Казанский государственный университет

Исторический факультет

Федорова Н.А.

МАТЕМАТИЧЕСКИЕ МЕТОДЫ В ИСТОРИЧЕСКОМ ИССЛЕДОВАНИИ

Курс лекций

Казань 1996 г.

ISBN 5-85264-013-1

Редактор - дли, проф., академик АН РТ **ИР.Тагиров.**

Рецензенты - К.И.Н., доц. Л.С.Тимофеева

(каф. современной отечественной истории);

К.И.Н., доц. А.А.Новиков

(каф. математической статистики).

Учебное пособие представляет собой курс лекций, читаемых на историческом факультете Казанского госуниверситета. Оно знакомит читателя как с историко-методологической основой применения математико-статистических методов в истории, так и с конкретными приемами исследования. Раскрываются правила оформления таблиц и графиков, смысл их использования в работе историка. Описанные в пособии методы не требуют привлечения сложной вычислительной техники, текст написан достаточно простым языком, материал проиллюстрирован разнообразными примерами.

Данное учебное пособие является начальной ступенью в овладении совокупностью математических методов, применяемых в-современной исторической науке. Оно рассчитано на студентов, аспирантов, преподавателей, научных работников и всех тех, кто интересуется приемами изучения исторических источников; на лиц, не владеющих специальными математическими знаниями.

ПРЕДИСЛОВИЕ.

На уровне обыденного сознания сохраняется устойчивое противопоставление истории и математики, мнение об их несовместимости. Однако контакты и довольно успешное сотрудничество специалистов этих наук начались очень давно.

Что может дать математику история? Ответ на этот вопрос удивительно прост - без истории математик не продвинулся бы в своей науке дальше элементарного счета предметов, оперируя, скорее всего, цифрами, соответствующими количеству пальцев. Почему? Да потому, что история - коллективная память человечества, а любое новое знание появляется только на основе уже достигнутого. В определенном смысле любая наука базируется, прежде всего, на истории - на сохранении, накоплении знаний, опыта.

Нужна ли историку математика? Здесь, по-моему, уместно вспомнить высказывание К. Маркса .о том, что "наука только тогда достигает совершенства, когда ей удается пользоваться математикой" (см,: Воспоминания о К. Марксе и Ф.Энгельсе. - М., 1956. - С. 66). Заявление максималистского характера, но посмотрите вокруг - математика сегодня проникла во все отрасли знания, дала жизнь новым научным направлениям, внедряется в искусство (вслед за пушкинским Сальери мы поверяем гармонию алгеброй). И в то же время науки не утрачивают своей специфики, а искусство остается искусством.

Какова же роль математики? Она является здесь средством, с помощью которого решаются многие сложнейшие задачи. Если смоделировать ситуацию, то можно спросить - чем удобнее отворить запертую на замок дверь: ломом или соответствующим ключом? Хочется надеяться, что читающий эти строки предпочтет ключ. Математика и является зачастую "ключом", способным раскрыть историкам новые факты, новые источники, создать концепцию, поставить точку в спорных вопросах, обобщить накопленную информацию, заставить более объективно взглянуть на пройденный человечеством путь, открыть новые перспективы и многое другое.

Но все замки одним ключом не откроешь. Как правильно подобрать ключ к замку? Какими математическими приемами следует воспользоваться в той или иной ситуации? Об этом и пойдет речь в данной книге.

Лекция 1.
МЕТОДОЛОГИЧЕСКИЕ ОСНОВЫ ПРИМЕНЕНИЯ МАТЕМАТИЧЕСКИХ МЕТОДОВ В ИСТОРИЧЕСКИХ ИССЛЕДОВАНИЯХ.

Процесс научного познания складывается из трех компонентов - методологии, методики и техники. Под методологией понимается совокупность основополагающих представлений и идей, принципов и приемов познания, которые являются теорией метода. Пути и способы их реализации, набор соответствующих правил и процедур составляют методику исследования. Для проведения любого исследования необходимы орудия, инструменты, образующие технику. Между этими составляющими существует диалектическая связь, т.е. активную роль может играть каждая, из перечисленных здесь, частей. В то же время они настолько взаимосвязаны, что существование их самостоятельно, в отрыве друг от друга невозможно, и все они подчинены главной цели - углублению и расширению наших знаний.

Современное состояние исторической науки характеризуется значительным расширением проблематики, связанным с необходимостью, с одной стороны, обобщить накопленный опыт и выйти на уровень фундаментальных работ, носящих теоретико-концептуальный характер. Например, требует комплексного подхода проблема сельской поземельной общины, существовавшей на Руси с V111 в. и до первой четверти XX в. Определенные ее элементы можно найти и в современных деревнях, в колхозах. Подобное исследование требует анализа и обобщения огромного объема источников, разных по характеру и формам выражения. С другой стороны, рухнувшая коммунистическая система открыла возможность обратиться ко многим, ранее запретным темам, расширила источниковую базу исследователя, сняв гриф секретности с ряда архивных и библиотечных комплексов. Это диктует потребность в детальном изучении определенных фактов, явлений, процессов. Кроме того, ряд исторических событий надо переосмыслить, сняв с их анализа идеологические догмы. История нуждается в повышении объективности своих выводов и наблюдений, в повышении точности.

Определенную помощь историку может оказать математика*. (Под математикой обычно понимается комплекс математических дисциплин и научных направлений, занимающихся изучением абстрактных структур и операциями над объектами общей природы, а значит и количественными характеристиками социальных явлений). В основе современных математико-статистических теорий лежит **понятие вероятности.** Под ней понимается объективная категория, выступающая мерой возможности того или иного результата, характеризующая с количественной определенностью возможность появления данного события. По классическому определению вероятность - это величина равная отношению числа возможных случаев, благоприятствующих данному событию, к числу всех равновозможных случаев. Предположим, что в студенческой олимпиаде участвуют 50 человек, из них 6 - студенты КГУ. В данном примере 50 - величина, характеризующая равновозможные шансы к победе, а 6 - шансы победы студентов КГУ. Следовательно, в 6 случаях из 50 возможных могут победить студенты КГУ; или 6:50 = 0,12, т.е. вероятность победы наших студентов равна 0,12 (или 12%).

Поддаются ли социальные явления вероятностному (с математической точки зрения) описанию?

Для вероятностных событий необходимо выполнение ряда условий:

1. Наблюдаемые явления либо могут быть **повторены** неограниченное число раз, либо сразу осуществимо наблюдение за одинаковыми событиями в большом количестве. Не надо лишний раз доказывать, что эксперимент, а значит бесчисленный повтор событий в истории невозможен. Однако осуществить наблюдение

за большим числом одинаковых событий можно при изучении массовых источников, массовых совокупностей однородных (однотипных по структуре) документов.

2. *Независимость* событий. Применительно к истории нельзя говорить о независимости исторических фактов, между ними существует причинно-следственная связь, но в данном случае речь идет о независимости документов. Каждый из них должен формироваться самостоятельно, а не списываться один с другого.

3. *Наличие постоянных условий* при создании источниковой базы.

Уход от идеи строгой детерминированности, обязательности происшедших исторических событий, введение в научный оборот комплексов массовых источников позволяет относить явления истории к вероятностным, а следовательно расширить методический арсенал введением в него математических методов.

Основной задачей изучения исторических явлений и процессов выступает раскрытие внутреннего механизма и всестороннее объяснение их сущности.

Конечная цель любого исторического исследования состоит в выявлении закономерностей. Одни проявляются в единичных случаях (динамические закономерности). Характер динамической закономерности устанавливает поведение каждого признака. Другие - только в массовых, т.е. в группе явлений, которая наряду с признаками, присущими индивидуальным явлениям, характеризуются и общими для всех (статистические закономерности).

Общественное явление складывается из массы индивидуальных и выявить историческую закономерность - значит найти повторяемость внутри всей массы явлений, где наряду с главными действует и множество второстепенных, неустойчивых, случайных факторов. Это приводит к тому, что в обществе нет строго определенных динамических закономерностей.

Использование в историческом исследовании методов изучения статистических закономерностей позволяет в массе случайных факторов выделить основные, главные тенденции, присущие в целом рассматриваемому явлению. Вместе с тем нельзя отбрасывать, упускать из поля зрения и второстепенные, малозначимые, а порой только нарождающиеся факторы, вызывающие те или иные скачки в основной линии развития общества.

Статистические закономерности теоретически базируются на **законе больших чисел,** суть которого в самом общем виде состоит в том, что только при большом числе наблюдений формируются и проявляются многие объективные закономерности общественных явлений. Влияние случайных факторов, случайных признаков тем меньше, чем больше рассмотрено единичных явлений. Так, например, среди студентов первого курса можно встретить человека в возрасте 28 лет. Закономерно ли это?

Статистическое обследование только одного вуза показало, что средний возраст первокурсника колеблется в пределах 18-20 лет, то же обследование в рамках рамках города дает возраст - 19 лет. Следовательно, 28 летний студент на 1 курсе - явление случайное, оно "растворилось" в массе наблюдений. Однако, если бы мы рассмотрели средний возраст на основе изучения всего 3-х студентов - 17, 20 и 28 лет, то наша средняя величина была бы 21,7 лет. Здесь в значительной мере сказалось бы влияние такого случайного фактора, как 28-летний возраст первокурсника.

Закон больших чисел означает, что случайные отклонения, присущие единичным явлениям, в большой массе не влияют на средний уровень изучаемой совокупности. Отклонения индивидуальных элементов как бы уравновешиваются, нивелируются в массе явлений одного типа и перестают зависеть от случайностей. Именно это свойство позволяет выйти на уровень статистической определенности, статистической закономерности. В законе больших чисел нашла свое выражение связь между необходимым и случайным..

Статистическая закономерность является количественным выражением определенной тенденции, но не всякая статистическая закономерность имеет исторический смысл. Можно обнаружить статистическую закономерность распространения культуры картофеля в России в годы крестьянской войны под предводительством Е.Пугачева. Однако весьма сомнительно влияние этой тенденции на ход исторических событий. Анализируя полученные данные, историк на основе содержательного, качественного подхода решает, отражает ли найденная статистическая закономерность историческое явление, какую степень обобщения несет, какие условия ее определили и т.п.

Таким образом, речь идет не о приобретении историей математической точности, а о расширении методического арсенала историка, о возможности получения новых сведений на более совершенном количественном и качественном уровне. Историческая наука не теряет своей специфики, т.к. математические приемы не заменяют качественный анализ и не затрагивает предмет исторической науки.

Не выработано математических методик, не связанных с качественной стороной работы. Не существует универсальных приемов исследования для всех исторических проблем, для всех исторических источников. Исходные теоретико-методологические принципы исторической науки определяют цели, пути **и** методы исследования. На их основе происходит отбор, анализ и обобщение фактического материала.

* * *

В процессе исследования соотношение количественного и качественного анализа происходит четыре этапа.

1. ***Постановка проблемы, выбор*** **источников** и ***определение существенных*** **признаков** происходит при преобладании содержательного, качественного анализа. Этот этап очень важен для всей последующей работы, т.к. от правильного выявления значимых признаков зависит выбор методов анализа.

Здесь происходит некоторая формализация источника. Все признаки по своей природе подразделяются на количественные (выражаемые числом) и качественные (определяемые словесно). Количественные признаки раскрывают меру определенных свойств объекта, а качественные (атрибутивные) - наличие этих свойств и их сравнительную интенсивность. Разновидностью качественных признаков выступают альтернативные, т.е. принимающие только два значения (классическим примером качественного альтернативного признака является "пол" - либо мужской, либо женский).

Велика роль математики при решении задач, связанных с повышением информативной отдачи источников. Современники, фиксируя те или иные аспекты исторических явлений, преследуют цель, отличную от исследовательской. Е связи с этим исследователь не всегда может найти в документах прямых сведений об интересующих аспектах явления. Практически любой источник содержит скрытую информацию, которая характеризует многообразные взаимосвязи, присущие историческим явлениям. Она выявляется в результате специальной обработки и анализа данных.

2. **Выбор** ***математических методов*** в зависимости от структуры источника, характера данных и сущности методов определяется в неразрывном единстве качественного и количественного анализа.

3. На третьем этапе наблюдается *относительная самостоятельность количественного анализа.* Происходит выяснение численных распределений значений признаков, количественных показателей меры зависимости между ними, определяются показатели интенсивности влияния группы факторов на изучаемую систему и тл. Идет расчет показателей по формулам.

Все явления без исключений характеризуются единством количества и качества. Сущность того или иного явления, которая выражает его качественную определенность, будет раскрыта только тогда, когда будет выявлена количественная мера данного качества.

4. ***Содержательная интерпретация полученных результатов*** и построение на их основе теоретических выводов требуют от исследователя знания предмета, его количественной и качественной стороны. Общей схемы для такой интерпретации не выработано. Здесь необходимо учитывать математический аспект интерпретации показателей, полученных в результате расчетов, исходя из сущности примененного метода. В тоже время нельзя упускать из вида содержательный смысл проблемы, отступать от исторической возможности и реальности обретенных показателей.

Между обозначенными здесь этапами существует теснейшая взаимосвязь. Каждый предыдущий этап влияет на последующий и наоборот. Так, характер источника определяет методику его анализа, в то же время сам метод влияет на выбор признаков.

Отмеченное выше единство качественных **и** количественных характеристик явления имеет большое значение при использовании математических методов и интерпретации их результатов. Изменение количественных параметров может происходить в рамках одного качества, а может приводить к приобретению явлением новой сущности, нового качества.

Так, например, увеличение значений такого количественного показателя, как размер землепользования, достигнув определенного уровня, приводит к смене социального статуса крестьянина (от бедняка к середняку, от середняка к кулаку ...), т.е. к появлению нового качества.

Различие значений признака у разных единиц совокупности в один и тот же период времени называется в статистике **вариацией.** Она является необходимым условием существования и развития массовых явлений. В общественной жизни каждой массовой совокупности, массовому процессу присуща специфическая мера вариации элементов, при которых данный процесс протекает нормально, не меняя своей качественной сущности.

* * *

Применение математико-статистических приемов в исторической науке имеет давние корни. Первые опыты в этом направлении в России начались в конце XIX в. на основе использования данных земской статистики. В работах А.Кауфмана, И.Лучицкого, Н.Любовича, Н.Нордмана, опубликованных в начале XX века, содержится не только пример использования статистических методов, но и первые попытки теоретического осмысления трудностей и преимуществ взаимодействия истории и математики. Эта традиция не была прервана революционными потрясениями 1917 г. и разнообразие методических подходов отличает труды историков 20-х гг. Интересные работы были созданы Г.Баскиным, Л..Крицманом, И.Росницким по проблеме социальной дифференциации, оригинальные гипотезы высказаны В.Анучиным, Л..Чижевским о цикличности исторических "всплесков" активности (восстаний, массовых забастовок, войн, межнациональных конфликтов и т.п.) в связи с солнечной активностью и др.

Превращение истории в классовую, партийную науку, выполняющую в значительной степени социальный заказ, поступавший от правящих структур, привело ее к описательности и подчинило концепции детерминированности исторического процесса, заложенной в "Кратком курсе истории ВКП(б)". Естественно, в этот период находили применение лишь те методы и приемы исследования, которые помогали в достижении идеологических целей. Дольше всего, пожалуй, математические приемы исследования в этих условиях продержались в археологии (см. работы А.Арциховского, М.Грязнова, П.Ефименко).

Новый этап наступил на рубеже 50-60-х гг. Он связан с появлением в СССР электронно-вычислительной техники. Особенностью этого времени является публикация работ, в большей мере посвященных демонстрации возможностей ЭВМ при обработке больших массивов информации, чем решению конкретно-исторических задач.

Внедрение ЭВМ дало возможность обратиться к массовым источникам, в которых историки тех лет видели путь преодоления описательности и субъективизма исторической науки. Среди наиболее значимых трудов этого периода - статьи и монографии В.Устинова, ЛКовальченко, Ю.Кахка и др.

Расширение крута проблем, решаемых с помощью математических методов и ЭВМ, постепенное накопление опыта в этой сфере, совершенствование приемов и техники обработки исторической информации позволило со второй половины 60-х гг. сосредоточиться на решении задач исторической науки. Здесь выделяются труды И.Ковальченко и Л.Милова по истории формирования Всероссийского аграрного рынка, В.Дробижева и А.Соколова по истории рабочего класса, К.Хвостовой по социально-экономическим явлениям средневековья, Г.Федорова-Давыдова по археологии и тд.

Застойный период ознаменовался критикой историков, оперирующих математическими приемами. Во-первых, это было связано с победой консервативного направления политики, а следовательно с усилением идеологического давления на все стороны жизни, в том числе и на развитие исторической науки. Во-вторых, критика имела под собой почву в лице историков-конъюнктурщиков, обратившихся к "модным" методам без должной необходимости и обоснованности. Все это вызвало к жизни работы популяризаторского характера, целью которых было доказательство важности и полезности для исторической науки сотрудничества с математикой. Наиболее рельефно эта тенденция проявилась в работах Б.Миронова, З.Степанова, Т.Славко, ряде историографических обзоров.

Однако именно в 60 - 80-е годы был накоплен огромный опыт применения математических методов и ЭВМ в исторической науке. С их помощью производится сравнительный анализ влияния различных факторов на исторический процесс, измеряется зависимость между признаками различных явлений, проверяется достоверность информации исторических источников, устанавливается их подлинность, доказывается авторство. Математика позволяет восстанавливать утраченную источниковую информацию, вводить в научный оборот новые документальные комплексы. На основе количественных приемов исследуется типология событий и социальных сил исторического процесса, его экономические характеристики. В связи с этим надо отметить труды Л.Ковальченко, Л.Бородкина, К.Литвака, Н.Селунской, Т.Славко, И.Гарсковой и ряда других современных исследователей.

В настоящее время историческая наука довольно широко пользуется математико-статистическими приемами, чему в значительной мере способствует компьютеризация рабочего места исследователя. В связи с этим наиболее актуальными считаются две проблемы - расширение математического инструментария за счет внедрения в историографию методов математической логики, теории информации, теории графов и т.д. Вторая проблема - хранение исторической информации при помощи ЭВМ, проблема создания баз и банков данных машиночитаемой информации по определенным *историческим* темам, периодам, регионам.

Задача данного учебного пособия сводится к ознакомлению студентов с теми математическими методами, которые они могут применить в своих учебных исследованиях, на уровне курсовых и дипломных работ без специальной математической подготовки, без привлечения сложной электронно-вычислительной техники, а также помогут в будущей профессиональной деятельности.

ДОПОЛНИТЕЛЬНО ПО ТЕМЕ ЧИТАЙТЕ:

1. Барг М.А. Категории и методы исторической науки. - М, 1984.

2. Бородкин Л.И. Методологические проблемы применения математических методов в историко-гуманитарных исследованиях//Математизация современной науки: предпосылки, проблемы, перспективы. - М., 1986. С.130-139.

3. Ковальченко И.Д. Методы исторического исследования.- М.,1987.

4. Математические методы в исследованиях по истории СССР. Библиографический указатель отечественной литературы 60-80-х гг. - Свердловск, 1989.

5. Миронов Б.Н., Степанов З.В. Историк и математика,- М.1975.

6. Славко Т.И. Математико-статистические методы в исторических исследованиях.- М., 1981.- С.3 - 29.

7. Устинов В.А., Фелингер А.Ф. Историко-социальные исследования: ЭВМ и математика.- М., 1973.

Лекция 2.
ГРУППИРОВКИ В ИСТОРИЧЕСКОМ ИССЛЕДОВАНИИ.

Познание человеком окружающей действительности начинается с конкретных вещей и явлений, которые представляются существенными сами по себе, независимо друг от друга. Углубление знаний раскрывает взаимосвязь предметов и явлений и в их массе обнаруживаются общие типы, общие законы путем "сглаживания" индивидуальных особенностей. Познание исторической реальности также начинается со знакомства с конкретными фактами, процессами, явлениями, которые первоначально кажутся сугубо индивидуальными и неповторимыми.

Каждый факт, каждое действующее в истории лицо характеризуется уникальным набором признаков, однако в процессе изучения выявляется общность в показателях. Либо повторяются или слабо различаются значения одних и тех же признаков разных явлений, либо повторяется или слабо различается набор признаков рассматриваемых явлений. Эта общность позволяет ряд явлений объединять в одну группу. Какую бы совокупность объектов мы не рассматривали, ее всегда можно разбить на группы по сходству признаков. Так, при всей неповторимости каждого человеческого лица в совокупности можно выделить типы (монголоидный, приветливый, овальный...)

Изучая явления прошлого по первичным статистическим данным, историк сталкивается с неупорядоченной последовательностью чисел, показателей, характеризующих тот или иной аспект явления или процесса. Одним из наиболее распространенных приемов представления совокупности разрозненных данных в удобной для восприятия форме выступает группировка. Она является основным начальным этапом обработки данных источника, фундаментом для большинства других приемов математико-статистического анализа.

Метод группировки заключается в разбиении исходной совокупности данных на группы, каждая из которых объединена общим» показателями. Различия между единицами одной группы должны быть меньше, чем различия между единицами разных групп.

Сгруппированные данные представляются в виде таблиц или графиков. Это позволяет охарактеризовать как в целом изучаемую совокупность, так и ее части; обнаружить и зафиксировать связи между признаками; обеспечить наглядность и компактность материала.

Имеющийся в распоряжении исследователя набор чисел называется статистической совокупностью. Количественные показатели, характеризующие рассматриваемый признак и принимающие различные значения - вариантами или переменными. Так, например, личные карточки студентов исторического факультета КГУ с указанием их возраста выступают в качестве статистической совокупности. Возраст - рассматриваемый признак, а конкретные его значения относительно каждого студента - варианты или переменные. Одна и та же варианта статистической совокупности может встречаться несколько раз. Величина, показывающая сколько раз (как часто) встречается то или иное значение переменной называется ее частотой. Допустим, в анализируемой совокупности 38 студентов в возрасте 23 лет. Это значит, что частота признака "возраст" при переменной "23" равна "38".

Здесь надо отметить, что не только сгруппированные данные оформляются в таблицы. На этапе формализации содержательной стороны источника, когда выделены интересующие исследователя признаки, их конкретные значения можно заносить в таблицу. Например, изучая агитационные листовки 60 - 70-х годов XX века, призывающие голосовать за того или иного рабочего-кандидата в депутаты, можно выделить следующие характеристики: пол; возраст; место рождения; стаж трудовой деятельности; уровень образования; партийность и др. Эти характеристики выступают признаками изучаемого явления (в данном случае - общественно-политической активности рабочего класса) и могут выполнить роль табличных граф. Заполняется такая таблица по мере поступления информации, по мере знакомства с историческим источником. Ее построение является первым этапом статистического изучения вариации признака (признаков).

Сведения источника, систематизированные в возрастающем или убывающем порядке и оформленные в виде таблицы называются *ранжированным* ***рядом.***

Для того, чтобы сведенные в таблицу данные не теряли своего значения, а использование таблицы имело смысл, необходимо соблюдать определенные правила при составлении (построении) таблиц.

1. Каждая таблица должна иметь свой заголовок. При минимальном количестве слов он должен полностью отражать внутреннюю структуру таблицы.

2. В одной таблице не должно быть много признаков. Важно помнить, что чем меньше признаков, характеристик сведено в одну таблицу, тем выше ее наглядность, проще анализ, представленных данных.

3. Не строить громоздких таблиц. Нет необходимости каждой варианте признака выделять отдельную графу таблицы. Целесообразно объединять несколько граф в одну под названием "прочие", при том, что эта графа не будет охватывать более 0,1 от общего числа наблюдений.

4. Не путать употребление "итого" и "всего". "Итого" выступает итогом для определенной части совокупности, а "всего" является итогом для совокупности в целом.

5. Громоздкие числа принято округлять.

Напомним арифметические правила: если округляется цифра больше 5, то округление идет в сторону увеличения числа: 2,27 при округлении - 2,3; если округляется цифра меньше 5, то

- в сторону уменьшения: 2,23 при округлении - 2,2; если округляется 5, то округление идет к четной цифре: 2,55 при округлении 2,6, а 2,45 при округлении - 2,4.

6. Каждая клеточка таблицы должна соответствовать определенному числу.

...		-

Если в распоряжении исследователя нет сведений по какому-то параметру, то рекомендуется ставить или прочерк (-) или троеточие (...).

0.0		

Если сведения есть, но выражены крайне малой величиной, то в таблицу вносится 0,0.

	(X)	

Если какое-либо значение получено исследователем, автором таблицы в результате приближенных, условных вычислений, то оно должно быть заключено в круглые скобки.

X?		

Если исследователь сомневается в достоверности значения того или иного параметра, взятого из источника, то рядом с сомнительным показателем ставится вопросительный знак.

Следует избегать включения в таблицу простых дробей. Они с трудом воспринимаются, плохо читаются. Целесообразно построить две таблицы - для числителей и знаменателей отдельно.

7. Таблицы сопровождаются сносками и примечаниями. Сноски относятся к части таблицы - строке, столбцу, клетке - и указывают на ограниченные обстоятельства, которые надо иметь в виду при чтении отмеченных фрагментов таблицы. Примечания относятся к таблице в целом. Чаще всего в них указывается источник информации. Если таблица авторская, следует указывать "Составлено по данным:..." Если таблица взята в готовом виде, то указывается источник информации.

* * *

Признаки, положенные в основу составления таблицы, могут быть дискретными, т.е. принимающими только целые значения и непрерывными, если отдельные их значения могут отличаться друг от друга на сколь угодно малую величину. Примером дискретного признака может служить количество детей в семье, а непрерывного - стаж работы.

В практике исторических исследований чаще используют таблицы с интервальной разбивкой признака, т.к. даже дискретные по сути признаки обладают таким количеством вариант, что составленная по ним таблица нарушает правило N 3, поскольку число групп в дискретном вариационном ряду должно определяться числом реально существующих значений признака.

Для того, чтобы не потерять информацию и в то же время составить компактную таблицу используют интервальные ряды. Здесь перед исследователем возникает проблема ***определения границ интервалов.*** Необходимо найти оптимальное число групп, количество интервалов признака и установить размер интервалов. Решение этой задачи зависит от степени однородности рассматриваемой совокупности.

В том случае, если совокупность однородна, рекомендуется брать равные интервалы. Необходимо помнить, что при описании тенденции в распределении переменных признака интервалы лучше укрупнить. В том случае, когда значение имеют конкретные данные относительно каждой группы, интервалы имеет смысл сделать небольшими. Таким образом, выбор интервалов зависит от свойств изучаемого процесса или явления и от цели работы, вопрос этот решается содержательным , качественным анализом и зависит от профессиональных навыков историка.

Однако, существует несколько формальных способов определения оптимальной величины интервала, т.е. такого его значения, при котором просматривалась бы специфика явления и в то же время группировка не была бы громоздкой. Наиболее проста в употреблении формула, предложенная Г.Стерджессом:

$$K = \frac{X_{max} - X_{min}}{1 + 3{,}2 \lg n}, \quad \text{где}$$

K - величина интервала;

Xmax - наибольшее значение признака;

Xmin - наименьшее значение признака;

п - число элементов совокупности.

Разберем применение формулы.

Дано: 100 рабочих со стажем от 1 до 42 лет. Определить оптимальную величину интервала для группировки рассматриваемой совокупности по стажу.

$$X_{max} = 42; \qquad X_{min} = 1; \qquad n = 100.$$

$$K = \frac{42 - 1}{1 + 3{,}2 \lg 100} \approx 5{,}5$$

Таким образом, оптимальной величиной интервала является 5,5 и группировка примет следующий вид:

стаж	1 -6,5	6,5 - 12	...

Пользуясь данными, группировка которых произведена статистиками-профессионалами, мы должны знать, что разбиение группировочного признака ими выполняется таким образом, чтобы распределение частот в каждой группе было примерно равным. В решении многих задач историку предпочтительнее пользоваться первичными, несгруппированными материалами, производя группировку и перегруппировку данных самостоятельно, в соответствии с целью своего исследования.

Границы интервалов для дискретных признаков устанавливаются без совпадения крайних показателей смежных интервалов. Так, например, группировка количества учащихся в классе должна выглядеть примерно так: 9 - 15; 16 - 22; 23 - 28...

Это объясняется свойствами самого признака, принимающего только целые значения. Значит, при совпадении крайних значений соседних интервалов один и тот же показатель может быть зафиксирован в двух интервальных группах. В нашем примере ясно, что подгруппа детей в количестве 15 человек относится только к первому интервалу (9 - 15).

Противоположное правило применяется для дробных (непрерывных) признаков - обязательное совпадение смежных границ интервалов. Предположим, необходимо произвести группировку признака "Стаж работы по найму" с вариантами от 1 месяца до 42 лет. Характер признака - непрерывный. Согласно правилу о совпадении границ интервалов распределение примет следующий вид: до 1 г.; 1 - 8; 8 - 20; 20 - 30; 30 - 40; 40 и более. Во вторую или в третью группу следует отнести стаж 8 лет?

Данное распределение предполагает наличие предлогов "от" и "до" и в полном виде выглядит так: до 1 г.; от 1 до 8; от 8 до 20; от 20 до 30 и т.д. Следовательно, стаж "8 лет" входит в третью группу.

Статистика различает закрытые и открытые интервалы. В первом случае указывается верхняя и нижняя границы интервала, во втором - определена только верхняя или нижняя граница (например, "до 1 года" или "20 десятин и более")

В случае неоднородной совокупности объектов изучения в основу построения группировки кладется качественный критерий, призванный выявить однородные типы. Этот принцип направлен на то, чтобы определить границы интервалов там, где количественное изменение признака приводит к появлению нового качества. В случае необходимости совокупность разбивается на однородные группы, внутри каждой строится своя шкала интервалов. Например, городское население России конца XIX в. выступает как качественно неоднородная совокупность. В основу ее группировки можно положить признак сословной принадлежности, а в случае необходимости каждую сословно-однородную группу можно подразделить в соответствии с задачами работы по какому-либо качественному или количественному признаку. Изучая купеческое сословие, количественный признак "размер оборотного капитала" может выявить качественное различие - купцов I гильдии и купцов II гильдии.

Формальные математико-статистические методы не могут оказать существенную помощь историку в выборе принципа группировки. Это прерогатива качественного анализа.

Метод группировки позволяет сложное явление представить через ряд более простых, что помогает прийти к анализу всей системы в целом. Метод способствует оценке информационного потенциала источника. Методом группировки характеризуются типы явлений в их взаимных отношениях, а также вскрывается причинная зависимость между отдельными факторами и общей тенденцией развития процесса.

В науке различают 3 основных вида группировок.

Типологические - расчленяют качественно - разнородную совокупность на однородные группы, на типы. В основу группировки закладывается качественный признак. Так, например, широко известно использование типологической группировки В.ИЛениным для анализа крестьянских хозяйств: пролетарских, собственно крестьянских и капиталистических (Поли. собр. соч. - Т.19. - С.326-330) или - бедняцких, середняцких и кулацких (Полн.собр.соч. - Т.3. - С .94-96, т.31. - С.426). Примером типологической группировки выступает распределение промышленности периода НЭП по социальным секторам - государственная, кооперативная, частная.

Структурные - представляют качественно-однородную совокупность в виде количественных групп. В основу этих группировок закладывается количественный признак. Примером может служить распределение рабочих по стажу; - по размерам заработной платы; - по возрасту и т.п.

Одним из важнейших моментов использования рассматриваемого метода и определения вида группировки является выбор группировочного признака. От приема группировки, от базового признака во многом зависят выводы, которые можно получить на основе одних и тех же материалов, анализируемых одним и тем же методом группировки. Главное требование к группировочному признаку - достоверность отражения структуры изучаемого явления в зависимости от времени и конкретно-исторических условий. Выполнение этого правила зависит от профессиональных навыков историка, т.е. от качественной стороны исследования.

Деление группировок на типологические и структурные относительно и зависит от характера задач, стоящих перед исследователем. Например, если задать границы землепользования, соответствующие определенным социальным группам крестьянства, то можно изучить и структуру крестьянских хозяйств по размерам землепользования и типы хозяйств относительно размеров землепользования. Таким образом, отличие заключается в подразделениях группировочного признака.

Третий вид - *аналитические* группировки. Они позволяют установить и на определенном уровне изучить взаимосвязь между признаками. В статистической литературе такие группировки еще называют факторными, при этом один из группировочных признаков рассматривается как результат, а другой - как фактор. Например, дана группировка малых предприятий по размерам прибыли и продолжительности оборота средств. Ясно, что при одном и том же сроке оборота капитала предприятия могут иметь разную прибыль. Следовательно признак "оборачиваемость средств" - фактор (иными словами - условие), а признак "прибыль" - результат. Чтобы установить связь между признаками, данные группируются по признаку-фактору.

Изучение и интерпретация данных аналитической группировки должны начинаться с предварительного выяснения принципиальной возможности существования связи между признаками.

Пример 2.1.:

(Здесь и далее первая цифра означает номер главы, вторая - номер примера):

Группировка рабочих по полу и месту рождения(в %).

Пол	Место рождения	
	Город	Село
мужской	7,5	92,5
женский	4,2	95,8

Исследователь обязан заинтересоваться, зависит ли пол человека от места рождения или определяется ли место рождения рабочего его полом? Ясно, что подобная группировка лишена смысла изучения взаимосвязи сведенных признаков.

Пример 2.2:

Группировка рабочих по стажу на данном предприятии и количеству детей.

Стаж	Количество детей		
	0	1	2 и более
до 1 года	39	8	14
1 - 5	19	2	6
5-10	3	15	9

В данном случае принципиальная связь между стажем работы на данном предприятии и количеством детей в семьях рабочих возможна. Следующим этапом работы можно считать построчное сравнение показателей. В рассматриваемом примере мы констатируем разброс количественных характеристик, что говорит об отсутствии связи между признаками.

Пример 2.3:

Группировка рабочих по стажу на данном предприятии и общему стажу работы по найму.

общий стаж	Стаж на данном предприятии					
	до 1	1-5	5-10	10-15	15-20	20 и более
до 1 г.	49	_	-	-	-	-
1-8	16	9	2	-	-	-
8-10	1	7	4	1	1	-
20-30	_	3	13	10	4	1
30-40	_	-	11	12	9	7
40 и более	-	-	7	5	12	6

Приведенный пример показывает, что данные концентрируются по диагонали таблицы (концентрация может выражаться также скоплением наиболее крупных значений признака около диагонали). Это говорит о наличии существенной, тесной связи между признаками. Направление диагонали свидетельствует о характере выявленной связи(таблица читается так же, как текст - слева направо, и наш взгляд, читая данные по строчкам постепенно движется из верхнего левого угла таблицы в нижний правый ее угол, т.е. идет по диагонали таблицы сверху вниз) Если диагональ идет сверху вниз (см. пример 2.3), то связь - прямая; снизу вверх - обратная.

Прямая связь означает, что увеличение значений одного признака (факторного) ведет за собой увеличение значений второго (результативного). При обратной связи росту значений факторного признака соответствует уменьшение значений результативного признака.

Метод группировки позволяет примерно определить роль того или иного признака в изучаемом явлении, силу его влияния на другие характеристики. Для этого используют безусловные распределения (структурные группировки значений интересуемого признака).

Пример 2.4:

Распределение служащих N-ского управления по возрасту.

возраст	20-25	25-30	30-35	35-40	40-45	45-50	50 и более
кол-во служ. в%	0,3	0,7	1,9	3,7	86,9	4,0	2,5

Данное распределение, при котором показатели сконцентрированы в одной группе, позволяет пренебречь рассмотренным признаком при дальнейшем изучении явления на уровне взаимосвязей, т.к. он не будет оказывать заметного, существенного влияния на другие. Однородный по возрасту (см. пример 2.4) коллектив может считаться гарантированным от конфликта "отцов" и "детей". Однородный по национальному составу коллектив не будет переживать межэтнических противоречий и т.д. Следовательно, в случае кризиса в данной группе работников причину надо искать вне рассмотренного таким образом признака. Если показатели в безусловном распределении более-менее равномерно охватывают изучаемую совокупность объектов, допустимо предположить, что этот признак будет оказывать активное влияние на анализируемое явление или процесс.

С помощью безусловного распределения можно выявить группу факторов, определяющих изучаемое событие и выделить признаки, имеющие место, но не играющие заметной роли.

• * *

Особое место среди группировок занимают ***динамические ряды,*** отражающие изменение явления во времени.

Если ввести хронологический показатель в уравнение 1 обезьяна + 1 обезьяна, то результат не обязательно будет равен 2 обезьяны. В зависимости от времени это может быть и целая стая, и человек, и одна обезьяна (вторая может умереть) и... бесчисленное множество вариантов.

Необходимость учета фактора времени раскрыта С.Маршаком в знакомом с детства стихотворении "Дама сдавала в багаж..." Исчезновение маленькой собачонки железнодорожники компенсировали рослой дворняжкой и объяснили фразой: "Однако за время пути собака могла подрасти". С формально-логической точки зрения ответ безупречен, а это значит, что без знания того, как менялись в течение этого времени характеристики изучаемого процесса или объекта, как менялась взаимосвязь его показателей, роль и значение каждого признака или их группы получить адекватную картину исторического процесса невозможно.

Для историка все это имеет особое значение, т.к. чаще всего мы изучаем процесс, имеющий определенную протяженность во времени.

Включение в группировку хронологического фактора обусловливает специфические требования к ее построению и методам анализа

Динамические ряды бывают моментные, **в** которых время задано в виде конкретных дат (моментов времени) и интервальные, где время задано в виде промежутков - лет, месяцев, суток... Показатели временного ряда называются **уровнями.**

Исследование динамических рядов начинается с доказательства нижеперечисленных требований предъявляемых к их построению.

1. Однородность явлений относительно каждой динамической группы, т.е. в один временной промежуток должны включаться одни и те же явления. Например, динамика обеспечения крестьянских дворов рабочим скотом потеряет смысл, если в одной графе - кони, в другой - волы или - в одном временном промежутке - уровень скота старше 3-х лет, а в другом - молодняка.

2. Неизменность территории, к которой относятся показатели. Это особенно важно проверить, т.к. история знает примеры, когда юридический статус территории остался прежним, а адми-

нистративно-территориальные границы изменились. Так, например, многократно изменялись границы Казанской губернии. Этим правилом пренебрегают, если цель исследования связана с изучением динамики тех или иных показателей в зависимости от изменения границ территории.

3. Единство методологии учета показателей. Уровни динамического ряда могут быть заданы либо абсолютными, либо относительными, либо средними величинами (соответственно подразделяются ряды динамики).

Историк, имеющий дело с относительными или средними величинами, обязан проверить историю их возникновения, т.е. выяснить на единицу чего подсчитывались имеющиеся показатели. Например, средняя урожайность крестьянских хозяйств в одни годы подсчитывалась с засеянной площади, а в другие - с убранной. Такие уровни несопоставимы и включать их в один динамический ряд неправомерно.

4. "Временной показатель, положенный в основу динамического ряда, в случае его интервальной разбивки должен иметь сопоставимые временные промежутки". (Славко Т.Н. Указ.соч., с.37). Вряд ли имело бы смысл сопоставление уровней динамического ряда, если в первой его графе указан период VI - VIII вв., во второй - IX век, а в третьей - 988 год. Изменения в показателях при такой хронологической разбивке нельзя было бы принимать за закономерные.

Показатели динамического ряда считаются сравнимыми при выполнении всех четырех условий его построения.

Метод группировки является необходимым начальным этапом анализа массовых источников. Он решает задачи установления статистических связей и закономерностей, выявления структуры изучаемой совокупности, определения и описания ее типов. С его помощью сложное явление или событие предстает в виде нескольких более простых. Подробный раздельный анализ последних приводит к получению целостного представления об исходной системе и полнее раскрывает информационный потенциал источника. Метод тесно связан с табличной и графической формой представления данных. Таблицы приводятся в целях обеспечения последующего анализа информационной базы и для иллюстрации полученных выводов. Таблицы необходимы для ясного описания существа проблемы.

ПО ТЕОРИИ МЕТОДА ДОПОЛНИТЕЛЬНО ЧИТАЙТЕ:

1. Елисеева И.И., Рукавишников В.О. Группировка, корреляция, распознавание образов. - М., 1977.

2. Елисеева И.И., Юзбашев М.М. Общая теория статистики.- М., 1995.- С.104-131.

3. Общая теория статистики. - М., 1985. - С.25-41.

4. Славко Т.И. Математико-статистические методы в исторических исследованиях. - М., 1981. - С.29-39.

5. Эренберг А. Анализ и интерпретация статистических данных. - М., 1981. - С.52-57.

Лекция 3.
ФОРМЫ ГРАФИЧЕСКОГО ИЗОБРАЖЕНИЯ.

Если в задачи исследования входит необходимость подчеркнуть какую-либо особенность данных, провести их сравнение, дать общее представление о закономерностях изменения в тех или иных показателях, обращаются к графикам. Они представляют собой условные изображения числовых величин и их соотношений с помощью линий, точек, геометрических фигур и т.п. Этим достигается наглядность, образность, эмоциональность, в определенной степени повышается эффективность восприятия материала

В исторической литературе используется несколько видов графиков, выбор которых зависит от цели работы и от характера представляемых данных. Графики классифицируются статистикой по способу построения. Историки чаще всего обращаются к диаграммам - полигонам распределения, гистограммам, кумулятам. В любом случае график (как и таблица) сопровождается заголовком (допустимо его размещение как над рисунком, так и под ним). Заголовок должен содержать информацию о характере изображенного показателя, единицах его измерения, территории и времени его определения

Гистограмма распределения. (Гистограмма - от греческого "гистос"- ткань; строение.) Это вид столбиковой диаграммы, применяемой для интервального ряда. На оси ОХ (абсцисс) откладываются интервалы значений варьирующего признака, а на оси OY (ординат) частоты признака, соответствующего масштаба.

Пример 3.1.:

Построим гистограмму распределения по данным таблицы: Распределение населения РФ по среднедушевому совокупному доходу в 1992 г.

среднедушевой доход в месяц (в тыс.руб.)	до 1	1-2	2-3	3-4	4-5	5-6	6-7	7-9	>9
кол-во чел. (в млн.)	7,0	32,6	34,2	25,2	20,0	9,8	6,3	7,0	6,6

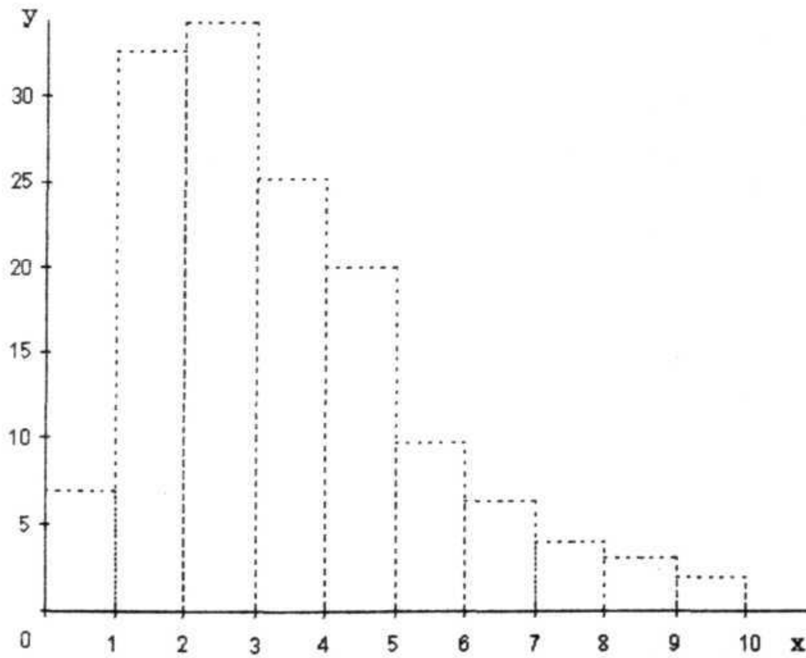

Рис.3.1. Гистограмма распределения населения РФ по среднедушевому совокупному доходу в 1992 г.

На оси абсцисс рисунка 3.1 отложены варианты признака (в нашем случае - среднедушевой доход в месяц в тыс. руб.), а на оси ординат его частоты (в нашем случае - количество человек в млн.). Каждое деление OY соответствует 5 млн. единиц, а OX - 1 тыс. На оси OX строим прямоугольники, высоты которых равны частотам соответствующих интервалов. Так мы поступаем, последовательно дойдя до интервала 7-9 (тыс. руб.). Все предыдущие интервалы были равны по величине 1 (тыс. руб.), а этот равен 2 (тыс. руб.). Здесь в действие вступает правило, согласно которому для частот больших интервалов на гистограмме распределения берется меньший масштаб (и наоборот - для частот меньших интервалов берется больший масштаб на графике). Следуя этому правилу, на графике мы должны построить прямоугольник, высота которого будет не 7 (млн. чел.) как указано в таблице примера 3.1, а 7/2, т.е. 3,5 (млн. чел.), т.к. величина данного интервала больше в 2 р. величины предыдущих интервалов.

Встречаются ситуации, затрудняющие выполнение этого правила. Например, когда все интервалы неравны по величине. Тогда на графике откладываются на оси ординат не частоты признака, а его плотности. **Плотность** - величина равная отношению частоты признака к величине соответствующего интервала, обозначается знаком "f".

$$f = P_i / h_i \quad \text{где}$$

f - плотность распределения;

P_i - частота признака;

h_i- величина интервала.

Соответственно высота прямоугольников на графике будет равна плотности варьируемого признака.

Выполнение этого правила помогает точнее характеризовать тенденцию изменения значений признаков, выявить характер распределения.

Полигон распределения. (Полигон - с греческого - многоугольник). Это вид линейного графика, представляющий собой замкнутую ломаную линию (с обязательными точками нулевых частот до первой и после последней вариант признака). Полигон распределения обычно используют для дискретного ряда. На оси абсцисс откладываются варианты признака, на оси ординат - его частоты. В системе координат наносятся точки, соответствующие заданным величинам. После чего эти точки соединяются ломаной линией, которая в начале и в конце должна прийти к нулевому значению.

Пример 3.2:

Построим полигон распределения по данным таблицы:

Распределение служащих по возрасту:

возраст	18	22	23	28	30	35
кол-во служащих	2	10	13	15	21	18

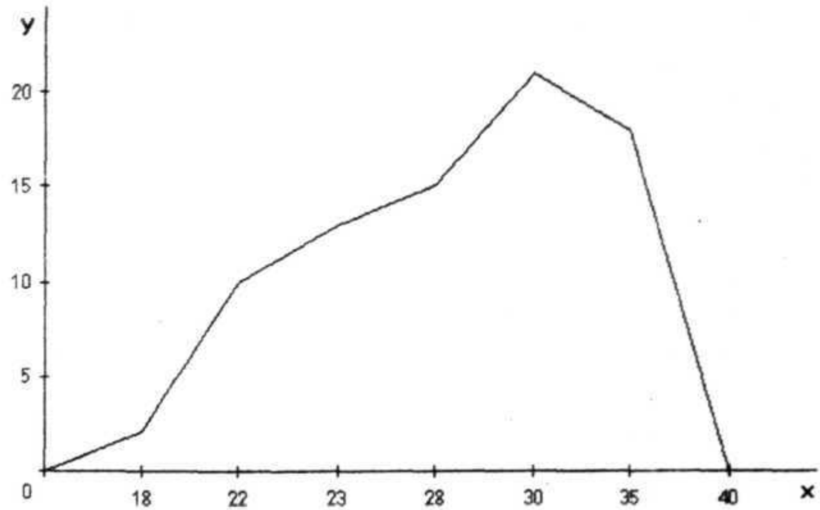

Рис.3.2. Полигон распределения служащих по возрасту.

Для того, чтобы фигура получилась замкнутой, мы на графике ввели дополнительно две варианты признака (16 и 36 лет), частоты которых равны нулю.

Полигон распределения можно получить и на основе интервального ряда. Для этого ломаной линией соединяют середины верхних оснований прямоугольников гистограммы распределения (см. Рис.3.3). Можно, не строя гистограмму, сразу на оси абсцисс отложить середины заданных интервалов, получив нужные точки на пересечении перпендикуляров, проведенных от середины интервалов на оси OX и из соответствующих значений частот на оси OY (см. Рис. 3.3).

Полигон распределения полезно получить в случае неравных интервалов. Он точнее характеризует закономерность изменения значений признака и решает проблему открытых интервалов. Кроме того, правильно построенный полигон распределения позволяет выявить тенденцию, скрытую табличной формой представления данных. Например, имеется структурная группировка грамотного сельского населения N-ского района по возрасту по данным 1928 года:

возраст	до 10	10-20	20-30	30-40	40-60	60-80	более 80
кол-во грамтн. в тыс.	16,7	41,2	59,0	47,4	56,6	38,2	10,3

Вполне закономерно выглядит увеличение числа грамотных людей по мере увеличения возраста до 30 лет, затем должно бы, исходя из исторических условий, наступать снижение уровня грамотности по мере старения населения. Однако в таблице выделяется V группа (56,6 тыс. грамотных в возрасте 40 - 60 лет). На графике этого "пика" грамотности не будет, т.к., следуя вышеуказанному правилу на оси ординат мы отложим не 56,6, а 56,6/2, т.е. 28,3 (тыс. чел.), поскольку V группа выделяется вдвое возросшим интервалом. Аналогично поступаем и с Y1 группой. Таким образом, можно сделать вывод о том, что наибольшее количество грамотного сельского населения было в возрасте 20 - 30 лет, а затем наблюдается явная тенденция к снижению количества грамотного населения по мере увеличения возраста.

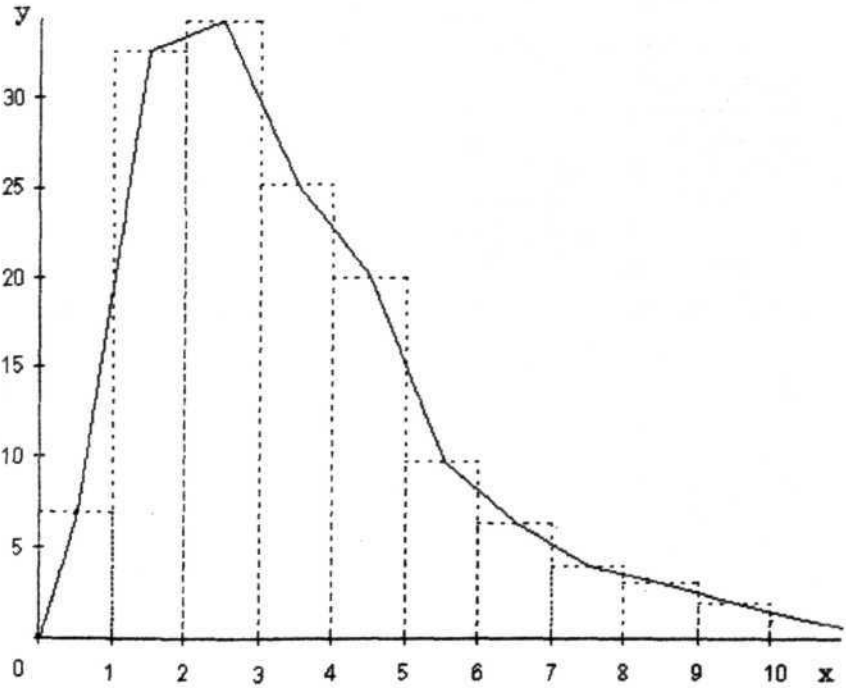

Рис. 3.3. Полигон распределения населения РФ по среднедушевому совокупному доходу в 1992 г.

Кумулята (от позднелатинского "скопление"). Это вид линейного графика, представляющий собой плавную кривую. Кумуляту используют как для дискретных, так и для интервальных вариационных рядов. На оси абсцисс откладывают значения рассматриваемого признака, а на оси ординат - накопленные частоты. Чтобы получить такой график необходимо предварительно преобразовать вариационный ряд в ряд накопленных частот (кумулятивный ряд). Он получается путем последовательного сложения частоты каждого класса с суммой предыдущих классов.

Пример 3.3:

Рассмотрим построение кумуляты по данным примера 3.1.

Распределение населения РФ по среднедушевому совокупному доходу в 1992 г.

среднедуш. доход в мес. (в тыср.)	до 1	1-2	2-3	3-4	4-5	5-6	6-7	7-9	>9
кол-во чел. (в млн.)	7,0	32,6	34,2	25,2	20,0	9,8	6,3	7,0	6,6
накопл. частоты	7,0	39,6	73,8	99,0	119,0	128,8	135,1	142,1	148,7

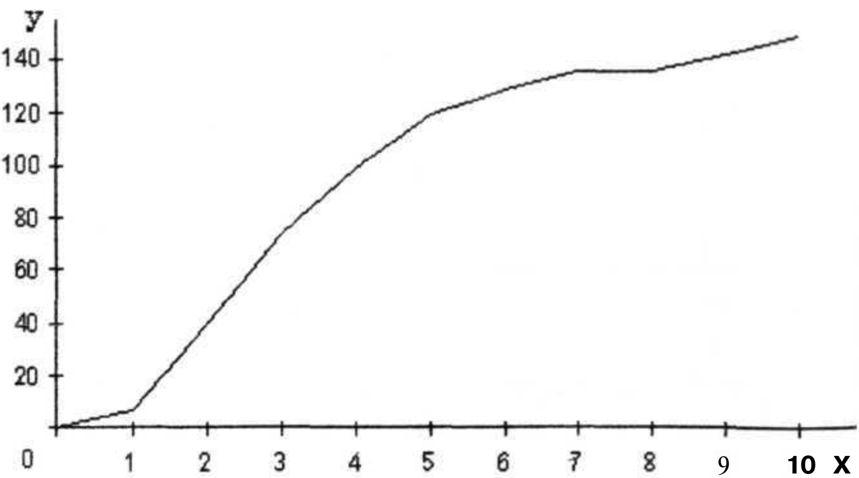

Рис. 3.4. Кумулята распределения населения РФ по среднедушевому совокупному доходу в 1992 г.

Тренд. При графическом изображении динамического ряда на оси абсцисс откладывается шкала времени, а на оси ординат - шкала уровней ряда. Показатели динамического ряда отмечаются точками в установленной системе координат. Затем эти точки соединяют прямыми линиями. Полученная в результате ломаная линия не должна быть замкнутой, не должна (если это **не**

предусмотрено показателями, данными источника) брать свое начало в нулевой точке и заканчиваться на нулевой отметке.

Пример 3А.:

Распределение численности населения СССР за 1975-1980 гг. (на 1 января).

год	1975	1976	1977	1978	1979	1980
кол-во насел, в млн.ч.	253,3	255,6	257,9	260,1	262,4	264,5

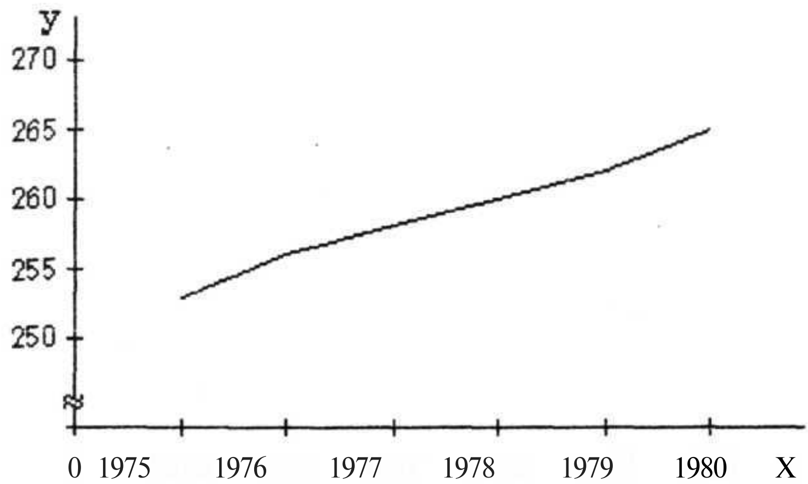

Рис 3.5. Динамика численности населения СССР в 1975-1980 гг.

Важно отметить, что ось ординат обязательно должна иметь начало в точке "нуль". Затем, если необходимо (как в нашем примере), то можно "разорвать" ось ординат двумя волнистыми линиями (см. Рис.3.5). Это делается в том случае, когда наличествуют значительные изменения в величинах рассматриваемого показателя или (как в нашем примере) колеблемость частот признака небольшая, но велики единицы его измерения.

Закономерность в изменениях уровней ряда иногда проявляется довольно ясно (см. Рис.3.5), а иногда она может затушевываться разного рода колебаниями. Тенденция динамики связана с действием долговременно существующих причин и условий развития. Колебания - с действием краткосрочных, случайных или циклических факторов, влияющих на отдельные уровни динамического ряда. Основная тенденция в изменении уровней называется трендом, выявление которого - одна из первых и важнейших задач изучения динамического ряда.

Обнаружить тренд можно различными методами - методом скользящей средней, наименьших квадратов, с помощью среднего прироста и т.д. Одним из приемов определения тренда выступает график. Для этого в системе координат уровни динамического ряда отмечают точками, а затем на основе зрительного анализа месторасположения точек проводят среднюю линию, которая называется трендом и отражает основную тенденцию развития.

Пример 3.6.

Динамика урожайности картофеля в 1883-1892 гг.

годы	1883	1884	1885	1886	1887	1888	1889	1890	1891	1892
урож. ц/га	145	168	146	177	176	190	186	176	211	170

Произвольное сравнение данных за отдельные годы приводит к выводам, значительно отличающимся друг от друга. Если сравнить урожайность 1891 г. с уровнем 1883, то получается, что за 8 лет она выросла на 66 ц. с 1 га, т.е. более, чем по 8 ц. с 1 га за год. В то же время, если сравнить урожайность 1892 г. с уровнем 1884 г., то увидим, что за 8 лет, из которых 7 лет те же, что и в предыдущем сравнении, урожайность возросла лишь на 2 ц. с 1 га. Помочь в этой ситуации может график (см. Рис.3.6).

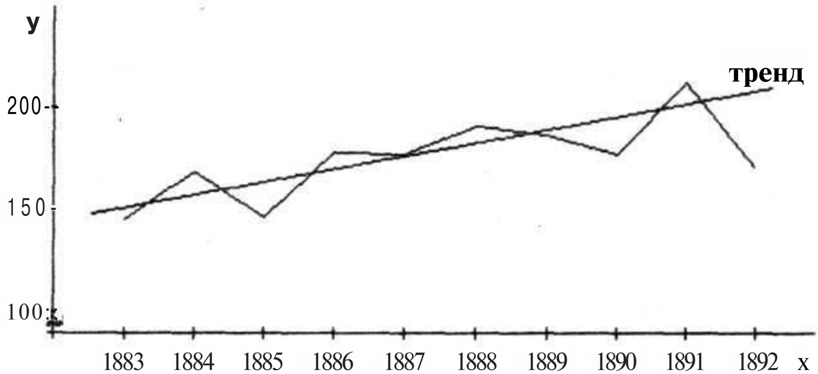

Рис.3.6. Динамика урожайности картофеля в 1883-1892 гг.

На Рис.3.6 хорошо видно, что в период 1883-1893 гг. наблюдается устойчивая тенденция к росту урожайности, колебания носят хаотический, скорее случайный характер.

Задачи графического метода не исчерпываются наглядностью. Графики позволяют приближенно получить средние характеристики - моду и медиану. Графиками проверяется характер и форма зависимости между признаками, что особенно важно для доказательства правомерности применения методов корреляционного анализа. На графике сразу видны пределы изменения показателей, их колеблемость, скорость изменения, выявляются и характеризуются закономерности.

Если исследователь испытывает необходимость получить изображение географической характеристики изучаемого явления, он может применить *картограмму.* На ней условными знаками, точками, штриховкой, цветом ... отмечается распределение изучаемого признака по определенной территории - стране, району, области, городу и т.п., обозначается интенсивность его проявления.

Картограммы помогают определить закономерность распространения какого-либо признака по территории, тенденцию его распределения между регионами. Так, например, картограммы широко используются для характеристики восстаний, забастовочного движения, миграционных связей, для анализа экономического развития отдельных территориальных образований.

В зависимости от задач исследования графики размещают в тексте работы или в приложении к ней. Чаще всего небольшие по формату рисунки иллюстративного характера, подтверждающие ранее полученные выводы, располагают по мере изложения материала в тексте исследования.

Вместе с тем, графический метод имеет свои ограничения. Во-первых, график не может включить столько данных, сколько может войти в таблицу. Во-вторых, на графике показываются всегда приблизительные, округленные значения, а значит пропадают детали, фиксируется только общая ситуация. В-третьих, построение графика, его точность во многом зависят от аккуратности исследователя.

На наш взгляд, эти замечания не снижают полезности и важности использования графического метода в исторических исследованиях.

ПО ТЕОРИИ МЕТОДА ДОПОЛНИТЕЛЬНО ЧИТАЙТЕ:

1. Елисеева И.И., Юзбашев М.М. Общая теория статистики.- М., 1995. - С.53-65, 87-89.

2. Герчук ЯП. Графики в математико-статистическом анализе. - М., 1972.

3. Количественные методы в исторических исследованиях.- М., 1984. - С.97-101.

4. Общая теория статистики. - М., 1984. - С.347-366, 188-190.

Лекция 4.
СРЕДНИЕ ВЕЛИЧИНЫ.

На каждый исторический факт, на каждое событие действует множество различных причин и сил, способствующих и препятствующих его появлению. Пытаясь классифицировать изучаемое явление, мы сталкиваемся с необходимостью выявления общих характеристик, относящихся как к любому элементу рассматриваемой совокупности, так и ко всей совокупности в целом. Такими общими характеристиками, раскрывающими определенные свойства и направление развития процесса, выступают средние величины.

Категория средней величины имеет одну из самых древних историй. Процесс становления абстрактных понятий связан с отбором общих черт некоего предмета или явления. При этом стираются, отбрасываются свойства, присущие исключительно отдельным объектам изучаемого явления. Так, обратившись к истории языка, можно заметить долговременное применение лексических единиц, выражающих понятие "снег" через его характеристики, через его проявления - "падающий с неба", "холодный", "мягкий", " мокрый", "чистый", "тающий" и т.д. В данном случае "снег" является обобщением, абстрактным понятием, вбирающем в себя все типичные признаки конкретного природного явления.

Практическое применение средние нашли в деле налогообложения в странах древнего мира. Оно основывалось на процедуре усреднения доходов разных социальных категорий граждан.

Теоретическое осмысление средних можно найти в трудах античных философов. Оно отражено в понятии гармонии, в процессе поисков общих закономерностей. В произведениях Аристотеля, Гераклита, Архимеда, Пифагора и других содержится понимание средней как равнодействующей всех определенных условий, которые участвуют в образовании рассматриваемой совокупности индивидуальных величин.

В истории науки один из первых, кто попытался придать средней величине статистический смысл был английский ученый В.Петти (1623 - 1687 гг.). Он раньше других ввел средние статистические показатели в разработку экономической теории. Спустя более 100 лет началось последовательное развитие теории самих средних. Ее родоначальником принято считать А.Кетле (1796 - 1874 гг.). Он, опираясь на философию французского позитивиста О.Конта, разрабатывал теорию всеобщих закономерностей, которые выступают в форме устойчивых во времени статистических результатов. Они лишены индивидуальности, это массовые закономерности. Однако А.Кетле прославился более не как философ-метафизик, а как социолог-математик. В его трудах теория средних величин опирается на математическую основу. Вплоть до настоящего времени категория "средняя величина" является важнейшим логическим узлом научного аппарата, и дискуссии по ее концептуальной ценности продолжаются.

Главное значение средних состоит в их обобщающей функции, т.е. в замене множества различных индивидуальных значений признака средней величиной, характеризующей всю совокупность явлений. Средняя отражает совокупный результат развития и является равнодействующей различных причин и сил, воздействующих на эти явления.

В исторической науке средние величины присутствуют давно, но не в полной мере. Для обработки массовых данных в статистике разработаны средние гармоническая, геометрическая, квадратическая, а также описательные средние - мода и медиана. Историки же традиционно обращаются, главным образом, к средней арифметической.

Использование средних предполагает следование определенным правилам.

1. До вычисления средних необходимо обеспечить качественную однородность совокупности.

Так, например, нельзя изучать среднюю землеобеспеченность по общим данным о наделах крестьян, мещан, дворян, купечества. Нарушение указанного принципа не позволит нам получить типическую характеристику признака в изучаемой совокупности.

2. Средние вычисляются по массовым данным, т.е. по данным достаточно большого числа единиц наблюдения.

Если обратиться к тому же примеру о средней землеобеспеченности, то согласно второму правилу нельзя изучать среднюю землеобеспеченность дворян по данным о размерах двух-трех имений. Мы обеспечили качественную однородность наблюдаемой совокупности, выделив группу дворян. Но для получения исторически реальной картины необходимо расширить число фактов. Это помогает снизить влияние недостоверной или нетипичной информации.

Средние рекомендуется вычислять по сведениям массовых источников, где действует закон больших чисел. Чем значительнее количество наблюдаемых фактов, тем бывает легче отделить случайное от необходимого.

В жизни чаще всего то общее и существенное, что свойственно всем явлениям одного вида скрыто их индивидуальными особенностями. Следовательно, невозможно вскрыть общее, рассматривая отдельные, малочисленные случаи. Чем больше единиц наблюдения, тем значительнее отвлекается средняя величина от специфических черт индивидуальных явлений.

3. Нельзя ограничиваться вычислением средней в целом по совокупности, не меньшее значение имеют средние характеристики и для каждого отдельного типа.

Используя тот же пример, можно предложить рассчитать средние величины землевладения дворян для разных губерний, для дворянства разных сословных групп (потомственного, личного, служилого), для дворянства разного экономического положения (безпоместного, малопоместного и др.) и так далее, в зависимости от цели и задач конкретного исследования.

На практике статистика использует средние величины, обобщающие явно неоднородные явления. Это особенно важно помнить при работе с уже сгруппированными данными и средними величинами, исчисленными до вас. В этом случае нужна проверка типичности средней величины по базовому группировочному признаку.

Средняя не сводится только к количественному выражению "индивидуальных уклонений". Одна из главных задач научного исследования - выявление закономерностей. Метод средних, игнорируя каждый отдельный случай, устанавливает их общее распределение в конкретных условиях места и времени. Средняя является специфической формой выражения содержания общего закона, который выступает в виде тенденции.

Средняя арифметическая $(\overline{X})$. - является самым распространенным видом средних величин. Если в исследовании автор не указывает вид примененного среднего показателя, подразумевается средняя арифметическая. Она исчисляется путем отношения суммы всех значений признака к общему числу наблюдений.

$$\overline{X} = \frac{X_1 + X_2 + X_3 + \ldots + X_n}{n},$$ где

$X_ь$ X_2, X_3... $X_п$, - варианты признака;

n - число единиц наблюдения.

Пример 4.1:

Даны сведения о заработной плате шести работников (в условных единицах) - 90, 120, 108, 206, 160, 184. Определить средний размер заработной платы данной совокупности работников.

$$\overline{X} = \frac{90 + 120 + 108 + 206 + 160 + 184}{6} \approx 144{,}67$$

Смысл $\overline{X}$ (в данном примере) сводится к показу, какой была бы заработная плата каждого работника, если бы они получали ее поровну. Согласно вычислению их средняя заработная плата равняется 144,67 условных единиц.

Если значения изучаемого признака в совокупности не повторяются (см. Пример 4.1), то любое значение этого признака оказывает на величину $\overline{X}$ одинаковое влияние, т.е. имеет одинаковый "вес".

Если изучаемая совокупность велика, исходная информация чаще всего представлена группировкой, где значения усредняемого признака встречаются по нескольку раз и частота их различна. Это значит, что любая варианта этого признака оказывает неодинаковое влияние на среднюю величину, которая должна представлять собой результат равномерного распределения значений признака. Для уравновешивания указанных влияний используют средневзвешенную величину, равную сумме произведений каждого значения признака на его частоту, деленной на сумму всех частот.

$$\overline{X} = \frac{\sum XP}{\sum P}, \qquad \text{где}$$

X - варианты признака;

P - частота вариант (в статистической литературе для обозначения частоты используют букву "f", в последние годы это обозначение проникает и в историческую литературу).

Пример 4.2:

Распределение футбольных матчей высшей лиги России по числу забитых мячей за игру в 1992 г.

число забит, мячей	0	1	2	3	4	5	6	7
число матчей	21	41	42	37	19	10	6	3

Определить среднее число забитых голов за одну игру.

$$\overline{X} = \frac{0\cdot 21 + 1\cdot 41 + 2\cdot 42 + 3\cdot 37 + 4\cdot 19 + 5\cdot 10 + 6\cdot 6 + 7\cdot 3}{21 + 41 + 42 + 37 + 19 + 10 + 63} \approx 2{,}34$$

Согласно вычислению в среднем за одну игру футбольных матчей высшей лиги в 1992 году забивалось 2.34 мяча. Обращает на себя внимание тот факт, что величина средней арифметической может принимать дробные значения даже для дискретных признаков. Об этом важно помнить при интерпретации результатов вычислений.

Если в группировке значения осредняемого признака заданы интервальным рядом, то при исчислении средней арифметической в качестве значения признака берутся середины интервалов. Условно предполагается, что единицы совокупности распределены равномерно по интервалу.

Для открытых интервалов значения признака определяются экспертным путем, качественным анализом, исходя из сущности и свойств природы признака. Исследователь не всегда имеет возможность провести подобную экспертизу. В этом случае можно использовать формальный способ прибавления единицы к максимальному определенному значению и вычитания единицы из минимального заданного значения признака.

Пример 4.3:

Распределение рабочих N-ского предприятия по возрасту.

возраст	до 20	20-30	30-40	40-50	страше 50
число рабочих	48	120	75	62	54

Определить средний возраст представленной группы рабочих

Мы можем предположить, что минимальный возраст рабочих - 17 лет (возраст получения общего среднего образования), а максимальный - 65 лет (по экспертной оценке - наиболее типичный возраст прекращения трудовой деятельности). Тогда первый интервал становится "17 - 20", а последний - "50 - 65", соответственно середины интервалов - 18,5 и 57,5 (лет).

Расчет проводится по формуле взвешенной средней арифметической, т.к. частоты вариант признака различны.

$$\overline{X} = \frac{18{,}5 \cdot 48 + 25 \cdot 120 + 35 \cdot 75 + 45 \cdot 62 + 57{,}5 \cdot 54}{48 + 120 + 75 + 62 + 54} \approx 34{,}56$$

Предположения наши достаточно условны. Произведем подсчет средней величины, применив формальный способ решения проблемы открытых интервалов. Вместо "до 20" берем 20-1, т.е. 19, а вместо "старше 50" - 50+1, т.е. 51. Формула приобретает следующий вид:

$$\overline{X} = \frac{19 \cdot 48 + 25 \cdot 120 + 35 \cdot 75 + 4 \cdot 62 + 51 \cdot 54}{48 + 120 + 75 + 62 + 54} \approx 33{,}65$$

Как видно из примера разница в показателях $\overline{X}$ несущественна, что иллюстрирует допустимость использования формальных методов.

В практике исторического исследования встречаются ситуации, когда индивидуальные значения осредняемого признака неизвестны. В распоряжении исследователя имеются некие суммарные значения объемных признаков. Средняя величина определяется отношением между имеющимися итоговыми данными.

Пример 4.4:

Известно, что с площади 145 десятин собран урожай в 2595,5 т какой-то продукции. Отношение 2595,5/145 показывает среднюю урожайность данной культуры с одной десятины. В данном примере она равна 17,9 тонн. Этот вид средней называется в статистике ***неявной формой средней.***

Встречаются случаи, когда в распоряжении исследователя имеются относительные показатели признака (доли, проценты, удельный вес и пр.). Общее определение средней арифметической сохраняет силу и в этом случае, но надо иметь в виду, что сумма таких показателей не является реальной величиной какого-либо признака.

Основное свойство средней арифметической состоит в представлении всех значений признака в распределении. Следовательно, ее величина подвержена влиянию как очень больших, так и очень малых вариант. В результате она перестает быть типичной. Особенно это чувствуется при асимметричном распределении. Так, в примере 4.3 при среднем возрасте рабочих 34 года

67,7% рабочих имеют возраст меньше или равный среднему значению и 32,3% были старше. Здесь видна явная асимметрия, обусловленная характером вариации признака. Для общественных явлений это естественно, строгая симметрия практически невозможна, а значит для их изучения мало знать среднюю арифметическую.

Мода. (Мо) представляет наиболее часто встречающееся значение признака в упорядоченной совокупности, наиболее типичное среднее значение.

В дискретном ряду Мо определяется без вычислений как значение признака с наибольшей частотой. Так, в примере 4.2 мода равна 2, т.к. этому значению признак соответствует наибольшая частота (42). Таким образом, чаще всего в 1992 г. за одну игру футбольных матчей высшей лиги России забивалось 2 мяча.

Если в вариационном ряду (в группировке) равная максимальная частота встречается у двух или нескольких значений признака, то он считается соответственно бимодальным или мультимодальным. Это говорит о неоднородности совокупности и, следовательно, надо проверить правильно ли составлена группировка.

Для вычисления моды в интервальном ряду сначала определяется модальный класс, т.е. интервал с наибольшей частотой. Затем Мо вычисляется по формуле:

$$Mo = X_o + K\frac{P_2 - P_1}{2P_2 - P_1 - P_3},$$ где

X_o - нижняя граница модального интервала;

K - величина интервала;

P_1 - частота интервала, предшествующего модальному;

P_2 - частота модального интервала;

P_3 - частота интервала, последующего за модальным.

Вычислим Мо по данным примера 4.3.

$$Mo = 20 + 10\frac{120 - 48}{2 \cdot 120 - 48 - 75} \approx 26{,}15$$

Получается, что наиболее типичный возраст рассматриваемой группы рабочих - 26,15 лет. Этот возраст наиболее часто встречается в данной группе рабочих.

Приближенное значение моды можно определить по графику. Для этого надо построить гистограмму распределения. Внутри "столбика" с наибольшей высотой проводят прямые линии, соединяющие его правый верхний угол с правым верхним утлом предшествующего "столбика", а левый верхний - с левым верхним утлом следующего "столбика". Абсцисса точки пересечения этих линий покажет моду. Проиллюстрируем сказанное графиком, построенным по данным примера 4.3 (см. Рис.4.1).

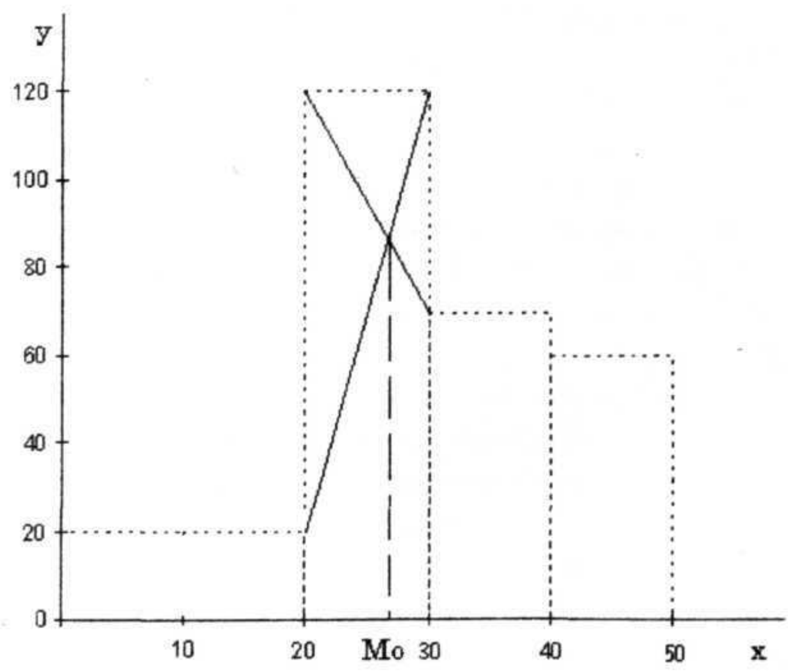

Рис.4.1. Гистограмма распределения рабочих N-ского предприятия по возрасту.

Графическое определение моды применяется во всех случаях, когда в задачу исследования не входит обязательное получение точного значения наиболее распространенной величины признака. Например, для проверки рабочей гипотезы, когда точная величина принципиальной роли не играет, или для повышения наглядности материала. По нескольким графикам можно провести приблизительное сравнение мод различных признаков, чего невозможно сделать по таблицам.

Медиана. (Ме) - величина, определяющая значение признака, находящегося в середине упорядоченной совокупности. Медиана делит изучаемую совокупность так, что число единиц с большим и меньшим, чем медиана значением признака, одинаково.

Чтобы определить Ме в дискретном ряду, надо построить ряд накопленных частот, затем поделить сумму всех частот пополам, а затем по накопленным частотам определить величину варианты, соответствующей той группе, в которой накопленная частота впервые превышает половину общей численности совокупности. В примере 4.2 ряд накопленных частот будет выглядеть так: 21; 62; 104; 141; 160; 170; 176; 179. Полусумма всех частот равна 179/2 = 89,5. Эта величина входит в третью из накопленных частот, т.е. в данном примере третья из накопленных частот своей величиной впервые превысила значение полусуммы всех частот. Следовательно, медиана равна 2, т.е. варианте признака, соответствующей третьей группе. Получив Ме, можно констатировать, что в половине футбольных матчей высшей лиги России в 1992 году забивалось в среднем по 2 мяча.

В интервальной группировке для вычисления Ме необходимо найти медианный интервал - интервал, которому соответствует первая из накопленных частот, превышающая половину суммы всех частот ряда распределения. Затем считают по формуле:

$$Me = X_o + K\frac{\sum P/2 - \sum_{m-1}}{P_m},$$ где

X_o - нижняя граница медианного интервала;

K - величина медианного интервала;

$\sum P$ - сумма частот;

$\sum_{m-1}$ - сумма частот интервалов, предшествующих медианному (накопленная частота в интервале, предшествующем медианному).

P_m - частота медианного интервала.

Определим Me по данным примера 4.3. Ряд накопленных частот принимает следующий вид: 48; 168; 243; 305; 359. Полусумма частот равна 359/2 = 179,5. Полученные данные говорят о том, что медианным является третий интервал, т.е. интервал "30 - 40". Подставляем имеющиеся показатели в формулу:

$$Me = 30 + 10\frac{179{,}5 - 168}{75} \approx 31{,}53$$

Величина Me свидетельствует, что половина рабочих рассматриваемой группы имеет средний возраст 31,53 года.

Примерное значение медианы можно определить с помощью графика. Для этого используют кумуляту, последнюю ординату которой делят пополам и через полученную точку проводят прямую, параллельную оси OX до пересечения ее с кумулятой. Перпендикуляр, восставленный из этой точки на ось абсцисс указывает значение медианы. Проиллюстрируем это положение графиком, построенным по данным примера 4.3. (см. Рис.4.2).

Значение и смысл графического определения медианы аналогичны графическому определению моды.

Обобщая три средних величины, рассчитанные по одним и тем же данным, видим существующую разницу. Средний возраст условной группы рабочих (7) - 33,6 лет, наиболее распространенный, часто встречающийся средний возраст (Mo) - 26,2 лет, при этом половина рассматриваемой группы имеет средний возраст

(Me) - 31,5 лет. Какой величине следует отдать предпочтение? Какой показатель считать наиболее достоверным и точным?

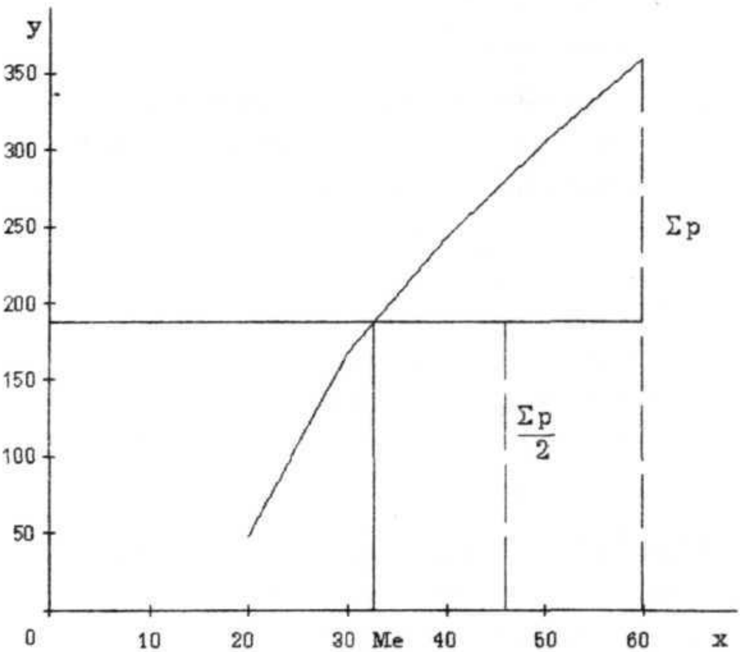

Рис.4.2. Кумулята распределения рабочих N-ского предприятия по возрасту.

<u>При решении этих вопросов надо помнить, что:</u>

1. Мода (Мо) имеет значение в том случае, когда ее величина расходится и с медианой (Ме), и со средней арифметической *(X),* им не следует пренебрегать. Это же можно сказать и о медиане. Так что для исследования полезно вычислять все три показателя.

2. Различие в значениях величин обусловлено асимметричным распределением. Средняя арифметическая подвержена влиянию каждой варианты, поэтому она смещается в направлении наибольших значений признака. На моду крайние (максимальные

и минимальные) варианты влияния не оказывают. Медиана зависит только от числа вариант, а не от их величины.

3. Медиана по своей математико-статистической природе является самой представительной средней. При больших колебаниях в значениях признаков или когда не определены крайние интервалы в группировках, лучше пользоваться медианой. При вычислении моды для интервальной группировки желательно, чтобы интервалы были равновеликими.

4. Мода чаще других величин применяется по отношению к качественным признакам. Если скопление частот возле моды составляет 10-15% их общего числа, особое значение приобретает медиана, представляя более достоверное значение среднего показателя.

* * *

Когда в разных совокупностях величина средней арифметической примерно одинакова - это еще не говорит об одинаковости или схожести самих совокупностей. За совпадением средних может скрываться разный размах вариаций признаков. Колеблемость показателей, т.е. разброс между максимальными и минимальными значениями признаков проверяется методом группировки и с помощью ***дисперсии*** (от латинского dispersio - рассеяние).

Простейшим способом изучения вариации признака в совокупности является размах вариации или ее амплитуда (R) Величина R определяется как разность между максимальным и минимальным значениями признака в изучаемой совокупности.

$$R = X_{max} - X_{min}$$

Так, по данным примера 4.3 размах вариации признака "возраст" равен:

$$R = 65 - 17 = 48(\text{лет}).$$

Амплитуда вариации определяет лишь наибольшее различие в значениях признака и обусловлена только двумя "крайними" величинами. Она не учитывает особенности распределения в целом, не учитывает все различия каждого значения признака. На практике чаще всего прибегают к изучению ***среднего квадратического отклонения*** (стандартного отклонения) конкретных значений признака от его средней величины. Оно обозначается σ (сигма) или S и позволяет определить границы, в которых изменяются конкретные значения признака. Величина, насколько в среднем каждое значение признака отличается от его средней арифметической, находится по формуле:

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2}{n}}, \qquad \text{где}$$

x- конкретные значения признака;

$\bar{x}$- средняя арифметическая ;

n- число наблюдений.

Стандартное квадратическое отклонение, возведенное в квадрат, называется дисперсией.

В том случае, если мы имеем дело с группировкой, с интервальным рядом, формула видоизменяется:

$$\sigma = \sqrt{\frac{\sum (x-\bar{x})^2 p}{\sum p}}, \qquad \text{где}$$

х- конкретные значения признака (для интервальной группировки - срединные значения признака);

$\bar{x}$- средняя арифметическая;

р- частота признака в группировке.

Рассмотрим вычисление среднего квадратического отклонения для данных примера 4.3.

$$\sigma = \sqrt{\frac{(18{,}5-34{,}56)^2\cdot 48+(25-34{,}56)^2\cdot 120+(35+34{,}56)^2\cdot 62+(57{,}5-34{,}56)^2\cdot 54}{48+120+75+62+54}} \approx 12{,}77$$

Полученное значение говорит о том, что в рассматриваемой совокупности рабочих N-ского предприятиях их возраст в среднем отклонялся на 12,77 лет от средней величины, равной 34,56 лет.

Достаточно просто вычисляется среднее квадратическое отклонение для определения размаха вариации качественных (альтернативных) признаков. Формула выглядит так:

$$\sigma = \sqrt{\frac{p_1 p_2}{n}},$$ где

P_1 - частота первой варианты признака;

P_2 - частота второй варианты признака;

n - число наблюдений.

Пример 4.4:

Даны сведения об успеваемости группы студентов в количестве 24 человек. После очередной экзаменационной сессии 6 человек имеют задолженности по тем или иным учебным дисциплинам, а 18 человек сдали успешно все экзамены. Среднее квадратическое отклонение в этом случае равно:

$$\sigma = \sqrt{\frac{6 \cdot 18}{24}} \approx 2{,}12$$

Дисперсии при интерпретации выражаются в тех же единицах, что и сами признаки. Это приводит, к тому, что будучи выражены в разных единицах измерения, средние квадратические отклонения несравнимы. То есть, нельзя сравнивать количество детей с земельной площадью. В случае необходимости пользуются коэффициентом вариации (V), определяемым как отношение стандартного отклонения к средней арифметической.

$$V = \sigma / \bar{x}$$

Полученную величину можно выразить в процентах. Сопоставление коэффициентов вариации нескольких признаков расширяет возможности исследователя при анализе и интерпретации распределений признаков - их равномерности, нормальности, колеблемости.

Показатели вариации раскрывают уровень репрезентативности (представительности) средней величины, степень ее точности, адекватности исторической реальности. При большом разбросе в значениях признака, при значительных показателях вариации средняя величина не является достаточно достоверной характеристикой изучаемой совокупности.

Какое место занимает дисперсия в исторических исследованиях? Во-первых, как мы уже отмечали выше, она является необходимым и обязательным дополнительным показателем при сравнении средних и сопоставлении различных группировок. Во-вторых, с ее помощью проверяется и обосновывается правомерность применения математических методов. Дисперсия служит своеобразным индикатором однородности изучаемой совокупности и нормальности ее распределения (см. лекцию 6). В-третьих, сравнение дисперсий различных признаков позволяет судить об их качественном значении в рассматриваемой системе. Дисперсии помогают не потерять сглаженное средними показателями своеобразие признаков изучаемого явления.

Средняя квадратпическая. Историки иной раз сталкиваются с проблемой определения средней земельной площади (площадь посевов; территория археологических раскопов; регион, охваченный восстанием и т.п.), но при этом известны не площади осредняемых участков, а только их линейные параметры. Путем качественного анализа проверяется возможность допущения, что каждый из рассматриваемых земельных участков имеет вид квадрата. Если проверка дала положительный результат, то для расчета средней площади имеющихся участков применяется формула средней квадратической:

$$\overline{X}_{кв} = \sqrt{\frac{X_1^2 + X_2^2 + X_3^2 + \ldots + X_n^2}{n}}$$

Пример 4.5:

Предположим, что имеются три участка земли. Протяженность одного -100 м, второго - 200 м, третьего - 300 м. Надо определить среднюю протяженность земельного участка. Величина средней арифметической = 200 м [(100+200+300)/3]. Ее реальность можно проверить, подсчитав площадь земельных участков, предположив, что они квадратной формы. Реальная площадь - $100^2+200^2+300^2$ = 140 000 $м^2$, а площадь трех участков со стороной 200 м - $3(200)^2$=120 000 $м^2$. Получилось, что мы "потеряли" в виду усреднения 20 000 $м^2$. Следовательно, средняя арифметическая нас не удовлетворяет.

Применим среднюю квадратическую:

$$\bar{X}_{кв} = \sqrt{\frac{100^2+200^2+300^2}{3}} \approx 216 м$$

Приняв 216 м за среднюю протяженность одного участка, площадь всех трех - 139 968 $м^2$. Как видим, "потери", вызванные усреднением значения признака, составили всего 32 $м^2$.

* * *

Средние показатели динамики. К средним показателям динамики относятся средний уровень ряда, средние абсолютные изменения и ускорения, средний темп роста. Все они выступают характеристиками тенденции.

Средний ***уровень*** $(\bar{y})$ интервального динамического ряда определяется как простая средняя арифметическая из уровней за равные промежутки времени или как средневзвешенная из уровней за неравные промежутки времени, длительность которых выступает в качестве "весов".

Пример ***4.6:***

Добыча нефти в СССР в 1976 - 1980 гг.

годы	1976	1977	1978	1979	1980
добыча нефти в млн.т	519,7	545,8	571,5	586,0	603,2

Как показывают данные таблицы, промежутки времени в примере 4.6. равные: по одному году. Значит мы должны применить здесь формулу простой средней арифметической для определения среднегодового уровня добычи нефти за 5 лет.

Средний уровень годовой добычи нефти за период 1976 - 1980 г.г. составил:

$$\overline{y} = \frac{519{,}7 + 545{,}8 + 571{,}5 + 586{,}0 + 603{,}2}{5} \approx 565{,}2 млн.\ т$$

Пример 4.7.

Распределение занятости учебно-вспомогательного персонала в приемной комиссии.

числа в июле	1-14	15-21	22-27	28-31
кол-во работн.	20	16	19	24

В примере 4.7. временные промежутки разные - 14 дней, 7 дней, 6 дней, 4 дня. В данном случае надо использовать формулу средневзвешенной.

$$\overline{y} = \frac{20 \cdot 14 + 16 \cdot 7 + 19 \cdot 6 + 24 \cdot 4}{31} \approx 19{,}4$$

Т.о. среднесписочное число работников учебно-вспомогательного персонала приемной комиссии в июле составило 19,4 человека.

В моментном ряду средняя величина характеризует обобщенное значение признака между начальным и конечным моментом наблюдения. Следовательно, начальный и конечный уровни лишь наполовину относятся к изучаемому отрезку времени, а на половину к прошлому и будущему периодам. Это обстоятельство определило формулу ***средней хронологической.***

$$\overline{y}_{хр} = \frac{y_1 / 2 + y_2 + y_3 + \ldots + y_n / 2}{n - 1}$$

Пример 4.8:

На 1 января 1924 г. в Средне-Волжском районе было зарегестрировано 47 546 переселенцев-мужчин (это не значит, что все они прибыли сюда 1 января), на 1 января 1925 г.- 46 725 мужчин, на 1 января 1926 Г.-64368 мужчин. Определить среднегодовое количество прибывших в Средне-Волжский район мужчин в 1924-26 гг.

$$\overline{У} = \frac{47546/2 + 46725 + 64368/2}{3-1} \approx 51341$$

В среднем ежегодно в период 1924-1926 гг. в Средне-Волжский район прибывали 51 341 мужчина.

В случае, если промежутки между датами моментного ряда не равны, хронологическая средняя вычисляется по формуле:

$$\overline{У}_{хр} = \frac{У_1 T_1 + У_2(T_1 + T_2) + У_3(T_2 + T_3) + \ldots У_n T_{n-1}}{2(T_1 + T_2 + T_3 + \ldots + T_{n-1})},$$ где

T - промежутки между датами;

$У_1, У_2, У_3 \ldots У_n$ - уровни ряда;

n - количество уровней.

Часть математиков считают проблему вычисления среднего уровня моментного ряда при неравных временных промежутках спорной. Однако в исторических исследованиях использование этой формулы возможно при тщательном контроле исходных данных и результатов вычисления качественным анализом.

Один из наиболее важных средних показателей динамического ряда - ***средний темп изменения*** (роста или сокращения). Он определяется по формуле средней геометрической. В литературе до сих пор нет универсального символа для обозначения этого показателя. Так, у Т.И.Славко используется G, у И.Д.Ковальченко - $K_р$, у И..И.Елисеевой - $\overline{k}$, у А.Я.Боярского - $\overline{T}$.

$$\overline{T} = \sqrt[n-1]{\frac{У_n}{У_1}},$$ где

$У_1$ - начальный уровень динамического ряда;

$У_п$ - последний уровень диинамического ряда;

n - число временных промежутков.

Вычислим среднюю геометрическую по данным предыдущего примера 4.8.

$$\overline{T} = \sqrt{\frac{64368}{47546}} \approx 1,16$$

В среднем ежегодно в период 1924-1926 гг. число иммигрантов-мужчин в Среднем Поволжье увеличивалось в 1,16 раз.

С помощью средней геометрической величины мы получили среднюю скорость изменения признака. В нашем распоряжении есть еще один средний показатель - **средний абсолютный прироста** (абсолютное значение). Он показывает на какую величину (в единицах измерения уровней ряда) показатель одного временного периода больше или меньше любого предшествующего. При возрастании уровней абсолютное изменение принимает положительное, а при уменьшении - отрицательное значения.

Он определяется по формуле:

$$\Delta\overline{У} = \frac{У_n - У_1}{n}, \quad \text{где}$$

$У_1$ - начальный уровень динамического ряда;

$У_n$ - конечный уровень динамичческого ряда;

n - число временных промежутков (число осредняемых отрезков времени).

Определение "начального" и "конечного" уровней динамического ряда в каждом вычислении зависит от задач исследования. По одной группировке можно определить несколько средних значений абсолютного прироста за разные временные промежутки.

Подсчитаем средний абсолютный прирост по данным примера 4.8.

$У_1$=47546; $У_п$= 64368. Чему равно n?

В нашем распоряжении сведения за 1924, 1925 и 1926 годы, а это значит, что мы имеем два осредняемых временных промежутка - с 1924 по 1925 гг. и с 1925 по 1926 гг. Следовательно, n=2, а не 3, как могло показаться вначале.

$$\overline{\Delta У} = \frac{64368 - 47546}{2} = 8411$$

Как видим, в 1924-1926 гг. в Среднее Поволжье ежегодно прибывали в среднем по 8411 мужчин. Интерпретация $\overline{\Delta У}$ и $\overline{T}$

сопровождается обязательным указанием двух хронологических единиц - периода, который характеризуется (в нашем случае - 1924-1926 гг.) и периода, на который рассчитан средний показатель (в нашем случае - выяснялся ежегодный показатель).

По среднему темпу роста можно без труда определить ***средний темп прироста,*** вычитая из значения $\overline{T}$ единицу. В нашем примере $\overline{T}$ = 1,16. Тогда средний темп прироста: 1,16-1 = 0,16. Полученное значение можно выразить в процентах, умножив его на 100% (у нас 0,16 100% = 16%).

Разделив абсолютный прирост на темп прироста (за соответствующий период) получим среднее абсолютное значение прироста ($\overline{\alpha}$):

$$\overline{\alpha} = \overline{\Delta y} / \overline{T}_{пр}$$

Встречаются ситуации, когда темпы роста и прироста, а также абсолютные приросты по годам снижаются, в то же время абсолютные значения 1% прироста возрастают. Может быть и обратный процесс.

Приведенные показатели служат основными характеристиками, применяемыми для анализа динамических рядов. Они позволяют судить об абсолютном и относительном изменениях уровней ряда. В заключение необходимо сделать несколько замечаний.

1. Все перечисленные показатели обладают высокой точностью и достоверностью при небольших колебаниях в значениях признака.

2. Средние хронологические особенно полезно вычислять при сравнительном анализе двух и более динамических рядов.

ПО ТЕОРИИ МЕТОДА ДОПОЛНИТЕЛЬНО ЧИТАЙТЕ:

1. Джини К. Средние величины. - М., 1970.

2. Елисеева И.И., Юзбашев М.М. Общая теория статистики. - М., 1995. - С.66-103, 257-306.

3. Измайлова М.О., Рахманкулов И.Ш. Категория "средняя величина" и ее методологическое значение в научном исследовании. - Казань, 1982.

4. Славко Т.И. Математико-статистические методы в исторических исследованиях. - М., 1981. - С.47-57.

5. Пасхавер И.С. Средние величины в статистике. - М., 1979.

6. Общая теория статистики. - М., 1984. - С.54-78, 94-104, 195-201.

Лекция 5.
МЕТОДЫ НЕСПЛОШНОГО НАБЛЮДЕНИЯ

Сегодня вряд ли кому-нибудь придет в голову испытывать на длительность горения все выпускаемые в стране электрические лампочки или на прочность - каждый изготавливаемый полиэтиленовый пакет. Ясно, что подобное испытание приведет к порче изделия, однако, покупая полиэтиленовую сумку-пакет, мы обнаруживаем на ней надпись, что она рассчитана на грузоподъемность "не более ... кг". Эти сведения базируются на проверке качества отдельных предметов, но выводы распространяются на всю партию, серию, вид изделия.

Если же обратиться к мировой истории, то использование результатов частичного обследования для решения ряда хозяйственных или социальных задач практиковалось еще в Древнем Египте и античной Греции при определении размеров налогообложения разных слоев общества, при решении некоторых военных вопросов и т.д.

Издревле еще до окончания уборочной страды крестьянин знал с достаточно высокой степенью точности объем и качество будущего урожая. Необходимую информацию давали выборочные обмолоты зерновых культур, проводимые как в крестьянских хозяйствах, так и в феодальных вотчинах. В России подобная практика известна с XVII в. Возможность судить о свойствах всей совокупности объектов на основе изучения ее некоторой части привлекала внимание как практиков, так и теоретиков.

Накопленный в статистике опыт несплошного наблюдения позволил расширить сферу его применения до решения проблем исторической науки. Это имеет очень большое значение, т.к. в своей работе историк сталкивается подчас с непреодолимыми сложностями недостатка или переизбытка данных, что приводит к иллюстративному характеру использования ценнейших источников.

В историческую науку из статистики пришло несколько методов несплошного наблюдения, различия между которыми, в основном, заключаются в системе отбора единиц для наблюдения из генеральной совокупности. (Генеральной совокупностью называется совокупность объектов, из которых производится отбор некоторой их части для обследования.).

1. *Монографический метод* предполагает всестороннее изучение и описание единичных объектов. Метод широко применяется в истории, однако он требует особой осторожности при использовании массовых исторических источников. Монографический метод может быть успешно использован только в том случае, когда, по замечанию А.Чупрова, у исследователя будет уверенность, что "единичные объекты, избранные для изучения, не выделяются какими-либо резкими отличиями из ряда других сходных предметов"(Цит. по Славко Т.И. Указ.соч.- С.60).

Это значит, что выводы, полученные на основании применения данного метода должны базироваться на заранее выявленных тенденциях и закономерностях. Монографический метод целесообразен для иллюстрации в качестве дополнительного приема.

2. *Метод основного массива* предполагает изучение той части единиц наблюдения, которая имеет по отношению ко всей совокупности в целом высокий удельный вес. Например, изучая развитие той или иной отрасли промышленности методом основного массива, мы должны будем сосредоточить все внимание на наиболее крупных предприятиях. На рубеже 20-30-х гг. XX в. на основе изучения нескольких "гигантов первой пятилетки" выводы о темпах индустриального развития СССР, о приросте промышленного производства, об уровне производительности труда были распространены на всю промышленность. Это дало основание пересмотреть показатели первого пятилетнего плана и в конечном итоге сорвать его результаты. Насколько правомерно распространение выводов, полученных на основании

применения метода основного массива на все единицы совокупности? Где его границы? Этот вопрос решается только профессиональной квалификацией исследователя.

Зачастую важные тенденции не проявляются в равной степени у всех единиц наблюдения; только зарождающиеся, набирающие силу факторы, которые возможно сыграют решающую роль в будущем, затушевываются в генеральной совокупности. Для их яркого показа используют метод основного массива. Он может быть рекомендован в дополнение к другим приемам частичного обследования для демонстрации наиболее важных, предварительно выявленных тенденций, наиболее передовых направлений в развитии общества.

3. ***Выборочный метод.*** "Под выборочным методом подразумевается такая система отбора единиц для наблюдения, при которой результаты, полученные на частичном объеме, отражают всю изучаемую совокупность, т.е. являются для нее репрезентативными" (Славко Т.И. Указ.соч. - С. 61).

Осуществление выборочного исследования складывается из трех основных этапов:

1. Определение объема выборочной совокупности.
2. Выбор способа отбора единиц для наблюдения.
3. Нахождение величины выборочной ошибки.

1. ***Определение объема* выборочной совокупности.** Поскольку мы имеем дело с частичными данными, надо помнить, что каждая выборка, каждый показатель, каждая варианта признака несет в себе некую погрешность, связанную с неполнотой единиц наблюдения. Следовательно, во многом точность и достоверность результатов исследования зависят от объема выборки.

Статистикой обнаружено, что при увеличении объема выборки до определенного уровня величина ошибки уменьшается быстро, а затем все медленнее. Разработаны специальные методики определения минимального объема выборки при сохранении достаточной репрезентативности показателей. Методы теории выборки предписывают определять необходимый объем относительно каждого количественного и качественного признака (на практике же это требование соблюдается редко).

Выборка должна быть такой, чтобы свойство репрезентативности было присуще каждому изучаемому признаку, поэтому численность выборки надо насчитывать многократно, исходя из допустимых ошибок различных показателей. Допустим, при анализе единоличных крестьянских хозяйств мы выделили четыре важнейших признака - размер землепользования, количество голов рабочего скота, найм рабочей силы и размер дохода на хозяйство. Вариации признаков различны, неодинакова и допустимая погрешность. Произведя необходимые вычисления, мы получили разные объемы выборки: 1150; 497; 720; 300. Исследование надо базировать на максимальной величине объема выборочной совокупности или создавать многоступенчатую выборку и специальную программу ее анализа (подробнее см. ниже).

Таким образом, выборочное изучение начинается с определения уровня точности будущих результатов. Он задается либо с помощью математических формул на основе предварительного изучения данных, либо по таблице достаточно больших чисел (см. Приложение 1). Второй способ относится к приближенным, но наиболее употребимым в исторических исследованиях методам определения выборочного объема. Таблица достаточно больших чисел (см. Приложение 1) рассчитана на определение выборочного объема для признаков, имеющих нормальное распределение (см. лекцию 6) или близкое к нему. Для остальных признаков точность снижается.

Исследователь задает желаемый уровень вероятности (Р) и возможную допустимую ошибку ($m_{доп}$) будущих результатов работы. В соответствии с этим на стыке строки и столбца находим искомую величину, которая означает объем выборки. Допустимая ошибка в историческом исследовании обычно не должна превышать 5% (т.е. до 0,05), а вероятность в пределах 0,95 - 0,99 обеспечивает высокую точность работы.

Предположим, мы задали P = 0,98 при $m_{доп}$ = 0,03, тогда оптимальный объем выборки определяется в 1503 единицы наблюдения (см. Приложение 1). Это значит, что .во-первых, мы должны изучить 1503 документа из генеральной совокупности; во-вторых, наши результаты в 98 случаях из 100 будут иметь ошибку, меньше 0,03 и только в двух случаях из каждых 100 ошибка может превысить этот уровень.

Объем выборки во многом зависит от цели работы. Для выявления общих тенденций изменения показателей достаточно иметь небольшую выборку. Для решения задач, связанных с необходимостью определения конкретных значений признаков объем выборки будет больше.

После нахождения объема выборочной совокупности, заданного уровня точности и вероятности переходят ко второму, не менее важному этапу работы - отбору единиц для наблюдения.

2. Выбор *способа отбора единиц для наблюдения.* Репрезентативность выборки обеспечивается объективностью отбора данных. Различают три способа - случайный отбор, выбор по определенной схеме и комбинация первого и второго способов. В зависимости от этого находятся виды выборки - собственно случайная, механическая, типическая и серийная (гнездовая).

При собственно случайном отборе в задачу исследователя входит обеспечение равных шансов для всех единиц генеральной совокупности попасть в выборку. Это можно сделать с помощью таблицы случайных чисел, в математике их разработано несколько (см. Приложение 2), путем жеребьевки. Например, нам нужна выборка в 320 единиц из 7000 (объем генеральной совокупности). Для этого мы должны пронумеровать все имеющиеся документы, а затем: - либо обратиться к таблице случайных чисел (см. Приложение 2), по которой с любого места, в любом порядке (по строкам или по столбцам) отбираем 320 случайных чисел, которые являются порядковыми номерами документов генеральной совокупности, составляющими выборку. Если встречается число превышающее своим значением величину генеральной совокупности (в данном примере - 7000 единиц) - оно пропускается. Так, начав отбор с третьего столбца пятой строки, двигаясь по вертикали, мы получаем следующие номера: 3371, 5323, 1796, 2105 и т.д. (см. Приложение 2). - либо проводим жеребьевку, составив колоду карточек с номерами от 1 до 7000, пронумеровав какие-нибудь шарики, палочки, ._ (любые однородные предметы). Тщательно тасуем эти номера и вытаскиваем, не глядя, 320 предметов, номера которых указывают номера документов, попадающих в выборку.

Механический отбор заключается в том, что генеральная совокупность делится на равные части, в зависимости от необходимого объема выборки, а затем из каждой части берется одна единица наблюдения (можно эти документы отбирать случайно, можно по определенному порядку, - каждый второй, пятый, одиннадцатый..). Например, получить механическую выборку в том же объеме в 320 единиц из той же генеральной совокупности в 7000 документов можно, поделив 7000 на 320, а потом из каждой подгруппы (их у нас будет 320) выберем по седьмому документу.

Механический отбор нежелателен, если элементы генеральной совокупности частично или полностью упорядочены (например, документы сложены в порядке возрастания значений признаков).

Типическая выборка формируется из генеральной совокупности, предварительно разделенной на качественно однородные группы, внутри которых производится случайный или механический отбор.

Типическую выборку еще называют районированной или стратифицированной. Пусть наше исследование единоличных крестьянских хозяйств охватывает большую территорию, различающуюся природно-климатическими условиями, или формами хозяйственной деятельности, или ... В этом случае для повышения точности результатов изучения мы предварительно выделяем типические группы, образованные по какому-либо качественному признаку: по районам, по формам хозяйственной деятельности, по национальности, по социальным категориям и т.д. в зависимости от задач работы. После чего, внутри каждой однородной группы проводим выборку. При этом возможен как пропорциональный отбор в соответствии с численностью единиц наблюдения в группе, так и непропорциональный. Для исторических исследований предпочтительней первый, т.к. повышается точность выводов и наблюдений, сделанных на его основе.

При серийной (гнездовой) выборке случайным образом определяются пункты (гнезда), внутри которых проводится сплошное наблюдение. Например, обследованию подвергаются не единичные крестьянские хозяйства, а целые деревни, села.

В математической статистике выборки делят на повторные и бесповторные. Допустим, вы пронумеровали все элементы генеральной совокупности, нанесли эти номера на карточки (шары, палочки или др. предметы жеребьевки) и начали отбор. Жребий, отобранный в выборку, может быть отложен в сторону (бесповторная выборка), а может быть возвращен в общую массу и иметь шанс вновь быть избранным (повторный отбор). В исторических и социально-экономических исследованиях не имеет смысла проводить повторную выборку и если специального указания в выборке нет - предполагается бесповторная выборка.

Объекты исторического исследования, как правило, имеют сложную структуру и разный разброс значений признаков, т.е. признаки имеют разную изменчивость. Например, признак "пол" имеет всего два варианта - мужской и женский, а признак "размер посева" - множество значений.

Организовать выборочное обследование бывает очень сложно. Исследователи обращаются к многоступенчатой (комбинированной) и многофазовой выборке. Сочетание разных способов и разных единиц отбора на разных этапах исследования создает многоступенчатую выборку. Например, типическим путем можно определить губернии, механическим - уезды, случайным - волости, далее провести отбор сел и дворов. Получится пятиступенчатая выборка.

Многофазовая выборка также предполагает несколько этапов исследования, отличающихся подробностью программы изучения. Для признаков, имеющих меньшую изменчивость можно сокращать объем выборки. Например, многофазовые выборки применялись земскими статистиками России в начале XX в. Так, пензенские статистики в 1913 г. провели сплошную перепись крестьянских хозяйств по сокращенной программе, каждое третье - по более полной краткой, каждое девятое - по полной хозяйственной, каждое двадцать седьмое - по полной специальной и 25 хозяйств каждого уезда подверглись детальному бюджетному описанию (см. Ленин В.И. Полн.собр.соч. - Т.24 - С.274-275).

В любом случае решение о способе отбора единиц для наблюдения, о виде выборки зависит от свойств объекта изучения, а следовательно предполагает обязательное предварительное знакомство с ним.

3. ***Нахождение величины выборочной ошибки.*** Нахождение величины выборочной ошибки связано с доказательством степени репрезентативности выборки, т.е. с выяснением насколько результаты, полученные на основе изучения выборочной совокупности, можно распространить на все единицы наблюдения. Ошибки выборки бывают случайными и систематическими. Систематические возникают при тенденциозном, неправильном отборе данных или при искаженных сведениях источника. Стал уже классическим пример ошибки сотрудников американского журнала "Literary digest" ("Литературное обозрение"), попытавшихся предсказать результаты президентских выборов в США в 1936 г., приняв телефонные справочники за генеральную совокупность. В результате опроса более, чем 2-х миллионов абонентов выходило, что на выборах победу одержит кандидат от республиканской партии. Прогноз журнала не оправдался - с весомым перевесом (рекордным для США) победил демократ Франклин Д. Рузвельт. Дело в том, что в условиях середины 30-х годов XX в. в США иметь квартирные телефоны могли себе позволить только состоятельные люди, а они, в большинстве своем, симпатизировали республиканцам, что и отразили итоги опроса. Таким образом, был нарушен принцип случайности отбора единиц для наблюдения, а значит выборка, несмотря на большой объем, не отражала политических настроений большинства американских избирателей.

Обнаружение и ликвидация систематических ошибок возможны только на основе прочных источниковедческих знаний, путем качественного анализа.

Случайные ошибки присутствуют в любом выборочном обследовании, даже когда соблюдены все правила выборочного метода. Они зависят от методов отбора единиц наблюдения (от вида выборки), от степени однородности генеральной совокупности, от изменчивости признаков, а также от используемых в дальнейшем методов обработки данных. Для каждого вида выборки в статистике разработаны свои способы вычисления случайных ошибок.

Оценить результаты выборочного исследования можно не только с помощью математических формул, но и путем привлечения дополнительной информации, сравнением с уже известными по другим источникам данными. Для этого надо иметь результаты изучения признаков относительно всей генеральной совокупности. Тогда путем сравнения средних величин и относительных показателей выборочной и генеральной совокупностей можно оценить репрезентативность выборки. Чем меньше разница в показателях, тем выше степень репрезентативности выборки. Считается, что выборочное обследование достаточно полно отражает исходную совокупность объектов, если разница в показателях не превышает 5%. Однако чаще всего исследователь лишен этой возможности и должен обращаться к сложным математическим вычислениям.

Сложнее дело обстоит тогда, когда историк встречается с частичными данными и не имеет представления об основных характеристиках генеральной совокупности, которую они представляют. Встает проблема использования неполных данных. Можно ли на основе их изучения делать обобщения и выходить на уровень тенденций и закономерностей, характерных для всего изучаемого явления, или надо ограничиться иллюстративным показом единичных объектов?

Если из 40 тысяч личных карточек рабочих какого-либо завода за период 1900 - 1941 гг. сохранились всего 280 штук, - можно ли их принять за выборку и для их изучения применить методы, разработанные для анализа выборочных совокупностей? Правомерно ли будет распространить полученные на их основе выводы на всю предполагаемую совокупность в 40 тысяч единиц?

В исторической литературе подобные неполные совокупности называются ***"естественными" выборками.*** Исследователь должен доказать, отражает ли стихийно образовавшаяся выборка некую генеральную совокупность, насколько частичные данные обладают свойствами массового источника.

Исходя из определения массового исторического источника, должны выполняться условия достаточности единиц наблюдения и их независимости, случайности имеющегося набора признаков. Первое условие достаточно большого объема в случае "естественных" выборок заменяется условием равномерности охвата частичными данными генеральной совокупности. Проверяется равномерность охвата частичными данными изучаемой территории и временного периода.

Взяв тот же условный пример, мы должны проверить охватывают ли сохранившиеся 280 документов все (или большинство) годы с 1900 по 1941 и имеются ли в них сведения относительно всех (или большинства) структурных подразделений завода.

Независимость признаков проверяется анализом происхождения и содержания совокупности документов. Надо доказать, что заполнение одного документа не влияло на заполнение другого, что текст одного документа не списывался с другого. Это достигается изучением истории формирования документов..

Случайность признаков определяется, во-первых, охватом каждым признаком всех (или большинства) возможных для генеральной совокупности значений и, во-вторых, случайностью различий в значениях признаков по отношению ко всем документам. Доказательство случайности признаков выступает одновременно и доказательством случайности выборки.

В математической статистике разработан ряд приемов, позволяющих проверить случайность выборки. Среди них метод серий, критерий знаков, метод последовательных разностей и др.

Суть метода серий состоит в проверке случайности расположения элементов в выборке. Рассмотрим его на примере признака возраст у группы рабочих в 16 человек: 20, 18, 25, 51, 33, 36, 28, 40, 42, 49, 32, 52, 29, 35, 34, 41. Первым делом вычислим средний возраст для данной группы рабочих по средней арифметической $(\bar{x})$. Он примерно равен 35,3 лет. Теперь сравниваем каждую варианту признака с его средним значением. Если оно меньше или равно средней величине - записываем "нуль", если больше - "единицу". Получаем такую последовательность: 0, 0, 0, 1, 0, 1, 0, 1, 1, 1, 0, 1, 0, 0, 0, 1. Сколько раз изменился показатель? Сколько раз происходила смена "0" на "1" и наоборот ? Подсчет показывает, что данная последовательность имеет 10 "серий". В том случае будет доказано случайное расположение элементов в выборке, если фактическое число "серий" будет находиться в теоретически допустимых пределах.

Теоретическое расчетное число "серий" (R_T) определяется в пределах:

$$\frac{n+1}{2} - \sqrt{n-1} \le R_T \le \frac{n+1}{2} + \sqrt{n-1}, \quad \text{где}$$

п - число единиц в совокупности.

В нашем случае п = 16.

Подставим это значение в формулу:

$$\frac{16+1}{2} - \sqrt{16-1} \le R_T \le \frac{16+1}{2} + \sqrt{16-1},$$

$$4{,}63 \le R_T \le 12{,}37$$

Фактическое число "серий" входит в теоретически определенный предел, а это значит, что можно считать доказанным случайность расположения значений признака в совокупности.

Достаточно активно историки с той же целью применяют метод критерия знаков. Он состоит в последовательном сравнении величины признака единицы совокупности с величиной этого же признака предыдущей единицы совокупности. Если разница между вариантами положительная - то это фиксируется знаком " + ", а если отрицательная - то " - ".

Проиллюстрируем применение метода критерия знаков на тех же данных о возрасте 16 рабочих, расположенных в той же последовательности, что и в источниковом комплексе. Результаты вычитания каждой последующей варианты из предыдущей дают следующий ряд плюсов и минусов:

20	18	25	51	33	36	28	40	42	49	32	52	29	35	34	41
	–	+	+	–	+	–	+	+	+	–	+	–	+	–	+

Далее подсчитаем число плюсов и минусов. При случайной выборке оно не должно отличаться друг от друга на значительную величину. Повысить точность нашего вывода о характере данных поможет обращение к специальным таблицам, определяющим критические границы и наименьшую погрешность в соответствии с объемом выборки. Если число плюсов вписывается в установленный по таблице интервал, то можно считать, что выборка достаточно полно отражает исходную совокупность по рассмотренному признаку.

Метод критерия знаков успешно применялся И.Д.Ковальченко при анализе состояния крестьянского хозяйства в первой половине XIX в. по сведениям подворных описаний (см. Ковальченко И.Д. Об опыте математико-статистической обработки выборочных данных о крестьянском хозяйстве в России XIX в. // Вестник МГУ - 1966 - N 11), Б.Н.Мироновым, изучавшим динамику хлебных цен в России в XVIII в. (см. Миронов Б.Н. Применение выборочного метода при анализе движения хлебных цен XVIII в. // Ежегодник по аграрной истории Восточной Европы. 1964. - Кишинев, 1966).

Кроме того, исследователь не должен упускать из вида качественный, традиционно-источниковедческий анализ при проверке стихийно образовавшейся выборки. Надо помнить, что если сбор и хранение документов, доставшихся историку, были преднамеренными, пристрастными, то данная совокупность не может считаться репрезентативной. Следовательно, получение надежных результатов на основе ее изучения весьма проблематично.

Только тогда уцелевшие документы можно принимать за стихийную выборку, когда у людей, от которых зависели их сбор и хранение, не было заинтересованности в том, чтобы какая-то часть сохранилась или погибла; и когда для наших предков критерий отбора документов для хранения не зависел от признака, исследуемого историком. Например, если в архиве отложились подворные описи только за урожайные годы, то они не могут признаваться за "естественную" выборку даже при соблюдении независимости и случайности признаков. Если же подворные описи сохранились из-за стойкости чернил - эту совокупность можно принять за "естественную" выборку и проверять ее на соответствие свойствам массового исторического источника.

* * *

В заключение можно обобщить, что любая выборочная совокупность будет репрезентативной при выполнении четырех основных требований:

1. Каждый элемент генеральной совокупности должен иметь равную возможность попасть в выборку.

2. Отбор документов должен быть произведен случайным образом, вне зависимости от цели и задач исследования, максимально объективно.

3. Проводить его требуется из однородных, по возможности, групп.

4. Выборочная совокупность должна иметь достаточно большой объем. В связи с этим встает проблема малых выборок, для которых математические теории, разработанные для выборочного метода, неприменимы. Объем малых выборок разными авторами определяется по-разному, но большинство сходится, что это совокупности менее 100 единиц наблюдения.

На сегодняшний день, несмотря на удачные и достаточно многочисленные примеры использования выборочного метода в исторических исследованиях, многие вопросы общей теории выборок нуждаются в дальнейшей методологической разработке.

ПО ТЕОРИИ МЕТОДА ДОПОЛНИТЕЛЬНО ЧИТАЙТЕ:

1. Дружинин Н.К. Выборочный метод и его применение в социально-экономических исследованиях. - М., 1970.

2. Елисеева И.И., Юзбашев М.М. Общая теория статистики. - М., 1995. - С.132-189.

3. Йейтс Ф. Выборочный метод в переписях и обследованиях. - М., 1965.

4. Количественные методы в исторических исследованиях. - М., 1984. - С.101-136.

5. Крылов В.Н. Выборочный метод в статистике. - М., 1957.

6. Миронов Б.Н., Степанов З.В. Историк и математика. - Л., 1975. - С. 34-52.

7. Общая теория статистики. - М., 1985. - С.263-299.

8. Рабочая книга социолога. - М., 1983. - С.200-235, 300-306.

9. Славко Т.И. Математико-статистические методы в исторических исследованиях. - М., 1981. - С.59-86.

Лекция 6.
КОРРЕЛЯЦИОННЫЙ АНАЛИЗ.

Изучая историю, нетрудно заметить, что существует взаимосвязь явлений и процессов, происходящих в природе и обществе, внутри общества, во времени и пространстве. Оценка исторического факта предполагает выявление факторов способствовавших и препятствовавших его появлению, а их оценка в историческом исследовании чаще всего бывает расплывчатой. Читаем - "сильное влияние...", "решающее значение..","определенное воздействие-." и т.п. Внести сюда количественную определенность помогает корреляционная связь, направленный на определение тесноты взаимосвязи признаков и степени воздействия различных факторов на изучаемый объект. Констатировать наличие связи между признаками позволяют аналитические группировки, но они не дают возможность количественно выразить силу взаимодействия одного признака с другим (парная корреляция) или же с совокупностью признаков (множественная корреляция).

Все связи, которые могут могут быть измерены, можно считать статистическими, частным случаем которых являются функциональные (жестко детерминированные). Они возможны лишь при условии, что на один из двух рассматриваемых признаков влияет только второй признак этой же пары и ничто больше. В реальной природе, а тем более в общественной жизни таких связей нет. На каждый исторический факт одновременно воздействует множество причин.

Другим частным случаем статистической связи выступает корреляционная связь, состоящая в том, что с изменением значений одного признака меняется среднее значение другого. В каждом отдельном случае изучаемый признак может принимать множество различных значений, в которых теряется закономерность. Так, например, если заводской цех оснастили новым прогрессивным оборудованием, - это не значит, что у каждого рабочего возрастет производительность труда, однако ее средний уровень обязательно изменится.

Термин корреляция употребляется в науке с конца XYIII века. Его ввел французский палеонтолог Жорж Кювье, основавший "закон корреляции", согласно которому череп с рогами обязательно принадлежал травоядному животному, обладавшему копытными конечностями; если же лапа имела когти, то животное было хищным, без рогов, но с крупными клыками. Об этом "законе" сохранился рассказ о неудачной шутке студентов, пытавшихся во время университетского карнавала напугать Кювье. Ряженный в шкуре и маске с рогами крикнул профессору: "Я тебя съем!" На что получил спокойный ответ, что рогатых хищников не бывает, а за незнание закона корреляции можно получить плохую оценку.

Корреляционная связь между признаками может возникать разными путями. Причинная зависимость предполагает, что один из пары рассматриваемых признаков выступает как фактор, второй - как результат. Например, качество почвы может рассматриваться фактором урожайности сельскохозяйственных культур.

Существует корреляционная связь и между двумя следствиями одной причины. Пример такой связи приводил крупнейший российский статистик начала XX в. А.А.Чупров. Рассматривались два признака - количество пожарных команд в городе и размер ущерба, причиненного городу от пожаров. Выходило, что, чем больше в городе пожарных, тем больше убытков от пожаров. Встал вопрос - не сократить ли пожарные команды?

В данном случае мы имеем дело не с причиной и следствием, а с двумя следствиями общей причины - размером города. Логично, что в крупных городах больше штат пожарных, т.к. чаще возникают пожары и ущерб огнем причиняется значительный.

Сложнее дело обстоит тогда, когда каждый из признаков является одновременно и причиной, и следствием. Здесь мы сталкиваемся со взаимосвязью, взаимозависимостью между признаками. Например, размер оплаты труда зависит от его производительности, но, в то же время, выступает в качестве стимула, а значит, фактора повышения уровня производительности труда.

Как и любая классификация, это деление носит достаточно условный характер, однако сделанные здесь замечания необходимо учитывать при интерпретации результатов корреляционного анализа.

Прежде, чем приступать непосредственно к корреляционному анализу, надо проверить правомерность его применения, надо проверить, будут ли его результаты реально отражать историческую картину. Признаки, исследуемые методом корреляции, должны быть нормально распределены и линейно зависимы между собой.

Признак обладает *свойством нормальности,* если его значения симметрично распределяются от "центра", которым считается его средняя арифметическая величина. Проще всего проверить нормальность распределения графическим методом. График нормально распределенного признака имеет колоколообразный вид с центром, совпадающим со значением средней арифметической (см. Рис. 6.1).

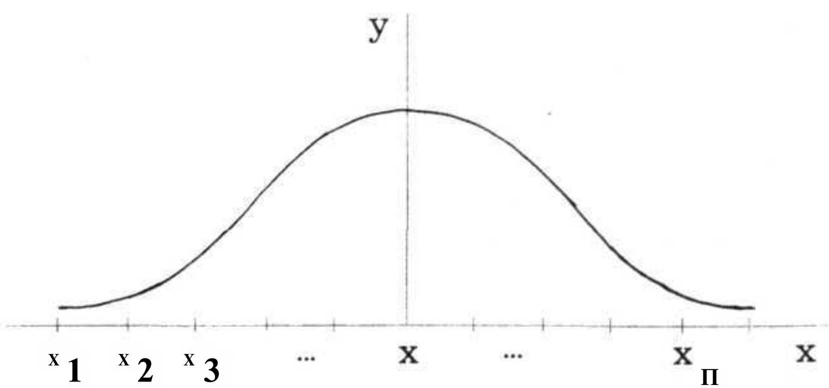

Рис.6.1. Графическое изображение нормального распределения.

В истории среди признаков, характеризующих развитие общества, нет строгой нормальности распределения. Практика использования математических методов в общественных науках доказала целесообразность относить к нормальным распределения с незначительно нарушенной симметрией, с перекосами в ту или иную сторону, с центром, совпадающим не со значением средней арифметической, а перенесенным в максимальное значение признака. К нормальным можно причислять и графики V-образной формы и "опрокинутые колоколы".

В статистике разработано много методов анализа распределений значений признаков, но, во-первых, они связаны со значительным объемом вычислений; во-вторых, в исторической науке нет пока необходимости получения точных данных об особенностях распределения тех или иных признаков. Это значит, что для проверки правомерности использования ряда математических методов, в частности корреляционного анализа, в исторических работах достаточно приближенного установления свойства нормальности, каким выступает график.

Свойство линейности в изучении взаимосвязи признаков также служит необходимым предварительным условием использования многих математических методов. Линейная зависимость между двумя признаками характеризуется условием, при котором с увлечением на единицу значений одного признака изменяются в ту или иную сторону значения второго.

Проверка формы зависимости проводится с помощью графического метода. В системе координат двух признаков точками отмечаются имеющиеся данные. Если пространство точек имеет вид прямой линии, то можно эту зависимость характеризовать как линейную, независимо от направления точечного скопления. (см.рис.6.2, 6.3, 6.4).

Примеры линейной зависимости между признаками.

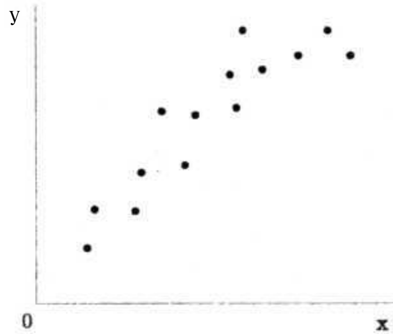

Рис. 6.2.

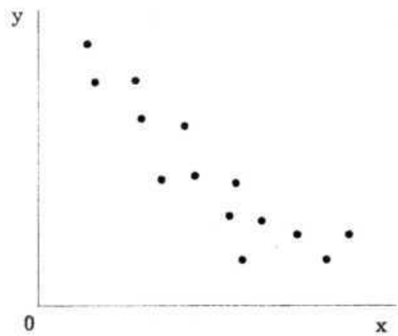

Рис. 6.3

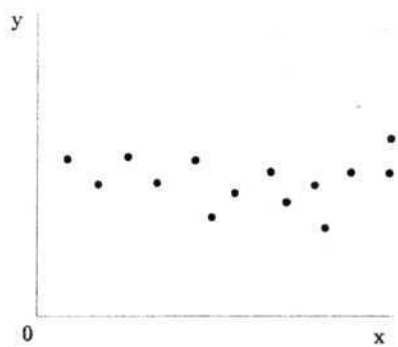

Рис.6.4

Так же, как и нормальности, строгой линейности в истории не существует. Достаточно приближенного выполнения данного свойства без привлечения более сложных специальных методик.

В заключение необходимо сделать несколько замечаний:

1. Проверка нормальности и линейности должна обязательно проводиться перед применением математических методов. От этого зависит степень исторической достоверности результатов математических вычислений.

2. Свойства нормальности и линейности выясняются по несгруппированным данным.

3. Нормальность и линейность определяются относительно каждого признака изучаемого явления.

4. Если признаки не отвечают свойствам нормальности и линейности - это еще не означает отказа от применения математико-статистических методов. Разработан ряд приемов, преобразующих значения признаков, существенно отклоняющихся от указанных свойств.

* * *

От характера исходных данных, от особенностей источника и задач исследования зависит выбор формулы коэффициента корреляции.

Чаще всего при изучении массовых источников применяют ***коэффициент линейной корреляции*** (r). Он вычисляется по формуле:

$$r = \frac{\sum x_i y_i - n\overline{xy}}{\sqrt{\sum x_i^2 - n\overline{x}^2} \cdot \sqrt{\sum y_i^2 - n\overline{y}^2}}, \quad \text{где}$$

X_i и y_i - значения рассматриваемых признаков;

$\overline{x}$ и $\overline{y}$ - средние арифметические величины признаков;

п - общее число наблюдений.

Пример 6.1.

Применение коэффициента линейной корреляции (r) рассмотрим по данным о возрасте и количестве детей двадцати пяти учителей. Необходимо определить тесноту связи между возрастом (х) и количеством детей (у) в выделенной группе учителей.

Сразу заметим, что возраст выступает в этом распределении как факторный признак, а количество детей - как результативный.

№ пп	возраст	кол-во дет	x_1y_1	x_1^2	y_1^2
1	28	2	56	784	4
2	25	1	25	625	1
3	28	1	28	784	1
4	40	2	80	1600	2
5	25	1	25	625	1
6	26	1	26	676	1
6	26	1	26	676	1
7	25	1	25	625	1
8	30	2	60	900	4
9	24	1	24	576	1
10	22	0	0	484	0
11	34	1	34	1156	1
12	27	1	27	729	1
13	25	1	25	625	1
14	27	1	27	729	1
15	24	0	0	576	0
16	23	0	0	529	0
17	26	0	0	676	0
18	25	0	0	625	0
19	25	0	0	625	0
20	38	3	114	1444	9
21	27	1	27	729	1
22	33	2	66	1089	4
23	25	0	0	625	0
24	26	0	0	676	0
25	26	1	26	676	1
$\sum$	684	23	695	19188	37

$$n=25; \qquad \bar{x} = \frac{\sum x_1}{n} = \frac{684}{25} \approx 27,4$$

$$\bar{y} = \frac{\sum y_1}{n} = \frac{23}{25} \approx 0,92$$

Подставляем полученные значения в формулу для нахождения r:

$$r = \frac{695 - 25 \cdot 27,4 \cdot 0,92}{\sqrt{19188 - 25(27,4)^2} \cdot \sqrt{37 - 25(0,92)^2}} \approx 0,821$$

Все коэффициенты корреляции изменяются в пределах от О до ±1. Чем ближе значение коэффициента к 0, тем меньше, слабее связь между признаками и чем ближе величина коэффициента к ±1, тем сильнее, значительнее, весомее связь между признаками. Если коэффициент корреляции принимает положительные значения - связь между признаками прямая, т.е. с увеличением значения одного признака - растет среднее значение второго. Если коэффициент корреляции имеет значение меньше О (т.е. отрицательное) - связь обратная.

При r больше или равным ±0,5 можно констатировать наличие существенной связи между признаками. Оценка значимости r во многом зависит от объема исследуемой совокупности. Если число наблюдений велико, то даже небольшая величина коэффициента линейной корреляции имеет определенную значимость, которой не следует пренебрегать. Это проверяется специальными статистическими таблицами, раскрывающими зависимость величины r от объема изучаемой совокупности.

В нашем примере - связь между признаками очень тесная и прямая, т.е. количество детей в семье в значительной мере зависит от возраста родителей и чем старше опрашиваемый, тем больше у него детей.

Применение коэффициента линейной корреляции имеет ряд ограничений. Во-первых, он исчисляется только для количественных признаков. Во-вторых, признаки, связь между которыми выявляется, должны быть нормально распределены. В-третьих, связь, сила которой должна быть измерена, должна быть линейной. До вычисления коэффициента следует проверить имеющиеся данные на соответствие, предъявляемым условиям.

Напомним, что нормальность и линейность проверяются графическим методом.

Приведенная формула определения величины r применяется только для первичных, несгруппированных данных.

При анализе исторических событий исследователи работают преимущественно с качественными признаками, разновидностью которых выступают альтернативные (здесь: принимающие только два значения). Для изучения силы их связи применяются ***коэффициент ассоциации*** **(Q)** и ***коэффициент сопряженности*** (Ф) или коэффициент контингенции (K_k). Их вычисление предваряется тем, что имеющиеся данные сводятся в таблицу четырех полей:

	X_1	X_1
У1	a	b
У2	c	d

а затем ведется расчет по формулам:

$$Q = \frac{ad - cb}{ad + cb}; \qquad \Phi = \frac{ad - cb}{\sqrt{(a+b)(b+d)(a+c)(c+d)}}$$

Пример 6.2:

Дано количество грамотного и неграмотного населения в городах и сельской местности Среднего Поволжья в середине 20-х гг. XX в.

место жительства	кол-во нас. в тыс.	
	грамотн.	неграмотн.
город	**740,4**	431,3
сельск. местн.	2993,4	6104,1

Таким образом, в распоряжении исследователя имются два признака - грамотность и место жительства. В данном распределении они приняли альтернативный характер, хотя в случае необходимости могут дробиться и принимать больше значений.

Определить уровень связи между признаками.

Подставим имеющиеся данные в формулы:

$$Q = \frac{740{,}4 \cdot 6104{,}1 - 431{,}3 \cdot 2993{,}4}{740{,}4 \cdot 6104{,}1 + 431{,}3 \cdot 2993{,}4} = 0{,}56$$

$$\Phi = \frac{740{,}4 \cdot 6104{,}1 - 4311{,}3 \cdot 2993{,}4}{\sqrt{(740{,}4 + 2993{,}3)(740{,}4 + 431{,}3)(431{,}3 + 6104{,}1)(29932{,}4 + 6104{,}1)}} = 0{,}2$$

Интерпретация полученных значений коэффициентов аналогична толкованию значений коэффициента линейной корреляции. Однако надо сделать несколько замечаний.

1. Оценивать связь между признаками как тесную, существенную можно при значении Ф не ниже ±0,3. (Некоторые исследователи считают, что коэффициент Ф дает более-менее точную характеристику при значениях, превышающих ±0,5.) Значение Q всегда несколько больше значения Ф.

2. Коэффициент ассоциации (Q) отражает одностороннюю связь между признаками, т.е. показывает степень влияния только одного признака на другой. Коэффициент сопряженности (Ф) раскрывает силу взаимосвязи между признаками при их обоюдном влиянии друг на друга.

3. Величина коэффициента сопряженности в определенной мере зависит от абсолютных значений признаков в таблице распределения. В связи с этим надо быть особенно осторожным при сравнении значений Ф, рассчитанных по разным исходным данным. Изменение его величины может в большей степени объясняться разницей абсолютных частот признаков, чем разницей си-

лы их взаимодействия. По возможности следует по одним и тем же данным вычислять оба коэффициента, особенно при необходимости их сопоставления.

Значения коэффициентов, полученных по данным нашего примера (см. Пример 6.2) говорят о том, что в условиях НЭП для населения региона Среднего Поволжья выбор места жительства в очень малой степени зависел от такого показателя, как грамотность. Что же касается взаимодействия этих характеристик, то оно почти отсутствует. Возможно это связано с общей низкой грамотностью населения, его культурной отсталостью.

Когда достаточно получить ориентировочное представление о тесноте связи между признаками можно обойтись без громоздких вычислений, обратившись к *коэффициенту совпадения знаков.* Метод, предложенный немецким психиатром Г.Т.Фехнером (1801-1887 гг.), основан на сравнении значений признаков с их средними величинами. Если значение признака больше его средней - оно фиксируется знаком "+", если меньше - знаком "_". Затем ведут подсчет по формуле:

$$k_c = \frac{a - b}{n},$$ где

a - число совпадений знаков;

b - число несовпадений знаков;

n - общее число случаев.

Определить уровень влияния затрат труда на валовой доход в сельском хозяйстве.

В нашем примере (см. Пример 6.3) из 10 случаев знаки отклонения значений признака от их средних величин совпали в 9 случаях и лишь в одном - не совпали (см. в таблице хозяйство N3).

Пример 6.3:

Дано распределение факторов валового дохода в сельском хозяйстве.

Nn/n	валовый доход (руб./га)	затраты труда (ч/га)	Знаки отклонения от средней величины	
			по доходу	по труду
1	704	265	+	-
2	293	193	+	-
3	346	229	-	+
4	420	193	–	-
5	691	225	+	+
6	679	255	+	+
7	457	201	-	-
8	503	208	–	-
9	314	170	-	-
10	803	276	+	+
сред, ариф.	521	221,5		

Следовательно, а = 9; в = 1; п = 10. Подставим полученные значения в формулу:

$$\kappa_c = \frac{9-1}{10} = 0{,}8$$

Коэффициент Фехнера измеряется и интерпретируется так же, как и коэффициент линейной корреляции. Полученное значение 0,8 говорит о тесной прямой связи между рассматриваемыми признаками: чем больше затраты труда, тем выше средний валовой доход на хозяйство.

Не всякий качественный признак можно превратить в альтернативный или заменить количественным. Достичь цель выявления и измерения влияния качественных признаков помогает ***коэффициент ранговой корреляции.***

Система ранжирования получила широкое распространение в исторических исследованиях. Суть ее состоит в предварительной экспертной оценке вариантов качественного признака и присвоении им количественного эквивалента, исходя из степени их интенсивности. Ранжировать изучаемые признаки обязательно в одном и том же порядке: либо по восходящей, либо по нисходящей линии. Ранжированию подвергаются как количественные, так и качественные признаки. Коэффициент корреляции рангов может быть вычислен и для изучения взаимосвязи между качественным и количественным признаками. Ранги чаще всего обозначаются порядковыми числительными - 1, 2, 3 ... Меру взаимосвязи между парой признаков, каждый из которых ранжирует изучаемую совокупность объектов, показывает коэффициент ранговой корреляции.

Одну из формул коэффициентов корреляции рангов предложил английский психолог Ч.Спирмен (1863-1945 гг.).

$$\rho = 1 - \frac{6\sum d^2}{n(n^2-1)} \qquad \text{или} \qquad 1 - \frac{6\sum d^2}{n^3 - n}$$

d - разность между парами рангов;

n - число сопоставляемых пар рангов, общее число наблюдений.

Коэффициент ранговой корреляции Спирмена измеряется и интерпретируется так же, как и другие корреляционные коэффициенты. При совпадении ранжированных рядов по обоим рассматриваемым признакам коэффициент примет значение 1, что говорит о максимально тесной прямой связи. Если объекты в одном ранжированном ряду прямо противоположны рангам второго признака, то на лицо максимально тесная обратная связь. В обоих этих случаях вычисления коэффициента не требуется, достаточно проанализировать взаимное расположение рангов.

Пример 6.4:

Дано распределение семейного состояния населения по среднедушевому доходу.

Семейное состоял, (х)	среднедуш. доход (у)	ранги		d	d^2
		х	у		
холостые	217	1	3	-2	4
разведен.	205	3	2	1	1
вдовые	220	4	4	0	0
семейные	200	2	1	1	1

Определить тесноту связи между рассматриваемыми признаками.

В приведенной таблице графы 3 и 4 показывают ранги рассматриваемых признаков. Они составлены для качественного признака (х) в восходящем порядке, исходя из хронологической поэтапности смены семейного состояния. Второй признак количественный - среднедушевой доход в месяц - (у) проранжирован также в восходящем порядке по степени интенсивности проявления. Пятая графа представляет разницу между парами рангов, а шестая - квадраты значений разности пар рангов. Полученные величины подставляем в формулу:

$$\rho = 1 - \frac{6 \cdot 6}{4^3 - 4} = 1 - \frac{36}{64 - 4} = 0{,}4$$

Значение коэффициента ранговой корреляции Спирмена в нашем примере свидетельствует о наличии прямой связи между рассматриваемыми признаками, но связь эта довольно невысока.

Приведенной формулой пользуются для сгруппированных данных или при малых выборках, т.е. тогда, когда каждый ранг встречается в исходной совокупности по одному разу. На практике гораздо чаще встречаются материалы, где значения признаков повторяются. В таких случаях формула коэффициента ранговой корреляции Спирмена имеет вид:

$$\rho = 1 - \frac{6(\sum d^2 + T_x + T_y)}{n^3 - n}$$

$$T_x = \frac{\sum (t_x^3 - t_x)}{12} \qquad T_y = \frac{\sum (t_y^3 - t_y)}{12}, \quad \text{где}$$

t_x - количество объединенных рангов первого признака;

t_y - количество объединенных рангов второго признака.

Пример 6.5:

Дано распределение студентов по полу и успеваемости.

успеваемость (x)	кол-во ст-ов.	кол-во ст-ок	ранг x
отличная	4	6	1
ударная	8	11	2
посредств.	10	8	3
неуспевающ.	1	2	4

пол (y)

женский - 1

мужской - 2

Определить тесноту связи между признаками.

Признак "успеваемость" (x) проранжирован по степени интенсивности проявления в убывающем порядке (первый ранг присвоен высшей успеваемости, последний - низшей). Второй признак - "пол" - носит альтернативный характер и также должен быть проранжирован в убывающем порядке. Для этого оценим количественную интенсивность проявления признака в рассматриваемой совокупности. Женщин больше мужчин. В соответствии с этим женщинам присваеваем первый ранг, а мужчинам - второй. Далее для простоты подсчетов сведем имеющиеся сведения в таблицу:

x	111111 1 1 1 1 2222222222222222222 333333333333333333444
y	111111 2 2 2 2 1111111111122222222 111111112222222222112
d	000000 -1 -1 -1 1 1111111111100000000222222221111111111332
d^2	000000 1 1 1 1 1111111111100000000 444444441111111111994

$$\sum d^2 = 79$$

$t_{x1}=10$; $t_{x2}=19$; $t_{x3}=18$; $t_{x4}=3$; $t_{y1}=27$; $t_{y2}=23$;

Подставляем полученные значения в формулы:

$$T_x = \frac{(10^3-10)+(19^3-19)+(18^3-18)+(3^3-3)}{12} = 1139$$

$$T_y = \frac{(23^3-23)+(27^3-27)}{12} = 2650$$

$$\rho = 1 - \frac{6(79+1139+2650)}{50^3-50} \approx 0{,}8$$

Полученное значение свидетельствует о прямой тесной связи между рассматриваемыми признаками, т.е. успеваемость во многом зависит от пола студента. Причем, положительное значение коэффициента говорит о более высокой успеваемости женской части учащихся.

В исторических исследованиях используются и другие коэффициенты ранговой корреляции (коэффициент Кендалла, коэффициент конкордации и др.), но общая теория статистики рекомендует пользоваться коэффициентом корреляции рангов Спирмена. Он менее трудоемок, достаточно представителен.

Сложнее обстоит дело при вычислении силы взаимодействия признаков, проявляющейся во времени, в развитии. В динамических рядах показатели могут быть обусловлены как случайными, так и детерминированными факторами, где каждое последующее явление обусловлено предыдущим. На изменение значений признака в динамическом ряду могут влиять сезонные колебания, цикличность процесса. Следовательно, прежде чем вычислять какой бы то ни было коэффициент корреляции, необходимо оценить характер признаков динамического ряда и факторы, определяющие изменения их значений. Анализ степени взаимодействия случайных признаков в динамическом ряду можно провести на основе уже рассмотренных корреляционных коэффициентов. Обязательно в данном случае должно присутствовать в тексте работы доказательство правомерности использования избранного приема исследования.

Когда предполагается, что компоненты динамического ряда могут быть связаны между собой, то прибегают к вычислению автокорреляции, раскрывающей силу зависимости между соседними уровнями динамического ряда. Она вычисляется по формуле линейного коэффициента корреляции. В качестве значений первого признака (х) берутся исходные уровни динамического ряда, за исключением последнего. В качестве значений второго признака (у) используются те же уровни динамического ряда, но без первого.

Пример 6.6:

Дано распределение социального состава РКП(б) в годы гражданской войны (в тыачел.)

год	рабочие	крестьяне	служащие	прочие
1918	65,4	16,7	25,8	7,1
1919	120,1	54,9	60,1	16,4
1920	188,8	108,4	104,7	29,5
1921	240,0	165,3	138,8	41,5

Подсчитаем автокорреляцию для категории "рабочих".

X	Y	XY	X^2	Y^2
65,4	120,1	7854,2	4277,2	14424,0
120,1	188,8	22674,8	14424,0	35645,4
188,8	240,0	45312,0	35645,4	57600,0

$$\sum x = 374{,}3; \qquad \sum y = 548{,}9; \qquad \sum xy = 75841{,}3;$$

$$\sum x^2 = 54346{,}6; \qquad \sum y^2 = 107669{,}4$$

$$\overline{X} = \frac{\sum x}{n} = \frac{374{,}3}{3} = 124{,}8; \qquad \overline{y} = \frac{\sum y}{n} = \frac{548{,}9}{3} = 183{,}0$$

$$r = \frac{\sum x_i y_i - n\overline{xy}}{\sqrt{\sum x_i^2 - n\overline{x}^2} \cdot \sqrt{\sum y_i^2 - n\overline{y}^2}} =$$

$$= \frac{75841{,}3 - 3 \cdot 124{,}8 \cdot 183{,}0}{\sqrt{54346{,}6 - 3(124{,}8)^2} \cdot \sqrt{107669{,}4 - 3(183{,}0)^2}} \approx 0{,}98$$

Полученное значение очень высоко, что свидетельствует о прямой зависимости динамики социального состава от его начального положения. Это может служить доказательством проведения направленной социальной политики в партийном строительстве, а следовательно, о детерминированности изучаемого процесса. Можно считать, что математически подтверждено, что изменение социального состава РКП(б) в годы гражданской войны нельзя считать стихийным процессом, это было управляемое и жестко контролируемое явление.

Коэффициент автокорреляции рассчитывается не только между соседними уровнями, но и между сдвинутыми на любое число единиц времени.

В математической статистике разработаны методы определения зависимости между динамическими рядами при помощи корреляционного анализа, однако они требуют дополнительных вычислений, связанных с исключением тренда, исключением автокорреляции.

* * *

Суммируя тему измерения взаимосвязей признаков, следует сделать несколько общих замечаний.

Во-первых, величины, подсчитанных разными методами коэффициентов корреляции несопоставимы между собой. Так, если величина Q оказалась больше величины r, то это не означает, что признаки, связь между которыми подсчитана по коэффициенту ассоциации, сильнее взаимодействуют, чем признаки, связь которых определена коэффициентом корреляции рангов.

Во-вторых, не рекомендуется сопоставлять силу связи признаков разной природы, т.к. их подсчет требует применения разных формул. К примеру, не рекомендуется выяснять в каком случае сильнее взаимодействие - между стажем и заработной платой работников или между национальностью и семейным состоянием. Прежде всего, исторически - это признаки разные (экономические, этнический, демографический), а с точки зрения применения математических методов - мы имеем здесь пару количественных и пару качественных признаков.

Если без комплексного анализа связей всех признаков обойтись нельзя, то лучше использовать меры зависимости, пригодные для номинального уровня измерения. При этом происходит "огрубление" исходной информации, уменьшается ее "точность", но повышается надежность показателей взаимодействия и связи признаков.

С помощью корреляционного метода можно выделить важнейшие факторы, влияющие на изучаемый признак или процесс, классифицировать признаки по степени влияния, исключить малозначимые признаки из системы.

При расчете коэффициента корреляции по выборочным данным необходима статистическая оценка степени надежности параметров корреляции. Она проводится по общим правилам проверки статистических гипотез.

Коэффициент корреляции не раскрывает степень воздействия факторного признака на результативный. Таким показателем служит ***коэффициент детерминации*** (D). Его значение определяет долю изменений, возникающих под влиянием факторного признака, в общей изменчивости результативного признака. Величина D измеряется значением коэффициента линейной корреляции, возведенным в квадрат и выраженным в процентах. Так, например, если коэффициент линейной корреляции двух признаков, из которых один является фактором, **а** второй - результатом, равен 0,7, то $D = (0,7)^2 100\% = 49\%$. Таким образом, только на 49% данный признак зависит от рассматриваемого фактора. Следовательно, на 51% его изменение обусловлено иными причинами.

Использованию корреляционного метода обязательно должен предшествовать качественный анализ, в ходе которого определяется принципиальная возможность существования связи между рассматриваемыми признаками, характер распределения их значений, доказывается правомерность использования формулы коэффициента корреляции. Оценка полученных результатов также должна вестись в тесном переплетении с качественной стороной работы. Только при взаимном дополнении теоретического и количественного методов можно достичь успеха в решении задач исторического исследования.

ДОПОЛНИТЕЛЬНО ПО ТЕОРИИ МЕТОДА ЧИТАЙТЕ:

1. Елисеева ЛИ. Статистические методы измерения связей.- Л, 1982.

2. Елисеева И.И., Юзбашев М.М. Общая теория статистики.- М., 1995.- С.190-256, 307-312.

3. Количественные методы в исторических исследованиях.- М., 1984- С.137-177, 201-203, 204-224.

4. Общая теория статистики. - М., 1985.- С.127-187, 214-227.

5. Рабочая книга социолога. - М., 1983 - С.172-192.

6. Славко Т.И. Математико-статистические методы в исторических исследованиях. - М., 1981. - С.87-115.

7. Ферстер Э., Ренц Б. Методы корреляционного и регрессионного анализа. - М., 1983.

8. Эренберг А. Анализ и интерпретация статистических данных. - М., 1981.

Лекция 7.
МАТЕМАТИЧЕСКИЕ МЕТОДЫ ИССЛЕДОВАНИЯ ТЕКСТОВ.

Наиболее обширную группу исторических источников составляют развернутые индивидуальные тексты. Источниковедение не относит нарративные источники к массовым, а значит они не соответствуют требованиям, предъявляемым к вероятностным событиям и к ним не применимы математические методы. Однако письменный текст имеет статистическую структуру и определенные характеристики, в нем содержащиеся, могут быть описаны с помощью вероятностных законов. Таким образом, в случае необходимости нарративный источник можно превратить в массовый путем частотных, классификационных преобразований, т. е. ***методом контент-анализа.***

Контент-анализ складывается из двух основных этапов. На первом, классификационном этапе, исходя из исследовательской цели и информативного потенциала источников, выделяется совокупность признаков, многократно встречающихся в документах. Здесь происходит формализация документа, вырабатывается некое подобие анкеты. Некоторые исследователи рассматривают процедуру контент-анализа состоящей из трех этапов (см., например, Миронов Б.Н. Историк и социология. - Л., 1984), подразделяя первый классификационный этап на два "действия" - качественного и количественного характера.

Сложность работы на классификационном этапе состоит не столько в выделении признаков, сколько в определении их градации, подразделений. Это связано со спецификой языка, особенностями словесных характеристик, системой их измерения. Например, один и тот же признак может встречаться в тексте, будучи выражен разными словами-синонимами, иносказаниями, развернутыми описаниями; иметь, в зависимости от контекста, различную эмоциональную и смысловую окраску; принимать и положительные, и отрицательные значения. В качестве смысловой единицы анализа текста может выступать тема, выраженная в целой статье, во фрагменте текста, ... до отдельных слов и даже слогов и букв.

В результате исследователь получает набор признаков, встречающихся в документах, в текстах достаточно большое число раз и принимающих переменные значения (в историографии полученные первичные понятия именуются символами). Далее вводятся категории - более общие, крупные показатели, являющиеся классами символов.

Следующим шагом первого этапа выступает подсчет часто ты появления в тексте каждого символа и частоты их смысловых связей. Если есть необходимость определеяется частота отношения автора (авторов) текста к полученным смысловым единицам.

Таким образом, можно сказать, что на первом этапе кон тент-анализа решается вопрос "что считать?". Работа здесь не может быть формализована, она определяется методологическими принципами исследователя, уровнем его профессиональной квалификации. В литературе в связи с этим высказано сомнение в объективности контент-анализа. Однако надо заметить, что метод позволяет проследить все шаги исследователя, все этапы его работы и воспроизвести их заново, проверить и перепроверить ее результаты. В этом смысле работы, выполненные на основе контент-анализа, выгодно отличаются от традиционного подхода, когда историк раскрывает свою концепцию, иллюстрируя ее отдельными фрагментами из текста источника. Здесь личная позиция автора в большей мере определяет характер результатов изучения текста и другой историк, обратившись к тому же источнику может получить совсем другие выводы и проиллюстрировать их другими фрагментами из того же текста.

На втором этапе решается вопрос "как считать?". В зависимости от характера количественных данных, от частотных классификаций, от группировок определяется процедура расчета показателей по разработанным математико-статистическим методам.

Рассмотрим конкретный пример применения контент-анализа в исторической науке. Д.В.Деопик с его помощью изучил памятник древнекитайской исторической традиции "Чуньцю" - "Вёсны и осени" (смДеопик Д.В. Опыт количественного анализа древней весточкой летописи "Чукыпо" //Математические методы в историко-экономических и историко-культурных исследованиях. - М., 1977.-С.144-190). Текст датируется при мерно YII-серединой Y в.в. до н.э. Он известен в многочисленных средневековых списках как часть трудов Конфуция. "Чуньцю" состоит из разрозненных фактов политической истории (и не только) государств бассейна Хуанхэ, он не содержит каких-либо выводов, оценок, мотивировок. Язык документа прост, изложение событий отличается краткостью, ясностью и устойчивостью формулировок. Это первая китайская летопись, поэтому она свободна от влияния традиций. Указанные аспекты упрощают системный анализ текста.

Объем "Чунцю" определялся количеством иероглифов (всего их 16257). Д.В.Деопик произвел расчет подробности текста по единицам времени (год, сезон, месяц указан не всегда). Это позволило сопоставить между собой все виды структурных единиц и оценить их через объем текста (количество иероглифов), приходящийся на каждую хронологическую единицу. Выяснилось, что объем информации, приходящейся на год растет от более ранних к более поздним периодам.

Более глубокий анализ распределения частот иероглифов по годам и по сезонам позволил неопровержимо доказать постепенность создания источника, доказать невозможность его написания "в один присест" на основе архивных материалов.

Переходя к анализу содержания "Чуньцю", автор в качестве главной смысловой единицы выделил элементарное действие, выраженное особым иероглифом. Более общим понятием системы описания является "действие", объединяющее небольшую группу (как правило до 4-х) микродействий. В источнике содержатся 64 действия, из них 20 - с высокой частотой (более 90% всех упоминаемых простейших событий).

Для каждого действия Д.В.Деопик получил 4 характеристики: 1) -распространенность (частота проявления в тот или иной период); 2) -особенности поведения тех или иных государств как субъектов; 3) -наличие географической специфики; 4) -наличие временных тенденций у субъектов и объектов исторического процесса.

Кроме того, 64 действия распределились по 7 группам - "внешняя политика", "военная история", "быт монархов", "внутренняя политика" "экономика", "природные явления", "сакральное". Каждая группа имела еще подгруппы.

Данное распределение и частотная классификация смысловых единиц текста. "Чуньцю" позволили составить компактные и наглядные таблицы, выявляющие как стабильные во времени, так и эволюционизирующие явления внутри- и внешнеполитической жизни, войн, дворцовых интриг. Контент-анализ дал возможность выявить специфику использованного источника, проанализировать даже самые "мелкие", казалось бы незначительные факты и раскрыть общие и частные тенденции в истории древнекитайских государств YIII-Y вв. до н.э.

Методика контент-анализа дала интересные результаты при обработке ответов на четыре хозяйственные анкеты 60-х гг. XYIII в. (см. Миронов Б.Н. Статистическая обработка от ветов на сенатскую анкету 1767 г. о причинах роста хлебных цен //Математические методы в исторических исследованиях.- М., 1972.- С.89-104). В анкету, составленную по предложению Екатерины II вошло 14 вопросов, разослана она была 18 губернаторам и 153 уездным воеводам. Всего анкетой 1767 г. было охвачено 176 уездов (56% всех уездов Российской империи).

В данном случае мы встречаемся с прямо противоположной ситуацией, чем в работе Д.ВДеопика. Содержание ответов на анкетные вопросы отличается большой расплывчатостью формулировок, значительным разнообразием вариантов. Это не позволяет применять статистический анализ к первоисточнику. В то же время традиционный качественный анализ затруднен большим количеством сохранившихся документов. Изучение содержания каждой анкеты ведет к неминуемой потере информации. Б.Н.Миронов составил возможные варианты ответов на каждый вопрос анкеты (от 5 до 31 варианта для разных вопросов). В результате содержание анкет теперь можно группировать по разработанным вариантам и определять их частоты. Проделанная работа внесла количественную определенность в вопрос о причинах повышения хлебных цен во второй половине XYIII в. и на некоторые другие вопросы социально-экономической истории России.

Не спадает интерес к проблемам социальной психологии, менталитета отдельных слоев общества. В научный оборот введены "приговоры" и петиции крестьянских сельских сходов в Совет министров, Государственную Думу, царю, Всероссийскому крестьянскому съезду. Они относятся к массовым источникам по общественному сознанию крестьянства.

С помощью контент-анализа О.Г.Буховец изучил содержание 72 приговоров и наказов политического характера, принятых крестьянами Самарской губернии в 1905-1906 гг. (См.Буховец О.Г. к методике изучения "приговорного" движения и его роли в борьбе крестьянства в 1905-1907 гг. //История СССР.- 1979, N 3). Классификационный этап завершился формированием 30 категорий требований. Подсчет частоты встречаемости каждой категории в тексте источника показал, что наиболее важными для крестьян были:

1) амнистия политическим заключенным (73% документов),

2) установление демократических свобод (58 %),

3) отмена смертной казни (43%),

4) упразднение земских начальников, полиции, стражников (39%) и т.д.

"Вековое чаянье" крестьянства - отмена частной собственности на землю и передача народу частновладельческих, казенных, монастырских, удельных и др. земель - оказалось только на 7-м месте. Это разрушает многие традиционные представления о крестьянском менталитете, о крестьянской борьбе, о их роли в революции.

Аналогичная обработка "приговоров" крестьян Воронежской губернии позволила затем сравнить структуру крестьянского "приговорного" движения в двух губерниях. Сделано это было О.Г.Буховцом с помощью метода ранговой корреляции. Он построил таблицы сопряжения парных требований по 200 документам двух губерний.

Особый интерес представляет возможность использования контент-анализа для выявления и изучения скрытой информации, содержащейся в источнике.

Известие, что любой исторический источник, как нарративный, так и статистический заключает в себе явную, выраженную или воспринимаемую информацию и скрытую., иначе именуемую связанной или структурной. Явная информация отражает главную цель автора (авторов) документа. Она фиксировалась создателями источника сознательно, была близка и понятна современникам, использовалась ими по прямому назначению. Скрытая информация попадала в документ ненамеренно, вне субъективной воли людей, причастных к сбору и фиксированию данных. Она является результатом стихийного отражения в источнике примет времени через взаимосвязи между процессами, явлениями, свойствами объектов.

Структурная информация обладает двумя важными особенностями. Во-первых, она выражается опосредованно, через события, составляющие явную информацию, через их характеристики. Во-вторых, имея колоссальный объем (т.к. взаимосвязи, как правило, сложны и многообразны), она бывает более объективной и достоверной в силу своего стихийного происхождения, чем те показатели, которые намеренно запечатлевались в документе.

Иногда скрытая информация появляется в тексте источника и сознательно. Авторы умышленно маскируют свои подлинные цели, идеи, убеждения, используют "эзопов" язык. Наиболее характерный пример представляет периодическая подцензурная печать. Известен факт проведения контент-анализа американской прессы во время второй мировой войны ГЛассуэлом. В результате им был неопровержимо доказан профашистский характер публикаций газеты "Истинный американец", что привело к ее закрытию. Иносказаниями, словесной "вуалью" широко пользовались российские журналисты - от Н.Новикова в XY111 в. до советских публицистов XX в.

В ряде случаев часть информации оказывается скрытой ввиду несистематизированности и обильности первичных материалов. Так, А.Сезько с помощью контент-анализа и ЭВМ превратил 5850 коммунистических листовок периода гражданской войны в 22 статистические таблицы, в которых отразилось все содержание этого источникового комплекса, выявилась динамика проблематики и многое другое.

Текст некоторых документов - речей, листовок, публикаций периодической печати, анкет, дневников, писем, мемуаров и т.д. - несет в себе много неясного в силу противоречивости авторской позиции. В этом случае только контент-анализ определит и количественно выразит степень противоречивости и меру неустойчивости идейных позиций автора.

Преимущество контент-анализа состоит в том, что с его помощью историк получает возможность перевода массовой текстовой информации в количественные показатели. Это значительно снижает субъективизм исследования, решается проблема его проверки и перепроверки.

В заключение надо сказать, что контент-анализ достаточно успешно применяется в истории для разнотипных нарративных источников. С его помощью обрабатывается информация периодической печати, дневников, мемуаров, переписки, авторских текстов (философские трактаты, записки, разного рода описания), летописей и т.д. Эффективность применения контент-анализа связана с возможностью реализации системного подхода к анализу содержания исторического источника. Этот метод помогает получить более объективную, обоснованную аргументацию выводов, а в ряде случаев получить и скрытую в источнике информацию.

На основе математических методов получены полезные результаты при дешифровке текстов (например, в случае тайнописи или неизвестного языка написания), в области атрибуции памятника письменности, установления авторства. Использование теории информации в текстологии позволяет оценить число промежуточных списков, предшествующих данному. С помощью математического инструментария можно установить оригинал среди многочисленных списков, определить место создания текста, датировать его. Возможна реконструкция источника, его ранее утраченных фрагментов, очищение оригинального текста от более поздних наслоений. Однако, несмотря на накопленный опыт, множество методологических и методических проблем, связанных с измерением нарративных источников остаются нерешенными.

Современный этап использования математико-статистических методов, в частности для анализа текстовых источников, связан с применением компьютерной техники. Еще в начале 60-х гг. появилась новая ветвь информатики - технология баз данных. Новый этап использования ЭВМ в исторических исследованиях наступил в 80-е гг. вместе с внедрением в практику персональных компьютеров. Сегодня историк выступает не только как пользователь исторических источников, но и как создатель новой исторической информации. Разрабатываются методические и методологические проблемы формализации источника, программного обеспечения исторических исследований, содержания и анализа баз машиночитаемой информации. В России действует ассоциация "История и компьютер", выпускающая свой информативный бюллетень, проводящая ежегодные конференции. Более 10 лет работает международная ассоциация "History and Computing", также ежегодно устраивающая свои научно-практические конференции, на которых в последние годы участвуют и российские ученые.

Новой тенденцией, с которой должна считаться современная историческая наука, является осуществление делопроизводства и работы ряда официальных и архивных учреждений в машиночитаемой форме. Архивы, библиотеки, музеи, исследовательские организации, статистические бюро, центры документации создают все больше машиночитаемых файлов для сохранения и вторичного использования информации.

Несмотря на то, что не всякий исторический источник нуждается в переводе его в машиночитаемую форму и мало еще разработано специальных программ, а стандартное программное обеспечение не может полностью удовлетворить историка, работа в этом направлении должна продолжаться и альтернативы в этом вопросе нет. Овладение математическими методами и компьютерными технологиями является свидетельством профессионализма современного историка.

ДОПОЛНИТЕЛЬНО ПО ТЕОРИИ МЕТОДА ЧИТАЙТЕ:

1. Гарскова И.М. Базы и банки данных в исторических исследованиях.- М., 1994.- 215 с.

2. Илизаров Б.С. Моделирование процессов автоэкспертизы письменных исторических источников методами документометрии (О методах количественной оценки источникового фонда страны).//Математические методы в социально-экономических и археологических исследованиях. - М., 1981.- С.196-221.

3. Информационный бюллетень ассоциации "История и компьютер". - Любой выпуск.

4. Математика в изучении средневековых повествовательных источников.- М., 1986.

5. Математические методы в историко-экономических и историко-культурных исследованиях.- М.,1977.- С.117-361.

6. Математические методы в исторических исследованиях.- М.,1972.- С.125-166.

7. Методы количественного анализа текстов нарративных источников. - М.,1983.

8. Миронов Б.Н. Историк и социология. - Л., 1984.

ПРИЛОЖЕНИЯ.

Приложения 1.

Таблица достаточно больших чисел (фрагмент).

p	0.10	0.09	0.08	0.07	0.06	0.05	0.04	0.03	0.02	0.01
0.75	33	40	51	67	91	132	206	367	827	3308
0.80	41	50	64	83	114	164	256	456	1026	4105
0.85	51	63	80	105	143	207	323	575	1295	5180
0.90	67	83	105	138	187	270	422	751	1690	6763
0.95	96	118	150	195	266	384	600	267	2400	9603
0.965	111	137	173	226	308	444	694	1234	2778	11112
0.970	117	145	183	240	327	470	735	1308	2943	11773
0.980	135	167	211	276	375	541	845	1503	3382	13529
0.990	165	204	259	338	460	663	1036	1843	4146	16587
0.991	170	210	266	348	473	682	1066	1895	4264	17057

Славко Т.И. Математико-статистические методыы в исторических исследованиях. - М., 1981. - С. 63.

Приложение 2.

Таблица случайных чисел (фрагмент).

	1	2	3	4	5	6	7	8
1	5489	5583	3156	0835	1988	3912	0938	7460
2	3522	0935	7877	5665	7020	9255	7379	7124
3	7555	7579	2550	2487	9477	0864	2349	1012
4	5759	3554	5080	9074	7001	6249	3224	6368
5	6303	6895	3371	3196	7231	2918	7380	0438
6	7351	5634	5323	2623	7803	8374	2191	0464
7	7068	7803	8832	5119	6350	0120	5026	3684
8	3613	1428	1796	8447	0503	5654	3254	7336
9	5143	4534	2105	0368	7890	2473	4240	8652
10	9815	5144	7649	8638	6137	8070	5345	4865

Количественные методы в исторических исследованиях. - М., 1984.- С.378.

ОГЛАВЛЕНИЕ:

Оригинал-макет Шавалеевой СМ.
Откопировано на ризографе в издательстве "Форт-Диалог"
Заказ № 82. ТиражJ$2. Бумага офсетная.
г.Казань, ул.Университетская, 17. Тел. (8432) 38-73-51