмотрено 100 или 400 клеток, истинная доля которых составляет 20% или 50%.

Ту же закономерность иллюстрирует рис. 15, на котором показаны кривые распределения результатов подсчета лейкоцитарной формулы для тех случаев, когда истинный процент клеток данного вида составляет 20% или 50%, а в мазке считалось 100 или 400 клеток.

Приведенные выше рассуждения и примеры молча предполагают, что заранее ничего не известно о клеточном составе данного мазка и любой результат равновероятен. На самом деле это не так, некоторые результаты встречаются значительно чаще, чем другие, это называется априорной вероятностью события. В частности, в большинстве лабораторий патологических результатов значительно меньше, чем нормальных. Поэтому, когда результат оказывается на границе нормы и патологии, более вероятно, что он туда слу-

чайно попал из зоны нормы, а не из зоны патологии. По этой же причине ошибочное выявление патологии (ложноположительный результат) более вероятно, чем ее ошибочное отклонение (ложноотрицательный результат). Такое положение в принципе устраивает медицину – лучше перестраховаться, чем пропустить болезнь. Но если результаты подсчета лейкоцитарной формулы имеют критическое значение и могут серьезно повлиять на принятие медицинского решения, надо не забывать и небольшой точности метода и увеличить число просматриваемых клеток.

Вопрос о том, насколько достоверно различие между результатами двух анализов лейкоцитарной формулы очень важен для оценки динамики заболевания и контроля за лечением. В этом случае сравниваются два экспериментальных результата, поэтому априорная вероятность каждого из них не имеет значения. Оценка достоверности различия основывается также на свойствах нормального распределения, аппроксимирующего биномиальное. Как уже отмечалось в этом случае среднее квадратичное отклонение $\sigma = \sqrt{nP(1-P)}$. Поскольку среднее квадратичное отклонение суммы или разности равно корню квадратному из суммы квадратов средних квадратичных отклонений слагаемых

$$\sigma_p = \sqrt{\sigma_1^2 + \sigma_2^2} = \sqrt{n P_1(1-P_1) + n P_2(1-P_2)}$$

Здесь σ_p среднее квадратичное отклонение разности, P_1 и P_2 число элементов данного вида в первом и во втором мазке, n число просмотренных в каждом мазке клеточных элементов, мы принимаем что оба раза было просмотрено одинаковое число клеток. Сравнивая найденное различие с σ_p , находим вероятность того, что различие случайно. Если она мала, – обычно меньше чем 0,05 (или 5%), считается, что различие достоверно.

Вероятность того, что при подсчете 100 элементов лейкоцитарной формулы обнаруженное различие случайно. (Строки соответствуют меньшим величинам, столбцы – разностям между большей и меньшей величинами)

Разность Значение	4	6	8	10	12	14	16
10	0,39	0,21	0,10	0,05	0,02	0,01	0,01
15	0,45	0,27	0,15	0,08	0,04	0,02	0,01
20	0,50	0,31	0,19	0,10	0,05	0,03	0,01
25	0,53	0,35	0,21	0,12	0,07	0,03	0,02
30	0,55	0,37	0,23	0,14	0,08	0,04	0,02
35	0,56	0,38	0,25	0,15	0,08	0,05	0,02
40	0,57	0,39	0,26	0,16	0,09	0,05	0,02
45	0,57	0,40	0,26	0,16	0,09	0,05	0,02
50	0,57	0,40	0,26	0,16	0,09	0,05	0,02
55	0,57	0,39	0,25	0,15	0,08	0,04	0,02
60	0,56	0,38	0,24	0,14	0,07	0,04	0,01
65	0,55	0,37	0,22	0,12	0,06	0,03	0,01
70	0,53	0,34	0,20	0,10	0,05	0,02	-

Для удобства в таблице приведены вероятности того, что обнаруженное различие случайно. Строки соответствуют меньшему результату, а столбцы увеличению числа элементов – т.е. разности между большим и меньшим результатами. Какой из них был получен по времени раньше, а какой позже, в данном случае не имеет значения. Так если результат первого исследования показал, что в мазке 50% клеток сегменты, а при повторном исследовании их стало 60% (увеличение на 10), по таблице вероятность такого изменения 0,14, это больше чем 0,05, поэтому изменение нельзя считать достоверным. Только, если число сегментов возросло до 64, сдвиг можно считать достоверным. Если изменение не очень достоверно, это отнюдь не означает, что данные надо выбросить в корзину, но со-

поставляя с другими симптомами болезни, врач должен проявить осторожность, понимая, что тут возможна ошибка.

Несмотря на кажущуюся простоту, разработанность и широкое распространение метода подсчета лейкоцитарной формулы под микроскопом, правильная и надежная оценка результатов представляет определенные трудности. Приведенное выше упрощенное изложение проблемы далеко не исчерпывают все возможные ситуации, для решения которых желательно в каждой лаборатории иметь специальную компьютерную программу. Такая программа должна помогать врачу делать статистически достоверные выводы, отстаивать профессиональные интересы и защитить от необоснованных нападок.

6. Заключение и практические рекомендации

Как видно из приведенных материалов, существует много способов оценки точности результатов лабораторной работы с использованием статистической обработки данных и других методов теории вероятностей. Дело руководства лаборатории выбрать тот, который больше всего подходит для конкретных условий. В любом случае эффективное использование статистических методов возможно только, если это предусмотрено конструкцией анализатора, либо данные автоматически поступают в лабораторный компьютер, снабженный соответствующей программой. Написать ее, руководствуясь описанными выше принципами, дело нелегкое, доступное практически любому грамотному программисту, главное это организовать ввод-вывод информации в соответствии с используемым в данной лаборатории типом аппаратуры. Существенно, чтобы программа работала автоматически, привлекая внимание врача-аналитика только тогда, когда это действительно нужно, т.е. когда соответствующий тест выявляет подозрение на возможное ухудшение качества. Ведь если лаборатория ежедневно использует несколько десятков показателей, человеку трудно все непрерывно держать в сфере своего внимания.

Метода контроля по данным пациентов (т.е. не требующие контрольных материалов) – дисперсионный анализ и критерий Колмогорова-Смирнова можно рекомендовать всем лабораториям, где ежедневно выполняется несколько десятков исследований данного вида. Основная особенность этих методов – необходимость большого объема материала. В каждой серии при дисперсионном анализе должно быть несколько сот проб, при использовании критерия Колмогорова-Смирнова не менее 50-100. Безусловное преимущество это дешевизна – обрабатываются только те анализы, которые все равно делаются. Характер распределения погрешностей не имеет значения – оба метода пригодны как для того случая, когда распределение нормальное, так и какое-либо другое. Недостаток очевиден – они позволяют оценить только воспроизводимость результатов и не дают непосредственной информации о правильности калибровки. Однако, в этом отношении последнее слово еще не сказано – можно думать, что средние результаты и характер распределения данных в популяции, при определенных условиях, должны быть одинаковыми во всех лабораториях. Это позволило руководителю предприятия «Аналика» М.И.Прищепе предложить осуществлять метрологический контроль лабораторной аппаратуры по средним цифрам результатов анализов, полученных в данной лаборатории. Реализация этого, очень интересного предложения сопряжена со значительными трудностями, но, в случае успеха, эффект оправдает все усилия.

Выше были описаны разные способы использования контрольных материалов для оценки качества, это отнюдь не исчерпывает все возможные подходы, но достаточно чтобы удовлетворить потребности разных лабораторий или аналитических методов. Выбор зависит от того, какой информацией мы располагаем о природе погрешностей.