

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Локтионова Оксана Геннадьевна

Должность: проректор по учебной работе

Дата подписания: 10.11.2023 03:18:59

Уникальный программный ключ:

0b817ca911e6668abb13a5d426d39e5f1c41eabbf73e943df4a4851fda56d089

## МИНОБРНАУКИ РОССИИ

Федеральное государственное бюджетное  
образовательное учреждение высшего образования  
"Юго-Западный государственный университет"  
(ЮЗГУ)

Кафедра биомедицинской инженерии

УТВЕРЖДАЮ

Проректор по учебной работе

О.Г.Локтионова

" 15 " 12



**УЧЕБНО-ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА**

Методические указания к проведению практических занятий для студентов  
специальности 30.05.03 - "Медицинская кибернетика"

Курск 2017

УДК 615.478

Составитель Н.А. Корневский, С.Н. Корневская, А.В. Быков.

Рецензент

Доктор технических наук, профессор *И.Е. Чернецкая*

Учебно-исследовательская работа: методические указания к поведению практических занятий / Юго-Зап. гос. ун-т; сост. Н.А. Корневский, С.Н. Корневская, А.В. Быков. - Курск, 2017. - 80 с.

Содержатся сведения, необходимые для выполнения практических занятий по учебно-исследовательской работе.

Методические указания по структуре, содержанию и стилю изложения материала соответствуют методическим и научным требованиям, предъявляемым к учебным и методическим пособиям.

Предназначены для студентов специальности 30.05.03 очной формы обучения.

Текст печатается в авторской редакции

Подписано в печать *15.12.17* Формат 60x84 1/16  
Усл. печ.л. 3,89. Уч. -изд.л. 3,53 Тираж 100 экз. Заказ. *2698* Бесплатно.  
Юго-Западный государственный университет  
305040, г.Курск, ул. 50 лет Октября, 94

---

## Введение

Предлагаемое методическое пособие ориентировано на формирование у обучающихся навыков в проведении научно-исследовательской работы в медицине.

В частности рассматриваются вопросы методологии планирования медико-биологических исследований, организации проведения клинических исследований и этические аспекты медицинских исследований. Значительное внимание уделяется методам обработки результатов исследования и формирования клинических выводов.

## Практическое занятие №1

### МЕТОДОЛОГИЯ ПЛАНИРОВАНИЯ НАУЧНЫХ ИССЛЕДОВАНИЙ В МЕДИЦИНЕ

**1. Цель работы:** сформировать у студентов представление о планировании научных исследований для медицинских приложений.

**2. Информационные материалы к занятию.**

Научное исследование включает в себя несколько блоков взаимосвязанных этапов. Первый — предплановые исследования, составление и утверждение плана НИР. Второй включает собственно процесс исследования (сбор материалов, характеризующих изучаемую проблему, накопление фактических данных о ней, их систематизация, выработка определенных представлений о проблеме). Третья часть исследования — оформление результатов научного поиска (интерпретация, отчет, публикация).

Первой и, возможно, важнейшей составной частью научно-исследовательской работы является предплановое исследование, которое включает выбор темы, выдвижение рабочей гипотезы, патентно-информационную проработку планируемой НИР, составление рабочего плана НИР, освоение методик.

Выбор темы исследования — крайне сложный творческий процесс, во многом обеспечивающий успех научной деятельности. Тема исследования должна быть актуальной, отличаться новизной, иметь научно-практическое значение. Поскольку выполнение научных исследований требует больших материально-технических и временных затрат, творческих усилий, выбор научной темы представляется важнейшим этапом НИР.

Выбор темы исследования целесообразно начинать с изучения опыта предшествующих поколений ученых. С этой целью необходимо изучить исследуемую проблему не только по источникам литературы, но и овладеть современной методической базой, знать, какие исследования проводятся в ведущих научных школах мира. Таким образом, на первом этапе планирования главное — исключить возможность дублирования и обеспечить достаточно высокий уровень собственных знаний по определенной проблеме. На втором этапе планирования следует найти оригинальный подход в постановке задачи исследования, а затем и в ее решении, используя новые методические подходы, области их использования и т.д.

В качестве приемов, позволяющих осуществлять поиск новых решений в научных исследованиях, можно использовать:

1. Ознакомление с текущей и ретроспективной информацией, преимущественно обзорно-аналитического характера.

2. Ознакомление с новейшими достижениями науки и техники в смежных областях знания и перенос в них методов и методических приемов из своей области знания или, наоборот, заимствование их из смежных областей.

3. Разработку новых, более совершенных методов исследования, технологий, приборов или композиционных материалов, открывающих широкие перспективы их использования в научных исследованиях.

4. Методы системного анализа, наукометрические методы и т. д.

Тема НИР должна отличаться: 1) новизной исследуемого вопроса и получаемых результатов; 2) актуальностью; 3) научно-практической значимостью; 4) доказательностью выдвигаемых исследователем положений, вытекающих из полученных результатов.

Таким образом, бездоказательные утверждения исследователя, равно как и отсутствие новизны, актуальности и научно-практической значимости полученных результатов, не могут расцениваться как результаты научных исследований.

Научно-исследовательские работы подразделяют на инициативные и заказные, индивидуальные и коллективные (комплексные). Инициативные исследования чаще всего носят поисковый характер и, как правило, финансируются за счет внутренних средств научного учреждения, полученного гранта и т. д. Заказная тематика входит в состав различных государственных программ или выполняется по договору с другими НИИ (ведомством) и соответственно финансируется Минздравом, научными фондами и другими заказчиками. Индивидуальные НИР, как правило, являются фрагментами кандидатских или докторских диссертаций, разделом комплексного исследования или хоздоговора. Комплексные (коллективные) исследования формируются в рамках программно-целевого планирования для решения крупных научных проблем и призваны ликвидировать мелкотемье, более рационально использовать имеющиеся ресурсы. Каждая научно-исследовательская работа подлежит обязательной государственной регистрации как при планировании, так и при ее завершении для исключения дублирования и параллелизма в работе.

Процесс выбора темы НИР зависит от того, является ли тема заказной или инициативной. В первом случае тема НИР, цели, задачи, сроки и ресурсы должны определяться заказчиком. Исполнитель и научный руководитель лишь представляют заказчику предложения по решению поставленной задачи с указанием своих квалификационных и ресурсных возможностей для проведения конкурсного отбора.

План НИР (может быть оформлен в виде договора между заказчиком и исполнителем) включает в себя обоснование, календарный план, калькуляцию стоимости, штатное расписание, отчет о проведенном патентно-информационном исследовании с заключением об исключении дублирования и целесообразности планирования НИР.

При составлении плана НИР и его экспертизе важно определить требуемую материально-техническую базу (объем предполагаемого эксперимента (наблюдения)), число исполнителей и их квалификационный состав, сроки выполнения темы. Все это в конечном счете определяет требуемый объем финансирования. В соответствии с этим в приложении к плану НИР (договору) — калькуляции сметной стоимости — перечисляются все планируемые затраты: заработная плата, начисления в фонд заработной платы и отчисления в фонд занятости, затраты на спецоборудование, материалы, командировочные и накладные расходы. Объективность сроков исследования определяется по календарному плану работы, который должен включать перечисление всех этапов и видов работы с указанием сроков и конкретных исполнителей. Срок выполнения можно рассчитать с помощью сетевого графика, обычно он не превышает 3 года.

Тема НИР, как правило, планируется на научную группу или лабораторию. Это и определяет численность исследователей и фонд заработной платы. Соотношение численности научного и вспомогательного состава зависит от характера планируемой работы и может колебаться от 1:1 до 1:4.

Тема НИР должна широко обсуждаться коллективом лаборатории, Ученого совета, экспертных Советов на всех этапах планирования, выполнения и завершения темы.

В процессе выбора темы НИР и ее планирования исполнитель должен критически оценивать свои реальные технические, кадровые, финансовые, квалификационные и иные возможности. Соответствие ресурсной, квалификационной, информационной и методологической базы исполнителя поставленным целям и задачам исследования позволяет выполнить план НИР в полном объеме и в срок.

Патентно-информационные исследования являются одним из основных разделов предплановых исследований. Они проводятся с целью определения научно-технического уровня и тенденций развития в планируемой области знания, патентоспособности предлагаемой к планированию НИР и исключения дублирования. Патентно-информационные исследования осуществляются на основе анализа патентной, научно-технической, коммерческой и других видов информации.

Патентно-информационные исследования осуществляются разработчиком совместно с сотрудниками патентно-информационного подразделения. Патентно-информационное исследование проводится в следующем порядке: определяется охраноспособность планируемой НИР, разрабатывается план и регламент исследования, осуществляются поиск, отбор, систематизация и анализ информации, составляется отчет о проведенном патентно-информационном исследовании с перечнем выявленных информационных источников-аналогов, оформляется заключение о целесообразности планирования (с обоснованием новизны, актуальности, технического уровня, отличительных особенностей по сравнению с аналогами) и исключении дублирования. По завершении темы НИР патентно-информационное исследование проводится в более углубленном виде: определяются новизна, технический уровень, патентная чистота разработанного объекта, целесообразность его правовой защиты за рубежом и продажи лицензий.

Особенность научных исследований в области медицины заключается в том, что исследователю часто доступна лишь косвенная информация о состоянии функций органов и систем человеческого организма, а любое инструментальное исследование может представлять определенную угрозу здоровью.

Поскольку жизнедеятельность человека определяется сложнейшими биохимическими и нейрофизиологическими процессами, зависящими от многочисленных эндогенных и социальных факторов, условий внешней среды, результаты искусственно воспроизводимых исследований и экспериментов носят порой сугубо индивидуальный характер и устанавливаются лишь посредством статистического их восприятия. Кроме того, в процессе медицинского эксперимента можно уяснить далеко не все стороны жизнедеятельности целостного организма. Комплексно оценить изучаемые экспериментальным путем явления такого рода возможно только путем накопления большого количества фактов, их сопоставления и статистической обработки, причем очень часто приходится пользоваться результатами других исследователей, опубликованных в научных изданиях.

И, наконец, многочисленные медицинские исследования и эксперименты невозможно проводить на человеке из-за их опасности для жизни и здоровья, а моделирование патологии человека на животных и эксперименты на них не всегда дают адекватные результаты при дальнейших клинических испытаниях.

Результаты НИР во многом определяются правильностью выбранных методов исследования. Наиболее широко применяемым в медицине эмпирическим методом является натурное наблюдение в естественных, клинических или лабораторных условиях. Объектами этого наблюдения могут быть здоровые или больные люди, продукты выделения и ткани живого организма, трупный материал, микроорганизмы и продукты их жизнедеятельности и т. д. Натурное наблюдение

включает всевозможные методики исследования — гистологические, психофизиологические и т. д. Отличительной особенностью методов натурального наблюдения является то, что они осуществляются только в отношении самого объекта научного изучения и только в обычных, естественных условиях его существования. Никакие модели, замещающие объект изучения, изменения условий его обитания не допускаются.

В отличие от натурального наблюдения экспериментальное исследование ставит перед собой задачу изучения явления в строго контролируемых и воспроизводимых условиях при активном воздействии экспериментатора на объект изучения.

Среди экспериментальных методов исследования следует выделить лабораторные опыты, психофизиологические исследования на людях, эксперименты на животных, клинические испытания, натурные испытания опытных образцов или объектов, математическое моделирование.

Приступая к любому экспериментальному исследованию, необходимо прежде всего выдвинуть рабочую гипотезу, на основе которой будет строиться эксперимент, четко определить цели и задачи исследования, составить план и разработать необходимую учетную документацию (протокол опыта, дневник и т. д.). Протокол исследования, как правило, ведется в пронумерованной и пронумерованной тетради, в которую вносят номер и дату проведения опыта, подробные сведения об объекте и методах исследования, полученные результаты. Для большей достоверности и иллюстрированности полученных результатов к протоколу опыта прилагаются рентгенограммы, кардиограммы, микрофотографии и т. д. Все они нумеруются в соответствии с номером опыта.

При экспериментах на животных важную роль играет правильный выбор объекта исследования, поскольку различные виды животных имеют различную реактивность на определенные внешние раздражители и по-разному моделируют патологические процессы, протекающие в человеческом организме. Так, например, собака наиболее часто используется при изучении различных патофизиологических процессов, токсикологические исследования чаще всего проводят на кошках, микробиологические — на белых мышах, крысах, морских свинках, кроликах. Самые ответственные исследования проводят на обезьянах, видовая специфичность которых наиболее близка к человеческому организму. Во всех случаях необходимо, чтобы подопытные животные были здоровыми, однородными (инбредные линии), однополыми, примерно одинаковой массы и возраста. Исключительно важное значение для достоверности полученных данных имеют условия содержания подопытных животных и уход за ними.

Клинические исследования — завершающий этап НИР по разработке новых лекарственных препаратов, методов диагностики и профилактики. Поскольку их проводят на больных и добровольцах, а также учитывая их особую важность, клинические испытания разрешаются в директивном порядке специально уполномоченными государственными органами. При организации клинических испытаний особое внимание должно уделяться подбору пациентов и аналогичной по составу и количеству контрольной группы. При клинических испытаниях новых методов лечения в качестве контроля должны использоваться больные, получающие общепринятое лечение, а не только плацебо.

В соответствии с "Международными этическими требованиями к биомедицинским исследованиям с участием человека" и Международной конвенцией по гражданским и политическим правам все медицинские исследования с участием человека должны строиться на трех этических принципах: уважение к личности, достижение пользы, справедливость. Во всех биомедицинских исследованиях с участием человека (больных или здоровых) исследователь должен получить информированное согласие субъектов, которые будут участвовать в испытании, а если субъект исследования (СИ) не способен дать его — информированное согласие близкого родственника или уполномоченного представителя. Информированное согласие означает согласие компетентного СИ, получившего всю необходимую информацию, адекватно ее понимающего и принимающего решение свободно, без избыточного влияния, побуждения или угрозы. СИ должен получить информацию о целях, методах, длительности исследования, ожидаемом риске или дискомфорте, альтернативных процедурах, степени обеспечения конфиденциальности, возможности в любой момент отказаться от исследования. Субъекту, принимающему участие в медицинском исследовании, должно быть оказано бесплатное лечение в случае повреждений, связанных с исследованием, а также выплачена компенсация за нетрудоспособность или инвалидность. В случае смерти в результате исследования компенсацию должны получить его родственники. СИ может получать плату за неудобства и потерянное время, компенсацию за расходы, понесенные в результате участия в исследовании, но эти выплаты не должны быть столь значительными, чтобы согласие участника испытаний целиком зависело от денежного вознаграждения.

Для оценки эффективности новых методов диагностики, профилактики и лечения, исключения ошибок и правильной интерпретации результатов клинических исследований их необходимо проводить в рамках рандомизированных контролируемых исследований, считающихся "золотым стандартом" для клинических сопоставлений.

Контролируемое клиническое исследование — это проспективное исследование, в котором сопоставляемые группы получают различные виды лечения: больные контрольной группы — стандартное (обычно лучшее по современным представлениям), а больные опытной группы — новое лечение. Важнейшее условие, обеспечивающее надежность контролируемого исследования, — однородность группы больных по всем признакам, которые влияют на исход заболевания (пол, возраст, наличие сопутствующих заболеваний, тяжесть и стадия основного заболевания и т. д.). Учитывая наличие множества взаимосвязанных факторов, определяющих прогноз, а также "скрытых" прогностических факторов, достичь сопоставимости групп наблюдения в наиболее полном объеме можно только при использовании метода случайного распределения пациентов на группы, т. е. рандомизации (random — случайный). Истинная рандомизация предполагает обязательное соблюдение непредсказуемого характера распределения больных на группы (исследователь не может предугадать, в какую группу попадает следующий больной, — "слепой отбор"). Для повышения эффективности рандомизации проводится предварительная стратификация — распределение вариантов лечения осуществляется в однородных группах больных, сформированных по ведущим прогностическим признакам (стратификационная рандомизация).

Однако следует учитывать, что несопоставимость групп наблюдения может возникнуть и после рандомизации (в результате отказа больных от лечения, серьезных осложнений и т. д.). Считается, что если более 80% включенных в исследование больных наблюдались до его окончания, результаты могут быть достаточно надежными. Выбывание из исследования более 20% больных дает основание сомневаться в полученных результатах.

Для снижения вероятности случайных или систематических ошибок опытная и контрольная группы должны включать достаточное число наблюдений. Достаточность определяется предполагаемыми различиями в эффекте лечения между опытной и контрольной группами (чем больше эффект, тем меньшими могут быть группы), видом статистического показателя эффекта терапии (при использовании средних величин исследования группы могут быть меньше), соотношением между числом больных в опытной и контрольной группах (при условии равенства по численности группы могут быть более малочисленными).

Определение оптимального числа случаев наблюдения составляет важный этап планирования эксперимента. Так, в случаях, когда результаты исследования будут выражаться качественно, требуется гораздо большее число наблюдений, чем при использовании количественных оценок, выраженных среднеарифметическими величинами. Кроме этого, следует помнить, что малое число исследований уменьшает их точность и достоверность, однако при большом числе исследований вместе с достоверностью возрастают и стоимость эксперимента, и сроки его

проведения. Для увеличения точности исследования в 2 раза необходимо увеличить число наблюдений в 4 раза. При этом число наблюдаемых случаев в контрольной и опытных группах не обязательно должно быть одинаковым. Число случаев, необходимых для проведения эксперимента, определяется при планировании НИР в каждом конкретном случае индивидуально по специальным формулам, описанным в ряде справочников по медицинской статистике.

В связи с тем что в одном учреждении нередко невозможно осуществить рассчитанное количество наблюдений, проводят кооперированные исследования — систему научных, научно-организационных и практических мероприятий, проводимых по единой программе и методике одновременно в нескольких учреждениях.

При планировании клинического исследования необходимо четко определить следующие положения: рационально ли определена "норма" — отсутствие болезни и обоснована ли "точка разделения" здоровых и больных; оценены ли воспроизводимость, чувствительность и специфичность используемых методов и т. д.

При клинической оценке новых методов диагностики необходимо прежде всего сопоставить предлагаемый метод с референсным, т. е. с самым надежным из соответствующих методов. При правильной организации исследования оценка результатов нового метода должна производиться "вслепую" — так, чтобы специалист не знал результатов других анализов у этого пациента и прежде всего результатов референсного метода. В противном случае неизбежны систематические ошибки типа "приближения к предполагаемому диагнозу", поскольку специалист, зная предполагаемый диагноз, непреднамеренно склонен к предвзятости при оценке результатов, особенно в случаях, когда возможно двоякое их толкование. Чтобы избежать таких ошибок, можно использовать двухэтапную оценку. На первом этапе специалист рассматривает материалы (ЭКГ, лабораторные препараты и т. д.) под номером исследования и лишь по необходимости проводит оценку с учетом дополнительных данных.

Методы исследования должны быть предварительно апробированы в ходе предплановых исследований. При этом следует убедиться в их чувствительности, специфичности, воспроизводимости, наличии реактивов на весь объем исследования, соответствии метода уровню мировых стандартов. Этим же требованиям должна отвечать и аппаратура. При отсутствии у исследователя дорогостоящей техники не обязательно ее закупать: можно ее арендовать или использовать в ходе кооперированных исследований.

При планировании НИР следует учитывать и заключительный этап научного исследования — анализ и интерпретацию полученных результатов с целью

выявления имеющихся закономерностей. Анализ статистических данных производят путем соответствующей математической обработки полученных результатов, приемы и способы которой подробно описаны в специальных руководствах по медицинской статистике. В последние годы статистическую обработку данных стали проводить на ПЭВМ с использованием специальных пакетов программ (например, Statgraph и др.), позволяющих быстро рассчитать средние величины и относительные коэффициенты, выявить характер и силу связи, степень достоверности, построить аналитические таблицы, диаграммы и графики.

В целом на процессы предплановых исследований, анализа и обобщения полученных результатов и их публикацию следует выделять не менее 10% планируемых ресурсов (временных, финансовых, кадровых), а на внедрение научных результатов в практику — не менее 25%. В то же время для многих НИР стадия внедрения планируется как самостоятельное исследование, а оформление научных результатов в виде публикаций, изобретений и т. д. осуществляется учеными уже после завершения НИР в процессе выполнения последующих исследований, нередко вытекающих из предыдущих результатов.

### **3. Порядок выполнения работы**

3.1. Перед практическим занятием ознакомьтесь с информационными материалами к занятию.

3.2. Преподаватель ставит вопрос по теме занятия и наполняет его конкретным содержанием по структуре данных и исследований патологии или способу лечения.

3.3. Выделяется время на подготовку вопроса и по желаниям студента или выбору преподавателя один из студентов раскрывает и конкретизирует суть вопроса.

### **4. Вопросы для самоконтроля**

1. Расскажите об основных этапах выбора темы исследования
2. Дайте классификацию научно-исследовательских работ
3. Как составляется план НИР?
4. В чем сущность проведения патентных исследований?
5. Охарактеризуйте контрольное клиническое исследование
6. В чем сущность кооперированных исследований?

## Практическое занятие №2

### Организация клинических исследований.

**1. Цель работы:** ознакомить студентов с организацией клинических исследований и сформировать навыки их проведения для типовых медицинских задач.

#### **2. Информационные материалы к занятию.**

Сведения, касающиеся правил проведения клинических исследований достаточно полно представлены в специализированной литературе, различных медицинских сайтах и сайтах законодательных органов, регулирующих юридические аспекты данного вопроса. Наиболее полную информацию о клинических исследованиях (их характеристика, правила организации и проведения, методы объективизации результатов и пр.) можно получить в следующих документах и руководствах:

- Федеральный закон от 12.04.2010 N 61-ФЗ «Об обращении лекарственных средств»;
- Федеральный закон от 22.08.2004 N 122-ФЗ «Страхование гражданской ответственности лиц, осуществляющих проведение клинических исследований лекарственных средств»;
- Национальный стандарт РФ ГОСТР 52379-2005 «Надлежащая клиническая практика»;
- Приказ МЗ РФ №266 от 19.06.2003 г. «Правила клинической практики в РФ»;
- Приказ МЗ РФ №267 от 19.06.2003 г. «Правила лабораторной практики в РФ»;
- Правила проведения исследований биоэквивалентности лекарственных средств.

Часть 1. Фармакокинетические аспекты.

Часть 2. Организационные аспекты испытаний на здоровых добровольцах.

• Основы клинической фармакологии и рациональной фармакотерапии./Под ред. Ю.Б. Белоусова, М.В. Леонова.- М. Литтерра.-2002.- 356с.

• Кобалава Ж.Д., Моисеев В.С. Клинические исследования по артериальной гипертонии.-М.- 2003.- 158с.

• Страус Ш.Е. и др. Медицина, основанная на доказательствах. – М., ГЭОТАР.- 2010.-320с.

В настоящих методических рекомендациях приведены основные положения, имеющиеся в вышеуказанных приказах и монографиях, касающиеся методологии проведения клинических исследований на примере работы с лекарственными средствами.

#### **2.1. Клинические исследования.**

**Клинические исследования** — это исследования, организованные для оценки различных вмешательств, условия проведения которых направлены на устранение влияния систематических ошибок на результаты.

**Цель клинических исследований** - оценка терапевтической или профилактической эффективности и переносимости нового лекарственного

средства (ЛС), установление наиболее рациональных доз и разработка схем его применения, сравнение с уже существующими препаратами.

Общепринятым во всем мире является выполнение требований качественной клинической практики при проведении любых клинических испытаний лекарственных средств.

Основными элементами Качественной клинической практики (GCP) являются:

защита прав субъекта исследования

- информированное согласие

- этический комитет

спонсор – монитор – исследователь

стандартные рабочие инструкции

сообщение побочных явлений

## **2.2. Фазы клинических исследований.**

В процессе клинических исследований новых лекарственных средств выделяют 4 взаимосвязанные фазы.

Фаза I - клинико-фармакологические исследования

Фаза I служит для изучения фармакологических свойств и подтверждения

безопасности нового ЛС у здоровых добровольцев. Ее задача - оценить

переносимость исследуемого препарата, установить наличие у него

терапевтического действия и создать рациональную основу для выбора доз и схем

применения. Исследования проводят на ограниченном числе добровольцев (5-10

человек).

### **Фаза II — пилотные и контролируемые исследования**

Цель фазы II - показать эффективность и безопасность ЛС на определенном контингенте больных, включающем 100—200 человек, и установить оптимальные режимы дозирования.

Фаза II включает 2 вида клинических исследований:

- пилотные исследования (фаза IIa);
- контролируемые исследования (фаза IIb).

**Пилотные исследования** проводятся с целью поиска дополнительных фармакологических свойств ЛС у больных. В ходе этих исследований выявляется необходимость дальнейших контролируемых исследований в этом направлении.

**Контролируемые исследования.** Они предусматривают наличие контроля или контрольной группы (контроль исходного состояния, плацебо-контроль, активный контроль), что позволяет избежать погрешностей связанных с влиянием различных зависимых или независимых факторов на результаты лечения. Основная и контрольная группы не должны различаться по полу, возрасту, тяжести заболевания и другим факторам, что достигается с помощью метода рандомизации.

Используется четыре типа контроля:

- контроль исходного состояния;
- плацебо-контроль;
- активный контроль;
- контроль по архивной статистике.

### **Контроль исходного состояния**

Контроль исходного состояния (baseline control) в той или иной форме используется при проведении всех клинических исследований. Клинические измерения до начала лечения (измерения исходного состояния) производятся у каждого испытуемого до проведения лечения с целью получения исходных данных, которые затем будут сравниваться с результатами после окончания лечения. Оценка исходного состояния может осуществляться либо с учетом безлекарственного периода лечения, либо, что предпочтительнее, с учетом периода лечения плацебо перед проведением рандомизации.

### **Плацебо-контроль**

Способ плацебо-контроля (placebo control), известный как технология "негатив-ного контроля" (negative control), заключается в назначении испытуемому плацебо - неактивного вещества, которое невозможно отличить от экспериментального ЛС ни по каким признакам (по внешнему виду, вкусу, запаху).

### **Активный контроль**

В активный (позитивный) контроль [active (positive) control], включается лечение с применением лекарственного средства, которое является эффективным относительно исследуемого показания (ЛС активного контроля). Как и в случае применения плацебо, ЛС активного контроля не отличается от изучаемого препарата.

### **Контроль по архивной статистике**

Контроль по архивной статистике, или исторический контроль (historical control), позволяет сравнить экспериментальный курс лечения с существующими данными об исходах конкретного заболевания.

При многолетних наблюдениях способ контроля по архивной статистике используется в том случае, когда не существует другого эффективного метода лечения известной патологии или редкого заболевания.

**Контролируемые исследования** часто носят сравнительный характер: сравнение эффективности и переносимости с другими препаратами, сравнение эффективности разных доз ЛС, сравнение переносимости и качества жизни и др. В качестве препарата сравнения может использоваться плацебо или другое ЛС, а также разные дозы одного препарата.

### **Фаза III - расширенные клинические исследования**

Главная цель расширенных клинических исследований - получить дополнительную информацию об эффективности и безопасности новых ЛС у больных в условиях, максимально приближенных к клинической практике. В ходе этих исследований изучаются особенности действия препарата у больных с сопутствующими заболеваниями, нарушениями кровообращения, функции печени и почек, оцениваются терапевтические преимущества, влияние на качество жизни, выявляются относительно редко встречающиеся побочные реакции и особенности взаимодействия нового препарата с другими ЛС, а также изучаются фармакоэкономические аспекты.

Фаза III клинических исследований завершается представлением препарата на регистрацию, в связи с чем она делится на две фазы:

- **фаза IIIa** охватывает период до представления заявки на регистрацию;
- **фаза IIIb** - период с момента подачи заявки на регистрацию и до окончательной регистрации препарата.

### **Фаза IV - пострегистрационные исследования**

После разрешения применения нового препарата в медицинской практике и его внедрения возможно проведение фазы IV - клинических исследований, целью которых является изучение возможностей для расширения показаний к применению ЛС, усовершенствование режимов назначения и схем лечения, а также длительное наблюдение (в течение многих лет).

Особое внимание обращается на сбор и анализ информации о побочных действиях изучаемых ЛС. Еще более важным представляется изучение отдаленной эффективности препаратов на выживаемость, которое проводится у многих сотен и тысяч больных.

Контролируемые исследования фазы IV могут включать разное число больных (от нескольких десятков до тысяч) и быть:

- ретроспективными,
- проспективными.

**Ретроспективные исследования** проводятся на основе прошлого опыта применения разных ЛС или видов терапии по данным историй болезни.

**Проспективные исследования** планируются на перспективу (до начала набора больных) и проводятся по общему протоколу в сбалансированных группах больных, что значительно повышает надежность полученных результатов.

Большинство серьезных научных исследований являются пострегистрационными, проспективными.

## **2.3. Группы клинических исследований и их модели**

Все клинические исследования делятся на две основные группы:

# I. Обсервационные - исследования без преднамеренного вмешательства

## 1. Описательные - без контрольной группы сравнения

- описание случаев - это описание одного или нескольких (менее 10) случаев из клинической практики, которые, по мнению исследователя, должны быть представлены вниманию коллег.

- описание серий случаев - это описание нескольких (10 и более) случаев одной патологии, которые, в силу своей схожести, могут быть сгруппированы.

- одномоментные (поперечные) исследования - это проведение оценки распространенности определенных исходов, показателей, факторов риска в определенный момент или в течение короткого промежутка времени.

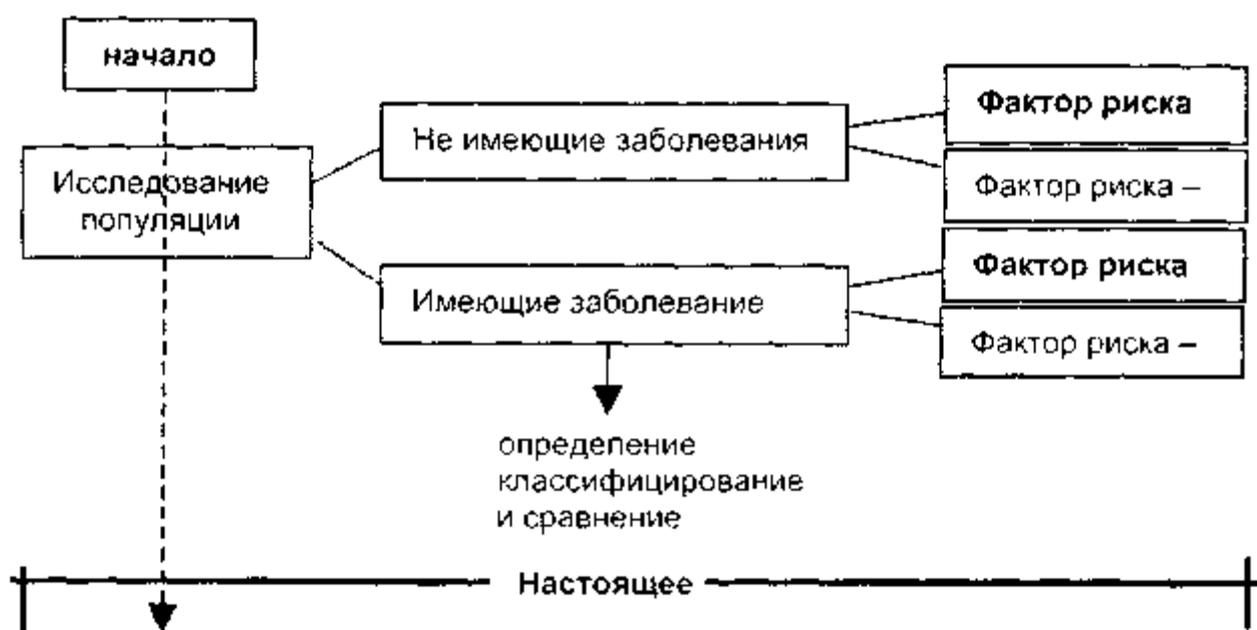
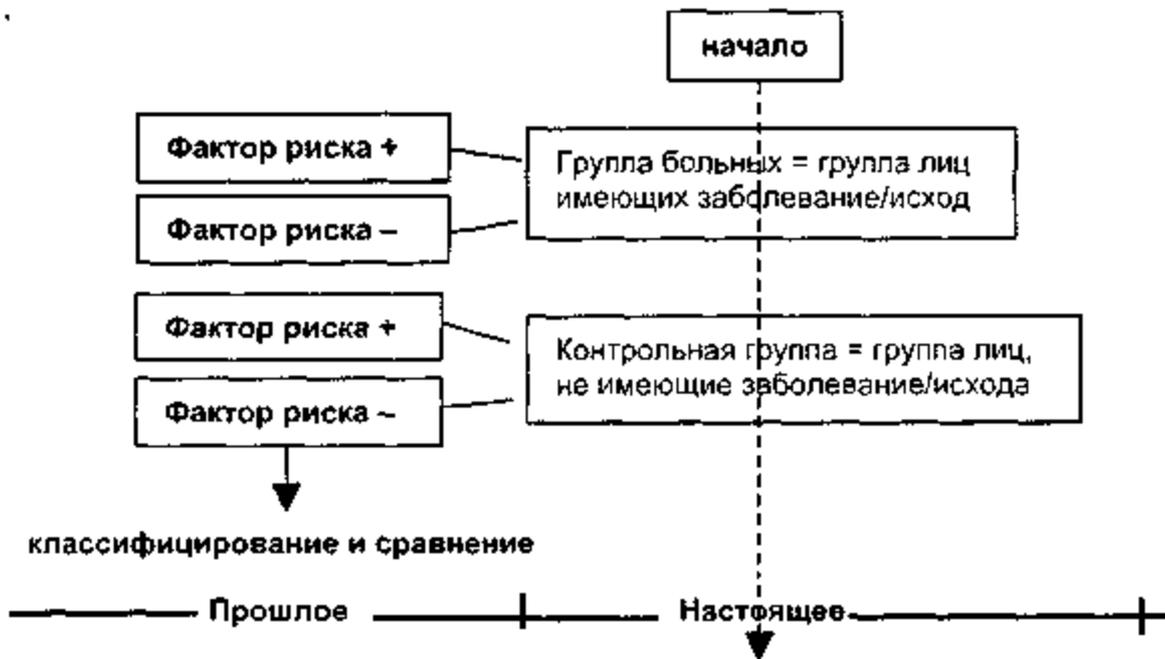


Схема дизайна одномоментного (поперечного) исследования

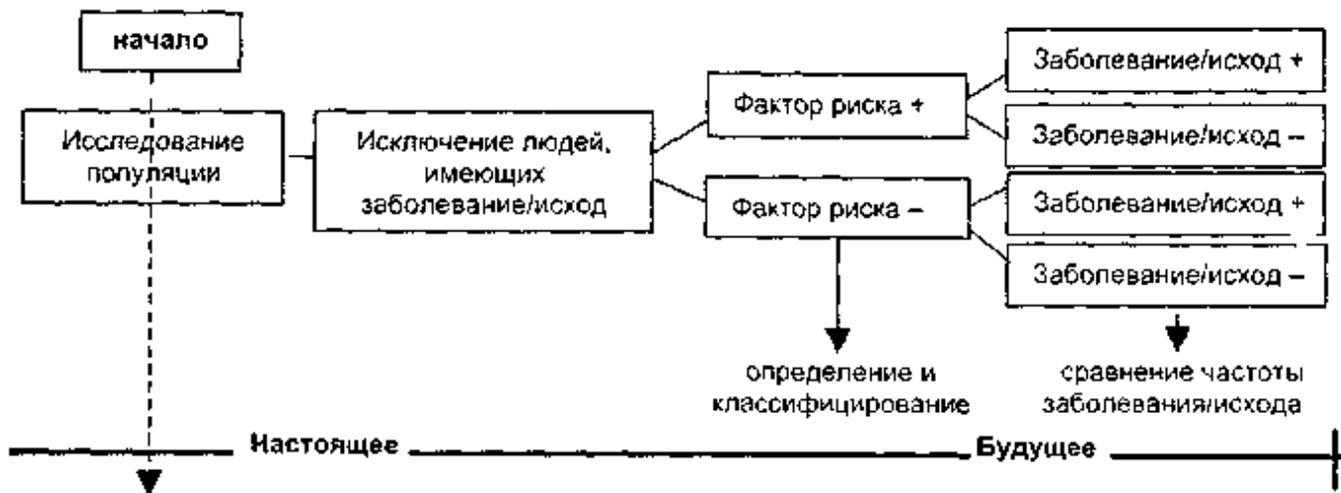
## 2. Аналитические - с формированием контрольной группы

- исследования случай-контроль – это исследование, структура которого предусматривает сравнение двух групп участников с развившимся и не развившимся клиническим исходом с целью выявления различий во влиянии определенных факторов (обычно факторов риска) на развитие этого исхода.



Схематическое изображение исследования случай-контроль

- когортное исследование – это исследование, структура которого позволяет проследить за группой (когортой) участников и выявить различия в частоте развития у них определенных клинических исходов.



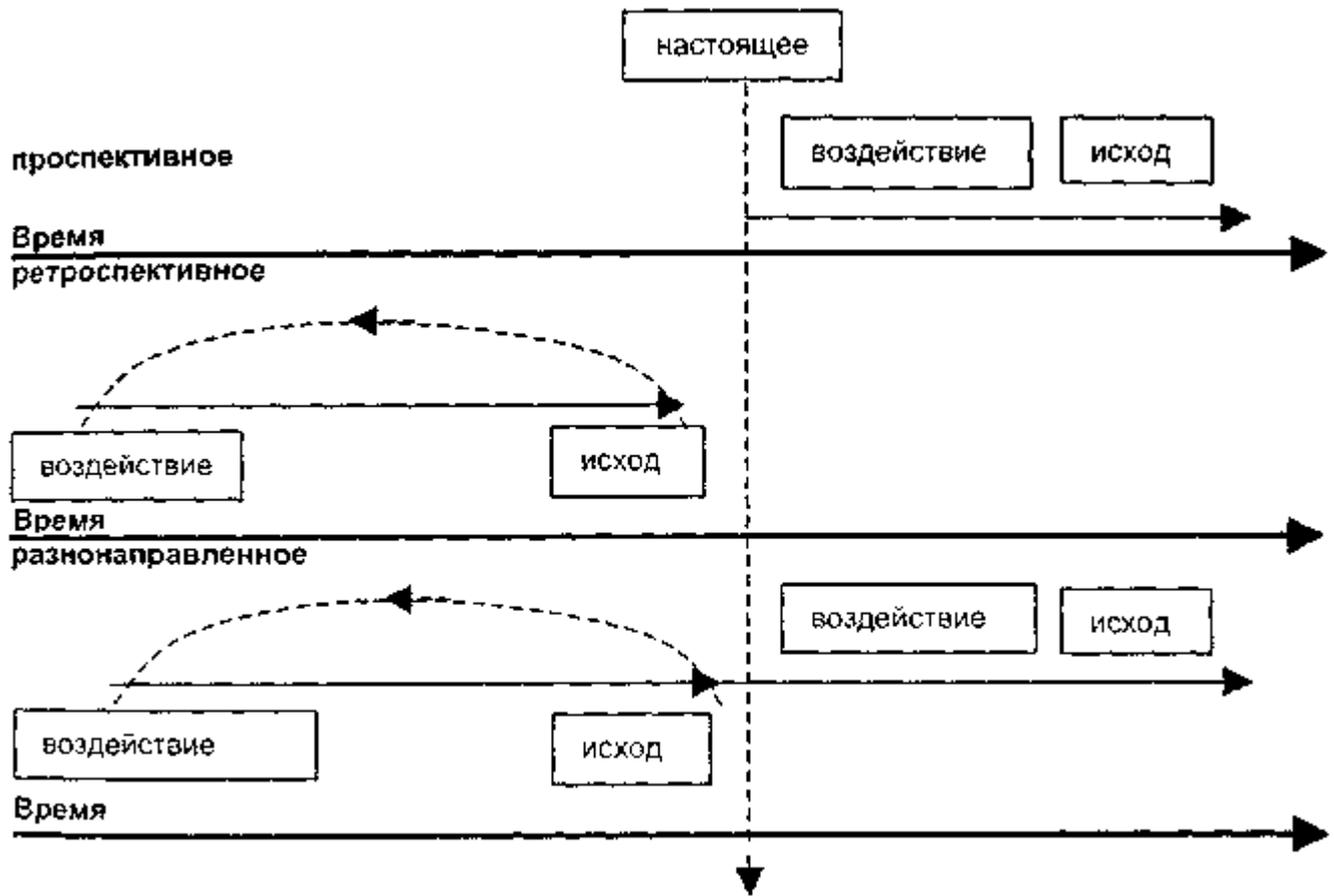
Схематическое изображение структуры когортного исследования (проспективного)

#### Типы когортных исследований

— *Проспективное* — группы пациентов сформированы в настоящее время и будут прослежены в будущем

— *Ретроспективное* — группы пациентов формируются по данным статистики прошедших лет (в условиях неизвестности их исходов), прослеживается последующая судьба (также по данным статистики)

— *Разнонаправленное* — формирование групп осуществляется по данным медицинской документации, наблюдение же больных происходит в настоящем времени (обычно применяется при учете двух исходов)



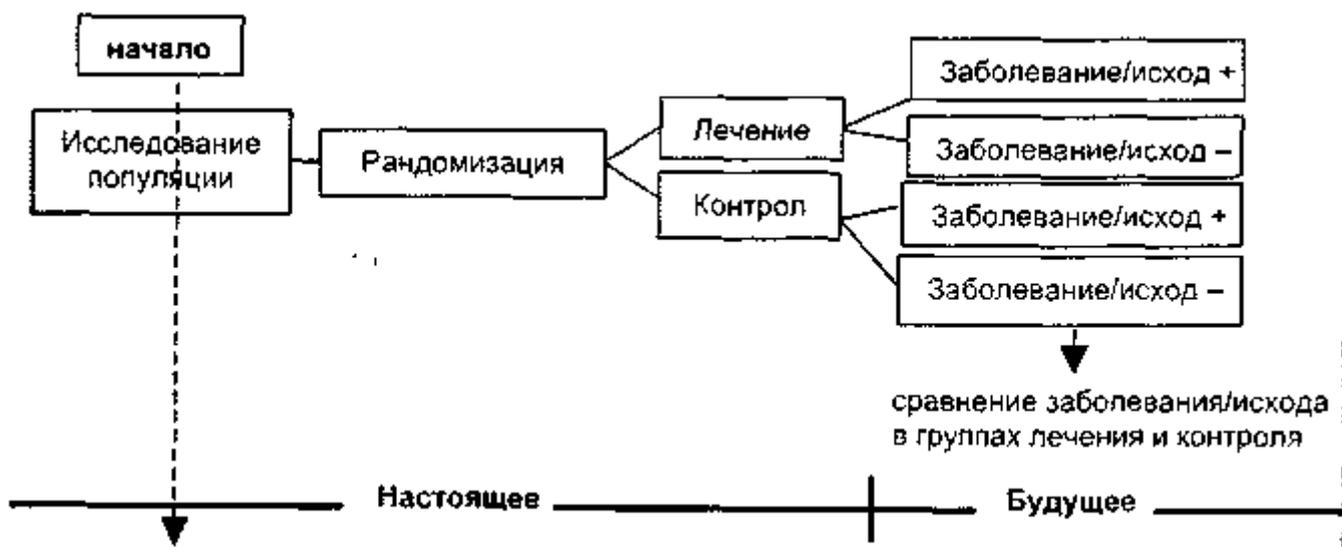
Схематическое изображение типов когортных исследований

**II Экспериментальные** - исследования, в котором изучаются влияния определенного вмешательства

1. *Рандомизированные* - проводится технически правильная рандомизация.

Рандомизированное контролируемое исследование — дизайн исследования, в котором участников в случайном порядке (рандомизированно) распределяют в, как минимум, две группы — одну основную (где применяется изучаемое вмешательство) и контрольную (где применяется плацебо или другое вмешательство).

2. *Нерандомизированные* - рандомизация не проводится или некачественно организована.



Схематическое изображение дизайна рандомизированного контролируемого исследования.

### **Типовые модели клинических исследований.**

Выделяют следующие типовые модели клинических исследований:

- исследования в одной группе (single group study);
- исследования в параллельных группах (parallel group study);
- исследования в группах "перекрестной" модели (crossover group study).

#### **Модель исследования в одной группе**

При проведении исследования в одной группе все испытуемые получают одно и то же экспериментальное лечение. Вместо сравнения результатов в пределах одной группы с результатами контрольной группы эта модель исследования направлена на то, чтобы сравнить результаты лечения по каждому испытуемому с его исходным состоянием до лечения или, по возможности, с результатами контроля по архивной статистике. Таким образом, испытуемых не рандомизируют по группам лечения, и отсутствует необходимость "маскировки" ЛС.

Модель одной группы может быть использована в фазе I клинических исследований, в которой принимают участие здоровые добровольцы, но обычно не используются в фазе III испытаний. Тем не менее такие модели полезны при проведении некоторых экспериментальных курсов лечения.



Схема модели исследований в одной группе

#### **Модель исследования в параллельных группах**

При проведении исследований в параллельных группах испытуемые двух или более групп получают различные курсы лечения или различные дозы ЛС. Для достижения статистической достоверности испытуемые распределяются по группам методом случайной выборки.

Модели исследований в параллельных группах считаются оптимальными для определения эффектов лечения и формулирования выводов на основе полученных результатов. В большинстве случаев клинические исследования проводятся в параллельных группах. Контролирующие органы чаще всего отдают предпочтение этой модели исследования ЛС, которая в основном применяется в фазе III испытаний.

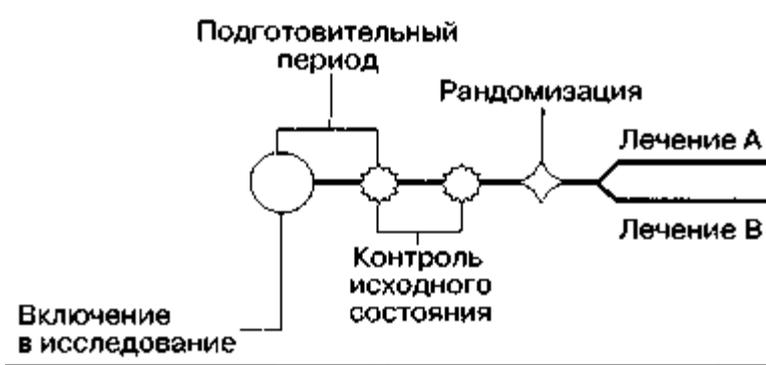


Схема модели исследования в параллельных группах

В определенных ситуациях клинического тестирования в параллельных группах могут быть использованы два варианта простых исследований:

- факториальные модели;
- неоднородные модели.

Модели реализуются в параллельных группах, и полученные данные обосновываются специфическими приемами сравнения

**Факториальная модель** - это модель исследований, в которых принимают участие несколько параллельных групп. Исследования на этой модели полезны, когда предстоит изучить действие комбинации различных препаратов и разработать схемы лечения

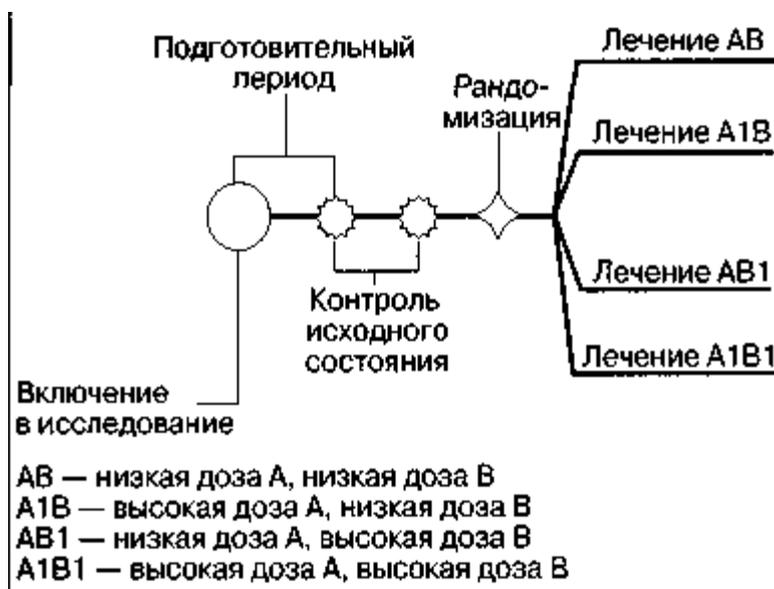


Схема факториальной модели исследований

**Неоднородная (прерываемая) модель "прекращения терапии"** [withdrawal (discontinuation) design] - это вариант исследований в параллельных группах, где все испытуемые вначале получают изучаемое лечение; затем для продолжения изучаемого лечения пациенты с соответствующими реакциями рандомизируются в группы с применением метода слепого исследования с двойным контролем или

использованием плацебо. Данная модель обычно используется для оценки эффективности экспериментального лечения путем прекращения приема препарата сразу после появления реакции и регистрации рецидива или ремиссии.

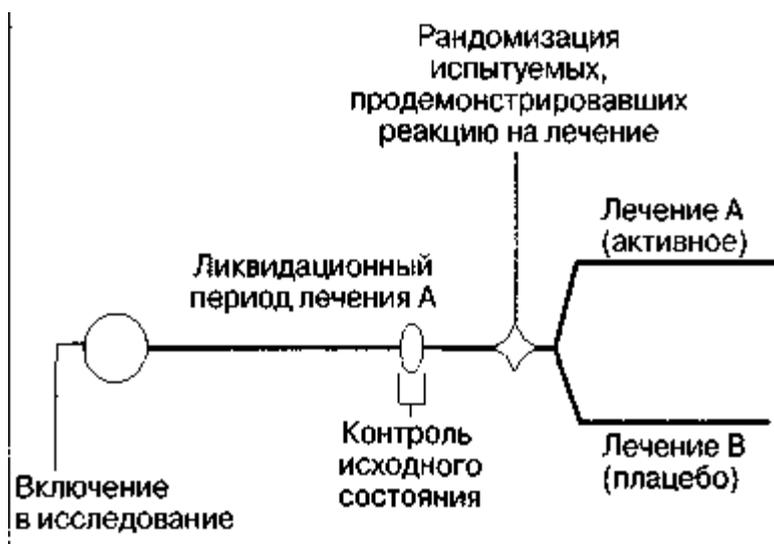


Схема неоднородной модели исследований

### Перекрестная модель

В отличие от исследований в параллельных группах, перекрестные модели позволяют оценить эффекты как изучаемых ЛС, так и сравнительных курсов лечения на одних и тех же испытуемых.

Испытуемых рандомизируют в группы, в которых проводят одинаковое курсовое лечение, но с различной последовательностью. Как правило, между курсами необходим ликвидационный период, чтобы пациенты вернулись к исходным показателям, а также чтобы исключить нежелательное влияние остаточных явлений предшествующего лечения на эффекты последующего. Ликвидационный период необязателен, если анализы индивидуальных реакций испытуемого ограничиваются их сравнением в конце каждого курса, а период лечения длится достаточно долго.

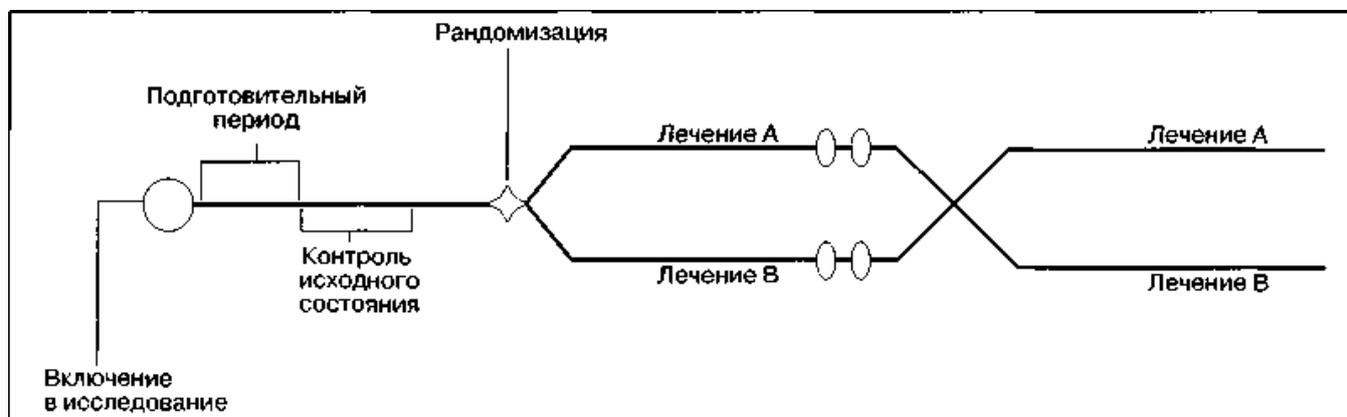


Схема "перекрестной" модели исследований

В некоторых перекрестных моделях используют предварительное "перекрещивание", которое означает, что пациенты, исключенные из исследований на

стадии лечения, могут быть переведены в группы альтернативного лечения раньше запланированных сроков.

Перекрестные модели обычно используют для изучения фармакокинетики и фармакодинамики, когда надо рассматривать изменчивость эффекта внутри популяции испытуемых. Кроме того, справедливо допущение, что эффекты первого курса не оказывают влияние на второй в фармакокинетических и фармакодинамических исследованиях, имеющих достаточный ликвидационный период.

Если имеется явная позитивная корреляция в группах пациентов при снятии характеристик ответных реакций, то перекрестные модели являются более экономичными по сравнению с моделями параллельных групп, поскольку в этом случае требуется меньшее количество испытуемых. При относительно постоянных клинических условиях перекрестная модель является эффективной и надежной. Поскольку все испытуемые получают изучаемое лекарственное средство, то "перекрестные" исследования также эффективны и для оценки безопасности.

## 2.4. Способы контроля погрешности.

**Погрешность** — это воздействие любого фактора, который систематически приводит к неправильным заключениям в отношении эффектов применяемого лечения.

Наиболее распространенным источником погрешностей являются ожидания исследователей и испытуемых. Погрешность в результате неоправданных ожиданий пациента или исследователя, а также погрешность по другим причинам уменьшается при использовании стандартных способов контроля. Методики контроля включают использование:

- контрольных групп;
- метода слепого исследования;
- рандомизации (со стратификацией или без нее).

**О применении контрольных групп** указано выше.

### **Слепое исследование**

Слепое исследование (blinding) предполагает, что испытуемый не имеет информации о методах лечения и, таким образом, исключает влияние на результаты лечения субъективных факторов. В отличие от слепого исследования, открытый метод подразумевает информирование пациента о характере назначенного лечения.

Исследование называется **простым слепым** (single-blind), когда только пациент "слеп" относительно назначенного лечения, а исследователь знает, какое лечение было назначено.

Если пациент и исследователь оба "слепы" относительно лечения, исследование называют **слепым с двойным контролем или двойным слепым** (double-blind). При проведении испытания с двойным контролем заказчик не должен раскрывать информацию о сущности лечения членам группы, которые осуществляют исследование, проводят мониторинг или обрабатывают данные.

Метод слепого исследования требует, чтобы все условия проведения испытаний были идентичны для всех испытуемых и все нюансы лечения были неразличимы.

### **Рандомизация**

Технология рандомизации предполагает назначение испытуемых в группы методом случайной выборки и одинаковое распределение переменных, которые могут оказать влияние на исходы лечения в группах. Рандомизированные контролируемые исследования сходны с когортными исследованиями, за исключением наличия такого важного момента, как рандомизация в отношении проводимого воздействия

#### *Преимущества рандомизации*

Исключает систематические ошибки при назначении лечения, что является более важным преимуществом рандомизированных контролируемых исследований. Исключает систематическую ошибку, связанную с выбором лечения для пациента. Исключает систематическую ошибку, вследствие воздействия вмешивающихся факторов. Обеспечивает непредсказуемость порядка

распределения участников в группы. Обеспечивает ослепление исследователей, участников и аналитиков, в том числе, и при назначении плацебо. Позволяет использовать теорию вероятностей для расчета правдоподобия того, что любые различия в исходах между группами обусловлены слу-чайностью.

#### *Методы рандомизации*

- Простая (неограниченная) рандомизация
- Ограниченная рандомизация
- Дизайн «подбрасывания монетки» со смещением
- Дизайн «урна»
- Стратифицированная рандомизация

#### *Простая (неограниченная) рандомизация*

Аналогична киданию кубика, подбрасыванию монеты, или вытаскиванию карты из колоды. На практике используют список случайных чисел или специально разрабо-танную компьютерную программу случайных чисел. Но при небольшом размере выборки при данном методе может произойти нарушение баланса между группами. При выборке более 200 человек эта возможность сводиться к минимуму

#### **Пример 1.**

Следуя по таблице случайных чисел, всем нечетным номерам вы определяете вмешательств А, всем четным - Б.

#### **Пример 2.**

Номера 00-49 могут определять вмешательство А, а 50-99 — вмешательство Б.

#### *Ограниченная рандомизация*

Обеспечивает случайное распределение участников в равные группы, в зависимости от общего количества включенных.

Достигается формированием блоков. Например, если блок равен 6, значит из каждых 6 участников, включенных подряд, 3-е будут распределены в группу вмешательства А, другие 3 — в группу вмешательства Б. Порядок же распределения участников по группам сохраняется случайным.

Рандомизация, при которой используются конверты, половина из которых приписывает вмешательство А, а другая — вмешательство Б, также является ограниченной

#### *Дизайн «смещенной монеты» со смещением*

Производится простая рандомизация участников между группами с определенной вероятностью (например, в 2-е группы с вероятностью 0,5/0,5). Когда различие между группами достигает определенного предела, вероятность распределения в группы оставшихся участников меняют (например, на 0,6/0,4).

#### *Дизайн рандомизации «урна»:*

Обозначается как UD ( $\alpha$ ,  $\beta$ ), где  $\alpha$  — число синих и зеленых шаров, изначально находящихся в урне, и  $\beta$  — число шаров, добавляемых в урну в процессе рандомизации.

Например, при рандомизации UD (2,1) в урне изначально содержится 2 синих шара и 2 зеленых шара — вероятность 0,5/0,5 ( $\alpha = 2$ ). Синий шар предписывает вмешательство А, зеленый — Б. Шары перемешиваются, один шар извлекается случайным образом, он определит группу вмешательства 1-ого участника. Если из урны был извлечен синий шар, то к ее изначальному содержимому добавляется 1 зеленый шар (дополнительный) ( $\beta = 1$ ). Следовательно, перед вторым изъятием в урне будут находиться 2 синих шара и 3 зеленых шара — вероятности 0,4/0,6. Если опять будет извлечен синий шар, тогда перед третьим выбором в урне будут 2 синих и 4 зеленых шара — вероятности 0,33/0,67. Этот процесс позволяет достигнуть баланса между формируемыми группами.

#### **Стратификация (стратифицированная рандомизация).**

Иногда формирование групп на основе метода случайной выборки стратифицируется. Стратификация (stratification) обеспечивает распределение испытуемых по группам лечения с учетом факторов, существенно влияющих на исход, например возраста, анамнеза, избыточного веса (когда предполагается, что вес пациента может быть связан с исходом заболевания). Стратификация гарантирует равномерное распределение указанных факторов в группах лечения.

#### **2.5. Параметры воздействия.**

Могут использоваться различные критерии оценки эффективности вмешательства:

##### **1. Окончательные, «жесткие» критерии:**

а) изменение показателей здоровья в группе, на которую направлено действие лекарственного средства или нелекарственной медицинской технологии (смертность, выживаемость, продолжительность жизни, инвалидизация, число сохраненных лет без инвалидности и т.п.);

б) изменение качества жизни, обусловленного здоровьем (например, число сохраненных лет качественной жизни (QALY)).

##### **2. Промежуточные, «суррогатные» критерии:**

а) прямые клинические эффекты (например, сдвиг физиологических и биохимических параметров, на изменение которых направлено действие лекарственного средства или нелекарственного метода лечения - снижение артериального давления, прирост гемоглобина, изменение симптомов заболевания; потеря или восстановление функций);

б) опосредованные клинические эффекты (снижение частоты осложнений, сокращение числа повторных госпитализаций и т.п.).

### **3. Порядок выполнения работы**

**3.1.** Перед практическим занятием ознакомьтесь с информационными материалами

**3.2.** В ходе первого занятия в режиме диалога и ответов у доски разбираются основные теоретические вопросы с конкретными примерами по разделам информационного материала

**3.3.** В конце первого занятия преподаватель предлагает темы рефератов и закрепляет их за конкретными студентами

**3.4.** В ходе последующих занятий в ходе диалога производится формирование разделов реферата с соответствующими расчетами и иллюстрациями.

**3.5.** На предпоследнем и последнем занятии проводятся публичные защиты рефератов и даются ответы на дополнительные вопросы собеседования.

#### **4. Вопросы для самоконтроля**

1. Дайте определение клинических исследований и определите их основные цели.
2. Расскажите о фазах исследования по новым лекарственным средствам
3. Что такое плацебо-контроль и активный контроль?
4. Что такое контроль по архивной статистике?
5. Что такое проспективное исследование?
6. Что такое обсервационные исследования?
7. Нарисуйте структуру когортного исследования
8. Определите модель исследования в параллельной группе и нарисуйте ее схему
9. Нарисуйте схему перекрестных исследований
10. Что такое слепое исследование?
11. Дайте определение рандомизации и назовите соответствующие методы
12. Расскажите о жестких критериях при выборе параметров воздействия
13. Что такое сурогатные выборы параметров воздействия?

## Практическое занятие №3

### ЭТИЧЕСКИЕ АСПЕКТЫ КЛИНИЧЕСКИХ ИССЛЕДОВАНИЙ

- 1. Цель работы:** научить студентов проведению организационных мероприятий обеспечивающих выполнение этических медицинских норм
- 2. Информационные материалы к занятию**

Научная ценность результатов, получаемых в ходе клинических исследований, не должна противопоставляться соблюдению этических норм, направленных на защиту каждого пациента, участвующего в исследовании. Этические стандарты, соответствующие современным требованиям к проведению клинических испытаний, являются результатом международных усилий по согласованию различных методов разработки терапевтических приборов и лекарственных средств (ЛС) и рационализации клинических исследований. Они затрагивают область защиты прав и безопасности испытуемых и должны применяться во всех исследованиях. Возможны лишь небольшие отклонения от этих норм, связанные с особенностями местного законодательства и их толкованием.

Международные научные и исследовательские сообщества совместно отказались от неэтичных и ненаучных методов ведения испытаний. Ряд документов содержит нормативы практики исследований, в которых принимают участие испытуемые, включая клинические испытания ЛС. В перечень этих документов входят:

- Хельсинкская декларация Всемирной медицинской ассоциации от 1964 г., дополненная в 1975, 1983, 1989 и 2000 гг.;
- Международные правила GCP (документы ВОЗ и Европейского экономического сообщества);
- Федеральный закон от 12.04.2010 N 61-ФЗ «Об обращении лекарственных средств»;
- Федеральный закон от 22.08.2004 N 122-ФЗ «Страхование гражданской ответственности лиц, осуществляющих проведение клинических исследований лекарственных средств»;
- Федеральный закон от 27.12.2002 N 184-ФЗ «О техническом регулировании»;
- Национальный стандарт РФ ГОСТР 52379-2005 «Надлежащая клиническая практика»;
- Приказ МЗ РФ №266 от 19.06.2003 г. «Правила клинической практики в РФ»;
- Приказ МЗ РФ №267 от 19.06.2003 г. «Правила лабораторной практики в РФ»;
- Федеральный закон от 27.07.2006 № 152-ФЗ «О персональных данных».

Согласно вышеперечисленным документам при проведении клинических исследований необходимо выполнение двух условий, которые играют важную роль в защите прав и обеспечении безопасности испытуемых:

- получение одобрения этического комитета (ЭК). Любое биомедицинское исследование с участием людей в качестве испытуемых может быть инициировано только после одобрения этим комитетом;

- получение информированного согласия. Пациенты могут включаться в научное исследование только после того, как они получили полную информацию о нем и дали осознанное и добровольное согласие на участие;
- должна обеспечиваться конфиденциальность информации, касающаяся состояния здоровья пациентов.

### 2.1. Принципы этической оценки клинических исследований.

Основные принципы этической экспертизы, реализуемой этическим комитетом, научных исследований на людях включают в себя:

- оценку и минимизацию рисков;
- оценку ожидаемой пользы,
- анализ соотношения риска и пользы;
- рассмотрение информированного согласия и процесса его получения, подбора испытуемых и их стимулирования к участию в исследованиях.

**Этический комитет** – это независимый компетентный орган, рассматривающий каждое клиническое исследование. Цель работы этического комитета - обеспечение защиты прав и состояния здоровья субъектов, подвергающихся клиническим исследованиям (здоровых добровольцев и больных). Без разрешения этического комитета начинать клиническое исследование нельзя.

**Документы, предоставляемые в региональный этический комитет:**

1. Заявление.
2. Протокол исследования (или аннотация диссертации), содержащий дизайн исследования.
3. Информированное согласие пациента.
4. Информация для пациента.
5. Инструкция по препарату или методу вмешательства, разрешающие документы федеральных надзорных органов на их использование.
6. Индивидуальная карта испытуемого.

**Дизайн исследования** – это схематичное представление последовательности действий исследователя по первичному обследованию и отбору пациентов, распределению их по группам (с указанием способа распределения), назначению различных схем терапевтического вмешательства, сроков лечения и методов контроля эффективности и безопасности терапии.

**Добровольное информированное согласие** (согласие информированного пациен-та) гарантирует, что будущие испытуемые понимают характер исследования и могут со знанием дела и добровольно принять решение о своем участии или неучастии. Эта гарантия защищает все стороны: как испытуемого, к самостоятельности которого проявляется уважение, так и исследователя, который в противном случае вступает в противоречие с законом.

### 2.2. Основные положения информированного согласия:

использовать можно только одобренную этическим комитетом версию информированного согласия.

информацию надо представить в письменном и устном виде.

предоставьте достаточное время больному или родственникам для принятия решения.

информированное согласие должно быть получено в письменном виде, раньше любых процедур, связанных с исследованием.

больной/доброволец и исследователь должны каждый собственной рукой подписать и датировать информированного согласия.

больной должен получить копию подписанного информированного согласия.

В последние годы уточнены требования к диссертационным работам о порядке проведения биомедицинских исследований у человека.

Экспертный совет по медицине ВАК Минобразования России обращает внимание руководителей диссертационных советов, что при принятии к рассмотрению и защите диссертаций, тематика которых связана с использованием лекарственных средств (как зарегистрированных, так и новых, находящихся на рассмотрении), методов диагностики и лечения у человека, необходимо проверять их соответствие международным и российским законодательным актам о юридических и этических принципах медико-биологических исследований у человека.

Любая научно-исследовательская работа, связанная с изучением действия лекарственных средств у людей, является клиническим исследованием и должна соответствовать предъявляемым к ней требованиям.

При планировании проведения биомедицинского исследования у людей законодательством регламентировано обязательное получение письменного информированного согласия лиц, участвующих в исследовании.

Все планируемые биомедицинские исследования у людей должны быть одобрены комитетом по этике. Основная задача комитета по этике – защита прав и здоровья испытуемых, а также гарантия их безопасности.

В том случае, если целью запланированной научной работы является изучение нового лекарственного средства, включая изучение новых показаний, режимов дозирования, путей введения зарегистрированных лекарственных препаратов, либо новых методов диагностики и лечения, необходимым является получение разрешения Минздрава России.

Без соблюдения всех вышеуказанных требований диссертация не может быть принята к рассмотрению диссертационными советами.

### **2.3. Отклонения от требований информированного согласия**

В некоторых случаях отказ от получения информированного согласия может быть обусловлен формой проведения исследования (т.е. необходимостью скрыть часть информации от пациента с целью повышения достоверности результатов).ЭК может одобрить такой отказ при следующих условиях.

- исследование связано с риском для испытуемых, не превышающим минимальный риск;
- отказ или изменение процедуры не окажет негативного влияния на права и благополучие испытуемых;
- исследование практически неосуществимо без отказа или изменения процедуры получения согласия;
- при первой удобной возможности испытуемым сообщат соответствующую дополнительную информацию об окончании их участия в исследовании. Такие ситуации возникают при проведении следующих видов исследований:

- эпидемиологические исследования;
- исследования, при которых требуется наблюдать за испытуемыми, которые не знают, что за ними наблюдают, либо не знают, что их поведение фиксируется для исследовательских целей;
- исследования с использованием плацебо, рандомизации и двойные слепые клинические исследования.

Исследователям, изучающим поведение людей, иногда требуется **наблюдать за испытуемыми**. Поскольку поведение испытуемых может измениться, если они будут знать, что за ними наблюдают, исследователи могут просить отмены требований по получению согласия.

Когда конкретный вид лечения для испытуемого будет определяться путем **рандомизации**, испытуемым объясняется процедура выбора понятным для них языком. Простого упоминания о том, что назначение на процедуру будет применяться путем произвольной выборки, математическими методами или при помощи лотереи может быть недостаточно.

Когда испытуемые привлекаются к клиническим испытаниям, в которых они могут получить **плацебо** вместо экспериментального лечения или когда они могут оставаться в неведении, какой именно из нескольких видов лечения они получат, можно говорить об элементах обмана. Большинство комментаторов в

настоящее время придерживается той точки зрения, что если испытуемых предупредить о возможном получении плацебо или о форме проведения исследования, можно говорить об отсутствии обмана в таком исследовании.

**В двойных слепых исследованиях** важно, чтобы испытуемые понимали, что ни они, ни их терапевт, ни исследователи не будут знать, какое лечение они получают. Испытуемые должны понимать, что хотя они могут в любой момент прервать свое участие в исследовании, никто не будет сообщать им никаких сведений о том, какой вид (виды) лечения лучше или хуже, до тех пор, пока исследование не будет закончено.

Одним из условий отклонения от требований по информированному согласию является **сообщение испытуемым всей информации при первом удобном случае после окончания исследования**. В каждом случае решается, должен ли состояться отчет исследователей перед испытуемыми после их невольного участия в исследовании или после их информированного участия в исследовании, включавшем от-дельные элементы обмана.

**Информация для пациента** содержит сведения об общей концепции проводимого исследования, методах и схемах лечебного вмешательства, объеме и частоте обследования, положительных и возможных отрицательных аспектах терапии, соотношении польза/риск, способах страхования пациента.

**Индивидуальная карта испытуемого** представляет собой формализованный бланк, в который вносятся паспортные данные о пациенте, результаты различных методов обследования, критерии контроля эффективности и безопасности проводимой терапии.

## Примеры форм информированного согласия для различных клинических исследований.

### СОГЛАСИЕ С ОБЩИМ ПЛАНОМ ОБСЛЕДОВАНИЯ И ЛЕЧЕНИЯ

Я, \_\_\_\_\_, находясь на лечении в \_\_\_\_\_ отделении ознакомлен лечащим врачом с намеченным мне планом обследования и лечения, включающие:

- 1) взятие крови из пальца и вены;
- 2) внутривенные и внутримышечные инъекции;
- 3) лучевые методы исследования;
- 4) консультации специалистов и возможные инструментальные методы осмотра.

Мне лично лечащим врачом были разъяснены цель, характер, ход и объем, а также способы планируемого обследования.

Я также ознакомлен с планом предлагаемого мне поэтапного медикаментозного лечения и действием данных лекарственных препаратов.

Я также ознакомлен с возможным изменением медикаментозной терапии в случае непереносимости мною тех или иных лекарственных препаратов, изменением состояния моего здоровья, требующего изменения тактики лечения, а также применения лекарственных препаратов и других методов лечения, которые могут быть назначены врачами-консультантами.

Кроме того, со мной были обсуждены последствия отказа от обследования и лечения. Я получил исчерпывающие и понятные мне ответы на вопросы и имел достаточно времени на обдумывание решения о согласии на предложенное мне обследование и лечение.

Подбор и осуществление медикаментозного и других видов лечения доверяю своему лечащему врачу.

\_\_\_\_\_  
\_\_\_\_\_  
(подпись пациента)

« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

СПИСОК  
ДИАГНОСТИЧЕСКИХ И ЛЕЧЕБНЫХ ИССЛЕДОВАНИЙ/МАНИПУЛЯЦИЙ,  
ДЛЯ ПРОВЕДЕНИЯ КОТОРЫХ НЕОБХОДИМО ЛЕЧАЩЕМУ ВРАЧУ  
ОФОРМИТЬ ПИСЬМЕННОЕ  
ИНФОРМИРОВАННОЕ СОГЛАСИЕ ПАЦИЕНТА

1. Установка катетера в периферические и центральные сосуды.
2. Установка катетера в мочевого пузыря и мочеточники.
3. Различные виды диагностических и лечебных пункций полостей и органов (стерильная, плевральная, люмбальная, брюшной полости, суставов).
4. Лучевые методы исследования органов, сосудов и суставов с использованием рентгенконтрастных веществ.
5. Диагностическое выскабливание матки и гистероскопия.
6. Все виды электрокоагуляций.
7. Переливание компонентов крови и кровезаменителей.
8. Подготовка и проведение эндоскопических исследований с возможной лечебно-диагностической биопсией.
9. Плановая электроимпульсная терапия.
10. Велоэргометрия, тредмил-тест.
11. Электроэнцефалография.
12. Все виды биопсий.
13. Все виды хирургических вмешательств.

ИНФОРМИРОВАННОЕ СОГЛАСИЕ  
НА ПРОВЕДЕНИЕ ГЕМОТРАНСФУЗИИ

Я, \_\_\_\_\_  
(фамилия, имя, отчество)

19 \_\_ года рождения, находясь на лечении в \_\_\_\_\_  
отделении \_\_\_\_\_, даю добровольное согласие на проведение  
гемотрансфузии.

Я знаю, что гемотрансфузия – это переливание компонентов крови.

Мне объяснены лечащим врачом суть и цели предстоящих мне  
гемотрансфузий.

Я знаю, что переливание всех компонентов донорской крови сопряжено с  
определенным риском осложнений и заражения вирусами гепатитов, ВИЧ и  
другими гемоконтактными инфекциями, несмотря на то, что каждый контейнер с  
компонентами крови прошел обследование на станции переливания крови на  
наличие возбудителей сифилиса, гепатитов, ВИЧ.

Я понимаю необходимость применения в моем лечении гемотрансфузии,  
которая производится сейчас и/или может возникнуть в дальнейшем до момента  
моей выписки, и даю свое согласие на переливание мне компонентов донорской  
крови в необходимом количестве, которое определяется лечащим врачом.

Я предупрежден о последствиях отказа от предлагаемой мне  
гемотрансфузии.

Я прочитал и понял все вышеизложенное, удовлетворен ответами лечащего  
врача на мои вопросы.

Подпись пациента \_\_\_\_\_

Подпись лечащего врача \_\_\_\_\_

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_ г.

ИНФОРМИРОВАННОЕ СОГЛАСИЕ  
НА ИССЛЕДОВАНИЕ/МАНИПУЛЯЦИЮ

Я,

\_\_\_\_\_

(Фамилия, Имя, Отчество)

19 \_\_ года рождения, находясь на лечении в \_\_\_\_\_ отделении \_\_\_\_\_, даю добровольное согласие на проведение исследования/манипуляции

\_\_\_\_\_,  
которая(ое) позволит уточнить диагноз, оценить состояние моего здоровья, определить ход дальнейшего лечения и прогноз.

Я информирован о характере предстоящего(ей) исследования/манипуляции и возможном развитии неприятных ощущений и осложнений: 1) болевых ощущениях; 2) повышенной кровоточивости, 3) тромбозе, эмболии (закупорке сосудов); 4) воспалительных явлениях, 5) обострении сопутствующих заболеваний; 6) аллергических реакциях, 7) колебаниях артериального давления; 8) механическом повреждении органа; 9) возможности неблагоприятного жизненного прогноза.

Я предупрежден о последствиях отказа от предлагаемого метода исследования/манипуляции.

Я прочитал и понял все вышеизложенное, удовлетворен ответами лечащего врача на мои вопросы.

Подпись пациента \_\_\_\_\_

Подпись лечащего врача \_\_\_\_\_

« \_\_\_\_ » \_\_\_\_\_ 20 \_\_ г.

## ИНФОРМИРОВАННОЕ СОГЛАСИЕ НА ОПЕРАЦИЮ

Я, \_\_\_\_\_  
(Фамилия, Имя, Отчество)

19 \_\_ года рождения, находясь на лечении в \_\_\_\_\_ отделении \_\_\_\_\_, информирован о состоянии моего здоровья и наличии у меня заболевания, требующего хирургического лечения, и что в связи с этим мне показано проведение оперативного вмешательства, так как другие методы лечения в данном случае недостаточно эффективны.

Мне планируется операция \_\_\_\_\_

Мне лично лечащим врачом были разъяснены цель и характер операции, возможные отклонения от намеченного плана хирургического вмешательства в зависимости от интраоперационной ситуации.

Я предупрежден о возможных негативных ощущениях и осложнениях этой операции, включая: 1) повышенную кровоточивость; 2) тромбоз, эмболию (закупорку сосудов); 3) развитие инфекционно-воспалительных осложнений; 4) обострение сопутствующих заболеваний; 5) аллергические реакции; 6) колебания артериального давления; 7) возможность неблагоприятного жизненного прогноза.

Я поставлен в известность, что оперирующим хирургом будет

\_\_\_\_\_  
(Фамилия, Имя, Отчество)

и даю свое согласие.

Я информирован лечащим врачом и оперирующим хирургом о степени риска, связанного с объемом оперативного вмешательства, вплоть до смертельного исхода.

Я предупрежден о режиме послеоперационного поведения и возможных последствиях при его нарушении, а также о возможных болевых ощущениях и методах обезболивания.

Мне объяснены возможные последствия отказа от предлагаемой мне операции.

Я прочитал и понял все вышеизложенное, удовлетворен ответами на все мои вопросы и даю добровольное согласие на операцию.

Подпись пациента \_\_\_\_\_

Подпись лечащего врача \_\_\_\_\_

Подпись оперирующего хирурга \_\_\_\_\_

Подпись зав. отделением \_\_\_\_\_

« \_\_\_\_\_ » \_\_\_\_\_ 20 \_\_ г.

## УВАЖАЕМЫЙ ПАЦИЕНТ!

Вам предстоит пройти обследование и лечение в нашем учреждении. Надеемся, что представленная информация заложит основу нашего дальнейшего взаимопонимания и тесного сотрудничества.

Вам будут предложены современные методы обследования, необходимые для оценки Вашего здоровья и выявления имеющихся заболеваний. Всю информацию, касающуюся результатов обследования, характера Вашего заболевания, возможных вариантов и результата лечения Вы сможете получить у своего лечащего врача, заведующего отделением или специалиста, который Вас обследовал.

В ряде случаев необходимые для Вас исследования процедуры и манипуляции носят инвазивный характер и могут вызвать развитие неприятных ощущений, побочных эффектов и осложнений. Все диагностические и лечебные мероприятия необходимые Вам будут осуществляться только с Вашего добровольного согласия. Вся необходимая информация для принятия Вами добровольного согласия будет предоставлена лечащим врачом.

Вы имеете право отказаться по своим соображениям от той или иной диагностической или лечебной процедуры. Ваш письменный отказ должен быть соответствующим образом оформлен в истории болезни. Он не будет иметь административных последствий, однако может повлиять на правильность определения заболевания, или явиться причиной снижения эффективности лечения.

Просим Вас указать, кому из близких Вам людей может быть предоставлена информация о состоянии Вашего здоровья:

\_\_\_\_\_

Желаем Вам успешного лечения и скорейшего выздоровления.

С данной информацией ознакомлен \_\_\_\_\_  
(подпись пациента)

« \_\_\_\_ » \_\_\_\_\_ 20\_\_ г.

### **3. Порядок выполнения работы**

- 3.1. Перед занятием ознакомьтесь с информационным материалом
- 3.2. В ходе первого занятия в процессе диалога и ответов у доски разбираются основные положения медицинской этики.
- 3.3. В конце первого занятия преподаватель формулирует темы для заполнения форм информированного согласия для выбранных пар пациент-врач
- 3.4. На последующих занятиях в режиме диалога обсуждается правильность заполнения бланков информированного согласия со сменой заданий и пар.
- 3.5. На предпоследнем и последнем занятии проводятся публичные защиты рефератов и дают ответы на дополнительные вопросы собеседования.
- 3.6. Вопросы для самоконтроля

### **4. Вопросы для собеседования**

1. Какими нормативными документами регламентируются клинические испытания?
2. Расскажите о принципах этической оценки клинических исследований
3. Перечислите перечень документов представляемых в этический комитет
4. Что такое информированное согласие и как оно оформляется?
5. В каких случаях допускаются отклонения от требований информированного согласия?
6. Защитите свой выбор формы информационного согласия

## **Практическое занятие №4**

# **СТАТИСТИЧЕСКАЯ ОБРАБОТКА РЕЗУЛЬТАТОВ ЭКСПЕРИМЕНТА**

**1. Цель работы:** ознакомить студентов с основами статистической обработки медицинских данных и научить выбирать критерии и производить обработку результатов экспериментов

## **2. Информационные материалы к занятию**

### **2.1. Статистический анонс**

В медицинских исследованиях наиболее часто используют две группы методов:

1) Критерии, позволяющие выявить различия между исследуемыми препаратами и изучаемыми группами больных по каким-либо параметрам.

2) Методы, устанавливающие наличие и степень взаимосвязи между двумя или несколькими признаками.

Реже используется третья группа, к которой относятся более сложные способы анализа данных - многомерные методы исследования (факторный анализ, кластерный анализ и т.д.).

Выбор статистических критериев определяется тем, какие параметры Вы собираетесь анализировать:

- Количественные (например: уровень артериального давления и т.д.), их можно упорядочить и производить над ними арифметические действия.

- Порядковые (например: степень тяжести заболевания – легкая, средняя, тяжелая) их можно упорядочить, но арифметические действия над ними производить нельзя.

- Качественные (например: окраска бактерий по Граму – красная или синяя), их нельзя упорядочить и над ними нельзя производить арифметические действия, их можно описать, подсчитав число наблюдений, имеющих одно и тоже значение, а затем оценить (сравнить) группы по количеству наблюдений, имеющих различные значения.

Вторым фактором, имеющим большее значение для количественных признаков, является характер распределения, наблюдаемых Вами объектов. Исходя из меркантильных соображений следует рассмотреть два вида распределений:

1) Нормальное

2) Ненормальное (которое включает ряд конкретных распределений, подробное рассмотрение которых не предусмотрено в настоящей работе )

Существуют несколько способов оценки принадлежности, полученного Вами распределения к нормальному:

1) Построение гистограммы частотного распределения признака и определение основных статистических параметров совокупности. Если Вы получите форму гистограммы близкую к той, что изображена на рисунке 1, то Вы имеете дело с нормальным распределением. При этом медиана близка или равна среднему, интервал  $M \pm 2\sigma$  включает не менее 95% наблюдений, асимметрия близка к нулю.

2) Использование нормальной вероятностной бумаги (функция реализована в компьютерных статистических программах). Если при этом, полученные значения признака располагаются почти на одной прямой, то распределение близкое к нормальному.

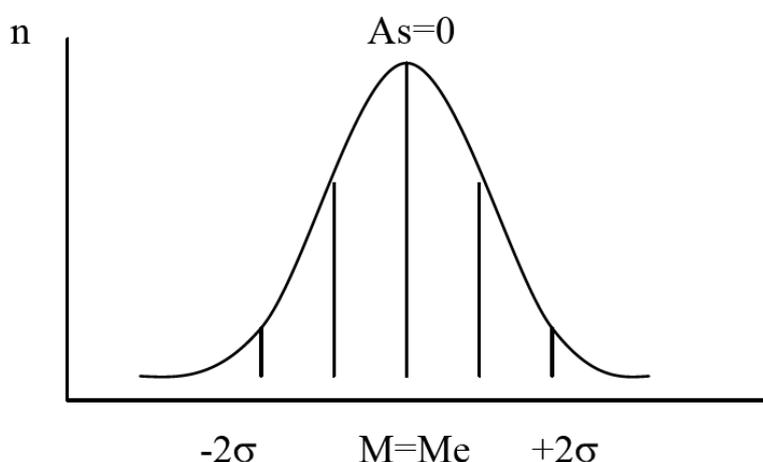


Рис.1 Графическое изображение нормального распределения.

Примечание.  $M$  – среднее арифметическое

$Me$  - медиана

$As$  - асимметрия

$\sigma$  - стандартное отклонение

3) Использование различных критериев (2, Колмогорова-Смирнова), которые оценивают степень отклонения Ваших данных от теоретических, имеющих нормальное распределение с той же средней и дисперсией, что и в Вашей группе. Если различия невелики, то Ваши данные имеют распределение близкое к нормальному.

Обращаем внимание, что оценка характера распределения должна проводиться обязательно, если, конечно, Вас интересует достоверность исследования.

К числу других факторов, влияющих на выбор конкретного критерия относятся:

- Количество анализируемых групп,

- Сравнение зависимых или независимых групп.

В случае нормального (или близкого к нему) распределения Ваших данных, для их описания следует использовать среднее и дисперсию. Для сравнения и оценки различий между группами следует использовать параметрические критерии, т.е. те, которые для выявления различий оперируют параметрами распределения (средним и дисперсией).

При отсутствии нормальности распределения используют непараметрические методы исследования, то есть те, которые заменяют реальные значения рангами и производят расчеты над рангами. Для корректного описания Ваших данных, имеющих ненормальное распределение, следует использовать медиану и квартили (а не среднее и дисперсию).

При анализе порядковых и качественных признаков приоритет выбора остается за непараметрическими методами.

Если Вы сомневаетесь в характере полученного Вами распределения, используйте непараметрические методы. Однако, следует помнить, что часто непараметрические методы имеют ограничения по объему анализируемой выборки (как правило не более 30 наблюдений).

## **2.2. Оценка различий между группами**

### **А. Количественные признаки.**

#### **1. Нормальное распределение**

##### *1. Сравнение независимых групп.*

###### а) Сравнение двух групп

ПРИМЕР 1. Оценить различия между мужчинами (первая группа) и женщинами (вторая группа) по уровню артериального давления

МЕТОД - Обычный двусторонний критерий Стьюдента (t-критерий) оценивает различия средних.

###### б) Сравнение трех и более групп.

ПРИМЕР 2. Оценить различия по уровню артериального давления в группах с различным возрастным цензом: 30-39 лет, 40-49 лет, 50-59 лет (соответственно 1-я, 2-я, 3-я группы).

МЕТОД - Дисперсионный анализ оценивает различия средних и дисперсии; устанавливает наличие отличий по всем группам (в ряду оцениваемых групп). Чтобы установить между какими двумя из трех групп существуют различия, а между какими отсутствуют, следует использовать критерий Стьюдента с поправкой Бонферрони (для исключения ошибки связанной с множественным сравнением).

Поясним изложенное. Число попарных сравнений равно 3 (1 и 2 группы, 2 и 3 группы, 1 и 3 группы). Допустим, что уровень значимости в двух сравнениях был равен 0,02, в третьем – 0,01 (без применения поправки Бонферрони). Ошибочное выявление различий возрастает пропорционально числу сравнений. Таким образом, истинный уровень значимости будет равен произведению полученной достоверности на число сравнений, т.е.  $0,02 \times 3 = 0,06$  и  $0,01 \times 3 = 0,03$ . Следовательно при уровне значимости принятом в медицине ( $p < 0,05$ ), достоверные отличия выявлены только в одном сравнении.

При числе сравнений больше 8 (т.е. при числе сравниваемых групп более 4) следует использовать критерий Ньюмена-Кейлса.

в) Сравнение нескольких групп с контрольной.

ПРИМЕР 3. Сравнить массу миокарда левого желудочка у здоровых лиц (контрольная группа) с массой миокарда у больных I, II, III степенью артериальной гипертонии (1, 2, 3 группы соответственно).

МЕТОД - Критерий Стьюдента с поправкой Бонферрони при числе групп не более 4. При большем количестве групп – критерий Даннета.

## *2 Сравнение зависимых групп (сравнение повторных измерений)*

а) Сравнение двух групп.

ПРИМЕР 4. Оценить достоверно ли изменился ударный объем у больных сердечной недостаточностью после лечения дигоксином.

МЕТОД - Парный критерий Стьюдента.

Оценивает различия средних разностей показателя. Требует нормального распределения разностей.

б) Сравнение трех и более групп.

ПРИМЕР 5. Оценить достоверность изменения уровня холестерина при лечении правастатином в различные сроки наблюдения – через 1, 3, 6, 9 месяцев.

МЕТОД - Дисперсионный анализ повторных измерений. Выявляет различия во всей последовательности сравнений. Для определения между какими двумя группами достоверны различия, а между какими – нет, используют критерий Стьюдента или Ньюмена-Кейлса для повторных измерений (отличием от аналогичных критериев для независимых групп является то, что в качестве оценки дисперсии берут остаточную вариацию, а при нахождении критического значения используют число степеней свободы остаточной вариации).

## **II Ненормальное распределение выборки**

### *1 Сравнение независимых групп*

Примеры задач являются аналогичными при нормальном распределении

а) Сравнение двух групп

МЕТОД - Критерий Манна-Уитни. Альтернативными методами могут быть: критерии Уайта, Ван дер Вадена, Колмогорова-Смирнова.

б) Сравнение трех и более групп.

МЕТОД - Критерий Крускала-Уоллиса. Выявляет различия в ряду изучаемых групп. Для оценки различий между двумя конкретными группами проводят попарное их сравнение с помощью непараметрических вариантов критериев Ньюмена-Кейлса или Даннета (при равном числе наблюдений в группах); критерий Данна (при неодинаковом количестве наблюдений в группах).

### *2 Сравнение зависимых групп (повторные измерения)*

а) Сравнение двух групп.

МЕТОД - Критерий Уилконсона (Вилконсона). Альтернативный метод – критерий знаков.

б) Сравнение трех и более групп.

МЕТОД - Критерий Фридмана . Выявляет различия между всеми группами. Для оценки различий между конкретными двумя группами проводят попарное сравнение с помощью адаптированных (непараметрических) вариантов критериев Ньюмена-Кейлса или Даннета.

### **Б. Порядковые признаки.**

Алгоритм выбора метода анализа такой же, как при ненормальном распределении количественных признаков.

### **В. Качественные признаки.**

*1 Сравнение независимых групп*

а) Сравнение двух групп

ПРИМЕР 6. Влияет ли назначение стрептокиназы на реканализацию коронарных артерий у больных острым инфарктом миокарда .Плацебо получали 40 больных, реканализация зарегистрирована у 8 человек. Стрептокиназу вводили 50 больным, реканализация отмечена у 20 человек.

МЕТОД - Критерий Z с поправкой Йетсена на непрерывность. Аналогичен по математическому выражению критерию Стьюдента, но используются долевые [  $p$  и  $(1-p)$  ] значения признака. Табличный вариант примера 6 выглядит следующим образом:

Препарат	Эффект есть	Эффекта нет
Плацебо	0,2	0,8
Стрептокиназа	0,4	0,6

Критерий Z применим, если произведение числа наблюдений на значение доли больше 5 в каждой из клеток таблицы. При несоблюдении последнего правила следует использовать точный критерий Фишера (в этом случае поле таблицы вводят абсолютное число наблюдений).

б) Сравнение трех и более групп.

ПРИМЕР 7. Условие то же, что и в примере 6, но добавлен третий вариант лечения – стрептокиназа + гепарин, при котором из 45 лечившихся, реканализация наблюдалась у 25 человек.

МЕТОД - Таблица сопряженности с оценкой различий с помощью критерия 2

В поле таблицы вводят абсолютные числовые значения признака. Критерий 2 обладает высокой надежностью, если во всех полях таблицы нет значений признака меньше 5. При этом оценивается наличие различий между всеми группами. Для выявления между какими двумя группами достоверны различия, а между какими – нет, следует строить несколько четырехпольных таблиц (содержащих результаты двух из трех методов лечения) с оценкой каждой из них по критерию 2 с поправкой Бонферрони.

2. Сравнение зависимых групп. (повторные измерения)

а) Сравнение двух групп.

ПРИМЕР 8. При лечении 40 больных АГ анаприлином эффект наблюдался у 25 человек, а при лечении этих больных гипотиазидом эффект был отмечен у 16 человек.

МЕТОД - Критерий Мак-Нимара.

Критерий 2 также может быть использован, но в этом случае он обладает меньшей чувствительностью.

б) Сравнение трех и более групп

ПРИМЕР 9. Аналогичен примеру 8, но добавляется третий препарат празозин – при лечении которым эффект наблюдался у 18 больных. Оценить достоверны ли различия между препаратами по числу больных с наличием эффекта.

МЕТОД - Критерий Кокрена. Альтернативным методом может служить таблица сопряженности с оценкой различий по 2.

Для выявления конкретных двух групп между которыми различия по числу больных с эффектом достоверны, общую таблицу следует разбить на несколько четырехпольных таблиц и сравнить их последовательно с помощью критерия 2.

### **2.3. Анализ зависимостей между признаками**

#### ***I Количественные признаки, нормальное распределение***

ПРИМЕР 10. Изучить наличие и характер взаимосвязи между ростом и весом взрослых мужчин.

МЕТОД - Регрессионный анализ.

Позволяет оценить характер связи (линейная, нелинейная) между изучаемыми признаками; построить уравнение регрессии, используя которое можно предсказать

степень изменения одного из изучаемых признаков (например вес) при изменении другого признака (роста) на определенную величину.

МЕТОД - Корреляционный анализ

Чтобы оценить силу связи между весом и ростом следует использовать корреляционный анализ, а именно определить коэффициент корреляции Пирсона. Данный метод кроме степени взаимосвязи показывает ее направленность (прямая или обратная).

#### ***II Количественные признаки, ненормальное распределение или порядковые признаки.***

ПРИМЕР 11. Определить, есть ли зависимость между степенью тяжести пневмонии и числом осложнений этого заболевания.

МЕТОД - Коэффициент корреляции Спирмена.

При анализе нескольких групп использует коэффициент корреляции Кенделла.

#### ***III Качественные признаки.***

ПРИМЕР 12. Оценить, есть ли зависимость между исходом заболевания (выздоровел, умер) и схемой лечения (традиционная, традиционная + новый препарат).

МЕТОД - Коэффициент сопряженности.

При четырехпольной таблице может использоваться коэффициент ассоциации Юла.

## ВЫБОР МЕТОДА СТАТИСТИЧЕСКОГО АНАЛИЗА В ЗАВИСИМОСТИ ОТ УСЛОВИЙ ЭКСПЕРИМЕНТА

Признак и распределение		Определение различий достоверности различий между признаками		Взаимосвязь признаков	
Независимые группы			Зависимые группы (повторные измерения)		
Две группы		Более двух групп	Одна группа до и после лечения		Одна группа, несколько видов лечения
Количественный, нормальное распределение	Критерий Стьюдента	Дисперсионный анализ. Затем критерий Стьюдента с поправкой Бонферрони или критерий Ньюмена-Кейлса	Парный критерий Стьюдента	Дисперсионный анализ повторных изменений, затем критерий Стьюдента или критерий Ньюмена-Кейлса для повторных изменений	Регрессионный анализ. Коэффициент корреляции Пирсона.
Количественный, ненормальное распределение или порядковый признак	Критерии Манна-Уитни, критерий Уайта, Ван дер Вадена, Колмогорова-Смирнова	Критерий Крускала-Уоллиса, затем непараметрические варианты критериев Ньюмена-Кейлса, Даннета или Данна.	Критерий Уилкоксона Критерий знаков	Критерий Фридмана, затем непараметрические варианты критериев Ньюмена-Кейлса или Даннета	Коэффициент ранговой корреляции Спирмена, коэффициент корреляции Кендалла
Качественный признак	Критерий 2 критерий Z, точный критерий Фишера	Критерий 2 затем четырехпольные таблицы с оценкой различий по 2	Критерий Мак-Нимара	Критерий Кокрена, затем четырехпольные таблицы с оценкой различий по 2	Коэффициент сопряженности. Коэффициент ассоциации Юла

## 2.4. Многомерные (многофакторные) методы анализа данных.

### *Дискриминантный анализ.*

Позволяет на основе наиболее информативных признаков, выявленных из совокупности наблюдаемых признаков, отнести объект к одному из нескольких классов.

### *Факторный анализ.*

Используют для выявления основных (главных) факторов, влияющих на изучаемые признаки какого-либо объекта.

### *Кластерный анализ*

Используют для проведения иерархической классификации данных или выделения в них кластеров (подгрупп).

### *Лог-линейный анализ*

Применяется для анализа сложных многоуровневых таблиц частот.

### *Анализ временных рядов и прогнозирование.*

Используйте этот метод, если Вам необходимо прогнозировать динамику изучаемого процесса

Весь ход статистического изучения заключается в следующем: производство, наблюдение, группировка материала, вычисление обобщающих показателей и их анализ. Два последних этапа и есть задачи математической статистики.

Оценка точности производимых измерений имеет большое экономическое и прикладное значение. Оценку точности результатов каждого из проводимых измерений делает сам экспериментатор. Поэтому знание практических методов оценки погрешностей результатов измерений необходимо материаловедам, участвующим в многочисленных технологических процессах и научных исследованиях.

Рассмотрим основные практические методы статистической оценки погрешностей результатов измерений.

Различают генеральную и выборочную совокупность измерений. Под генеральной совокупностью подразумевают всё множество возможных значений измерений  $x_j$ .

Для выбора совокупности число измерений  $x_j^N$  ограничено. Пусть имеется набор (выборка) экспериментальных данных  $x_1, x_2, \dots, x_N$ . Обработку этих данных производят обычно в следующем порядке:

1. Определяют точечные оценки математического ожидания  $\bar{X}$  дисперсии  $S_x^2$  и среднего квадратичного отклонения  $S_x$ :

$$\bar{X} = \frac{1}{N} \sum X_i \quad (2.1)$$

$$S_x^2 = \frac{1}{N-1} \sum (X_i - \bar{X})^2 \quad (2.2)$$

$$S_x = +\sqrt{S_x^2} \quad (2.3)$$

где  $\bar{X}$  - среднее значение случайной величины или точечная оценка математического ожидания  $M$ ;  $S_x^2$  - точечная оценка дисперсии  $\sigma^2$ ;  $S_x$  - точечная оценка среднего квадратичного отклонения  $\sigma$ .

Точечные оценки стремятся к соответствующим параметрам распределения ( $M$ ,  $\sigma^2$ ,  $\sigma$ ) при числе измерений  $N$  стремящемся к бесконечности.

Оценка дисперсии характеризует однородность изменений. Чем выше оценка дисперсии, тем больше разброс значений.

Коэффициент вариации  $K_V$  характеризует изменчивость измерений. Чем выше коэффициент вариации, тем больше изменчивость измерений относительно средних значений.  $K_V$  оценивает также разброс значений при сравнении нескольких выборок:

$$K_V = \frac{S_x}{\bar{X}} \cdot 100\%. \quad (2.4)$$

2. Анализируют ряд в целях обнаружения грубых ошибок и промахов. Появление этих ошибок вполне вероятно, а наличие их ощутимо влияет на результат измерений. Однако прежде чем исключить то или иное измерение, нужно убедиться, что это действительно грубая ошибка, а не отклонение вследствие статистического разброса. Известно множество методов определения грубых ошибок статистического ряда. Наиболее простой метод - правило трех сигм: разброс случайных величин от среднего значения не должен превышать:

$$X_{\max, \min} = \bar{X} \pm 3\sigma = \bar{X} \pm 3S_x. \quad (2.5)$$

Более достоверными являются методы, базирующиеся на использовании доверительного интервала. При наличии грубых ошибок критерии их появления вычисляются по формулам

$$\begin{aligned} \beta_1 &= (X_{\max} - \bar{X}) / S_x \sqrt{(N-1)/N}; \\ \beta_2 &= (\bar{X} - X_{\min}) / S_x \sqrt{(N-1)/N}. \end{aligned} \quad (21.6)$$

3. Находят точечные оценки очищенного ряда.
4. Определяют интервальные оценки математического ожидания и дисперсии с помощью доверительной вероятности.

Точечные оценки (особенно для выбора малого объема) могут значительно отличаться от истинного значения оцениваемого параметра. Для того, чтобы дать представление о точности и надежности такой оценки используют доверительные интервалы и доверительные вероятности.

Доверительной вероятностью (достоверностью) измерения называется вероятность того, что истинное значение измеряемой величины попадает в данный

доверительный интервал  $a \leq X_g \leq b$ .

Доверительная вероятность  $P_g$  описывается выражением:

$$P_g = P[a \leq X_g \leq b] = \frac{1}{2} \left[ \frac{\varphi(b - \bar{X})}{S_x} - \frac{\varphi(a - \bar{X})}{S_x} \right], \quad (2.7)$$

где  $X_g$  - действительное (истинное) значение случайной величины;  $\varphi(t)$  интегральная функция Лапласа, определяемая выражением:

$$\varphi(t) = \frac{2}{\sqrt{2\pi}} \int_0^t e^{-t^2/2} dt. \quad (2.8)$$

Значение  $a = (1 - P_g)$  называют уровнем значимости. Значения интегральной функции Лапласа приводятся в соответствующих таблицах.

Аргументом этой функции является отношение половины доверительного интервала  $\mu$  к среднеквадратичному отклонению  $S_x$ , т.е.:

$$t = \mu \sqrt{N} / S_x; \mu = b - \bar{X}; \mu = -(a - \bar{X}). \quad (2.9)$$

Если на основе определенных данных установлена доверительная вероятность  $P_g(0,90; 0,95)$ , то устанавливается точность измерений (доверительный интервал  $2\mu$ ) на основе соотношения  $P_d = \varphi(\mu / S_x)$ . Половина доверительного интервала равна:

$$\mu = \frac{S_x \cdot t}{\sqrt{N}}, \quad (2.10)$$

где  $t$  - аргумент функции Лапласа (при  $N > 30$ ) или функции Стьюдента (при  $N \leq 30$ ). Требуемую точность измерений можно определить для разных уровней доверительной вероятности, приняв значение  $t$  по таблице 2.1 при числе степеней свободы  $f = N - 1$ .

Зная величину доверительного интервала, можно получить интервальную оценку математического ожидания:

$$\bar{X} - \mu < M < \bar{X} + \mu \text{ или } X_D = \bar{X} \pm \mu. \quad (2.11)$$

Подставив значение  $\mu$  из (2.10), получим:

$$\bar{X} - \frac{S_x \cdot t}{\sqrt{N}} < M < \bar{X} + \frac{S_x \cdot t}{\sqrt{N}} \text{ или } X_{\text{Д}} = \bar{X} \pm \frac{S_x \cdot t}{\sqrt{N}}. \quad (2.12)$$

Следует иметь в виду, что определяемый указанным образом доверительный интервал характеризует абсолютную погрешность оценки исследуемой величины за счет случайных ошибок. Однако в эксперименте может присутствовать и систематическая погрешность связанная с погрешностями средств измерений. При этом возможны три случая:

Если приборная погрешность  $\Delta$  по сравнению со случайной ошибкой  $\mu$  мала, то величиной  $\Delta$  можно пренебречь.

Если приборная погрешность велика, то пренебрегают случайной ошибкой  $\mu$ .  
Если погрешность серии измерений соизмерима с погрешностью прибора, то границы доверительного интервала:

$$\mu = \sqrt{\frac{S_x^2 t^2}{\sqrt{N}} + \left[ \frac{t(\infty)}{3} \right]^2}. \quad (2.13)$$

Формулой (2.13) следует пользоваться при  $tS_x / \sqrt{N} \leq 3\Delta$ . Если же  $tS_x / \sqrt{N} > 3\Delta$  то доверительный интервал вычисляют по формуле (2.10). Для дисперсии  $\sigma^2$  интервальная оценка находится с помощью  $\chi^2$  - распределения по формуле

$$fS_x^2 / \chi_{1f; 1-\alpha/2}^2 < \sigma^2 < fS_x^2 / \chi_{2f; \alpha/2}^2, \quad (2.14)$$

где  $\chi_{1f; 1-\alpha/2}^2$  - значение критерия  $\chi^2$  при числе степеней свободы  $f=N-1$  и уровне значимости  $1-\alpha/2$ ;  $\chi_{2f; \alpha/2}^2$  - значение критерия  $\chi^2$  при числе степеней свободы  $f=N-1$  и уровне значимости  $\alpha/2$ . Значение  $\chi^2$  приводится в таблице 2.2. Интервальная оценка  $S_x$  находится по формуле:

$$\sqrt{fS_x^2 / x_1^2} \leq \sigma \leq \sqrt{fS_x^2 / x_2^2}. \quad (2.15)$$

Таблица 2.1 – Процентные точки одностороннего распределения Стьюдента

f	Доверительная вероятность $P_{\text{Д}}$				
	<b>0,750</b>	<b>0,900</b>	<b>0,950</b>	<b>0,975</b>	<b>0,99</b>

1	1,000	3,078	6,314	12,70	31,82
2	0,816	1,886	2,920	4,303	6,965
3	0,765	1,638	2,353	3,182	4,541
4	0,741	1,533	2,132	2,776	3,747
5	0,727	1,476	2,015	2,571	3,365
6	0,718	1,440	1,943	2,447	3,143
7	0,711	1,415	1,895	2,365	2,998
8	0,706	1,397	1,860	2,306	2,896
9	0,703	1,383	1,833	2,262	2,821
10	0,700	1,372	1,812	2,228	2,764
11	0,697	1,363	1,796	2,201	2,718
12	0,695	1,356	1,782	2,179	2,681
13	0,694	1,350	1,771	2,160	2,650
14	0,692	1,345	1,761	2,145	2,624
15	0,691	1,341	1,753	2,131	2,602
16	0,690	1,337	1,746	2,120	2,583
17	0,689	1,333	1,740	2,110	2,567
18	0,688	1,330	1,734	2,101	2,552
19	0,688	1,328	1,729	2,093	2,539
20	0,687	1,325	1,725	2,086	2,528

5. Оценивают относительную погрешность результатов серии измерений при заданной доверительной вероятности:

$$\varepsilon = \frac{\mu}{\bar{X}} \cdot 100\%. \quad (2.16)$$

В случае необходимости определяют минимальное количество измерений при заданной точности. Для проведения опытов с заданной точностью и достоверностью необходимо знать минимальное, но достаточное число измерений для данных условий. Задача сводится к установлению минимального объема выборки при заданных значениях доверительного интервала и доверительной вероятности. Зная точность измерений (или относительную погрешность) при  $N_{\min} = N$  из (2.14), подставив выражение для  $\mu$ , получим:

$$N_{\min} = \frac{S_x^2 \cdot t^2}{\varepsilon^2 (\bar{X})^2} \quad \text{или} \quad N_{\min} = \frac{K_B^2 \cdot t^2}{\varepsilon^2}. \quad (2.17)$$

Расчеты по описанному выше алгоритму реализуются на ПЭВМ в табличном процессоре Excel пакета Microsoft Office 2007.

### 3. Порядок выполнения работы

1. Получить выборку экспериментальных данных. Представить результаты эксперимента в виде таблицы.
2. Провести статистическую обработку данных по формулам (2.1-2.17).
3. Провести обработку результатов эксперимента на ПЭВМ, используя лист Microsoft Excel 2007 (или более новую версию).
4. Проанализировать полученные результаты и сделать выводы. Подробнее последовательность выполнения показана на конкретном примере.

Пример выполнения задания

Для полученной выборки из 18 (N=18) измерений провести статистическую обработку результатов эксперимента:

7	7	8	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

1. Проанализируем ряд в целях обнаружения грубых ошибок и промахов. Точечная оценка математического ожидания вычисляется по формуле (2.1):

$$\bar{X} = 74,28.$$

С помощью формул (2.2, 2.3) вычисляют точечную оценку среднеквадратичного отклонения:

$$S_x = 6,36.$$

Пользуясь правилом трех сигм, по формуле (2.5), вычисляем допустимый разброс случайных величин:

$$X_{\max, \min} = 74,28 \pm (3 \cdot 6,36) = 93,36 \dots 55,2,$$

т.е. грубой ошибки в ряде нет. Однако значение 92 следует проверить с помощью критерия  $\beta$ . По формулам (2.6), (коэффициент  $\beta_1$  используют для максимального значения в выборке):

$$\beta_1 = \frac{92 - 74,22}{6,2 \sqrt{17/18}} = 2,87.$$

При доверительной вероятности  $P_D = 0,99$  и  $N = 18$   $\beta_{\max} = 2,96$  (таблица 2.3).

Так как  $2,87 < \beta_{\max}$ , измерение 92 не является грубым промахом. В технических расчетах обычно принимают доверительную вероятность  $P_g = 0,95$ , то проверим соответствующее такой вероятности значение  $\beta_{\max}$ .  $P_D = 0,95$ ,  $N = 18$ , то  $\beta_{\max} = 2,62$  в этом случае  $2,87 > \beta_{\max}$  и значение 92 следует исключить как грубых промах. Запишем очищенный ряд (N=17):

7	7	8	8	9	0	1	2	3	4	5	6	7	8	9	0	1
---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---	---

2. По формулам (2.1-2.3) найдем точечные оценки очищенного ряда:

$$\bar{X} = 73,24; S_x^2 = 22,19; S_x = 4,71.$$

По формуле (2.4) вычисляем коэффициент вариации:

$$K_B = \frac{4,71}{73,24} \cdot 100\% = 6,43\%.$$

3. Найдем интервальные оценки очищенного ряда. Поскольку  $N < 30$ , ряд следует отнести к малой выборке, и доверительный интервал вычисляется с применением коэффициента Стьюдента  $t$ .

При  $P_D = 0,95$  и  $f = 17 - 1 = 16$  коэффициент Стьюдента  $t = 1,746$  (таблица 2.1). Доверительный интервал вычисляется по формуле (2.10) при  $N = 17$ :

$$\mu = \pm \left( \frac{4,71 \cdot 1,746}{\sqrt{17}} \right) = 1,99.$$

Относительная погрешность результатов измерений вычисляются по формуле (2.16):

$$\varepsilon = \frac{1,99}{73,24} \cdot 100\% = 2,72\%.$$

Запишем истинное значение измеряемой величины:

$$X_D = \bar{X} \pm \mu = 73,24 \pm 1,99.$$

Найдем интервальную оценку дисперсии  $\sigma^2$  по формуле (2.14) при уровне значимости  $\alpha = 1 - 0,95 = 0,05$ :

$$12,5 \leq \sigma^2 \leq 49,87.$$

Найдем интервальную оценку среднеквадратичного отклонения по формуле (2.15):

$$\sqrt{12,5} \leq \sigma < \sqrt{49,87}$$

$$3,53 \leq \sigma < 7,06.$$

4 Пусть необходимая точность измерений составляет 5%. Определим минимальное количество измерений для достижения заданной точности. По формуле (2.17):

$$N_{\min} = \frac{6,43^2 \cdot 1,746^2}{5^2} = 5,045.$$

Округляем полученное значение  $N_{\min}$  до большего целого,  $N_{\min} = 6$ . Для достижения заданной точности достаточно сделать 6 измерений.

Таблица 2.2 – Процентные точки  $\chi^2$  - распределения

f / P <sub>д</sub>	0,025	0,975
1	0,00098	5,024
2	0,0506	7,378
3	0,2158	9,348
4	0,4844	11,14
5	0,312	12,83
6	1,2373	14,45
7	1,690	16,01
8	2,180	17,53
9	2,700	19,02
10	3,247	20,48
11	3,816	21,92
12	4,404	23,34
13	5,009	24,74
14	5,629	26,12
15	6,262	27,49
16	6,908	28,85
17	7,564	30,19
18	8,231	31,53
19	8,907	32,85
20	9,591	34,17

Таблица 2.3 – Критерий появления грубых ошибок

N	При P <sub>д</sub>		
	0,90	0,95	0,99
1	1,41	1,41	1,41
2	1,64	1,69	1,72
3	1,79	1,87	1,96
4	1,89	2,00	2,13
5	1,97	2,09	2,26
6	2,04	2,017	2,37
7	2,10	2,24	2,46

8	2,15	2,29	2,54
9	2,19	2,34	2,61
10	2,23	2,39	2,66
11	2,26	2,43	2,71
12	2,30	2,46	2,76
13	2,33	2,49	2,80
14	2,35	2,52	2,84
15	2,38	2,55	2,87
16	2,40	2,58	2,90
17	2,43	2,60	2,93
18	2,45	2,62	2,96

*Составление программы для статистических расчетов в табличном процессоре Excel пакета Microsoft Office*

В таблице 2.4 приводятся номера использованных формул и соответствующие им выражения в строке Excel. Не приводятся логические операции для удобства работы вычислений и приемы работы с массивами, т.к. их легко можно реализовать с помощью встроенной справки в Excel.

Таблица 2.4 – Расчет формул для статистической обработки в табличном процессоре Excel пакета Microsoft Office

	Выражение
	=ОКРУГЛ((СРЗНАЧ("Диапазон выборки"));2)
	=ОКРУГЛ((ДИСП("Диапазон выборки"));2)
	=ОКРУГЛ((СТАНДОТКЛОН("Диапазон выборки"));2)
	=СТАНДОТКЛОН("Диапазон выборки")/СРЗНАЧ("Диапазон выборки")
	Максимум: =ОКРУГЛ((СРЗНАЧ("Диапазон выборки")+(3*СТАНДОТКЛОН("Диапазон выборки")));2); Минимум: =ОКРУГЛ((СРЗНАЧ("Диапазон выборки")-3*СТАНДОТКЛОН("Диапазон выборки"));2)
	Бэта 1: =ОКРУГЛ(((МАКС("Диапазон выборки")-СРЗНАЧ("Диапазон выборки"))/(СТАНДОТКЛОН("Диапазон выборки")*КОРЕНЬ(((СЧЁТ("Диапазон выборки")-1)/СЧЁТ("Диапазон выборки")))));2) Бэта 2: =ОКРУГЛ(((СРЗНАЧ("Диапазон выборки")-МИН("Диапазон выборки"))/(СТАНДОТКЛОН("Диапазон выборки")*КОРЕНЬ(((СЧЁТ("Диапазон выборки")-1)/СЧЁТ("Диапазон выборки")))));2)
0	=ОКРУГЛ(((СТАНДОТКЛОНф5:Ш)*(СТЮДРАСПОБР((2*(1-"Доверительная вероятность"));СЧЁТ("Диапазон выборки")-1)))/КОРЕНЬ(СЧЁТ("Диапазон выборки")));2)

4	<p>Меньшее:  <math>=\text{ОКРУГЛ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3"!C3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 2)); 2)</math></p> <p>Большее:  <math>=\text{ОКРУГЛ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3"!C3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 1)); 2)</math></p>
5	<p>Меньшее:  <math>=\text{ОКРУГЛ}((\text{КОРЕНЬ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3"!C3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 2))))); 2)</math></p> <p>Большее:  <math>=\text{ОКРУГЛ}((\text{КОРЕНЬ}((\text{СЧЁТ}(\text{"Диапазон выборки"}) * \text{ДИСП}(\text{"Диапазон выборки"})) / (\text{ИНДЕКС}(\text{"Таблица 3"!C3:D22; СЧЁТ}(\text{"Диапазон выборки"}); 1))))); 2)</math></p>
6	$=\text{СТАНДОТКЛОН}(\text{"Диапазон выборки"}) * (\text{СТЮДРАСПОБР}((2 * (1 - \text{"Доверительная вероятность"})); \text{СЧЁТ}(\text{"Диапазон выборки"} - 1))) / \text{КОРЕНЬ}(\text{СЧЁТ}(\text{"Диапазон выборки"})) / \text{СРЗНАЧ}(\text{"Диапазон выборки"})$
7	$=\text{ОКРУГЛВВЕРХ}(((\text{СТЕПЕНЬ}((\text{СТАНДОТКЛОН}(\text{"Диапазон выборки"}) * 100 / \text{СРЗНАЧ}(\text{"Диапазон выборки"}); 2) * \text{СТЕПЕНЬ}((\text{СТЮДРАСПОБР}((2 * (1 - \text{"Доверительная вероятность"})); \text{СЧЁТ}(\text{"Диапазон выборки"} - 1)); 2)) / \text{СТЕПЕНЬ}(\text{"Заданная точность"; 2})); 0)$
<p>Примечание: Формат ячеек - Процентный устанавливается в формуле 4 и 16, все остальные ячейки - Общий.</p>	

## Задание

Получить статистический ряд и провести экспериментальные статистические процедуры. Статистический ряд можно получить с помощью пакета MathCad.

### 4. Содержание отчета

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов; результаты вычислений на ПЭВМ.
4. Выводы.

### 5. Контрольные вопросы

1. В чем разница между понятиями «дисперсия» и «оценка дисперсии»?
2. Какие оценки параметров распределения вы знаете?
3. Что называют коэффициентов вариации?

4. Что такое точность измерений, достоверность измерений?
5. В каких случаях при определении доверительного интервала следует пользоваться интегральной функцией Лапласа?

## Практическое занятие №5

### «СТАТИСТИЧЕСКИЕ ХАРАКТЕРИСТИКИ СЛУЧАЙНЫХ ВЕЛИЧИН»

1. **Цель работы:** научиться вычислять оценки основных числовых характеристик по экспериментальным данным.

#### 2. Теоретические сведения

Математическим ожиданием дискретной случайной величины  $X$ , принимающей конечное число значений  $x_i$  с вероятностями  $p_i$ , называется сумма:

$$M(X) = x_1 \cdot p_1 + x_2 \cdot p_2 + x_3 \cdot p_3 + \dots + x_n p_n.$$

Свойства математического ожидания:

1.  $M(c \cdot X) = c \cdot M(X)$ ,  $c \in R$ ,
2.  $M(X + Y) = M(X) + M(Y)$ ,  $X, Y \in E$ ,
3.  $M(X \cdot Y) = M(X) \cdot M(Y)$ , для независимых случайных величин  $X$  и  $Y$ .

Дисперсией случайной величины  $X$  называется число:

$$D(X) = M\{[X - M(X)]^2\} = M(X^2) - [M(X)]^2.$$

Свойства дисперсии:

1.  $D(c \cdot X) = c^2 \cdot D(X)$ ,  $c \in R$ ,
2.  $D(X + Y) = D(X) + D(Y)$ , для независимых случайных величин  $X$  и  $Y$ .

Среднеквадратичное отклонение определяется по формуле:

$$\sigma = \sqrt{D(X)}.$$

Статистическая функция СЧЕТ используется для определения числа значений (рисунок 2.1).

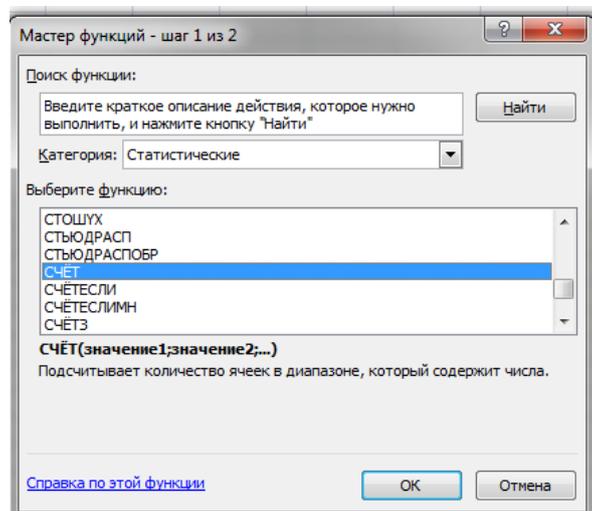


Рисунок 2.1 – Статистическая функция СЧЁТ

Диапазон ячеек указывается адресами первой и последней ячейки данных, записанными через двоеточие, например В6:В15 (рисунок 2.2).

Функция СРЗНАЧ рассчитывает среднее значение выборки.

Функция СТАНДОТКЛОН - стандартное отклонение выборки. Аргументами этих функций служит все тот же диапазон ячеек.

Статистическая функция СТЬЮДРАСПОБР используется для нахождения коэффициента Стьюдента. Этот коэффициент зависит от вероятности ошибки (при обычно задаваемой надежности 95% вероятность ошибки составляет 5%) и от числа степеней свободы n-1. Также это значение можно найти по таблице критических значений t-критерия (приложение 2).

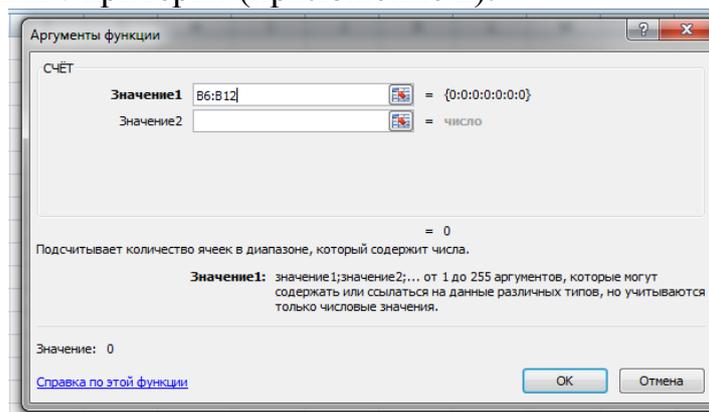


Рисунок 2.2 – Диапазон ячеек

Для нахождения доверительного интервала используется обычная формула умножения «=СТЮДРАСПОБР\* $X_{cp}$ ».

### 3. Порядок выполнения работы

В результате эксперимента были получены значения переменной X:

i	1	2	3	4	5	6	7	8	9	10
X	14,84	14,80	14,84	14,81	14,63	14,81	14,80	14,85	14,84	14,80

Данные введем в столбец В (рисунок 2.3). В столбцах D и E - подсказки характеристик, которые мы будем рассчитывать. Используя приведенные выше функции, в столбец F поместим результаты. По найденному значению ячейки F11 окончательный результат доверительного интервала можно записать так: с 95%-ной надежностью  $X = 14,81 \pm 0,046$ . В заключение вычислим относительную ошибку определения доверительного интервала:  $S = ДИ / X_{cp}$  (формула: «=F11/F7»). Значение относительной ошибки обычно выражают в процентах, в нашем случае 0,3 %.

	A	B	C	D	E	F
1						
2						
3	Данные эксперимента				Обработка	
4						
5	№	X				
6	1	14,85		Число значений n	СЧЁТ	10
7	2	14,8		Среднее значение $\bar{X}_{ср}$	СРЗНАЧ	14,803
8	3	14,84		Станд. Отклонение S	СТАНДОТКЛОН	0,0643
9	4	14,81		Ст. откл. Среднего $S_{ср}$	$S/\text{КОРЕНЬ}(n)$	0,0203
10	5	14,63		К. Стюд (5%, n-1) t	СТЮДРАСПОБР	2,2622
11	6	14,81		Доверит. Интервал ДИ	$t^*S_{ср}$	0,046
12	7	14,8		Относит. ошибка s	ДИ/ $\bar{X}_{ср}$	0,0031
13	8	14,85				
14	9	14,84				
15	10	14,8				

Рисунок 2.3 – Расчет данных

Получить статистический ряд и определить основные характеристики случайных величин:

- число значений;
- среднее значение выборки;
- стандартное отклонение выборки;
- коэффициент Стьюдента;
- доверительный интервал;
- относительную ошибку.

Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

#### 4. Содержание отчета

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов; результаты вычислений на ПЭВМ.
4. Выводы.

#### 5. Контрольные вопросы

1. Что такое математическое ожидание?
2. Перечислите свойства математического ожидания.
3. Что такое дисперсия?
4. Перечислите свойства дисперсии?
5. Что такое среднеквадратичное отклонение?

## Практическое занятие №6 «ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ УРАВНЕНИЯ ЛИНЕЙНОЙ РЕГРЕССИИ»

**1. Цель работы:** научиться вычислять коэффициенты уравнения линейной регрессии по пространственной выборке.

### 2. Информационные материалы к занятию

Выбрав вид функции регрессии, т.е. вид рассматриваемой модели зависимости  $Y$  от  $X$ , например, линейную модель  $Y_x = a + bx$ , необходимо определить конкретные значения коэффициентов модели.

При различных значениях  $a$  и  $b$  можно построить бесконечное число зависимостей вида  $Y_x = a + bx$ , т.е. на координатной плоскости имеется бесконечное количество прямых, нам же необходима такая зависимость, которая соответствует наблюдаемым значениям наилучшим образом. Таким образом, задача сводится к подбору наилучших коэффициентов.

Линейную функцию  $a+bx$  ищем, исходя лишь из некоторого количества имеющихся наблюдений. Для нахождения функции с наилучшим соответствием наблюдаемым значениям используем метод наименьших квадратов.

Обозначим:  $Y_i$  - значение, вычисленное по уравнению  $Y_i = a + bx_i$ ;  $y_i$  - измеренное значение,  $\varepsilon_i = y_i - Y_i$  - разность между измеренными и вычисленными по уравнению значениям,  $\varepsilon_i = y_i - a - bx_i$ .

В методе наименьших квадратов требуется, чтобы  $\varepsilon_i$ , разность между измеренными  $y_i$  и вычисленными по уравнению значениям  $Y_i$ , была минимальной. Следовательно, находим коэффициенты  $a$  и  $b$  так, чтобы сумма квадратов отклонений наблюдаемых значений от значений на прямой линии регрессии оказалась наименьшей:

$$S = \sum_{i=1}^n \varepsilon_i^2 = \sum_{i=1}^n (y_i - Y_i)^2 = \sum_{i=1}^n (y_i - a - bx_i)^2 \rightarrow \min. \quad (2.1)$$

Исследуя на экстремум эту функцию аргументов  $a$  и  $b$  с помощью производных, можно доказать, что функция принимает минимальное значение, если коэффициенты  $a$  и  $b$  являются решениями системы:

$$\begin{cases} a \sum_{i=1}^n x_i + b \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \\ na + b \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \end{cases} \quad (2.2)$$

Если разделить обе части нормальных уравнений на  $n$ , то получим:

$$\begin{cases} a \frac{\sum_{i=1}^n x_i}{n} + b \frac{\sum_{i=1}^n x_i^2}{n} = \frac{\sum_{i=1}^n x_i y_i}{n} \\ a + b \frac{\sum_{i=1}^n x_i}{n} = \frac{\sum_{i=1}^n y_i}{n} \end{cases}$$

Учитывая, что

$$\frac{\sum_{i=1}^n x_i}{n} = \bar{x}, \quad \frac{\sum_{i=1}^n y_i}{n} = \bar{y}, \quad \frac{\sum_{i=1}^n x_i^2}{n} = \overline{x^2}, \quad \frac{\sum_{i=1}^n x_i y_i}{n} = \overline{xy}. \quad (2.3)$$

Получим

$$\begin{cases} a\bar{x} + b\overline{x^2} = \overline{xy} \\ a + b\bar{x} = \bar{y} \end{cases},$$

отсюда  $a = \bar{y} - b\bar{x}$ , подставляя значение  $a$  в первое уравнение, получим:

$$\begin{aligned} \bar{x}(\bar{y} - b\bar{x}) + b\overline{x^2} &= \overline{xy} \Rightarrow \overline{x\bar{y}} - b\bar{x}^2 + b\overline{x^2} = \overline{xy} \Rightarrow b(\overline{x^2} - \bar{x}^2) = \overline{xy} - \bar{x}\bar{y}, \\ b &= \frac{\overline{x\bar{y}} - \bar{x}\bar{y}}{\overline{x^2} - \bar{x}^2} = \frac{\text{cov}(x, y)}{S_x^2}, \end{aligned} \quad (2.4)$$

где  $\text{cov}(x, y) = \overline{x\bar{y}} - \bar{x}\bar{y}$  - выборочная ковариация;  $S_x^2$  - выборочное значение дисперсии величины  $x$ , определяемой по формуле

$$S_x^2 = \overline{x^2} - \bar{x}^2. \quad (2.5)$$

$$a = \frac{\overline{x^2}\bar{y} - \bar{x}\overline{xy}}{\overline{x^2} - \bar{x}^2}. \quad (2.6)$$

При этом  $b$  называют коэффициентом регрессии;  $a$  называют свободным членом уравнения регрессии и вычисляют по формуле:

$$a = \bar{y} - b\bar{x}. \quad (3.7)$$

Полученная прямая является оценкой для теоретической линии регрессии. Имеем:

$$Y = a + bx = \bar{y} - b\bar{x} + bx = \bar{y} + b(x - \bar{x}).$$

Итак,  $Y = \bar{y} + b(x - \bar{x})$  является уравнением линейной регрессии.

Регрессия может быть прямой ( $b > 0$ ) и обратной ( $b < 0$ ). Прямая регрессия означает, что при росте одного параметра, значения другого параметра тоже увеличиваются. А обратная, что при росте одного параметра, значения другого параметра уменьшаются.

### 3. Порядок выполнения работы

Для определения зависимости между сменной добычей угля на одного рабочего (переменная  $Y$ , измеряемая в тоннах) и мощностью угольного пласта (переменная  $X$ , измеряемая в метрах) на 10 шахтах были проведены исследования, результаты которых представлены в таблице 3.1.

Таблица 3.1 – Результаты исследования

i	1	2	3	4	5	6	7	8	9	10
X	8	11	12	9	8	8	9	9	8	12
Y	5	10	10	7	5	6	6	5	6	8

Вычислим эти коэффициенты, используя табличный процессор Excel. На рисунке 3.1 показан фрагмент документа Excel, в котором: а) размещены выборочные данные таблицы 1; б) запрограммировано вычисление коэффициентов системы

$$\begin{cases} b_0 + b_1 \cdot \bar{x} = \bar{y} \\ b \cdot \bar{x} + b_1 \cdot \bar{x}^2 = \overline{xy} \end{cases};$$

в) запрограммировано вычисление  $b_0$ ,  $b_1$  по формулам

$$b_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{S_X^2} = \frac{m_{XY}}{S_X^2};$$

$$b_0 = \bar{y} - b_1 \cdot \bar{x}$$

соответственно.

	A	B	C	D	E	F	G	H
1		Исходные данные			=B3^2		=B3*C3	
2		$x_i$	$y_i$	$x_i^2$	$x_i \cdot y_i$			
3		8	5	64	40	=(E13-B13*C13)/(D13-B13^2)		
4		11	10	121	110			
5		12	10	144	120	$b_1$	1,016	
6		9	7	81	63	$b_0$	-2,75	
7		8	5	64	40			
8		8	6	64	48		=C13-G5*B13	
9		9	6	81	54			
10		9	5	81	45			
11		8	6	64	48			
12		12	8	144	96			
13	Средние значения	9,4	6,8	90,8	66,4			
14								
15		=СРЗНАЧ(В3:В12)			=СРЗНАЧ(Е3:Е12)			

Рисунок 3.1 - Вычисление коэффициентов линейной регрессии

Заметим, что для вычисления средних значений используется функция Excel СРЗНАЧ(диапазон ячеек).

Вычислим значения, используя данные таблицы 3.2.

$$\bar{x} = 9,4; \bar{y} = 6,8; \overline{x^2} = 90,8; \overline{xy} = 66,4.$$

Таблица 3.2 – Расчет результатов исследования

№ п/п	x	y	xy	$x^2$	$y^2$	$(y - \bar{y})$	$(y - \bar{y})^2$	$\hat{y}_i$	$(y - \hat{y})$	$(y - \hat{y})^2$
1	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
2	8	5	40	64	25	-1,8	3,24	5,38	0,38	0,1429
3	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
4	8	6	48	64	36	-0,8	0,64	5,38	-0,62	0,3869
5	9	7	63	81	49	0,2	0,04	6,39	-0,61	0,3672
6	9	6	54	81	36	-0,8	0,64	6,39	0,39	0,1552
7	9	5	45	81	5	-1,8	3,24	6,39	1,39	1,9432
8	11	10	110	121	100	3,2	10,24	8,43	-1,57	2,4775
9	12	10	120	144	100	3,2	10,24	9,44	-0,56	0,3114
10	12	8	96	144	64	1,2	1,44	9,44	1,44	2,0794
Сумма	94	68	664	908	496	0	33,6	68	0	8,3934
Среднее	9,4	6,8	66,4	90,8	49,6					

Коэффициент a, b вычислим по формулам (3.7) и (3.4) соответственно:

$$a = -2,75; b = 1,02.$$

а само уравнение регрессии  $\hat{y} = f(x) = a + bx$  примет вид:

$$\hat{y} = f(x) = -2,75 + 1,02x.$$

## **Задание**

Получить статистический ряд из 15 значений и вычисление коэффициентов линейной регрессии.

Статистический ряд можно получить с помощью пакета MathCad (приложение 1).

### **4. Содержание отчета**

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

### **5. Контрольные вопросы**

1. Приведите формулу линейной регрессии.
2. Что такое регрессия?
3. На каком методе основан классический подход к оцениванию параметров линейной регрессии?
4. Что такое прямая регрессия?
5. Что такое обратная регрессия?

**Практическое занятие №7**  
**«ПОСТРОЕНИЕ НЕЛИНЕЙНОЙ РЕГРЕССИИ»**

**1. Цель работы:** научиться построению уравнения нелинейной регрессии и вычислению индекса детерминации  $R^2$ .

**2. Информационные материалы к занятию**

В случае нелинейных зависимостей с помощью некоторых преобразований можно перейти к линейным. Приведем некоторые из нелинейных функций и их модификацию к линейной функции – таблица 2.1.

Таблица 2.1 – Модификация нелинейных функций

№ п/п	Исходная спецификация	Преобразование $x \rightarrow x^*$	Преобразование $y \rightarrow y^*$	Вычисление b по $b^*$	Вычисление a по $a^*$
1	$y = a + \frac{b}{x} + \varepsilon$	$x^* = \frac{1}{x}$	$y^* = y$	$b = b^*$	$a = a^*$
2	$y = \frac{1}{a + bx + \varepsilon}$	$x^* = x$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
3	$y = \frac{x}{a + bx + x\varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \frac{1}{y}$	$b = a^*$	$a = b^*$
4	$y = ae^{bx + \varepsilon}$	$x^* = x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
5	$y = ae^{\frac{b}{x} + \varepsilon}$	$x^* = \frac{1}{x}$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$
6	$y = \frac{1}{a + be^{-x} + \varepsilon}$	$x^* = e^{-x}$	$y^* = \frac{1}{y}$	$b = b^*$	$a = a^*$
7	$y = ax^b e^\varepsilon$	$x^* = \ln x$	$y^* = \ln y$	$b = b^*$	$a = e^{a^*}$

**3. Порядок выполнения работы**

В таблице 3.1 приведены значения независимой переменной X (товарооборот, тыс. р.) и значения зависимой переменной Y (уровень издержек обращения по отношению к товарообороту, %).

Таблица 3.1 - Значения независимой переменной X и значения зависимой переменной Y

X	7	10	15	20	30	45	60	120
Y	10	9	7,2	6	6,3	5,8	5,4	5

На рисунке 1.1 изображены данные в виде диаграммы. Из этого рисунка видно, что зависимость между показателями нелинейная. Будем искать

регрессионную зависимость в виде степенной:  $y = ax^b$ ; переходя к логарифмам (модификация таблица 2.1), получим линейную зависимость  $y^* = a^* + bx^*$ , где  $y^* = \ln y$ ,  $x^* = \ln x$ .

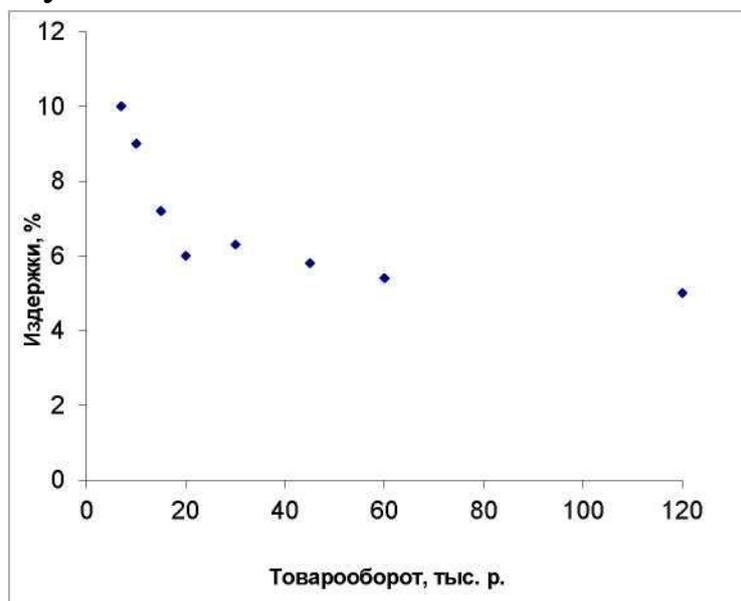


Рисунок 3.1 – Диаграмма товарооборота и уровня издержек обращения по отношению к товарообороту

Определяем коэффициенты  $a^*$ ,  $b$  для модификации данных. Исходные данные и основные вычисления приведены в таблице 3.2.

Таблица 3.2 – Исходные данные

№ п/п	x	$x^* = \ln x$	$y^* = \ln y$	$x^2$	$y^2$	$(y - \bar{y}_x)^2$	$\hat{y}_x$	$(y - \hat{y}_x)^2$
1	7	1,95	2,30	3,79	5,30	10,00	9,12	0,77
2	10	2,30	2,20	5,30	4,83	4,68	8,37	0,40
3	15	2,71	1,97	7,33	3,90	0,13	7,59	0,15
4	20	3,00	1,79	8,97	3,21	0,70	7,08	1,17
5	30	3,40	1,84	11,57	3,39	0,29	6,42	0,02
6	45	3,81	1,76	14,49	3,09	1,08	5,82	0,00
7	60	4,09	1,69	16,76	2,84	2,07	5,43	0,00
8	120	4,79	1,61	22,92	2,59	3,38	4,60	0,16
Итого		26,04	15,16	91,14	29,15	22,32	-	2,67

Определим коэффициенты  $a^*$  и  $b$ :  $a^* \approx 2,68$ ;  $b \approx -0,24$ . По величине  $a^* \approx 2,68$  потенцированием находим исходный коэффициент  $a$ :  $a = e^{2,68} \approx 14,58$ .

Таким образом, уравнение регрессии имеет вид:  $y = 14,58x^{-0,24}$ . Изобразим эту кривую на диаграмме с исходными данными (рисунок 3.2).

Вычислим

$$Q_e = 2,67; Q_r = Q - Q_e = 22,32 - 2,67 = 19,65.$$

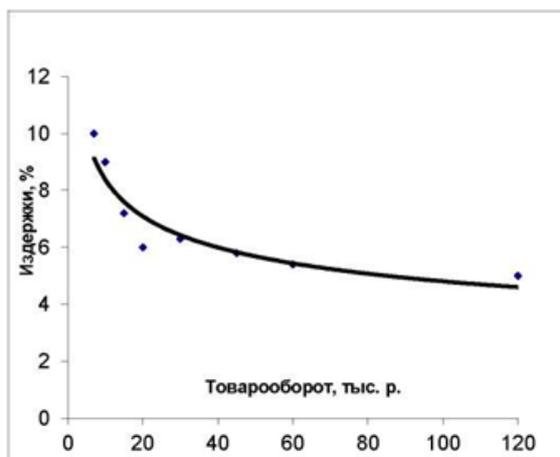


Рисунок 3.2 – Уравнение регрессии на диаграмме с исходными данными

Индекс детерминации равен

$$R^2 = 1 - \frac{Q_e}{Q} = 1 - \frac{2,67}{22,32} = 1 - 0,12 = 0,88.$$

Вычисленное значение говорит о том, что уравнением регрессии объясняется 88% дисперсии результативного признака, а на долю прочих факторов приходится лишь 12%.

### Задание

Построить уравнение нелинейной регрессии и вычислить индекс детерминации  $R^2$  по исходным данным:

X	14*N	20*N	30*N	40*N	60*N	90*N	120*N	240*N
Y	20*N	18*N	14,2*N	12*N	12,6*N	11,6*N	10,8*N	10*N

где N – номер варианта студента по списку.

### 4. Содержание отчета

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

### 5. Контрольные вопросы

1. Что такое параболическая зависимость величины  $X$  от величины  $Y$ ?
2. Что такое степенная зависимость величины  $X$  от величины  $Y$ ?
3. Что такое логарифмическая зависимость величины  $X$  от величины  $Y$ ?
4. Перечислите алгоритм построения диаграммы в среде Excel.
5. Что такое индекс детерминации?

**Практическое занятие №8**  
**«ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ**  
**МНОЖЕСТВЕННОЙ РЕГРЕССИИ»**

**1. Цель работы:** научиться вычислять коэффициент линейной множественной регрессии с помощью MS Excel.

**2. Информационные материалы к занятию**

Вектор коэффициентов, найденный методом наименьших квадратов, является решением следующей системы уравнений:

$$X^T X b = X^T Y,$$

где  $X$  – матрица размерности  $10 \times 3$ , первый столбец которой составлен из 1, а другие два столбца составлены из значений  $X_{i1}, X_{i2}$ , т.е. матрица  $X$  имеет структуру

$$X = \begin{pmatrix} 1 & 8 & 5 \\ 1 & 11 & 8 \\ \dots & \dots & \dots \\ 1 & 12 & 7 \end{pmatrix},$$

а  $Y$  – вектор, составленный из 10 значений  $Y_i$ , т.е.

$$Y = \begin{pmatrix} 5 \\ 10 \\ \dots \\ 8 \end{pmatrix}.$$

Матрица  $X^T X$  имеет обратную матрицу  $(X^T X)^{-1}$ , и тогда вектор коэффициентов равен

$$b = (X^T X)^{-1} (X^T Y). \quad (7.1)$$

Для реализации этой матричной формулы необходимо выполнить следующие операции: транспонирование; умножение матриц (частный случай - умножение матрицы на вектор); вычисление обратной матрицы. Все эти операции можно реализовать с помощью матричных функций MS Excel категории функций Ссылки и массивы. Для работы с этими функциями можно либо обратиться к

Мастеру функций и выбрать нужную категорию функций, затем указать имя функции и задать соответствующие диапазоны ячеек, либо непосредственно ввести с клавиатуры имя функции и задать соответствующие диапазоны ячеек.

Функция ТРАНСП(диапазон ячеек) осуществляет транспонирование матрицы, где параметр диапазон ячеек задает все элементы транспонируемой матрицы (или вектора).

Функция МУМНОЖ(диапазон\_1; диапазон\_2) осуществляет умножение матриц в категории функций Математические. Параметр диапазон\_1 задает элементы первой из перемножаемых матриц, а параметр диапазон\_2 - элементы второй матрицы. При этом перемножаемые матрицы должны иметь соответствующие размерности (если первая матрица размерности  $n \times k$ , вторая -  $k \times m$ , то результатом будет матрица размерности  $n \times m$ ).

Функция МОБР(диапазон ячеек) осуществляет обращение матрицы (вычисление обратной матрицы) в категории функций Математические. Параметр диапазон ячеек задает все элементы обращаемой матрицы, которая должна быть квадратной и невырожденной.

При использовании этих функций необходимо соблюдать следующий порядок действий:

1. Выделить фрагмент ячеек, в которые будет занесен результат выполнения матричных функций (при этом надо учитывать размеры исходных матриц);
2. Ввести арифметическое выражение, содержащее обращение к матричным функциям MS Excel;
3. Одновременно нажать клавиши [Ctrl], [Shift], [Enter]. Если этого не сделать, то вычисляется только один элемент результирующей матрицы или вектора.

### 3. Порядок выполнения работы

Данные о сменной добыче угля на одного рабочего (переменная  $Y$ , измеряемая в тоннах), мощности пласта (переменная  $X_1$ , измеряемая в метрах) и об уровне механизации работ в шахте (переменная  $X_2$ , измеряемая в процентах), характеризующие процесс добычи угля в 10 шахтах, приведены в таблице 3.1.

Таблица 3.1 – Исходные данные

Номер шахты $i$	$X_1$	$X_2$	$Y$
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Предполагая, что между переменными  $Y$ ,  $X_1$ ,  $X_2$  существует линейная зависимость, необходимо найти аналитическое выражение для этой зависимости, т.е. построить уравнение линейной регрессии, вычислить коэффициенты

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix} \text{ уравнения регрессии.}$$

Сформируем матрицу  $X$  и вектор  $Y$ . Затем выполним формирование матрицы  $X^T X$ , вектора  $X^T Y$  и вычисление вектора  $b = (b_0, b_1, b_2)^T$  по формуле (7.1).

$$b = \begin{pmatrix} -3,5393 \\ 0,8539 \\ 0,3670 \end{pmatrix}$$

В результате получим вектор коэффициентов и тогда уравнение регрессии  $\hat{y}(x_1, x_2) = b_0 + b_1 x_1 + b_2 x_2$  имеет вид

$$\hat{y}(x_1, x_2) = -3,5393 + 0,8539x_1 + 0,367x_2.$$

На рисунках 3.1, 3.2 представлен результат вычисления линейной множественной регрессии.

	A	B	C	D	E	F	G
1			Исходные матрица A и вектор x				
2		2	2	6			
3		5	4	8			5
4	матрица A =	7	3	5		вектор x =	6
5		5	1	3			7
6		6	5	2			
7	Транспонирование и умножение матриц						
8			=ТРАНСП(B2:D6)				
9							
10		2	5	7	5	6	
11	матрица $A^T =$	2	4	3	1	5	
12		6	8	5	3	2	
13							
14		44	66	50			
15	матрица $A^T A =$	66	105	87			
16		50	87	83			
17							
18		64		=МУМНОЖ(B2:D6;B10:F12)			
19	вектор $A \cdot x =$	105					
20		88		=МУМНОЖ(B2:D6;G3:G5)			

Рисунок 3.1 – Матричные функции MS Excel

	A	B	C	D	E	F
1		<b>1</b>	<b>8</b>	<b>5</b>		<b>5</b>
2		<b>1</b>	<b>11</b>	<b>8</b>		<b>10</b>
3		<b>1</b>	<b>12</b>	<b>8</b>		<b>10</b>
4		<b>1</b>	<b>9</b>	<b>5</b>		<b>7</b>
5	<b>X =</b>	<b>1</b>	<b>8</b>	<b>7</b>	<b>y =</b>	<b>5</b>
6		<b>1</b>	<b>8</b>	<b>8</b>		<b>6</b>
7		<b>1</b>	<b>9</b>	<b>6</b>		<b>6</b>
8		<b>1</b>	<b>9</b>	<b>4</b>		<b>5</b>
9		<b>1</b>	<b>8</b>	<b>5</b>		<b>6</b>
10		<b>1</b>	<b>12</b>	<b>7</b>		<b>8</b>
11						
12		<b>10</b>	<b>94</b>	<b>63</b>		<b>68</b>
13	<b>X<sup>T</sup> · X =</b>	<b>94</b>	<b>908</b>	<b>603</b>	<b>X<sup>T</sup> y =</b>	<b>664</b>
14		<b>63</b>	<b>603</b>	<b>417</b>		<b>445</b>
15	=МУМНОЖ(ТРАНСП(B1:D10);B1:D10)					
16	=МУМНОЖ(ТРАНСП(B1:D10);F1:F10)					
17		<b>4,0201</b>	<b>-0,323</b>	<b>-0,14</b>		<b>-3,5393</b>
18	<b>(X<sup>T</sup> · X)<sup>-1</sup> =</b>	<b>-0,323</b>	<b>0,054</b>	<b>-0,029</b>	<b>b =</b>	<b>0,8539</b>
19		<b>-0,14</b>	<b>-0,029</b>	<b>0,0653</b>		<b>0,3670</b>
20						
21	=МОБР(B12:D14)		=МУМНОЖ(B17:D19;F12:F14)			

Рисунок 3.2 – Вычисление коэффициентов регрессии

### Задание

По исходным данным построить уравнение линейной регрессии и вычислить

$$b = \begin{pmatrix} b_0 \\ b_1 \\ b_2 \end{pmatrix}$$

вектор коэффициентов уравнения регрессии.

Исходные данные

$X_1$	$X_2$	Y
8*N	5*N	5*N
11*N	8*N	10*N
12*N	8*N	10*N
9*N	5*N	7*N
8*N	7*N	5*N
8*N	8*N	6*N
9*N	6*N	6*N
9*N	4*N	5*N
8*N	5*N	6*N
12*N	7*N	8*N

где N – номер варианта студента по списку.

## 4. Содержание отчета

1. Цель работы.

2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

## **5. Контрольные вопросы**

1. Что осуществляет функция ТРАНСП (диапазон ячеек) в Excel?
2. Что осуществляет функция МУМНОЖ (диапазон\_1; диапазон\_2) в Excel?
3. Что осуществляет функция МОБР (диапазон ячеек) в Excel?
4. Что такое коэффициент линейной множественной регрессии?
5. Приведите уравнение линейной множественной регрессии?

**Практическое занятие №9**  
**«ВЫЧИСЛЕНИЕ КОЭФФИЦИЕНТОВ ЛИНЕЙНОЙ**  
**МНОЖЕСТВЕННОЙ РЕГРЕССИИ И ПРОВЕРКА ЗНАЧИМОСТИ В**  
**РЕЖИМЕ «РЕГРЕССИЯ»»**

**1. Цель работы:** научиться вычислять коэффициент линейной множественной регрессии и проверять значимость экспериментальных данных с помощью MS Excel.

**2. Информационные материалы к занятию**

Табличный процессор MS Excel содержит модуль Анализ данных. Этот модуль позволяет выполнить статистический анализ выборочных данных (построение гистограмм, вычисление числовых характеристик и т. д.). Режим работы Регрессия этого модуля осуществляет вычисление коэффициентов линейной множественной регрессии с  $k$  переменными, построение доверительных интервалов и проверку значимости уравнения регрессии.

После вызова режима Регрессия на экране появляется диалоговое окно (рисунок 2.1), в котором задаются следующие параметры:

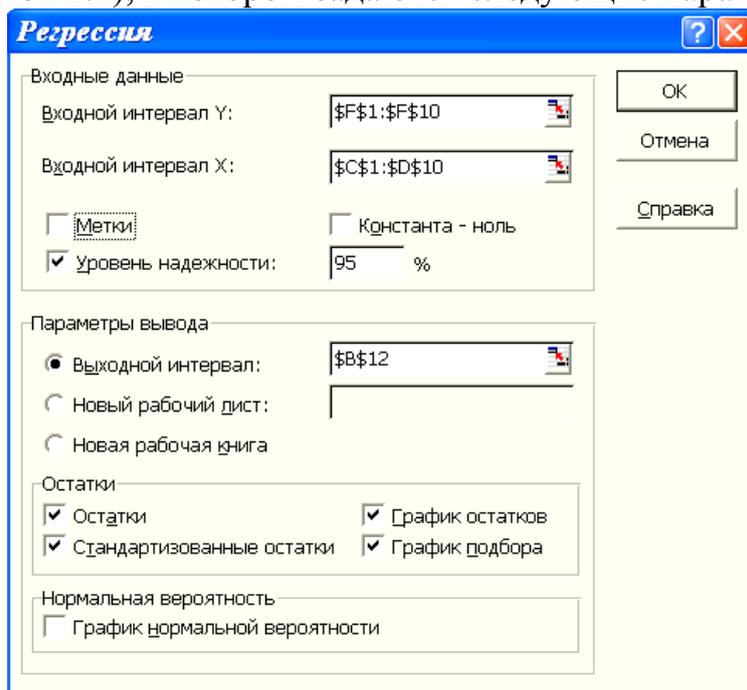


Рисунок 2.1 – Диалоговое окно режима Регрессия

1. Входной интервал  $Y$  - вводится диапазон адресов ячеек, содержащих значения  $Y_i$  - (ячейки должны составлять один столбец).
2. Входной интервал  $X$  - вводится диапазон адресов ячеек, содержащих значения независимых переменных. Значения каждой переменной представляются одним столбцом. Количество переменных - не более 16.
3. Метки - включается, если первая строка во входном диапазоне содержит заголовок. В этом случае автоматически будут созданы стандартные названия.
4. Уровень надежности - при включении этого параметра задается надежность  $Y$  при построении доверительных интервалов.

5. Константа-ноль - при включении этого параметра коэффициент  $b_0 = 0$ .
6. Выходной интервал - при включении активизируется поле, в которое необходимо ввести адрес левой верхней ячейки выходного диапазона, который содержит ячейки с результатами вычислений режима Регрессия.
7. Новый рабочий лист - при включении этого параметра открывается новый лист, в который, начиная с ячейки A1, вставляются результаты работы режима Регрессия.
8. Новая рабочая книга - при включении этого параметра открывается новая книга, на первом листе которой, начиная с ячейки A1, вставляются результаты работы режима Регрессия.
9. Остатки - при включении вычисляется столбец, содержащий невязки  $y_i - \hat{y}_i, i = 1, \dots, n$ .
10. Стандартизованные остатки - при включении вычисляется столбец, содержащий стандартизованные остатки.
11. График остатков - при включении выводятся точечные графики невязки в зависимости от значений переменных  $x_j, j = 1, \dots, k$ . Количество графиков равно числу  $k$  переменных  $x_j$ .
12. График подбора - при включении выводятся точечные графики предсказанных по построенной регрессии значений  $y_i$  от значений переменных  $x_j, j = 1, \dots, k$ . Количество графиков равно числу  $k$  переменных  $x_j$ .

### 3. Порядок выполнения работы

Данные о сменной добыче угля на одного рабочего (переменная  $Y$ , измеряемая в тоннах), мощности пласта (переменная  $X_1$ , измеряемая в метрах) и об уровне механизации работ в шахте (переменная  $X_2$ , измеряемая в процентах), характеризующие процесс добычи угля в 10 шахтах, приведены в таблице 3.1.

Таблица 3.1 – Исходные данные

Номер шахты $i$	$X_1$	$X_2$	$Y$
1	8	5	5
2	11	8	10
3	12	8	10
4	9	5	7
5	8	7	5
6	8	8	6
7	9	6	6
8	9	4	5
9	8	5	6
10	12	7	8

Первоначально введем в столбец С десять значений первой переменной  $X_1$  в столбец D - десять значений второй переменной  $X_2$ , а в столбец F - десять значений зависимой переменной Y.

После этого вызовем режим Регрессия и в диалоговом окне зададим необходимые параметры (рисунок 3.1). Результаты работы приведены на рисунке 3.2 - 3.3. Заметим, что из-за большой «ширины» таблиц, в которых выводятся результаты работы режима Регрессия, часть результатов помещена в другие ячейки.

Дадим краткую интерпретацию показателям, значения которых вычисляются в режиме Регрессия. Первоначально рассмотрим показатели, объединенные названием Регрессионная статистика (рисунок 3.2).

ВЫВОДИТОГОВ			
<i>Регрессионная статистика</i>			
Множественный R	0,9009		
R-квадрат	0,8116		
Нормированный R-квадрат	0,7578		
Стандартная ошибка	0,9509		
Наблюдения	10		
<i>Дисперсионный анализ</i>			
	<i>df</i>	<i>SS</i>	<i>MS</i>
Регрессия	2	27,2704	13,635
Остаток	7	6,3296	0,904
Итого	9	33,6000	
		<i>F</i>	<i>Значимость F</i>
		15,0794	0,0029

Рисунок 3.2 – Результат работы режима Регрессия

Множественный R - корень квадратный из коэффициента детерминации. R-квадрат - коэффициент детерминации  $R^2$ .

Нормированный R-квадрат - приведенный коэффициент детерминации

$$\hat{R}^2 = 1 - \frac{(n-1)Q_e}{(n-k)Q} = 1 - \frac{n-1}{n-k}(1 - R^2).$$

Стандартная ошибка – оценка S для среднеквадратического отклонения  $\sigma$ .

Наблюдения - число наблюдений n.

Перейдем к показателям, объединенным названием Дисперсионный анализ (рисунок 3.2).

Столбец df - число степеней свободы. Для строки Регрессия показатель равен числу независимых переменных  $k_1 = k - 1$ ; для строки Остаток он равен  $k_2 = n - (k + 1) = n - k$ ; для строки Итого  $k_1 + k_2 = n - 1$ .

Столбец  $SS$  - сумма квадратов отклонений. Для строки Регрессия показатель равен величине  $Q_r$  (формулы (3.2.2)); для строки *Остаток* - равен величине  $Q_e$ ; для строки Итого  $Q = Q_r + Q_e$ .

Столбец  $MS$  - дисперсии, вычисленные по формуле  $MS = \frac{SS}{df}$ , т.е. дисперсия на одну степень свободы.

Столбец  $F$  - значение  $F_c$ , равное F-критерию Фишера, вычисленному по формуле

$$F_c = \frac{Q_r / k_1}{Q_e / k_2}.$$

Столбец *Значимость F* - значение уровня значимости, соответствующее вычисленной величине f-критерия и равное вероятности  $P(F(k_1, k_2) \geq F_c)$ , где  $F(k_1, k_2)$  - случайная величина, подчиняющаяся распределению Фишера с  $k_1, k_2$  степенями свободы. Если вероятность меньше уровня значимости  $\alpha$  (обычно  $\alpha = 0,05$ ), то построенная регрессия является значимой.

Перейдем к следующей группе показателей, объединенных в таблице 3.2.

Таблица 3.2 – Продолжение результатов работы режима Регрессия

	<i>Коэффициенты</i>	<i>Стандартная ошибка</i>	<i>t-статистика</i>
У-пересечение	-3,539	1,907	-1,8564
Переменная X 1	0,854	0,221	3,8726
Переменная X 2	0,367	0,243	1,5108
	<i>P-Значение</i>	<i>Нижние 95%</i>	<i>Верхние 95%</i>
	0,1058	-8,0477	0,9690
	0,0061	0,3325	1,3753
	0,1746	-0,2074	0,9415

Столбец Коэффициенты - вычисленные значения коэффициентов  $b_0, b_1, \dots, b_k$ , расположенных сверху вниз.

Столбец Стандартная ошибка - значения  $S_{bj}$ ,  $j = 0, \dots, k$ , вычисленные по формуле

$$S_{bj} = \sqrt{S^2 \left\{ (X^T X)^{-1} \right\}_{jj}}$$

Столбец t-статистика - значения статистик  $T_{bj}$ .

Столбец P-значение - содержит вероятности случайных событий  $P(t(n-k) \geq T_{bj})$ , где  $t(n-k)$  - случайная величина, подчиняющаяся распределению Стьюдента с  $n-k$  степенями свободы.

Если эта вероятность меньше уровня значимости  $\alpha$ , то принимается гипотеза о значимости соответствующего коэффициента регрессии.

Из таблицы 3.2 видно, что значимым коэффициентом является только коэффициент при  $x_1 : b_1 = 0,854$ , так как P-значение при  $x_1$  равно 0,0061, что меньше  $\alpha$ .

Столбцы Нижние 95 % и Верхние 95 % - соответственно нижние и верхние интервалы для оцениваемых коэффициентов  $\beta_j$ .

Перейдем к следующей группе показателей, объединенных в таблице 3.3.

Столбец Наблюдение - содержит номера наблюдений.

Столбец Предсказанное  $Y$  - значения  $\hat{y}_i$ , вычисленные по построенному уравнению регрессии.

Столбец Остатки - значения невязок  $y_i - \hat{y}_i$ .

Таблица 3.3 - Продолжение результатов работы режима Регрессия

ВЫВОД ОСТАТКА			
<i>Наблюдение</i>	<i>Предсказанное Y</i>	<i>Остатки</i>	<i>Стандартные остатки</i>
1	5,127	-0,127	-0,152
2	8,790	1,210	1,443
3	9,644	0,356	0,424
4	5,981	1,019	1,215
5	5,861	-0,861	-1,027
6	6,228	-0,228	-0,272
7	6,348	-0,348	-0,415
8	5,614	-0,614	-0,732
9	5,127	0,873	1,041
10	9,277	-1,277	-1,523

В заключение рассмотрения результатов работы режима Регрессия приведем график невязок - остатков  $Y_i - \hat{Y}_i$  при заданных значениях только второй переменной (рисунок 3.3). Наличие чередующихся положительных и отрицательных значений невязок является косвенным признаком отсутствия систематической ошибки (неучтенной независимой переменной) в построенном уравнении регрессии.

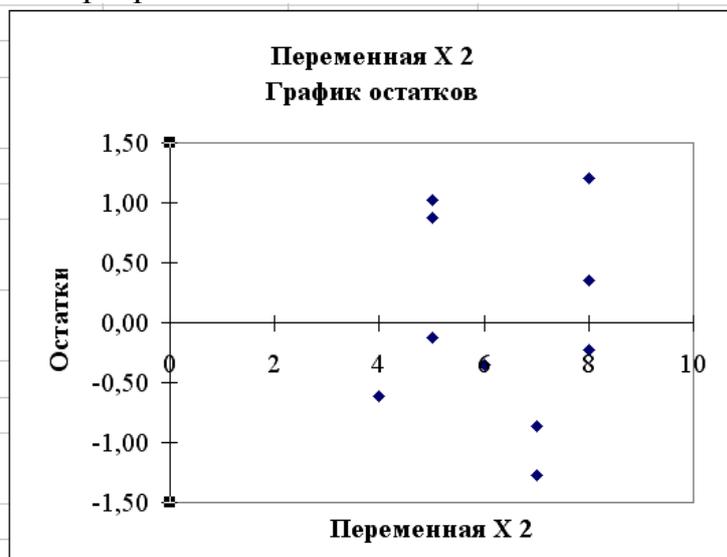


Рисунок 3.3 – График невязок как функция переменной  $X_2$

### Задание

По исходным данным, используя режим Регрессия, вычислить вектор коэффициентов уравнения регрессии.

Исходные данные

$X_1$	$X_2$	$Y$
8*N	5*N	5*N
11*N	8*N	10*N
12*N	8*N	10*N
9*N	5*N	7*N

8*N	7*N	5*N
8*N	8*N	6*N
9*N	6*N	6*N
9*N	4*N	5*N
8*N	5*N	6*N
12*N	7*N	8*N

где N – номер варианта студента по списку.

#### **4. Содержание отчета**

1. Цель работы.
2. Краткие теоретические положения.
3. Технология выполнения работы и полученные результаты: задание; результаты измерений; формулы для получения оценок выборки и результаты расчетов.
4. Выводы.

#### **5. Контрольные вопросы**

1. Перечислите возможности режима Регрессия.
2. Какие параметры задаются после появления диалогового окна Регрессия?
3. Что такое Метки в диалоговом окне Регрессия?
4. Что такое Остатки в диалоговом окне Регрессия?
5. Перечислите алгоритм обработки данных, используя режим Регрессия.