Основные формы, виды и способы статистического наблюдения

2.Тема 2. Статистическое наблюдение

2.1.Понятие о статистическом наблюдении

Любое экономико-статистическое исследование начинается со статистического наблюдения. Статистическое наблюдениеСтатистическое наблюдение - это предварительная стадия статистического исследования, которая представляет собой планомерный, научно организованный учет (сбор) первичных статистических данных о массовых социально-экономических явлениях и процессах.

Не всякий сбор данных можно назвать статистическим наблюдением. Наблюдение будет статистическим, во-первых, когда оно сопровождается регистрацией изучаемых фактов в соответствующих учетных документах для дальнейшего их обобщения, во-вторых - когда носит массовый характер. Это обеспечивает охват значительного числа случаев проявления того или иного процесса, необходимого и достаточного для того, чтобы получить данные, которые касаются не только отдельных единиц совокупности, но и всей совокупности в целом.

Статистическое наблюдение должно отвечать ряду важнейших требований:

а) проводиться непрерывно и систематически;

б) учет массовых данных должен быть таким, чтобы не только обеспечивалась полнота данных, но и учитывалось их постоянное изменение;

в) данные должны быть максимально достоверны и точны;

г) исследуемые явления должны иметь не только научную, но и практическую ценность.

Сбор статистических данных может проводиться как органами государственной статистики, научно-исследовательскими институтами, другими государственными структурами, так и экономическими службами банков, бирж, предприятий, фирм. Только в этом случае исследователи получают достоверную и достаточно разнообразную статистическую информацию, позволяющую всесторонне изучать социально-экономические явления.

2.2.Этапы, формы, виды и способы статистического наблюдения

Статистическое наблюдениеСтатистическое наблюдение (сбор первичного статистического материала) состоит из трех основных этапов:

подготовка статистического наблюдения;

организация и производство наблюдения;

контроль полученных первичных данных.

На этапе подготовки статистического наблюдения определяется цель, устанавливаются объект и единица наблюдения, разрабатываются инструментарий и программа наблюдения. Общей Статистическое наблюдение: целью статистического наблюдения является получение достоверной информации о тенденциях развития явлений и процессов для последующего принятия управленческих решений. Она должна быть конкретной и четкой. Нечетко поставленная цель может привести к сбору не тех данных, которые необходимы для решения конкретной задачи.

Цель определяет объект статистического наблюдения. Статистическое наблюдение: Объект наблюдения есть некоторая исследуемая статистическая совокупность или физических лиц (население, работники), или юридических лиц (предприятия, фирмы, учебные заведения), или физических единиц (производственное оборудование, средства передвижения и транспортировки, жилые дома), т.е. исследуемая статистическая совокупность состоит из отдельных единиц.

Единица наблюдения - это первичный элемент объекта статистического наблюдения, который является носителем признаков, подлежащих регистрации. Указание важнейших признаков позволяет установить границы исследуемой совокупности. Скажем, если необходимо провести исследование рентабельности полиграфических предприятий, то необходимо определить формы собственности этих предприятий, организационно-правовые основы, количество работников предприятия, объем реализации продукции, т.е. то, что отличает как государственные и негосударственные предприятия, так и малые и крупные предприятия. Только в этом случае мы получим достоверную статистическую информацию.

Единицу наблюдения следует отличать от отчетной единицы. Отчетной называют такую единицу, от которой поступают отчетные данные. Она может совпадать или не совпадать с единицей наблюдения.

Обоснование цели, выбор единиц наблюдения, отчетных единиц, отбор существенных признаков, период времени проведения статистического наблюдения, формы отчетности излагаются в программе статистического наблюдения. Обычно Программа наблюдения называют перечень вопросов, которые подлежат регистрации при проведении наблюдения. Чтобы программа наблюдения была научно обоснована и правильно составлена, к ней предъявляются следующие требования:

четкая и конкретная формулировка главной цели наблюдения;

определение места и времени наблюдения, где определяются критический момент (дата или интервал времени, по состоянию на который проводится регистрация признаков) и срок (период заполнения статистического формуляра);

выделение ряда наиболее существенных признаков объекта наблюдения;

комплексное определение типа, основных черт и свойств изучаемого явления;

вопросы, сформулированные в программе, не должны носить двусмысленный характер;

соблюдение логического принципа последовательности вопросов;

включение в программу вопросов контрольного характера для проверки собираемых статистических данных;

сочетание закрытых и открытых вопросов программы.

Программа оформляется в виде документа, так называемого статистического формуляра, который обеспечивает единообразие получаемых сведений от каждой отчетной единицы. Формуляр имеет титульную часть (сведения о тех, кто проводит наблюдение) и адресную часть (адрес и подчиненность отчетной единицы). Программа имеет приложение - инструкцию (Статистическое наблюдение: инструментарий статистического наблюдения), которая определяет порядок проведения наблюдения и порядок заполнения формы отчетности.

На втором этапе решаются важнейшие организационные вопросы статистического наблюдения. Они заключаются в том, чтобы выбрать соответствующие целям и задачам конкретного статистического наблюдения организационные формы наблюдения, виды наблюдения и способы получения статистической информации.

Все многообразие форм, видов и способов наблюдения можно представить следующим образом.

По форме организации статистического наблюдения: отчетность; специально организованное статистическое обследование - перепись; регистры.

По видам статистического наблюдения: а) по времени регистрации фактов (текущее или непрерывное; прерывное - периодическое, единовременное); б) по охвату единиц совокупности (сплошное; несплошное - основного массива, выборочное, монографическое).

По способам получения статистической информации: непосредственное наблюдение; документальный способ; опрос - экспедиционный, анкетный, явочный, корреспондентский, саморегистрация.

Основной формой статистического наблюдения является отчетность. Если первичный учет (первичный учетный документ) регистрирует различные факты, то отчетность является обобщением первичного учета.

Отчетность - официальный документ, который скрепляется подписями лиц, ответственных за предоставление и достоверность собранных сведений, и утверждается органами государственной статистики. Кроме годовой может иметь место ежедневная, недельная, двухнедельная, месячная и квартальная отчетность. Отчетность может быть представлена по почте, телеграфу, телетайпу, факсу.

К специально организованному статистическому наблюдению можно отнести перепись. На практике проводится перепись населения, материальных ресурсов, зеленых насаждений, незавершенных строительных объектов, оборудования и т.д.

Перепись - наблюдение, повторяющееся через равные промежутки времени, задачей которого является не только определение численности и состава исследуемой совокупности, но и анализ количественных изменений в период между двумя обследованиями. Из всех переписей наиболее известны переписи населения.

Формой непрерывного статистического наблюдения является регистровое наблюдение (регистр), объектами которого являются долговременные процессы, имеющие фиксированное начало, стадию развития и фиксированное время завершения. Регистр основан на системе отслеживания состояния переменных и постоянных показателей. В статистической практике различают регистры населения и регистры предприятий. В настоящее время в России существует Единый государственный регистр предприятий всех форм собственности (ЕГРПО), информационный фонд которого содержит: регистровый код, сведения о территориальной и отраслевой принадлежности, форме подчиненности, виде собственности, справочные сведения и экономические показатели (среднесписочная численность работников; средства, направляемые на потребление; остаточная стоимость основных средств; балансовая прибыль или убыток; уставный фонд). При закрытии предприятия в десятидневный срок ликвидационная комиссия информирует об этом службу ведения регистра.

Рассмотрим коротко виды статистического наблюдения по времени регистрации фактов. Статистическое наблюдение: Непрерывное (текущее) статистическое наблюдение - это систематическая регистрация фактов или явлений по мере их поступления с целью изучения их динамики. Например, регистрации актов гражданского состояния (рождения, браки, смерти), регистрация страховыми компаниями всех несчастных случаев и других неблагоприятных событий по мере их возникновения.

Видами Статистическое наблюдение: Прерывное наблюдения являются единовременное и периодическое. Первое есть разовое сплошное наблюдение для сбора количественных характеристик явления или процесса в момент его исследования. Периодическое наблюдение проводится через определенные промежутки времени по схожим программе и инструментарию. Например, периодическое исследование пассажиропотоков в общественном транспорте, периодическая регистрация цен производителей по отдельным товарам (один раз в месяц или в квартал).

По охвату единиц совокупности статистическое наблюдение бывает сплошным и несплошным. Статистическое наблюдение: Сплошное наблюдение охватывает все единицы исследуемой совокупности (например, общая перепись населения). В свою очередь, Статистическое наблюдение: несплошное наблюдение охватывает только часть исследуемой совокупности. В зависимости от того, как выбрана эта часть, несплошное наблюдение можно подразделить на выборочное (основано на принципе случайного отбора), метод основного массива (исследуются самые существенные или наиболее крупные единицы изучаемой совокупности) и так называемое монографическое наблюдение (подробное исследование отдельных единиц изучаемой совокупности с целью выявления намечающихся тенденций).

Что касается способов получения статистической информации (способов статистического наблюдения), то здесь выделяют три основных способа: непосредственное наблюдение, документальное наблюдение и опрос.

Достаточно надежным источником данных является Наблюдение непосредственное наблюдение, когда можно установить факт, подлежащий регистрации. Но данный способ требует значительных затрат труда и наличия всех необходимых условий. Чаще всего он используется при наблюдении за вводом в действие строительных объектов.

Другой надежный способ - Наблюдение документальный, основанный на использовании в качестве источника информации различных документов учетного характера (счета, рекламации и т.д.) и способствующий получению точной информации.

Способ наблюдения, при котором источником сведений являются слова респондентов, называют опросом. Его разновидности: устный (экспедиционный), анкетный, корреспондентский, явочный опрос и саморегистрация.

устный опрос может быть как прямым (непосредственное общение счетчика с респондентом), так опосредованным (например, по телефону).

При анкетном способе определенное число респондентов получают специальные вопросники либо лично, либо через средства печати. Данный вид опроса применяется в исследованиях, где нужны ориентировочные результаты, не претендующие на высокую точность (изучение общественного мнения).

Явочный способ используется в сплошном наблюдении, когда необходимо личное присутствие (регистрация браков, разводов, рождений и т.д.).

При корреспондентском способе сведения сообщаются штатом добровольных корреспондентов, в силу чего полученный материал не всегда носит качественный характер.

Наконец, при способе саморегистрации формуляры заполняются самими респондентами, а счетчики консультируют и собирают формуляры. В статистической практике различные виды статистических наблюдений могут сочетаться, дополняя друг друга.

На третьем этапе собранный статистический материал должен пройти контроль. Как показывает практика, даже при четко организованном статистическом наблюдении встречаются погрешности и ошибки, которые требуют исправления. Поэтому целью этого этапа является как счетный, так и логический контроль полученных первичных данных. Расхождение между расчетным и действительным значениями исследуемой величины в статистике называют ошибкой наблюдения. В зависимости от причин возникновения различают ошибки регистрации и ошибки репрезентативности.

Ошибки регистрации могут быть случайными и систематическими. Ошибки регистрации: Случайные ошибки не имеют определенной направленности и возникают под действием случайных факторов (перестановка цифр, смещение строк и граф при заполнении статистического формуляра). При обобщении массового материала эти ошибки взаимопогашаются.

Ошибки регистрации: Систематические ошибки регистрации имеют определенную направленность, могут либо завышать, либо занижать конкретное значение показателя, что в итоге приводит к искажению действительного положения. Примерами систематической статистической ошибки при регистрации служат округление возраста населения на цифрах, заканчивающихся на 5 и 0, преуменьшение доходов в документации для налоговых органов, элементы недостоверности, которые вносят предприятия в те характеристики, от которых зависит расчет с кредиторами, и т.д.

Для выявления ошибок используется счетный контроль, особенно для проверки итоговых сумм. Помимо счетного используется и логический контроль, который может поставить под сомнение правильность полученных данных, поскольку основан на логической взаимосвязи между признаками. Например, при переписи населения полученный факт, что пятилетний ребенок имеет среднее образование, ставится под сомнение и в этом случае ясно, что при заполнении формуляра допущена ошибка.

Если ошибки регистрации свойственны любому Статистическое наблюдению (сплошному и несплошному), то Ошибки репрезентативности - только несплошному наблюдению. Они характеризуют расхождения между значениями показателя, полученного в обследуемой совокупности, и его значением по исходной (генеральной) совокупности. Ошибки репрезентативности также могут быть случайными и систематическими. Ошибки репрезентативности: Случайные ошибки возникают, если отобранная совокупность не полностью воспроизводит все признаки генеральной совокупности и величину этих ошибок можно оценить. Ошибки репрезентативности: Систематические ошибки репрезентативности могут возникать, если нарушен сам принцип отбора единиц из исходной совокупности. В этом случае проводятся проверка полноты собранных данных, арифметический контроль точности информации на предмет ее достоверности, проверка логической взаимосвязи показателей.

Контрольной проверкой собранных данных завершается статистическое наблюдение.

Способы распостранения данных

Распространение выборочных данных на генеральную совокупность является конечной задачей выборочного наблюдения.

Обычно применяется два способа такого распространения: **способ прямого пересчета и способ коэффициентов.**

Способ прямого пересчета состоит в том, что средняя величина признака, найденная посредством выборки, умножается на число единиц генеральной совокупности.

Например, необходимо определить средний процент брака в партии консервов, состоящей из 10000 банок. Для выборочного наблюдения в случайном порядке было отобрано 900 банок. Анализ качества отобранных банок консервов показал, что средний процент брака в данной совокупности составил 1,5%. Среднее квадратическое отклонение равно 0,3%. Максимальная ошибка выборочного наблюдения с вероятностью 0,997 равна 0,3%.

Таким образом, средний процент брака в генеральной совокупности находится в пределах 1,5% ± 0,3%, т.е. колеблется от 1,2% до 1,8%.

Имея данные об общей величине партии, определяем общее количество бракованных банок, которое будет колебаться в пределах 1,8-1,2% от 10000, или 180-120 единиц. Можно пределы не указывать, а пользоваться средней выборочной как генеральной средней. Тогда среднее количество бракованных банок в генеральной совокупности составит 1,5% от 10000, т.е. 150 единиц.

Второй способ, или способ коэффициентов применяется тогда, когда выборочное обследование проводится в целях проверки данных сплошного наблюдения.

Сущность этого метода заключается в том, что на основании сопоставления данных сплошного и данных выборочного наблюдений устанавливают процент расхождений (процент недоучета), который и служит коэффициентом поправки, налагаемой на данные сплошного наблюдения.

Например, имеются данные о количестве скота, находящегося в личном пользовании согласно переписи, а также согласно контрольному обходу (табл. 6.2).

Таблица 6.2

Количество скота, находящегося в индивидуальном пользовании населения

Группа скота	Учтено во всех хозяйствах по переписи	Учтено в хозяйствах, подвергнутых контрольному обходу		За время, прошедшее от переписи до контрольного обхода в хозяйствах, подвергнутых контрольному обходу	
		по переписи	при контрольном обходе	прибыло	убыло
Коровы	9200	850	863	6	2
Нетели и телки,	1200	140	144	4	1

рожденные в прошлом году и старше					
Телки, рожденные в этом году	800	80	87	2	–
ИТОГО	10200	1070	1094	12	3

Чтобы определить процент недоучета, нужно найти разность между данными контрольного обхода и данными сплошного наблюдения, а затем полученную величину разделить на данные сплошного наблюдения.

При переписи недоучтено:

коров: 863 - 850 - 6 + 2 = 9;

нетелей: 144 - 140 - 4 + 1 = 1;

телок: 87 - 80 - 2 + 0 = 5.

Данные контрольного обхода о количестве телок сопоставляют с данными переписи.

Отсюда коэффициент недоучета коров равен $\frac{9 \cdot 100}{850} = 1{,}06\%$.

Коэффициент недоучета нетелей — $\frac{1 \cdot 100}{140} = 0{,}72\%$.

Коэффициент недоучета телок — $\frac{5 \cdot 100}{80} = 6{,}25\%$.

Полученные результаты выборочного наблюдения (проценты недоучета) распространяются на всю совокупность.

Для этого поправочные коэффициенты (проценты недоучета) умножаем на данные сплошного наблюдения, полученные в результате переписи скота во всех хозяйствах (табл. 6.3).

Таблица 6.3

Расчет фактического количества поголовья скота при помощи поправочных коэффициентов (процент недоучета)

Группы скота	Поправные коэффициенты, %	Учтено во всех хозяйствах	Количество скота с поправкой на данные выборочного обследования
Коровы	1,06	9200	9298
Нетели	0,72	1200	1209
Телки	6,25	800	850
Итого	–	11200	11357

Условия, обеспечивающие точность данных статистического наблюдения

Организационный план статистического наблюдения - это составная часть общего плана наблюдения, в которой изложен порядок его организации и проведения. В нем даются разъяснения программно-методологических и организационных вопросов. К первым относятся формулирование цели и задач наблюдения, определения его объекта и единицы, разработка программы. К организационным вопросам относятся: фиксация места, времени и сроков наблюдения, указание на то, кто его проводит, как оно проводится и как осуществляется поставка статистическим формулярами лиц, выполняющих наблюдения, способы доставки заполненных формуляров в соответствующие статистические органы. Сюда относят также ряд специфических подготовительных работ, таких, как подбор и обучение кадров, привлекаемых к проведению наблюдения, подготовка графического материала и др..

В организационном плане статистического наблюдения конкретизируются права и обязанности отдельных учреждений и организаций, участвующих в мероприятиях наблюдения.

При планировании наблюдения прежде всего определяют органы наблюдения - организаторов и исполнителей работ, а также права и обязанности каждого соисполнителя.

Программу и план статистического наблюдения разрабатывают органы государственной статистики на уровне Министерства Государственного комитета статистики Украины. Низовые органы государственной статистики преимущественно выполняют работу по сбору статистических данных, их первоначальном контроля и возведению по определенной программе.

В плане указывают срок проведения наблюдения, т.е. время начала и окончания сбора сведений. Это время нельзя отождествлять со временем наблюдения, то есть временем, к которому относятся сведения. Статистические показатели характеризующие изучаемое явление или за определенный период, либо на определенный момент времени. Например, данные о количестве произведенной продукции можно взять только за период (день, декаду, месяц, квартал, год), а показатели запасов материальных ценностей могут быть представлены на определенный момент времени (на начало месяца, на начало квартала, на начало или конец года и т.д..).

В плане должно быть точно определена территория, на которой осуществляется наблюдение, а также лица и организации, ответственные за проведение подготовительных работ, сбор, проверку и обработку информации по отдельным участкам территории.

Среди организационных вопросов значительное место в плане отводится проведению подготовительных работ. Прежде всего надо составить список отчетных единиц, то есть тех, которые будут обследованы. Этот список (коллективных предприятий, государственных, арендных предприятий и т.д.). Необходим для проверки полноты сведений, поступающих а также определение объема работ и расчета необходимого количества рабочих для проведения статистического наблюдения.

Важным подготовительным мероприятием является расчет потребности в кадрах для проведения наблюдения, их подбор и инструктаж. Необходимо заранее отпечатать и разослать бланки документов и инструкции по их заполнению. Инструктаж считается одним из важнейших подготовительных работ статистического наблюдения. Успех проведения последнего во многом зависит от уровня подготовленности кадров.

При подготовке сложных статистических наблюдений, как правило, в плане предусматривается проведение пробных наблюдений с целью проверки на практике проекта плана и программы основного наблюдения. Материалы таких пробных наблюдений используют для уточнения, дополнения и конкретизации программы и плана наблюдения, а также инструкций.

Среди подготовительных работ видное место должно принадлежать (особенно это касается переписи населения) пропаганде наблюдения среди населения. Средства массовой информации должны вести разъяснительную работу относительно задач и цели наблюдения, в значительной мере способствовать повышению эффективности решения планово-организационных вопросов наблюдения. Разъяснительная работа о целях и задачах проведения наблюдения осуществляется через прессу, радио, телевидение и другие средства массовой информации.

Таким образом, организационный план статистического наблюдения предусматривает: определение времени наблюдения, времени и места его проведения, порядок передачи материалов наблюдения, комплекс подготовительных работ, меры, обеспечивающие точность (достоверность) статистических данных. Каждой из названных выше категорий и этапов статистического наблюдения можно дать следующие определения.

Принципы формирования выборок

Формирование выборки – это процесс определения целевой аудитории и отбора из всей совокупности потенциальных респондентов группы, имеющей все свойства совокупности. В данном разделе рассматриваются некоторые наиболее часто используемые способы формирования выборки.

Существуют два основных метода построения выборки: вероятностный и детерминированный.

Вероятностная выборка – такая, которая получается на основе вероятностных законов. Это значит, что любой из объектов генеральной совокупности может быть отобран в выборку с определенной вероятностью.

Детерминированная выборка – такая, когда отбор в выборку производится на основе каких-либо принятых дополнительных условий, ограничивающих круг вероятных респондентов. Их применяют, когда невозможно ограничить круг потенциальных респондентов по какому-либо формальному свойству или признаку.

Вероятностные методы формирования выборки

Вероятностные выборки предполагают, что в основе метода лежит принцип случайности, когда каждый элемент множества имеет равные шансы попасть в число опрашиваемых респондентов. Случайность в данном случае подразумевается не как чистая случайность, а как некая система организации отбора. И система состоит в том, что именно случай определяет окончательный выбор респондента.

Основой вероятностных выборок в маркетинговых исследованиях является строгое требование равновероятного доступа к любому респонденту. На практике это требование не всегда удается соблюсти. Для обеспечения равной вероятности участия в опросе любого возможного респондента исследователю, как правило, необходимо иметь явный или опосредованный список всех потенциальных респондентов.

Вероятностная выборка – фактически единственный способ получить объективные данные, которые будут иметь необходимую точность. Поскольку каждый респондент равновероятно может быть выбран, это дает возможность оценить вероятность того, что параметры, полученные для выборки, находятся в пределах соотношений, характерных для всей генеральной совокупности.

Рассмотрим способы формирования выборок.

Простая случайная выборка

В целевой группе каждый респондент имеет равный шанс быть выбранным для участия в опросе. Для отбора применяются несколько разновидностей схем.

А. Можно использовать принцип жеребьевки, когда составляется перепись всей генеральной совокупности и фиксируется на бумажных либо электронных носителях. Затем производится жеребьевка. Это аналог лотереи. Такой подход дает наиболее точные результаты, но в то же время является самым трудоемким.

Б. Можно использовать генераторы случайных чисел, которые “выдают” номера из всего перечня объектов. На практике способ используется нечасто, так как прежде чем выбирать все объекты, необходимо каким-то образом пронумеровать, а после из пронумерованных вытаскивать как лотерею. Способ используется при телефонном опросе, когда имеется телефонная книга или некий список абонентов.

В. Использование таблицы случайных чисел. Можно, начав с произвольно выбранного в таблице места, отбирать определенное количество последующих чисел, и они будут определять номера респондентов, которых нужно опросить.

Пример. Рассмотрим, как можно построить случайную выборку для последующего опроса, скажем, зарегистрированных клиентов некоего предприятия. Пусть у предприятия имеется 1 тыс. клиентов. Нам необходимо опросить 100 из них. Всех клиентов необходимо пронумеровать, можно составить полный список. Далее сделать 1000 одинаковых листков, написать на них все числа от 1 до 1000 и, используя принцип лотереи, вытаскивать по одному из барабана. Или можно воспользоваться несложной компьютерной программой генератора случайных чисел и взять 100 чисел, используя в них только первые три разряда или последние три. Аналогично применяется таблица случайных чисел, когда, начиная со взятого наугад, используются последующие 100 значений, ограничиваясь необходимым количеством разрядов.

Стратифицированная (или расслоенная) выборка

При организации стратифицированной выборки генеральная совокупность разбивается на несколько групп (слоев), отличающихся друг от друга по каким-либо признакам. Например, взрослое население можно разбить на группы 18-25 лет, 26-35 лет, 36-45 лет и т.д. Стратификация – это дифференциация на основе определенных критериев, таких как профессия, доход, образование, потребительские возможности, социальный статус и др.

Главное в этом способе заключается в наличии возможности проводить исследования, выбирая респондентов из каждого слоя (группы) на основе простой случайной выборки.

Например, в маркетинговом исследовании можно всех респондентов разделить на две страты – мужчины и женщины. Предположим, респонденты распределены в соотношении 50 х 50. Если в сформированной выборке был обеспечен отбор половины из одной страты и половины из другой, такая выборка будет являться стратифицированной (расслоенной) с пропорциональным распределением. Из каждой группы (слоя) выбирается количество респондентов в соответствии с существующими требованиями.

В том случае, если из имеющихся страт исследователя интересует одна, то можно в выборке предусмотреть необходимое большее количество респондентов из интересующей страты и значительно меньшее количество – из второй. Такая выборка называется стратифицированной с оптимальным распределением. Количество респондентов из каждой страты отбирается в соответствии с некоторыми установленными пропорциями, характерными для всей совокупности. Если основные покупатели жвачки – это молодежь от 18 до

25 лет, то оптимальным распределением будет такое, когда из этой группы отберется 50 % респондентов, а из других возрастных групп – по 10 %.

Стратифицированная выборка является более точной по сравнению с простой случайной. Метод отбора предпочтителен, когда генеральная совокупность очень неоднородна по качественному составу, когда случайный отбор неэффективен. Однако стратифицированная выборка может быть использована только при наличии дополнительной информации о всей совокупности либо статистик предыдущих исследований.

Кластерная выборка

Термин “кластер” произошел от английского слова *“cluster”,* означающего группу или кучку (поэтому иногда называется групповой выборкой). Кластерная выборка производит отбор респондентов из групп, объединенных по специфическим неповторяющимся признакам. Например, аптеки в городе – это кластер.

Группы отбираются на основе сформулированного исследователем принципа, а внутри выбранных кластеров производится сплошной или случайный опрос.

Кластерная выборка наиболее часто применяется в маркетинговых исследованиях, особенно в корпоративных. Она наиболее трудоемкая и наиболее эффективная. Имеет несколько вариаций, которые рассматриваются ниже.

Одноступенчатая кластерная выборка – сначала отбирается кластер, а дальше из этого кластера по методу простой случайной выборки отбираются респонденты.

Двухступенчатая кластерная выборка – отбор респондентов производится из выбранного кластера в два приема. Этот процесс аналогичен совершению международного звонка, когда сначала набирается код страны, потом код города, а уж потом номер абонента.

В практике маркетинговых исследований часто встречается разновидность кластерной выборки территориальная, построенная на принципах географической привязки к определенному региону, городу, республике, жилых кварталов, домов.

Преимущество кластерной выборки состоит в том, что она проще в организации, экономит средства и время. Но при этом необходимым требованием является достаточно большое количество элементов во всей совокупности, чтобы соблюсти формально принцип случайности. Для кластерной выборки требуются большие по размеру выборки по сравнению со случайными.

Детерминированные методы формирования выборки

Детерминированные методы формирования выборки, которые иногда называют неслучайными, невероятностными или предопределенными, представляют собой способы отбора, основанные на принципах, отличных от случайных.

На практике применение детерминированных методов может быть обусловлено следующими причинами:

1) невозможность проведения случайной выборки из-за:

– ограниченности ресурсов в широком смысле слова (денежных средств, времени, отсутствия полного списка генеральной совокупности и т.п.),

– этических проблем, связанных с конфиденциальностью получаемой информации или спецификой изучаемой тематики,

– категорического отказа от участия в опросе;

2) отсутствие жесткой необходимости проведения случайного опроса.

Данные методы являются альтернативой вероятностным методам. Существует большое количество разнообразных вариантов составления детерминированных выборок. В процентном соотношении в маркетинговых исследованиях такие выборки применяются значительно чаще, чем вероятностные, из-за простоты формирования и удобства.

Классификацию детерминированных методов составления выборки можно представить следующим образом.

Квотная выборка. Квотный метод выборки предполагает предварительное наличие некоторого объема сведений о характере генеральной совокупности. Однако эти сведения не используются для определения объема выборки, а отбор производится по определенным формальным признакам респондентов. В основе квот лежит суждение исследователя и его видение изучаемой группы потребителей. Точность исследования зависит во многом от выбора интервьюера, который производит отбор респондентов по заданным критериям.

Квотная выборка моделирует в малом масштабе всю генеральную совокупность. Предположим, изучается проблема продажи мебели, а из предыдущих исследований нам известно, что ключевыми факторами покупки новой мебели является принадлежность к некоторой социально-экономической прослойке и возраст покупателя. Тогда задание на отбор для опроса может быть сформулировано, чтобы выбирать в магазине только людей по выбранным характеристикам.

Систематическая выборка – повторяющаяся процедура выбора респондентов из списка на основе системы выбора. Изначально необходимо наличие некоего списка, из которого отбираются пронумерованные респонденты, например, через каждого десятого по списку. В этом способе существует формальная проблема, с какого номера начинать отсчет.

Строго говоря, систематическая выборка могла бы быть отнесена к вероятностному классу, если бы каждый элемент совокупности имел равную вероятность попасть в нее, а элементы в списке располагались бы совершенно случайно. Однако ни один список не может удовлетворять такому условию.

Систематическая выборка находит широкое применение в маркетинговых исследованиях из-за простоты реализации. В социологии на основе таких выборок строится статистика семейных бюджетов. По систематическому признаку отбираются для исследования населенные пункты, регионы, кварталы и дома. Способ прост и удобен, дает экономию времени.

Удобная выборка применяется в том случае, когда выбор респондентов ограничен и для исследования не представляется возможным определить формальные ограничения на отбор. Скажем, опрашиваются только те посетители выставки, которые обращаются на наш стенд и являются представителями компаний.

Предрешенная выборка – выборка, в которую отбирают только тех, кто удовлетворяет определенно сформулированным требованиям. Например, отбираются только те, кто является собственником коттеджа, или все те, кто зашел в магазин с 15.00 до 16.00 ч. Если маркетолога интересуют пользователи горных велосипедов, то есть два места, где их можно обнаружить: это ассоциация любителей данного вида спорта или места проведения соревнований.

“Снежный ком”. В некоторых случаях группы респондентов бывают настолько специфическими, что о составлении списков не может быть никакой речи, а поиск представителей, интересующих маркетологов, представляет серьезную проблему. Часто исследователям приходится использовать этот метод при работе с *VIP*-персонами, когда “с улицы” практически бесполезно к ним обращаться с вопросами. Тогда можно в кругу знакомых найти одного из таких представителей и просить рекомендации для обращения к их знакомым. Или другой случай: как найти людей, имеющих дома определенную марку принтеров? Метод “снежного кома” позволяет решить задачу, так как часто люди обмениваются со знакомыми особенностями работы с такой техникой, задают вопросы в определенных сложных ситуациях. Раскручивая такие цепочки знакомств, вполне возможно решить задачу проведения исследования.

Перечисленные способы формирования детерминированных выборок не исчерпывают весь возможный арсенал. Приведенные схемы являются основными.

В практике маркетинговых исследований много споров ведется о сути методов формирования выборок. К любому из них можно “придираться” по существу. Однако реальность такова, что как вероятностный подход, так и детерминированный имеют право на существование и признание исследователями в качестве адекватного исследовательского инструмента. На практике в исследовании обычно стремятся выбрать наиболее подходящий и экономичный способ формирования выборки, которая будет адекватно отражать конкретную генеральную совокупность и даст уверенность исследователю в точности получаемых выводов.

В табл. 5.30 приведен сравнительный анализ преимуществ и недостатков для перечисленных выше методов формирования выборок.

Таблица 5.30.

Сравнительный анализ методов выборки

Метод	**Преимущества**	**Недостатки**
Вероятностная выборка		
Простая случайная	Проста для понимания, результаты можно отнести к генеральной совокупности	Сложно построить основу для выборки, невысокая точность, большие затраты
Стратифицированная	Высокая точность, включает все важные страты	Сложность подбора переменных для стратификации, высокие затраты на проведение

Кластерная	Проста в применении, недорогая	Невысокая точность, сложно оценить результат
Детерминированная выборка		
Квотная	Может удобно регулироваться исследователем	Необъективность отбора, нет гарантий репрезентативности
Систематическая	Позволяет увеличить репрезентативность, проста в применении	Может снизить репрезентативность
Удобная, предрешенная	Низкая стоимость, удобство, быстрота	Нерепрезентативна
"Снежный ком"	Работает там, где другие неприменимы	Необъективность отбора, нерепрезентативна

В завершение раздела еще раз хотелось бы подчеркнуть, на что следует обращать внимание, используя шкалы при проведении измерений. Прежде всего необходимо соблюдать требование достоверности полученных результатов измерений. Лучший способ соблюдения этого требования – провести одно или несколько повторных измерений. При их сравнении будут видны все нестыковки, неточности ответов и неискренность в ответах опрашиваемых.

Однако практически невозможно измерить абсолютно точно то, что относится к маркетинговой деятельности. Причины этого могут быть очень разные. Перечислим следующие потенциальные источники ошибок в измерении:

1) характеристики индивида, влияющие на оценку (интеллект, уровень образования);

2) краткосрочные или временные факторы (здоровье, эмоции, усталость);

3) ситуационные факторы (присутствие других людей, шум и факторы, отвлекающие внимание);

4) форма шкалы, используемой в измерении;

5) неточность инструкций для заполнения анкеты;

6) технические факторы (перенасыщенность пунктами в анкете, неудачный дизайн анкеты);

7) работа интервьюеров и т.д.

Разложение дисперсии и его роль в статистическом анализе

Дисперсия в статистике находится как среднее квадратическое отклонение индивидуальных значений признака в квадрате от средней арифметической. В зависимости от исходных данных она определяется по формулам простой и взвешенной дисперсий:

1. Простая дисперсия (для несгруппированных данных) вычисляется по формуле: $\delta^2 = \overline{x^2} - \bar{x}^2$

2. Взвешенная дисперсия (для вариационного ряда): $\delta^2 = \frac{\sum (x-\bar{x})^2 n}{\sum n}$

где n - частота (повторяемость фактора X)

Свойство 1. Дисперсия постоянной величины равна нулю.

Свойство 2. Уменьшение всех значений признака на одну и ту же величину А не меняет величины дисперсии $\sigma^2_{(x-A)} = \sigma^2_x$. Значит, средний квадрат отклонений можно вычислить не по заданным значениям признака, а по отклонениям их от какого-либо постоянного числа.

Свойство 3. Уменьшение всех значений признака в K раз уменьшает дисперсию в K2 раз, а среднее квадратическое отклонение в K раз $\sigma^2_{(x/K)} = \sigma^2_x / K^2$. Значит, все значения признака можно разделить на какое-то постоянное число, например, на величину интервала ряда, исчислить среднее квадратическое отклонение, а затем умножить его на постоянное число: $\sigma_x = \sigma_{x/K} \cdot K$.

Свойство 4. Если вычислить средний квадрат отклонений от любой величины А, в той или иной степени отличающейся от средней арифметической (), то он всегда будет больше среднего квадрата отклонений, вычисленного от средней арифметической $\sigma^2_A > \sigma^2_x$. Средний квадрат отклонений при этом будет больше на величину (– А)2 : $\sigma^2_A = \sigma^2_x + (\bar{x}-A)^2$, или $\sigma^2_x = \frac{\sum (x_i - A)^2 \cdot f_i}{\sum f_i} - (\bar{x}-A)^2$.

Значит, дисперсия от средней величины всегда меньше дисперсий, вычисленных от любых других величин, т.е. она имеет свойство минимальности.

На этих математических свойствах дисперсии основываются способы, которые позволяют упростить ее вычисление. Например, расчет дисперсии по способу моментов или способу отсчета от условного нуля применяется в вариационных рядах с равными интервалами. Расчет производится по формуле: $\sigma^2 = \frac{\sum \left(\frac{x_i - A}{K}\right)^2 \cdot f_i}{\sum f_i} \cdot K^2 - (\bar{x} - A)^2$,

где K – ширина интервала;

А – условный нуль, в качестве которого удобно использовать середину интервала, обладающего наибольшей частотой;

$\frac{\sum\left(\frac{x_i - A}{K}\right)^2 \cdot f_i}{\sum f_i}$ – момент второго порядка.

Правило сложения дисперсии заключается в том, что общая дисперсия равна сумме межгрупповой и средней из внутригрупповых дисперсий

$\sigma^2 = \delta^2 + \overline{\sigma_i^2}$, где δ^2 — межгрупповая дисперсия; $\overline{\sigma_i^2}$ — средняя из внутригрупповых дисперсия.

Виды связей между социально-экономическими явлениями

Исследование объективно существующих связей между явлениями – важнейшая задача общей теории статистики. В процессе статистического исследования зависимостей вскрываются причинно-следственные отношения между явлениями, что позволяет выявлять факторы (признаки), оказывающие основное влияние на вариацию изучаемых явлений и процессов. Причинно-следственные отношения – это связь явлений и процессов, когда изменение одного из них – причины. Ведет к изменению другого – следствия.

Особое значение при исследовании причинно-следственных связей имеет выявление временной последовательности: причина должна всегда предшествовать следствии, однако не каждое предшествующее событие следует считать причиной, а последующее – следствием.

В реальной социально-экономической действительности причину и следствие необходимо рассматривать как смежные явления, появление которых обусловлено комплексом сопутствующих более простых причин и следствий. Между сложными группами причин и следствий возможны многозначные связи, когда за одной причиной будет следовать то одно, то другое действие или одно действие имеет несколько разных причин. Каждое явление может выступать в одних случаях как причина, а в других как следствие.

Но чем сложнее изучаемые явления, тем труднее выявить причинно-следственные связи между ними. Взаимное переплетение различных внутренних и внешних факторов неизбежно приводит к некоторым ошибкам в определении причины и следствия. Социально-экономические явления представляют собой результат одновременного воздействия большого числа причин. Поэтому при изучении этих явлений необходимо выявлять главные, основные причины, абстрагируясь от второстепенных.

На первом этапе статистического изучения связи проводят качественный анализ изучаемого явления, связанный с анализом природы социального или экономического явления при помощи экономической теории, социологии.

Второй этап – построение модели связи. Он базируется на методах статистики: группировках, средних величинах, таблицах и т.д. Третий, последний этап – интерпретация результатов – вновь связан с качественными особенностями изучаемого явления.

Статистика разработала множество методов изучения связей, выбор которых зависит от целей исследования и поставленных задач. Признаки по их значению для изучения взаимосвязей делятся на два класса. Признаки, обуславливающие изменения других, связанных с ними признаков, называются *факторными* или просто *факторами.* Признаки, изменяющиеся под действием факторных признаков, являются *результативными.*

Связи между явлениями и их признаками классифицируются по степени тесноты связи, направлению и аналитическому выражению.

В статистике различают функциональную связь и статистическую зависимость. *Функциональной* называют такую связь, при которой определенному значению факторного признака соответствует одно значение результативного признака. Функциональная связь проявляется во всех случаях наблюдения и для каждой единицы исследуемой совокупности.

Если причинная зависимость проявляется не в каждом отдельном случае, а в общем, среднем при большом числе наблюдений, то такая зависимость называется *статистической*. Частным случаем связи является *корреляционная* связь, при которой изменение среднего значения результативного признака обусловлено изменением факторных признаков.

По степени тесноты связи, в зависимости от величины коэффициента корреляции, различают следующие критерии оценки тесноты связи: связь практически отсутствует, слабая, существенная, тесная.

По направлению выделяют связь прямую и обратную. При *прямой связи* с увеличением или уменьшением значений факторного признака происходит увеличение или уменьшение значений результативного. Так, рост производительности труда способствует увеличению уровня рентабельности производства. В случае *обратной связи* значения результативного признака изменяются под воздействием факторного, но в противоположном направлении по сравнению с изменением последнего. Так, с увеличением уровня фондоотдачи снижается себестоимость единицы производимой продукции.

По аналитическому выражению выделяют связи *прямолинейные* (или просто *линейные*) и *криволинейные* (*нелинейные*). Если статистическая связь между явлениями может быть приближенно выражена уравнением прямой, то ее называют *линейной связью*; если же она выражается уравнением какой-либо кривой (параболы, гиперболы, степенной, показательной, экспоненциальной и т.д.), то такую связь называют *нелинейной* или *криволинейной.*

По количеству факторов, действующих на результативный признак, связи различают *однофакторные* (один фактор) и *многофакторные* (два и более факторов). Однофакторные (простые) связи обычно называют парными (так как рассматривается пара признаков). Например, корреляционная связь между прибылью и производительностью труда. В случае многофакторной (множественной) связи имеют ввиду, что все факторы действуют комплексно, т.е. одновременно и во взаимосвязи, например, корреляционная связь между производительностью труда и уровнем организации труда, автоматизации производства, квалификации рабочих, производственным стажем, простоями и другими факторными признаками.

2. Основные статистические методы выявления корреляционной связи

Для выявления наличия связи, ее характера и направления в статистике используются следующие методы: анализ параллельных рядов; аналитические группировки; графический метод; корреляционный и регрессионный анализ.

Метод сопоставления параллельных рядов основан на сопоставлении двух или нескольких рядов статистических величин. Такое сопоставление позволяет установить наличие связи и получить представление о ее характере и направлении. Для этого факторы, характеризующие результативный признак, располагают в возрастающем или убывающем порядке, а затем прослеживают изменение величины результативного признака. Сопоставление и анализ расположенных таким образом рядов значений изучаемых величин позволяют установить наличие связи и ее направление. Зависимость между факторами и показателями может прослеживаться во времени.

До исследования методом параллельных рядов необходимо провести анализ сопоставляемых явлений и установить наличие между ними причинных связей (а не просто сопутствия). Например, только потом, что между урожайностью и себестоимостью продукции сельского хозяйства имеется причинная связь, становится возможным сопоставление параллельных рядов этих показателей.

К недостатку метода взаимозависимых параллельных рядов следует отнести невозможность определения количественной меры связи между изучаемыми признаками. Однако он удобен и эффективен, когда речь идет о необходимости установления связей между показателями и факторами, характеризующими экономический процесс.

Графический метод. Графически взаимосвязь двух признаков изображается с помощью *поля корреляции.* В системе координат на оси абсцисс откладывают значения факторного признака, а на оси ординат – результативного. Каждое пересечение линий, проводимых через эти оси, обозначают точкой. При отсутствии тесных связей имеет место беспорядочное расположение точек на графике (рис. 1). Чем сильнее связь между признаками, тем теснее будут группироваться точки вокруг определенной линии, выражающей форму связи.

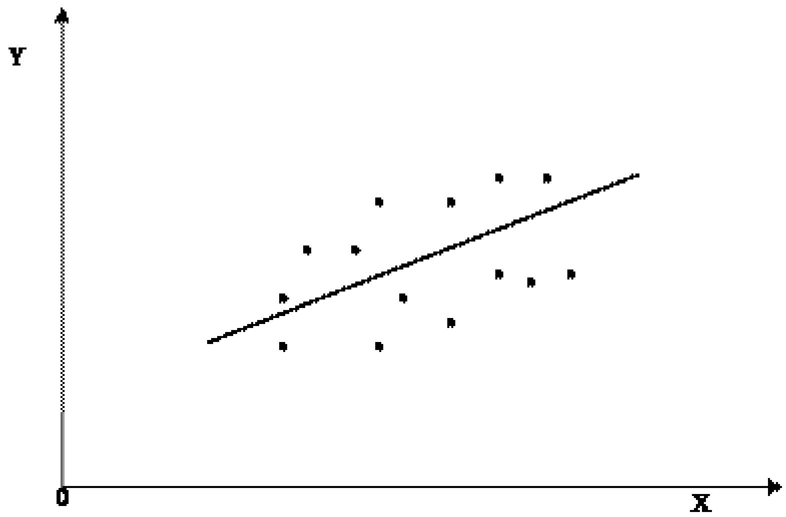

Рис.1 График корреляционного поля

Метод аналитических группировок. Статистическая связь будет проявляться отчетливее, если применить для ее изучения аналитические группировки. Чтобы выявить зависимость с помощью этого метода, нужно произвести группировку единиц совокупности по факторному признаку и для каждой группы среднее и относительное значение результативного признака. Сопоставляя затем изменения результативного признака по ере изменения факторного, можно выявить направление, характер и тесноту связи между ними с помощью эмпирического корреляционного отношения. Однако метод группировок не позволяет определить форму (аналитическое выражение) влияния факторных признаков на результативный.

Методы изучения связи не количественных переменных

Корреляционно-регрессионный метод применим только к количественным признакам. Однакозадача измерения связи ставится перед статистикой и по отношению к таким признакам, как пол,образование, занятие, семейное состояние человека, отрасль, форма собственности предприятия,т. е. признакам, не имеющим количественного выражения.

Учеными разных стран за последние сто лет разработано несколько методов измерения связейтаких признаков. Отметим прежде всего уже рассмотренный ранее коэффициент корреляциирангов Спирмена, применимый и к количественным, и неколичественным, но поддающимсяранжированию признакам. Так, например, можно при помощи одной группы экспертовпроранжировать кандидатов на занятие какой-либо должности по степени профессиональнойподготовленности, а другую гру ппу экспертов просить проранжировать тех же кандидатов поличностным и эти ческим качествам, а затем измерить связь между рангами.

Важным частным случаем задачи является измерение связи при альтернатив ной вариации двухпризнаков, один из которых имеет характер причины, а друг ой - следствия. Например, присоциологическом обследовании 1000 жителей гор ода были поставлены два вопроса:
1. Считаетели вы, что ваши доходы позволяют обеспечивать удовлетворение ос новных потребностей?
2.Удовлетворяет ли вас деятельность мэра города? Можно предположить, что п ричинойотрицательного ответа на второй вопрос у части населения является не удовлетворенность ихпотребностей доходами, т.е. имеется связь между ответам и на оба вопроса. Для измерения этойсвязи составляют двухмерное (дихотомич еское) распределение ответов 2х2, приведенное в табл. 8.15.

Таблица 8.15

Взаимосвязь между ответами на два вопроса социологического обследования

Ответы на 1-й вопрос		Ответы на 2-й вопрос		Итого
		Да, *а*	Нет, *в*	
Да	*A*	170	80	$\Sigma A = 250$
Нет	*B*	230	520	$\Sigma B = 750$
Итого		$\Sigma a = 400$	$\Sigma в = 600$	$N = 1000$

Если бы все, ответившие «да» на 1-й вопрос, отвечали бы «да» на 2-й вопрос, и так же совпадалиответы «нет», то связь была бы предельно тесной, функциональной. Но на самим делераспределение ответов на оба вопроса не со впадает. Большая часть ответивших «да» на 1-йвопрос ответила «да» и на 2-й вопрос, но часть ответила «нет». То же относится к ответившим«да» на 2-й вопрос. Связь есть, но неполная, типа корреляционной, и нужно измерить тес нотуэтой связи.

К. Пирсон предложил показатель, названный коэффициентом ассоциации. В числителе этогоотносительного показателя разность произведения чисел с оди наковыми ответами на обавопроса: да-да и нет-нет и произведения чисел с неодинаковыми ответами: да-нет И нет-да. Взнаменателе коэффициента ассоциации - корень квадратный из произведен ия всех четырехчастных итогов. В буквенных обозначениях по табл. 8.13 имеем:

$$K_{ассоциации} = \frac{Aa \cdot Bb - Ab \cdot Ba}{\sqrt{\sum A \sum B \sum a \sum b}}; \qquad (8.48)$$

$$K_{ассоциации} = \frac{170 \cdot 250 - 80 \cdot 230}{\sqrt{250 \cdot 270 \cdot 400 \cdot 600}} = 0{,}330$$

Свойства коэффициента ассоциации такие же, как и у коэффициента коррел яции: коэффициентассоциации обращается в нуль, если оба произведения в чис лителе точно уравновешиваются (чтокрайне маловероятно); он равен плюседин ице, если отсутствуют оба гетерогенных сочетания Ab иBa; равен минус единиц е, если отсутствуют гомогенные сочетания ответов Aa и Bb.

Другой метод измерения связи по «четырехклеточной таблице» предложен а нглийскимистатистиками Эдни Дж. Юлом (1871-1951) и Морисом Дж. Кендэлом (1907). Числитель этогокоэффициента, называе мого коэффициентам контингенции, совпадает с числителемкоэффициента ассо циации Пирсона, а в знаменателе - сумма тех же произведений, разностькоторы х стоит в числителе:

$$K_{контингенции} = \frac{Aa \cdot Bb - Ab \cdot Ba}{Aa \cdot Bb + Ab \cdot Ba}; \quad K_{контигенации} = \frac{170 \cdot 520 - 80 \cdot 230}{170 \cdot 520 + 80 \cdot 230} = 0{,}655$$

Как видим, коэффициент Юла-Кендэла значительно выше, чем коэффициент Пирсона. Крупныйнедостаток да нного коэффициента в том, что уже при равенстве нулю только одного из двухг етерогенных сочетаний - либо Ab, либо Ba коэффициент Юла - Кендэла обраща ется в единицу.Можно сказать, что этот показатель очень либерально оценивает тесноту связи, завышает ее.

Наконец, вполне возможно предложить показатель тесноты связи в форме о тношения избыткасуммы гомогенных сочетаний над их пропорциональной сум мой к предельно возможному избытку.

Для этого необходимо вначале вычислить, каковы были бы пропорциональн ые числа гомогенныхсочетаний Aa и Bb? Пропорциональные числа - это доли о т общей численности совокупности «N»,которые были бы получены при полно м отсутствии взаимосвязи группировок по двум признакам(ответам на два вопр оса), т. е. числа (ΣA·Σa:N) и (ΣB·Σb:N), составляющие по данным табл. 8.13:

$$Aa' = \frac{250 \cdot 400}{1000} = 100 \quad \text{и} \quad Bb' = \frac{750 \cdot 600}{1000} = 450.$$

При отсутствии связи на первой диагонали таблицы в сумме было бы 100 + 450 = 550 единицсовокупности, а на самом деле их 170 + 520 = 690. Избыток, образовавшийся ввиду прямой связимежду ответами, составил 69 0—550 = 140.

Предельно возможный избыток был бы в том случае, если бы не было гетер огенных сочетаний, т.е. Ab и Ba. Он составляет 140+80 + 230 = 450. Сам же показатель тесноты связи - отношениефактического излишка к пре дельному: 140 : 450 =0,311. Как видим, этот показатель близок ккоэффициенту ассоциации, но обла дает чрезвычайно логичной и ясной интерпретацией: связьсоставляет 0,311 или 31,1%, от предельно возможной функциональной. Этот показатель - аналогне к оэффициента корреляции, а коэффициента детерминации. Поэтому правомерно обозначитьего как R2 или ?2 . Он имеет вид:

$$\eta^2 = \frac{Aa + Bb - \left|Aa' + Bb'\right|}{N - (Aa + Bb)}, \qquad (8.49)$$

где

$$Aa' = \frac{\sum A \sum a}{N}; \qquad Bb' = \frac{\sum B \sum b}{N}.$$

Подставляя эти выражения в (8.49), получим:

$$\eta^2 = \frac{Aa + Bb - \frac{\sum A \sum a + \sum B \sum b}{N}}{N - \frac{\sum A \sum a + \sum B \sum b}{N}} = \frac{N(Aa + Bb) - (\sum A \sum a + \sum B \sum b)}{N^2 - (\sum A \sum a + \sum B \sum b)}.$$

(8.50)[1]

При наличии не двух, а более возможных значений каждого из взаимосвязан ных признаков такжеразработаны разные методы измерения тесноты связи.

Рассмотрим некоторые из этих мер на примере изучения влияния религиозн ой принадлежности наформирование супружеских пар. Воспользуемся данным и ФРГ, где такой учет ведется постоянно.Статистический ежегодник Федератив ной Республики Германии приводит распределениеживорожденных младенцев по религиозной принадлежности отца и матери. При этом выделены 5групп по религиозной принадлежности граждан: евангелическая (в России их чаще имен уютпротестантами); 2) римско-католическая; 3) прочие христиане (включая и православных); 4)других религий; 5) неверующие или не указавшие религиозную принадлежность (табл. 8.16).

Таблица 8.16
Распределение новорожденных в ФРГ по религиозной принадлежности отца и матери в 1993 г.
(тыс. чел.)

Религия отца	Религия матери					Итого $\sum f_i$
	Евангелическая	Римско-католическая	Прочие христиане	Другие религии	Неверующие и не указавшие	
Евангелическая	146,1	57,6	1,1	0,5	8,8	214,1
Римско-католическая	57,3	195,9	1,1	0,7	5,2	260,2
Прочие христиане	1,3	1,4	10,5	0,1	0,3	13,6
Другие религии	1,8	2,0	0,1	62,8	1,1	67,8
Неверующие и не указавшие	29,1	16,1	0,7	0,8	77,7	124,4
Итого $\sum f_i$	235,6	273,0	13,5	64,9	93,1	680,1

Источник: Statistisches Jahrbuch für die BRD. – 1995. – С. 74.

В табл. 8.16 представлена «решетка» 5 × 5, и все ее клетки не пусты: встречаются детные бракимежду лицами любых вероисповеданий. Но при этом наибольшие числа располагаются вдоль«главной диагонали», т. е. явно преобладают случаи, когда и отец и мать ребенка принадлежат к одной и той же религии. Такое явление ярче всего проявляется среди лиц «других религий» – магометан, иудеев, буддистов, индуистов. Среди них в 92,6% оба родителя относятся к той же вере. Среди лиц евангелического вероисповедания только 68,2% родительских пар относятся к той же вере.

Гипотеза о связи заключается в том, что существует предпочтение к заключению брака между лицами одинаковой религиозной принадлежности.

Соответствующие этой гипотезе частоты f_{ij} расположены вдоль первой диагонали таблицы (из верхнего левого угла в правый нижний).

Пирсон предложил для измерения связи в распределениях со множеством групп по обоим признакам показатель

$$C = \sqrt{\frac{\varphi^2}{1+\varphi^2}}, \qquad (8.51)$$

где

$$\varphi^2 = \left(\sum_{i=1}^{k_1}\sum_{j=1}^{k_2}\frac{f_{ij}^2}{f_i f_j}\right) - 1, \qquad (8.52)$$

где k_1 и k_2 – число групп по признакам I и II.

$$\varphi^2 = \frac{146{,}1^2}{214{,}1 \cdot 235{,}6} + \frac{57{,}6^2}{214{,}1 \cdot 273} + \ldots + \frac{77{,}7^2}{124{,}4 \cdot 93{,}1} - 1 = 2{,}1364;$$

$$C = \sqrt{\frac{2{,}1364}{3{,}1364}} = 0{,}825.$$

Недостаток коэффициента Пирсона в том, что он не достигает единицы и при полной связи признаков, а лишь стремится к единице при увеличении числа групп. Полезно поэтому провести корректировку коэффициента Пирсона, разделив его величину на предельно возможное значение, которое легко получается при подстановке в (8.52) значений $f_{ij} = f_i = f_j$, что имеет место при полной связи признаков. Имеем

$$C_{\max} = \sqrt{\frac{k-1}{k}}$$

и вычисляем значения $C_{\max}$ (табл. 8.17).

Таблица 8.17

Предельные значения коэффициента Пирсона

K	2	3	4	5	6	7	8	9	10
$C_{\max}$	0,707	0,816	0,866	0,894	0,913	0,926	0,935	0,943	0,949

В рассмотренной табл. 8.16 при 5 группах корректированный показатель связи Пирсона составит:

$$C_{корр} = 0{,}8253 : 0{,}894 = 0{,}923.$$

Русским статистиком А. А. Чупровым (1874 – 1926) предложен другой показатель связи T.

$$T = \sqrt{\frac{\varphi^2}{\sqrt{(k_1-1)(k_2-1)}}}. \tag{8.53}$$

Так как при полной связи $\varphi^2 = k - 1$, то при равенстве $k_1 = k_2$ показатель Чупрова будет равен единице. Для случая, когда $k_1 \neq k_2$, шведский математик и статистик Г. Крамер в 1946 г. предложил дополнение:

$$T_{kp} = \sqrt{\frac{\varphi^2}{k_{\min}-1}}. \tag{8.54}$$

С одной стороны, действительно, любая диагональ в таблице $k_1 \cdot k_2$ при $k_1 \neq k_2$ может содержать только $k_{\min}$ число клеток. Но, с другой стороны, вовсе не обязательно связывать гипотезу о полной связи признаков в случае $k_1 \neq k_2$ только с частотами на какой-либо диагонали. Случай этот сложен и подлежит рассмотрению только по каждой конкретной задаче. Далее поэтому рассматривается только случай, когда $k_1 = k_2$. По данным табл. 8.16 показатель Чупрова составил:

$$T = \sqrt{\frac{2{,}1364}{\sqrt{(5-1)\cdot(5-1)}}} = 0{,}7308.$$

Как правило, показатель Чупрова гораздо строже оценивает тесноту связи, чем показатель Пирсона, слишком быстро приближающийся к единице.

Модифицируя для любого числа групп ранее предложенный для двух групп способ и формулы (8.49) и (8.50) с учетом принятых в табл. 8.16 обозначений частот, получаем:

$$\eta^2 = \frac{\sum_{i=1}^{k} f_{ij}(i=j) - \sum_{i=1}^{k} f'_{ij}(i=j)}{\sum_{i=1}^{k}\sum_{j=1}^{k} f_{ij} - \sum f'_{ij}(i=j)}, \qquad (8.55)$$

где

$$f'_{ij}(i=j) = \frac{f_i \cdot f_j(i=j)}{\sum_{i=1}^{k}\sum_{j=1}^{k} f_{ij}},$$

т. е. частоты в клетках первой диагонали при отсутствии связи признаков. Подставив значения f'_{ij} в (8.55), получаем формулу, аналогичную (8.50):

$$\eta^2 = \frac{\left(\sum_{i=1}^{k}\sum_{j=1}^{k} f_{ij}\right)\sum_{i=1}^{k} f_{ij}(i=j) - \sum_{i=1}^{k} f_i \cdot f_j(i=j)}{\left(\sum_{i=1}^{k}\sum_{j=1}^{k} f_{ij}\right)^2 - \sum_{i=1}^{k} f_i f_j(i=j)}, \qquad (8.56)$$

где

$$f'_{ij}(i=j) = \frac{\sum f_i \cdot \sum f_j(i=j)}{\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} f_{ij}}.$$

т. е. $f'_{ij}(i{=}j)$ – это такие частоты в клетках первой диагонали таблицы, которые были бы при отсутствии связи: пропорциональные частоты.

По данным табл. 8.16 имеем:

$$\sum_{i=1}^{k_1}\sum_{j=1}^{k_2} f'_{ij}(i=j) =$$

$$= \frac{(214,1 \cdot 235,6) + (260,2 \cdot 273) + (13,6 \cdot 13,5) + (67,8 \cdot 64,9) + (124,4 \cdot 93,1)}{680,1} = 202,17 \qquad \sum_{i=1}^{k_1} \sum_{j=1}^{k_2} f'_{ij} (i = j) =$$

146,1+195,9+10,5+62,8+77,7=493,0 .

$$\eta^2 = \frac{493 - 202,17}{680,1 - 202,17} = 0,6085 \; ; \quad \eta = 0,780$$

Таким образом, за счет предпочтения браков между лицами одинаковых рел игий на главнуюдиагональ «собралось» 60,85% возможных родительских пар сверх равномерногораспределения: связь составила 60,85% предельно тесной. Итак, все способы измеренияпоказали, что влияние религии на формирование супружеских пар в ФРГ в 1993 году былозн ачительное.

Если кроме количественных факторов при многофакторном регрессионном анализе включается инеколичественный, то применяют следующую методику: наличие неколичественного фактора уединиц совокупности обозначают единиц ей, его отсутствие - нулем. Если таких факторов, илиградаций неколичественного фактора нескольк о, в уравнение регрессии вводятся несколько такназываемых «фиктивных пере менных», принимающих значения либо единицы, либо нуля.Например, пусть и меется три количественных фактора урожайности (x1, x2, x3) и три природныхзоны. В ЭВМ вводятся переменные в порядке их принадле жности к той или иной зоне (табл. 8.18).

Линейное уравнение регрессии будет иметь вид:

y? = a +b1x1 + b2x2 + b3x3 + b4u1 + b5u2 (8.57)

Величина коэффициента b4 означает, что все единицы II зоны при тех же зн аченияхколичественных факторов, как и единицы I зоны, будут в среднем имет ь значение y? на b4 больше(или меньше, если b4 < 0), чем единицы совокупности I зоны. Величина b5 озна-

чает то же для единиц совокупности III зоны. Иначе говоря, мы получаем сразу три зональных регрессионных модели:

I зона: $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 \quad (u_1 = 0;\ u_2 = 0);$

II зона: $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_4u_1 \quad (u_2 = 0);$

III зона: $\hat{y} = a + b_1x_1 + b_2x_2 + b_3x_3 + b_5u_2 \quad (u_1 = 0).$

Число фиктивных переменных должно быть на единицу меньше числа градаций качественного (неколичественного) фактора. С помощью данного приема можно измерять влияние уровня образования, местожительства, типа жилища и других социальных или природных не измеряемых количественно факторов, изолируя их от влияния количественных факторов.

Таблица 8.18

Зоны	Результативный признак y_i	Количественные факторы			Фиктивные переменные	
		x_1	x_2	x_3	u_1	u_2
I	y_1	x_{11}	x_{21}	x_{31}	0	0
	y_2	x_{12}	x_{22}	x_{32}	0	0
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	y_{n_1}	x_{n_1}	x_{2n_1}	x_{3n_1}	0	0
II	y_{n_1+1}	x_{1n_1+1}	x_{2n_1+1}	x_{3n_1+1}	1	0
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	$y_{n_1+n_2}$	$x_{1n_1+n_2}$	$x_{2n_1+n_2}$	$x_{3n_1+n_2}$	1	0
III	$y_{n_1+n_2+1}$	$x_{1n_1+n_2+1}$	$x_{2n_1+n_2+1}$	$x_{3n_1+n_2+1}$	0	1
	.	.	.	.	.	.
	.	.	.	.	.	.
	.	.	.	.	.	.
	$y_{n_1+n_2+n_3}$	$x_{1n_1+n_2+n_3}$	$x_{2n_1+n_2+n_3}$	$x_{3n_1+n_2+n_3}$	0	1

Анализ тенденции динамического ряда

Изменение социально-экономических явлений во времени изучается статистикой методом построения и анализа динамических рядов. **Ряды динамики** - это значения статистических показателей, которые представлены в определенной хронологической последовательности.

Каждый динамический ряд содержит две составляющие:

1) **показатели периодов времени** (годы, кварталы, месяцы, дни или даты);

2) **показатели, характеризующие исследуемый объект** за временные периоды или на соответствующие даты, которые называют **уровнями ряда**.

Уровни ряда выражаются как абсолютными, так и средними или относительными величинами. В зависимости от характера показателей строят динамические ряды абсолютных, относительных и средних величин. Ряды динамики из относительных и средних величин строят на основе производных рядов абсолютных величин. Различают интервальные и моментные ряды динамики.

Динамический интервальный ряд содержит значения показателей за определенные периоды времени. В интервальном ряду уровни можно суммировать, получая объем явления за более длительный период, или так называемые накопленные итоги.

Динамический моментный ряд отражает значения показателей на определенный момент времени (дату времени). В моментных рядах исследователя может интересовать только разность явлений, отражающая изменение уровня ряда между определенными датами, поскольку сумма уровней здесь не имеет реального содержания. Накопленные итоги здесь не рассчитываются.

Важнейшим условием правильного построения динамических рядов является **сопоставимость уровней рядов**, относящихся к различным периодам. Уровни должны быть представлены в однородных величинах, должна иметь место одинаковая полнота охвата различных частей явления.

Для того, чтобы избежать искажения реальной динамики, в статистическом исследовании проводятся предварительные расчеты (смыкание рядов динамики), которые предшествуют статистическому анализу динамических рядов. Под **смыканием рядов динамики** понимается объединение в один ряд двух и более рядов, уровни которых рассчитаны по разной методологии или не соответствуют территориальным границам и т.д. Смыкание рядов динамики может предполагать также приведение абсолютных уровней рядов динамики к общему основанию, что нивелирует несопоставимость уровней рядов динамики.

9.2.

Показатели изменений уровней динамических рядов

Для характеристики интенсивности развития во времени используются статистические показатели, получаемые сравнением уровней между собой, в результате чего получаем систему абсолютных и относительных показателей динамики: абсолютный прирост, коэффициент роста, темп роста, темп прироста, абсолютное значение 1% прироста. Для характеристики интенсивности развития за длительный период рассчитываются средние показатели: средний уровень ряда, средний абсолютный прирост, средний коэффициент роста, средний темп роста, средний темп прироста, среднее абсолютное значение 1% прироста.

Если в ходе исследования необходимо сравнить несколько последовательных уровней, то можно получить или сравнение с постоянной базой (базисные показатели), или сравнение с переменной базой (цепные показатели).

Базисные показатели характеризуют итоговый результат всех изменений в уровнях ряда от периода базисного уровня до данного (i-го) периода.

Цепные показатели характеризуют интенсивность изменения уровня от одного периода к другому в пределах того промежутка времени, который исследуется.

Абсолютный прирост выражает абсолютную скорость изменения ряда динамики и определяется как разность между данным уровнем и уровнем, принятым за базу сравнения.

Абсолютный прирост (базисный)

$\Delta_{i(Б)} = y_i - y_0$, (9.1)

где y_i - уровень сравниваемого периода; y_0 - уровень базисного периода.

Абсолютный прирост с переменной базой (цепной), который называют скоростью роста,

$\Delta_{i(Ц)} = y_i - y_{i-1}$, (9.2)

где y_i - уровень сравниваемого периода; y_{i-1} - уровень предшествующего периода.

Коэффициент роста K_i определяется как отношение данного уровня к предыдущему или базисному, показывает относительную скорость изменения ряда. Если коэффициент роста выражается в процентах, то его называют темпом роста.

Коэффициент роста базисный

$K_{i(Б)} = \frac{y_i}{y_0}$. (9.3)

Коэффициент роста цепной

$K_{i(Ц)} = \frac{y_i}{y_{i-1}}$. (9.4)

Темп роста

$T_р = K \cdot 100\%$. (9.5)

Темп прироста $T_П$ определяется как отношение абсолютного прироста данного уровня к предыдущему или базисному.

Темп прироста базисный

$T_{п(Б)} = \frac{y_i - y_0}{y_0} \cdot 100\%$. (9.6)

Темп прироста цепной

$T_{п(Ц)} = \frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%$. (9.7)

Темп прироста можно рассчитать и иным путем: как разность между темпом роста и 100 % или как разность между коэффициентом роста и 1 (единицей):

1) $T_п = T_р - 100\%$; 2) $T_п = K_i - 1$. (9.8)

Абсолютное значение одного процента прироста A_i . Этот показатель служит косвенной мерой базисного уровня. Представляет собой одну сотую часть базисного уровня, но одновременно представляет собой и отношение абсолютного прироста к соответствующему темпу роста.

Данный показатель рассчитывают по формуле

$$A_i = \frac{y_i - y_{i-1}}{T_{\text{п}i/(i-1)}} = \frac{y_i - y_{i-1}}{\frac{y_i - y_{i-1}}{y_{i-1}} \cdot 100\%} = \frac{y_{i-1}}{100} = 0{,}01 y_{i-1} \,. \qquad (9.9)$$

Для характеристики динамики изучаемого явления за продолжительный период рассчитывают группу средних показателей динамики. Можно выделить две категории показателей в этой группе: а) средние уровни ряда; б) средние показатели изменения уровней ряда.

Средние уровни ряда рассчитываются в зависимости от вида временного ряда.

Для интервального ряда динамики абсолютных показателей средний уровень ряда рассчитывается по формуле простой средней арифметической:

$$\bar{y} = \frac{\sum y}{n}, \qquad (9.10)$$

где n - число уровней ряда.

Для моментного динамического ряда средний уровень определяется следующим образом.

Средний уровень моментного ряда с равными интервалами рассчитывается по формуле средней хронологической:

$$\bar{y} = \frac{\frac{y_1}{2} + y_2 + y_3 + \ldots + y_{n-1} + \frac{y_n}{2}}{n-1}, \qquad (9.11)$$

где n - число дат.

Средний уровень моментного ряда с неравными интервалами рассчитывается по формуле средней арифметической взвешенной, где в качестве весов берется продолжительность промежутков времени между временными моментами изменений в уровнях динамического ряда:

$$\bar{y} = \frac{\sum y \cdot t}{\sum t}, \qquad (9.12)$$

где t - продолжительность периода (дни, месяцы), в течение которого уровень не изменялся.

Средний абсолютный прирост (средняя скорость роста) определяется как средняя арифметическая из показателей скорости роста за отдельные периоды времени:

$$\bar{\Delta} = \frac{\sum \Delta}{n-1}, \text{ или } \bar{\Delta} = \frac{y_n - y_1}{n-1}, \qquad (9.13)$$

где y_n - конечный уровень ряда; y_1 - начальный уровень ряда.

Средний коэффициент роста ($\bar{K}_\text{р}$) рассчитывается по формуле средней геометрической из показателей коэффициентов роста за отдельные периоды:

$$\bar{K}_\text{р} = \sqrt[n-1]{K_{\text{р}1} K_{\text{р}2} \ldots K_{\text{р}\,n-1}}, \qquad (9.14)$$

где $К_{р1}$, $К_{р2}$, ..., $К_{р\ n-1}$ - коэффициенты роста по сравнению с предыдущим периодом; n - число уровней ряда.

Средний коэффициент роста можно определить иначе:

$$\bar{K}_р = \sqrt[n-1]{\frac{Y_n}{Y_1}} . \qquad (9.15)$$

Средний темп роста, %. Это средний коэффициент роста, который выражается в процентах:

$$\bar{T}_р = \bar{K}_р \cdot 100 . \qquad (9.16)$$

Средний темп прироста $\bar{T}_р$, %. Для расчета данного показателя первоначально определяется средний темп роста, который затем уменьшается на 100%. Его также можно определить, если уменьшить средний коэффициент роста на единицу:

$$\bar{T}_п = \bar{T}_р - 100; \quad \bar{T}_п = \left(\bar{K}_р - 1\right) \cdot 100 . \qquad (9.17)$$

Среднее абсолютное значение 1% прироста можно рассчитать по формуле

$$\bar{A} = \frac{\bar{\Delta}}{\bar{T}_п} . \qquad (9.18)$$

9.3.
Способы обработки динамического ряда

В ходе обработки динамического ряда важнейшей задачей является выявление основной тенденции развития явления (тренда) и сглаживание случайных колебаний. Для решения этой задачи в статистике существуют особые способы, которые называют методами выравнивания.

Выделяют три основных способа обработки динамического ряда:

а) укрупнение интервалов динамического ряда и расчет средних для каждого укрупненного интервала;

б) метод скользящей средней;

в) аналитическое выравнивание (выравнивание по аналитическим формулам).

Укрупнение интервалов - наиболее простой способ. Он заключается в преобразовании первоначальных рядов динамики в более крупные по продолжительности временных периодов, что позволяет более четко выявить действие основной тенденции (основных факторов) изменения уровней.

По интервальным рядам итоги исчисляются путем простого суммирования уровней первоначальных рядов. Для других случаев рассчитывают средние величины укрупненных рядов (**переменная средняя**). Переменная средняя рассчитывается по формулам простой средней арифметической.

Скользящая средняя - это такая динамическая средняя, которая последовательно рассчитывается при передвижении на один интервал при заданной продолжительности периода. Если, предположим, продолжительность периода равна 3, то скользящие средние рассчитываются следующим образом:

$$\bar{y}_1 = \frac{y_1 + y_2 + y_3}{3} \; ; \; \bar{y}_2 = \frac{y_2 + y_3 + y_4}{3} \; ; \; \bar{y}_3 = \frac{y_3 + y_4 + y_5}{3} \text{ и т.д.} \qquad (9.19)$$

При четных периодах скользящей средней можно центрировать данные, т.е. определять среднюю из найденных средних. К примеру, если скользящая исчисляется с продолжительностью периода, равной 2, то центрированные средние можно определить так:

$$\bar{y}_1^1 = \frac{\bar{y}_1 + \bar{y}_2}{2}\ ;\ \bar{y}_2^1 = \frac{\bar{y}_2 + \bar{y}_3}{2}\ ;\ \bar{y}_3^1 = \frac{\bar{y}_3 + \bar{y}_4}{2} \text{ и т.д.} \qquad (9.20)$$

Первую рассчитанную центрированную относят ко второму периоду, вторую - к третьему, третью - к четвертому и т.д. По сравнению с фактическим сглаженный ряд становится короче на (m - 1)/2, где m - число уровней интервала.

Важнейшим способом количественного выражения общей тенденции изменения уровней динамического ряда является **аналитическое выравнивание ряда динамики**, которое позволяет получить описание плавной линии развития ряда. При этом эмпирические уровни заменяются уровнями, которые рассчитываются на основе определенной кривой, где уравнение рассматривается как функция времени. Вид уравнения зависит от конкретного характера динамики развития. Его можно определить как теоретически, так и практически. Теоретический анализ основывается на рассчитанных показателях динамики. Практический анализ - на исследовании линейной диаграммы.

Задачей аналитического выравнивания является определение не только общей тенденции развития явления, но и некоторых недостающих значений как внутри периода, так и за его пределами. Способ определения неизвестных значений внутри динамического ряда называют интерполяцией. Эти неизвестные значения можно определить:

1) используя полусумму уровней, расположенных рядом с интерполируемыми;

2) по среднему абсолютному приросту;

3) по темпу роста.

Способ определения количественных значений за пределами ряда называют **экстраполяцией**. Экстраполирование используется для прогнозирования тех факторов, которые не только в прошлом и настоящем обусловливают развитие явления, но и могут оказать влияние на его развитие в будущем.

Экстраполировать можно по средней арифметической, по среднему абсолютному приросту, по среднему темпу роста.

При аналитическом выравнивании может иметь место **автокорреляция**, под которой понимается зависимость между соседними членами динамического ряда. Автокорреляцию можно установить с помощью перемещения уровня на одну дату. Коэффициент автокорреляции вычисляется по формуле

$$r_a = \frac{\overline{y_t y_{t-1}} - \bar{y}_t \cdot \bar{y}_{t-1}}{\sigma_{y_t} \sigma_{y_{t-1}}}\ . \qquad (9.21)$$

Автокорреляцию в рядах можно устранить, коррелируя не сами уровни, а так называемые остаточные величины (разность эмпирических и теоретических уровней). В этом случае корреляцию между остаточными величинами можно определить по формуле

$$r = \frac{\sum(x-\bar{x}_t)(y-\bar{y}_t)}{\sqrt{\sum(x-\bar{x}_t)^2\sum(y-\bar{y}_t)^2}} . \quad (9.22)$$

Анализ рядов динамики предполагает и исследование **сезонной неравномерности** (**сезонных колебаний**), под которыми понимают устойчивые внутригодовые колебания, причиной которых являются многочисленные факторы, в том числе и природно-климатические. Сезонные колебания измеряются с помощью **индексов сезонности**, которые рассчитываются двумя способами в зависимости от характера динамического развития.

При относительно неизменном годовом уровне явления **индекс сезонности** можно рассчитать как процентное отношение средней величины из фактических уровней одноименных месяцев к общему среднему уровню за исследуемый период:

$$И_с = \frac{\bar{y}_i}{\bar{y}_0} \cdot 100 . \quad (9.23)$$

В условиях изменчивости годового уровня индекс сезонности определяется как процентное отношение средней величины из фактических уровней одноименных месяцев к средней величине из выровненных уровней одноименных месяцев:

$$И_с = \frac{\bar{y}_i}{\hat{y}_i} \cdot 100 . \quad (9.24)$$

Особенности изучения взаимосвязи рядов динамики

Временной ряд (он же ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Показателя временного ряда называются уровнями ряда динамики. Каждый уровень ряда динамики формируется под воздействием целого комплекса факторов.

Реальные данные временного ряда могут складываться при одновременном влиянии трех компонент. Итак, факторы уровней временного ряда по характеру воздействия можно условно разбить на три группы:

4) факторы, формирующие тенденцию ряда (T);

5) факторы, формирующие циклические колебания ряда (S);

6) случайные факторы (E).

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма компонент, называется *аддитивной.* Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной.*

Изучение взаимосвязи экономических переменных по данным временных рядов осложнено тем, что в этих рядах может быть тенденция. Если в ряду динамики переменной y и в ряду динамики x есть компонента «Т», то в результате мы получим тесную связь между y и x. Однако из этого факта еще нельзя делать вывод о том, что изменение x есть причина изменения y, то есть что между этими изменениями есть причинно-следственная связь.

Например, за последние 10 – 15 лет в Российской Федерации сократилось поголовье КРС и увеличилось число крестьянских (фермерских) хозяйств. Коэффициент корреляции между уровнями этих рядов динамики высок по величине; знак указывает на обратную связь. Однако это не означает, что рост численности фермерских хозяйств явился фактором снижения поголовья. Чтобы выявить причинно-следственную зависимость между переменными, необходимо устранить *ложную корреляцию* между ними, вызванную наличием тенденции.

Существует несколько способов исключения тенденции в рядах динамики. Первый способ называется *метод отклонений от тренда.* Пусть имеется $y_t = T + e$ и $x_t = T + e$. Проводится аналитическое выравнивание каждого ряда: $y_t = T_y$ и $x_t = T_x$, где T_y и T_x – это оценки трендовых компонент. Затем определяется остаток в каждом наблюдении $y_t - \tilde{y}_t = e_y$ и $x_t - \tilde{x}_t = e_x$, так как остаточная компонента не содержит тенденции. Далее изучается зависимость между самими остатками $e_y = f(e_x)$. Если между переменными есть связь, то она проявится в согласованном изменении остатков. Недостатком данного способа является то, что содержательная интерпретация параметров такой модели затруднительна. Однако модель может быть использована для прогнозов и, кроме того, коэффициент парной корреляции между остатками отразит связь переменных.

Второй способ преодоления тенденции в рядах динамики – это *метод последовательных разностей.* Если временной ряд содержит ярко выраженную линейную тенденцию, то для ее устранения можно заменить исходные уровни разностями первого порядка, то есть цепными абсолютными приростами: $\Delta_y = y_t - y_{t-1}$ и $\Delta_x = x_t - x_{t-1}$.

Далее прирост y рассматривается как функция прироста x: $\Delta_y = f(\Delta_x)$.

Недостатком второго способа является потеря информации (приростов на единицу меньше, чем уровней), что в условиях малого числа наблюдений крайне нежелательно. Достоинством является возможность интерпретации параметров. Коэффициент регрессии b покажет изменение прироста результата при единичном изменении прироста фактора.

Третьим способом является *включение в модель регрессии фактора времени*: $y_t = a + b_1 x_1 + b_2 t$. В данном случае коэффициенты чистой регрессии легко интерпретируются, имеют естественные единицы измерения. Коэффициент b_1покажет на сколько единиц изменится результат при единичном изменении фактора при условии существования неизменной тенденции; коэффициент b_2 отразит влияние всех прочих факторов, формирующих тенденцию, кроме x_1. Однако данный способ построения регрессионной модели требует большего объема наблюдений, так как в модели появляется еще один параметр.

Если тренды признаков являются экспонентами (или показательными функциями), то вместо корреляции абсолютных отклонений от трендов можно применить метод корреляции цепных темпов роста уровней, поскольку именно темпы роста – основной параметр экспоненциальных и показательных трендов.

Законы распределения выборочных характеристик в нормальной генеральной совокупности

Определение статистической оценки

Пусть требуется изучить количественный признак генеральной совокупности. Допустим, что из теоретических соображений удалось установить, какое именно распределение имеет признак. Возникает задача оценки параметров, которыми определяется это распределение. Например, если известно, что изучаемый признак распределен в генеральной совокупности по нормальному закону, то необходимо оценить математическое ожидание и среднеквадратическое отклонение, так как эти два параметра полностью определяют нормальное распределение. Если имеются основания считать, что признак имеет распределение Пуассона, то необходимо оценить параметр λ, которым это распределение определяется. Обычно имеются лишь данные выборки, полученные в результате n наблюдений: $x_1, x_2, \ldots, x_n$. Через эти данные и выражают оцениваемый параметр. Рассматривая $x_1, x_2, \ldots, x_n$ как значения независимых случайных величин $X_1, X_2, \ldots, X_n$ можно сказать, что найти статистическую оценку неизвестного параметра теоретического распределения означает найти функцию от наблюдаемых случайных величин, которая и дает приближенное значение оцениваемого параметра.

Точечные статистические оценки

Статистической оценкой неизвестного параметра теоретического распределения называют функцию от наблюдаемых случайных величин. Статистическая оценка неизвестного параметра генеральной совокупности одним числом называется ***точечной***. Рассмотрим следующие **точечные оценки**: смещенные и несмещенные, эффективные и состоятельные.

Для того чтобы статистические оценки давали хорошие приближения оцениваемых параметров, они должны удовлетворять определенным требованиям. Укажем эти требования. Пусть Θ^* есть статистическая оценка неизвестного параметра Θ теоретического распределения. Допустим, что по выборке объема n найдена оценка Θ_1^*. Повторим опыт, т. е. извлечем из генеральной совокупности другую выборку того же объема и по ее данным найдем оценку Θ_2^* и т. д. Получим числа $\Theta_1^*, \Theta_2^*, \ldots, \Theta_k^*$, которые будут различаться. Таким образом, оценку Θ^* можно рассматривать как случайную величину, а числа $\Theta_1^*, \Theta_2^*, \ldots, \Theta_k^*$ — как возможные ее значения.

Если оценка Θ^* дает приближенное значение Θ с избытком, то найденное по данным выборок число Θ $(k = 1, 2, \ldots, n)$ будет больше истинного значения Θ. Следовательно, и математическое ожидание (среднее значение) случайной величины Θ^* будет превышать Θ, то есть $M(\Theta^*) > \Theta$. Если Θдает приближенное значение Θ с недостатком, то $M(\Theta^*) < \Theta$.

Использование статистической оценки, математическое ожидание которой не равно оцениваемому параметру, приводит к систематическим ошибкам. Поэтому нужно потребовать, чтобы математическое ожидание оценки Θ было равно оцениваемому параметру. Соблюдение требования $M(\Theta^*) = \Theta$устраняет систематические ошибки.

Несмещенной называют статистическую оценку Θ^*, математическое ожидание которой равно оцениваемому параметру Θ, то есть $M(\Theta^*) = \Theta$.

Смещенной называют статистическую оценку Θ^*, математическое ожидание которой не равно оцениваемому параметру.

Однако ошибочно считать, что несмещенная оценка всегда дает хорошее приближение оцениваемого параметра. Действительно, возможные значения Θ^* могут быть сильно рассеяны вокруг своего среднего значения, т. е. дисперсия величины Θ^* может быть значительной. В этом случае найденная по данным одной выборки оценка, например Θ^*, может оказаться удаленной от своего среднего значения $\overline{\Theta^*}$, а значит, и от самого оцениваемого параметра Θ. Приняв Θ_1^* в качестве приближенного значения Θ, мы допустили бы ошибку. Если потребовать, чтобы дисперсия величины Θ^* была малой, то возможность допустить ошибку будет исключена. Поэтому к статистической оценке предъявляются требования эффективности.

Эффективной называют статистическую оценку, которая (при заданном объеме выборки n) имеет наименьшую возможную дисперсию. При рассмотрении выборок большого объема к статистическим оценкам предъявляется требование состоятельности.

Состоятельной называют статистическую оценку, которая при $n \to \infty$ стремится по вероятности к оцениваемому параметру. Например, если дисперсия несмещенной оценки при $n \to \infty$ стремится к нулю, то такая оценка оказывается также состоятельной.

Рассмотрим вопрос о том, какие выборочные характеристики лучше всего в смысле несмещённости, эффективности и состоятельности оценивают генеральную среднюю и дисперсию.

Пусть изучается дискретная генеральная совокупность относительно количественного признака. **Генеральной средней** называется среднее арифметическое значений признака генеральной совокупности. Она вычисляется по формуле

$$\overline{x}_{\text{g}} = \frac{1}{N}\sum_{i=1}^{N} x_i \quad \text{или} \quad \overline{x}_{\text{g}} = \frac{1}{N}\sum_{i=1}^{k} x_i m_i$$

где x_i — значения признака генеральной совокупности объема N; m_i — соответствующие частоты, причем

$$\sum_{i=1}^{k} m_i = N.$$

Пусть из генеральной совокупности в результате независимых наблюдений над количественным признаком извлечена выборка объема n со значениями признака $x_1, x_2, \ldots, x_n$. **Выборочной средней** называется среднее арифметическое значений признака выборочной совокупности и вычисляется по формуле

$$\overline{x}_{\mathrm{v}} = \frac{1}{n}\sum_{i=1}^{n} x_i \quad \text{или} \quad \overline{x}_{\mathrm{v}} = \frac{1}{n}\sum_{i=1}^{k} x_i m_i$$

где x_i — значения, признака в выборочной совокупности объема n; m_i — соответствующие частоты, причем

$$\sum_{i=1}^{k} m_i = n.$$

Если генеральная средняя неизвестна и требуется оценить ее по данным выборки, то в качестве оценки генеральной средней принимают выборочную среднюю, которая является несмещенной и состоятельной оценкой. Отсюда следует, что если по нескольким выборкам достаточно большого объема из одной и той же генеральной совокупности будут найдены выборочные средние, то они будут приближенно равны между собой. В этом состоит свойство **устойчивости выборочных средних**.

Если дисперсии двух совокупностей одинаковы, то близость выборочных средних к генеральным не зависит от отношения объема выборки к объему генеральной совокупности. Она зависит- от объема выборки: чем больше объем выборки, тем меньше выборочная средняя отличается от генеральной.

Для того чтобы охарактеризовать рассеяние значений количественного признака X генеральной совокупности вокруг своего среднего значения, вводят сводную характеристику — генеральную дисперсию. **Генеральной дисперсией** D_{g} называется среднее арифметическое квадратов отклонений значений признака генеральной совокупности от их среднего значения $\overline{x}_g$, которое вычисляется по формуле

$$D_g = \frac{1}{N}\sum_{i=1}^{N}(x_i - \overline{x}_g)^2 \quad \text{или} \quad D_g = \frac{1}{N}\sum_{i=1}^{k}(x_i - \overline{x}_g)^2 m_i$$

Для того чтобы охарактеризовать рассеяние наблюденных значений количественного признака выборки вокруг своего среднего значения хв, вводят сводную характеристику — выборочную дисперсию. **Выборочной дисперсией** D_v называется среднее арифметическое квадратов отклонений наблюденных значений признака от их среднего значения $\overline{x}_v$, которое вычисляется по формуле

$$D_v = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x}_v)^2 \quad \text{или} \quad D_v = \frac{1}{n}\sum_{i=1}^{k}(x_i - \overline{x}_v)^2 m_i$$

Кроме дисперсии для характеристики рассеяния значений признака генеральной (выборочной) совокупности вокруг своего среднего значения используют сводную характеристику — среднее квадратическое отклонение. **Генеральным средним квадратическим отклонением** называют квадратный корень из генеральной дисперсии: $\sigma_g = \sqrt{D_g}$. **Выборочным средним квадратическим отклонением** называют квадратный корень из выборочной дисперсии: $\sigma_v = \sqrt{D_v}$.

Пусть из генеральной совокупности в результате n независимых наблюдений над количественным признаком X извлечена выборка объема n. Требуется по данным выборки оценить неизвестную генеральную дисперсию D_g. Если в качестве оценки генеральной дисперсии принять выборочную дисперсию, то эта оценка приведет к систематическим ошибкам, давая заниженное значение генеральной дисперсии. Объясняется это тем, что выборочная дисперсия является смещенной оценкой D_g. Другими словами, математическое ожидание выборочной дисперсии не равно оцениваемой генеральной дисперсии, а равно $M(D_v) = \frac{n-1}{n}D_g$.

Легко исправить выборочную дисперсию так, чтобы ее математическое ожидание было равно генеральной дисперсии. Для этого нужно умножить D_v на дробь $\frac{n}{n-1}$. В результате получим исправленную дисперсию s^2, которая будет несмещенной оценкой генеральной дисперсии:

$$s^2 = \frac{1}{n-1}\sum_{i=1}^{k}(x_i - \overline{x}_{\text{v}})^2 m_i$$

Интервальные оценки

Наряду с точечным оцениванием, статистическая теория оценивания параметров занимается вопросами интервального оценивания. Задачу интервального оценивания можно сформулировать так: по данным выборки построить числовой интервал, относительно которого с заранее выбранной вероятностью можно сказать, что внутри него находится оцениваемый параметр. Интервальное оценивание особенно необходимо при малом количестве наблюдений, когда точечная оценка малонадежна.

Доверительным интервалом $\left(\tilde{\Theta}_n^{(1)}; \tilde{\Theta}_n^{(2)}\right)$ для параметра Θ называется такой интервал, относительно которого с заранее выбранной вероятностью $p = 1 - \alpha$, близкой к единице, можно утверждать, что он содержит неизвестное значение параметра Θ, то есть $P\left\{\tilde{\Theta}_n^{(1)} < \Theta < \tilde{\Theta}_n^{(2)}\right\} = 1 - \alpha$. Чем меньше для выбранной вероятности число $\left|\tilde{\Theta}_n^{(1)} - \tilde{\Theta}_n^{(2)}\right|$, тем точнее оценка неизвестного параметра Θ. И, наоборот, если это число велико, то оценка, проведенная с помощью данного интервала, малопригодна для практики. Так как концы доверительного интервала зависят от элементов выборки, то значения $\tilde{\Theta}_n^{(1)}$ и $\tilde{\Theta}_n^{(2)}$ могут изменяться от выборки к выборке. Вероятность $p = 1 - \alpha$ принято называть **доверительной** (надежностью). Обычно надежность оценки задается наперед, причем в качестве p берут число, близкое к единице. Выбор доверительной вероятности не является математической задачей, а определяется конкретной решаемой проблемой. Наиболее часто задают надежность, равную 0,95; 0,99; 0,999.

Доверительный интервал для генеральной средней при известном значении среднего квадратического отклонения и при условии, что случайная величина (количественный признак X) распределена нормально, задается выражением

где p — наперед заданное число, близкое к единице, а значения функции $\Phi(t)$ приведены в таблице прил. 2.

Смысл этого соотношения заключается в следующем: с надежностью p можно утверждать, что доверительный интервал покрывает неизвестный параметр $\overline{x}_{\text{g}}$, точность оценки $\delta = \frac{t\sigma}{\sqrt{n}}$. Число t определяется из равенства $2\Phi(t) = p$, или $\Phi = \frac{p}{2}$. По прил. 2 находят аргументt, которому соответствует значение функции Лапласа, равное $\frac{p}{2}$.

Пример 1. Случайная величина X имеет нормальное распределение с известным средним квадратическим отклонением $\sigma = 3$. Найти доверительные интервалы для оценки неизвестной генеральной средней по выборочным средним, если объем выборок $n = 36$ и надежность оценки $p = 0{,}95$.

Решение. Найдем t. Из соотношения $2\Phi(t) = 0{,}95$ получим, что $\Phi = 0{,}475$. По прил. 2 находим $t = 1{,}96$. Найдем точность оценки $\delta = \frac{t\sigma}{\sqrt{n}} = \frac{1{,}96 \cdot 3}{\sqrt{36}} = 0{,}98$. Доверительные интервалы будут таковы: $(\overline{x}_{\text{v}} - 0{,}98; \overline{x}_{\text{v}} + 0{,}98)$. Например, если $\overline{x}_{\text{v}} = 4{,}1$, то доверительный интервал имеет следующие доверительные границы: $\overline{x}_{\text{v}} - 0{,}98 = 4{,}1 - 0{,}98 = 3{,}12$; $\overline{x}_{\text{v}} + 0{,}98 = 4{,}1 + 0{,}98 = 5{,}08$;. Таким образом, значения неизвестного параметра $\overline{x}_{\text{g}}$, согласующиеся с данными выборки, удовлетворяют неравенству $3{,}12 < \overline{x}_{\text{g}} < 5{,}08$.

Доверительный интервал для генеральной средней нормального распределения признака при неизвестном значении среднего квадратического отклонения задается выражением

Отсюда следует, что с надежностью p можно утверждать, что доверительный интервал покрывает неизвестный параметр $\overline{x}_{\text{g}}$.

Существуют таблицы (прил. 4), пользуясь которыми, по заданным tp и n находят вероятность p и, наоборот, по заданным p и n находят tp.

Пример 2. Количественный признак X генеральной совокупности распределен нормально. По выборке объема $n = 16$ найдены выборочная средняя $\overline{x}_{\text{v}} = 20{,}2$ и исправленное среднеквадратическое отклонение $s = 0{,}8$. Оценить неизвестную генеральную среднюю с помощью доверительного интервала с надежностью $p = 0{,}95$.

Решение. Найдем tp. Пользуясь прил. 4 по $p = 0{,}95$ и $n = 16$ находим $tp = 2{,}13$. Найдем доверительные границы:

Итак, с надежностью $p = 0{,}95$ неизвестный параметр $\overline{x}_{\text{g}}$ заключен в доверительном интервале $19{,}774 < \overline{x}_{\text{g}} < 20{,}626$.

Основные способы организации выборки

Достоверность статистических выводов и содержательная интерпретация результатов зависит от **репрезентативности**выборки, т.е. полноты и адекватности представления свойств генеральной совокупности, по отношению к которой эту выборку можно считать представительной. Изучение статистических свойств совокупности можно организовать двумя способами: с помощью **сплошного** и **несплошного** наблюдения **. Сплошное наблюдение** предусматривает обследование всех **единиц** изучаемой **совокупности**, а **несплошное (выборочное) наблюдение** — только его части.

Существуют пять основных способов организации выборочного наблюдения:

1. **простой случайный отбор**, при котором n объектов случайно извлекаются из генеральной совокупности N объектов (например с помощью таблицы или датчика случайных чисел), причем каждая из возможных выборок имеют равную вероятность. Такие выборки называются **собственно-случайными**;

2. **простой отбор с помощью регулярной процедуры** осуществляется с помощью механической составляющей (например, даты, дня недели, номера квартиры, буквы алфавита и др.) и полученные таким способом выборки называются **механическими**;

3. **стратифицированный** отбор заключается в том, что генеральная совокупность объема N подразделяется на подсовокупности или слои (страты) объема $N_1, N_2, ..., N_r$ так что $N_1 + N_2 + ... + N_r = N$. Страты представляют собой однородные объекты с точки зрения статистических характеристик (например, население делится на страты по возрастным группам или социальной принадлежности; предприятия — по отраслям). В этом случае выборки называются**стратифицированными** (иначе, **расслоенными, типическими, районированными**);

4. методы **серийного** отбора используются для формирования **серийных** или **гнездовых выборок**. Они удобны в том случае, если необходимо обследовать сразу "блок" или серию объектов (например, партию товара, продукцию определенной серии или население при территориально-административном делении страны). Отбор серий можно осуществить собственно-случайным или механическим способом. При этом проводится сплошное обследование определенной партии товара, или целой территориальной единицы (жилого дома или квартала);

5. **комбинированный** (ступенчатый) отбор может сочетать в себе сразу несколько способов отбора (например, стратифицированный и случайный или случайный и механический); такая выборка называется **комбинированной**.

Виды отбора

По **виду** различаются индивидуальный, групповой и комбинированный отбор. При **индивидуальном отборе** в выборочную совокупность отбираются отдельные единицы генеральной совокупности, при **групповом отборе** — качественно однородные группы (серии) единиц, а **комбинированный отбор** предполагает сочетание первого и второго видов.

По **методу** отбора различают **повторную и бесповторную** выборку.

Бесповторным называется отбор, при котором попавшая в выборку единица не возвращается в исходную совокупность и в дальнейшем выборе не участвует; при этом численность единиц генеральной совокупности **N** сокращается в процессе отбора. При **повторном** отборе **попавшая** в выборку единица после регистрации возвращается в генеральную совокупность и таким образом сохраняет равную возможность наряду с другими единицами быть использованной в дальнейшей процедуре отбора; при этом численность единиц генеральной совокупности **N** остается неизменной (метод в социально-экономических исследованиях применяется редко). Однако, при большом **N (N → ∞)** формулы для**бесповторного** отбора приближаются к аналогичным для **повторного** отбора и практически чаще используются последние (**N = const**).