

В.П. Гранков

**ВЫБОРОЧНОЕ
НАБЛЮДЕНИЕ**

СОДЕРЖАНИЕ

	<i>Стр</i>
1. Общее понятие о выборочном наблюдении	3
2. Причины и условия применения выборочного наблюдения	6
3. Значение теории вероятностей в обосновании выборочного метода	8
4. Разработка теоретических основ выборочного метода в нашей стране	11
5. Генеральная совокупность и выборка из нее	13
6. Способы отбора, обеспечивающие репрезентативность выборки	23
7. Ошибки выборочного наблюдения (средняя ошибка репрезен- тативности случайной выборки)	26
8. Возможная ошибка случайной выборки	30
9. Определение необходимой численности выборки	36
10. Типическая (районированная) выборка	41
11. Механическая выборка	50
12. Серийная (гнездовая) выборка	52
13. Применение выборочных наблюдений в советской статистике . .	56

Василий Павлович Гранков — Выборочное наблюдение

Редактор *Е. М. Шенцис*

Художник *Э. Г. Кусс*

Техн. редактор *А. А. Капралова*

Корректор *К. М. Мешкова*

Сдано в набор 12/VII 1955 г.

Подписано к печати 28/XI 1955 г.

Бумага 60×92¹/₁₆ Объем 4,75 печ. л. Уч.-изд. л. 4,72 Тираж 5 000

Л129082 Госстатиздат, Москва, ул. Разина, 3, пом. 69. Заказ 3332

Цена 2 р. 35 к.

Типография № 5 Углетехиздата
Москва, Южно-портовый 1-й проезд, д. 17

В. П. ГРАНКОВ

ВЫБОРОЧНОЕ НАБЛЮДЕНИЕ

Государственное статистическое издательство
Москва 1955

1. Общее понятие о выборочном наблюдении

Статистика представляет собой особую отрасль общественной науки. Она изучает количественную сторону массовых общественных явлений в неразрывной связи с их качественной стороной, исследует количественное выражение закономерностей общественного развития в данных конкретных исторических условиях, изучает количественную сторону общественного производства в единстве производительных сил и производственных отношений, а также явлений культурной и политической жизни.

Всякое статистическое исследование распадается на три этапа:

1) Статистическое наблюдение — собирание массовых статистических сведений о социально-экономических явлениях. 2) Разработка по определенной программе собранных массовых статистических данных (сводка и статистическая группировка). 3) Анализ и обобщение полученных на основе сводки и группировки статистических цифр.

Статистическое наблюдение представляет собой регистрацию по установленной программе сведений о социально-экономических явлениях и заполнение первичных статистических документов: отчетов, переписных и опросных листов, списков, учетных карточек и т. д. В процессе статистического наблюдения обследованию могут подвергаться все единицы данной совокупности или только некоторая их часть.

Таким образом, по охвату единиц обследуемой совокупности статистические наблюдения можно подразделить на два вида:

1) *Сплошное наблюдение*. В этом случае обследованию подвергаются все без исключения единицы данной совокупности. Такими сплошными статистическими наблюдениями у нас в СССР являются: перепись населения, перепись скота, заключительный учет посевных площадей, перепись вагонов, пятидневные отчеты колхозов о ходе сева, месячные отчеты колхозов о состоянии животноводства (форма № 24) и т. д.

2) *Несплошное наблюдение*. В этом случае обследованию подвергаются не все единицы данной совокупности, а только некоторая часть их. Несплошные статистические наблюдения проводятся различными способами. Наиболее распространены следующие

виды сплошного наблюдения: анкетный, корреспондентский способ, монографическое обследование, выборочное наблюдение.

При анкетном способе для получения необходимых сведений рассылают или раздают специально разработанные вопросники (анкеты). Так как заполнение анкет основано на принципе добровольности, то некоторая часть разосланных анкет не возвращается и в разработку не поступает. При этом полученные сведения, как правило, носят односторонний характер. Это означает, что полученные данные не являются репрезентативными, т. е. неправильно, недостаточно точно отображают изучаемое явление или признак в данной совокупности. Примерами анкетного обследования могут служить: анкетное обследование пассажиров железнодорожного транспорта (возвращается около 5% раздаваемых анкет); обследование клиентов почтово-телеграфной связи с точки зрения проверки качества работы местных органов связи; проверка качества продукции массового потребления и культтоваров путем раздачи анкет покупателям в момент покупки этих товаров.

Корреспондентский способ несколько напоминает анкетный способ. Отличие заключается в том, что в этом случае анкеты рассылают специально подобранным лицам, так называемым корреспондентам, которые систематически представляют в статистические органы необходимые сведения. Как правило, корреспонденты работают за определенную плату. Система таких корреспондентов находит широкое распространение в капиталистических странах и в первую очередь в США, где господствует частная буржуазная статистика. Корреспонденты подбираются из числа «благонадежных» лиц, и работают они по заказу и в угоду капиталистам. Представляемые ими статистические сведения носят односторонний, предвзятый характер. Такие сведения тоже не являются репрезентативными для всей совокупности. Сеть добровольных корреспондентов из числа зажиточных крестьян была в дореволюционной России, а также в первый период Советской власти, когда еще не была налажена система статистической отчетности по социалистическим предприятиям. Представляемые этими корреспондентами статистические данные были в значительной степени искажены. На основе специальных дополнительных наблюдений и разработок к ним была установлена система поправок. Исключительно большие искажения корреспондентская сеть стала давать в период массовой коллективизации крестьянских хозяйств, в период ликвидации кулачества как класса. В 1929 г. сеть корреспондентов была ликвидирована и заменена системой отчетности социалистических предприятий.

Монографическое обследование заключается в том, что обследованию и подробному описанию подлежат отдельные, специально отобранные единицы данной совокупности. В условиях СССР монографическому обследованию подвергаются лучшие, передовые предприятия с целью подробного изучения и распространения опыта работы этих предприятий. Монографические

обследования фабрик, заводов, совхозов, МТС и колхозов имеют большое значение в нашей советской статистике. Однако статистические сведения, собранные по передовым предприятиям, не отображают работу всех предприятий. Поэтому перед статистикой наряду с изучением опыта передовых предприятий стоит и другая задача — собрать обобщенные данные, характеризующие работу всех предприятий. Для этой цели применяют выборочное наблюдение.

Выборочное наблюдение — наиболее совершенный и научно обоснованный способ несплошного наблюдения, рассчитанный на то, чтобы на основе обследования какой-то части совокупности судить о всей данной совокупности. В данном случае исключается преднамеренный отбор единиц из общей совокупности: отбор единиц для обследования производится в случайном порядке. Выборочное наблюдение характеризуется тем, что обследованию подвергается заранее обусловленная часть данной совокупности ($\frac{1}{20}$, $\frac{1}{10}$, $\frac{1}{5}$ и т. д.). Все намеченные (отобранные) для обследования единицы подвергаются обследованию в обязательном порядке, без единого исключения и без замены одних единиц другими.

К выборочному наблюдению в силу случайности отбора единиц применимы закон больших чисел, теоремы теории вероятностей, дающие возможность характеризовать предел ошибок, допущенных при выборочном обследовании.

В. И. Ленин неоднократно указывал на необходимость и целесообразность применения выборочных обследований. В своем письме управляющему ЦСУ П. И. Попову 1 сентября 1921 г. В. И. Ленин в *«необходимейшие работы»* включает выборку «для изучения, *небольшого* числа типичных предприятий (фабрик, совхозов) и *учреждений...»*.

Необходимость научной организации выборочных обследований подчеркивалась в 1934 г. в указаниях директивных органов по вопросу о статистике животноводства, где было указано на недостаточный объем контрольных обходов после переписи скота (1,5—2%), что не давало возможности установить поголовье скота по мелким территориальным единицам, например по административным районам.

Выборочное наблюдение находит широкое применение в нашей государственной статистике. В СССР проводятся следующие выборочные статистические обследования: контрольные обходы после переписи скота; регистрация цен на колхозных рынках; бюджетные обследования семей рабочих, служащих и колхозников и др.

Выборочное наблюдение находит широкое применение в промышленных и сельскохозяйственных предприятиях, например при определении качества продукции: процента жирности молока, всхожести семян, натуре зерна, крепости пряжи, долговечности электро- и радиоламп и т. д.

2. Причины и условия применения выборочного наблюдения

Формы и методы организации и проведения статистического наблюдения зависят от задач и конкретной обстановки, определяются условиями целесообразности, рациональным использованием сил и средств и другими факторами, связанными с проведением данной статистической работы.

Некоторые статистические исследования можно организовать любым способом — путем сплошного обследования или выборочным методом. Примерами таких работ являются: определение средней выработки и средней заработной платы рабочих, средней урожайности, удельного веса породного скота в районе и т. п. В практике встречаются такие случаи, когда величину изучаемого явления можно установить только сплошным обследованием. Например, общую численность населения в стране, посевную площадь, поголовье скота и т. п. можно определить только на основе сплошного учета (переписи).

Могут быть и противоположные случаи, когда средний размер изучаемого признака устанавливается только на основе несплошного обследования. Примером может служить лабораторное определение продолжительности горения электрических и радиоламп. Для определения их среднего срока службы только некоторая часть электро- или радиоламп направляется в лабораторию. Там их испытывают под током, фиксируя начало испытания и момент их перегорания. Ясно, что нельзя все электро- или радиолампы направить на такое лабораторное испытание.

В том случае, когда статистическое исследование можно произвести на основе сплошного или выборочного наблюдения, часто отдается предпочтение выборочному наблюдению, так как оно имеет ряд преимуществ по сравнению со сплошным наблюдением, а именно: сокращаются сроки проведения работы; значительно уменьшаются материальные и денежные затраты; требуется меньшее количество обследователей (счетчиков, учетчиков и регистраторов); при выборочном обследовании иногда можно добиться более точных результатов по сравнению со сплошным наблюдением; при прочих равных условиях в выборке можно значительно расширить программу статистического наблюдения по сравнению со сплошным наблюдением. Все эти факторы имеют важное, а иногда и решающее значение при организации и проведении статистического наблюдения.

В период Великой Отечественной войны решающее значение имели сроки статистического наблюдения. Если в нормальных условиях на проведение той или иной переписи требовалось несколько месяцев, то в условиях военной обстановки эта работа проводилась за несколько дней.

Для того чтобы быстро собрать и разработать необходимые статистические данные, прибегают к организации выборочного обследования, если заранее известно, что сплошное наблюдение

не обеспечит представление необходимых данных в установленный срок. Так, например, на основе отчетных данных от всех совхозов и колхозов можно получить точные данные о количестве собранного зерна, а эти данные поступают после сбора и обмола всего урожая. Но в условиях нашего социалистического хозяйства для целей планирования и перераспределения хлеба надо возможно раньше иметь данные об урожайности.

Вопрос о сокращении средств при проведении статистических работ, о рациональном и целесообразном их использовании также имеет чрезвычайно большое значение. Значительного сокращения расходов можно добиться путем замены сплошного наблюдения выборочным. В выборочном обследовании требуется меньше бланков, сокращается заработная плата регистраторов, инструкторов-контролеров и персонала, занятого разработкой собранных статистических материалов, сокращаются почтово-телеграфные и транспортные расходы, оплата командировок и т. д.

Объем выборочного обследования значительно меньше по сравнению со сплошным обследованием. Следовательно, для его проведения требуется меньшее количество обследователей, учетчиков, регистраторов. Поэтому для такой работы можно подобрать более квалифицированных работников, лучше их подготовить, проинструктировать и легче проконтролировать их работу. В итоге повышается качество работы, точность, достоверность статистических сведений. Поэтому при выборочном наблюдении иногда можно добиться даже более точных данных, чем при сплошном обследовании.

Пример. Ежегодно в СССР проводятся переписи скота. Это — сплошное статистическое наблюдение. Сразу же после окончания переписи, с целью уточнения полученных на основе переписи данных о поголовье скота, для внесения поправок на возможный недоучет скота проводится более тщательно выборочное наблюдение, так называемые контрольные обходы — проверка и подсчет скота в натуре по каждому хозяйству в некоторых отобранных на выборку населенных пунктах. Таким образом, выборочное наблюдение оказывается дополнением к сплошному наблюдению: на основе контрольных обходов вносят уточнения, правки на результаты переписи.

При одних и тех же материальных и денежных затратах, за одни и те же сроки, но при помощи выборки можно расширить программу наблюдения, т. е. собрать более подробные данные, по большему количеству показателей и признаков, более детально, всесторонне характеризовать изучаемую совокупность.

Иногда при наличии большого объема работ сплошное наблюдение осуществить невозможно. Например, для определения натуре зерна (среднего веса 1 л зерна) сплошным наблюдением потребовалось бы установить для всего зерна в совхозе или колхозе, или по району в целом два показателя: общий вес зерна в килограммах и объем его в кубических дециметрах (литрах). Еще

большие трудности были бы при определении на основе сплошного обследования процента сорности зерна, определении общих запасов древесины по всем лесным массивам СССР, запасов полезных ископаемых в том или ином районе и т. д. Во всех подобных случаях прибегают к выборочному наблюдению. Наконец, как было уже отмечено выше, сплошное обследование непригодно в тех случаях, когда в результате обследования происходит уничтожение, приведение в негодность обследуемых единиц данной совокупности, например: испытание продолжительности горения электро- и радиоламп, лабораторное определение всхожести семян, испытание крепости пряжи на разрыв, дегустация продукции пищевой и вкусовой промышленности (папирос, чая, вина, консервов и т. д.). Во всех подобных случаях применяется только выборочное наблюдение.

3. Значение теории вероятностей в обосновании выборочного метода

Почему же данные выборочного наблюдения иногда оказываются более точными, чем результаты сплошного наблюдения? Чтобы ответить на этот вопрос, мы должны установить понятие об ошибках статистического наблюдения. Ошибки статистического наблюдения подразделяются на два вида: 1) ошибки регистрации, 2) ошибки репрезентативности.

Ошибки регистрации и могут быть как в сплошном, так и в несплошном статистическом наблюдении. Ошибки регистрации — это расхождение между данными, записанными (зарегистрированными) в первичных статистических документах, и фактическими данными, существующими в данной совокупности. Ошибки регистрации происходят по следующим причинам:

а) небрежное и невнимательное отношение к работе регистраторов и отсутствие у них достаточных и твердых знаний и навыков по данному статистическому наблюдению;

б) неточность и неисправность измерительных приборов, например наличие неточных рулеток, метров, двухметровок при измерении площадей земельных участков, неисправных весов при взвешивании продукции и т. д.;

в) умышленное искажение статистических данных, сообщение неправильных сведений регистраторами или лицами, дающими эти сведения.

Характерная особенность нашей советской статистики, ее принципиальное отличие от буржуазной статистики заключается в том, что все статистические цифры у нас в СССР должны быть абсолютно точными, объективными. Поэтому на всех этапах статистической работы, и в первую очередь в процессе статистического наблюдения, ведется самая решительная борьба против всякого рода ошибок, извращений и искажений статистических данных.

Применяются следующие меры борьбы против ошибок реги-

страции. Для проведения статистического наблюдения подбирают добросовестных и наиболее квалифицированных работников, знакомых с учетно-статистической работой. Кадры регистраторов тщательно готовят, обучают, инструктируют, а в процессе статистического наблюдения контролируют и на месте исправляют обнаруженные недочеты в их работе. В статистических работах пользуются исправными и тщательно проверенными измерительными приборами. Проводят широкую массовую работу среди населения: разъясняют цели статистической работы и задачи получения точных статистических данных, необходимых для планирования народного хозяйства и управления социалистическим государством.

Поскольку в выборочном наблюдении меньший объем работ, то при наличии более квалифицированных кадров, лучшей их подготовки и более тщательного контроля за их работой ошибки регистрации меньше, чем при сплошном наблюдении. Этим и объясняется то, что при выборочном наблюдении, например при контрольных обходах после переписи скота, данные получаются более точные по сравнению со сплошным наблюдением.

Ошибки репрезентативности — это ошибки отображения, ошибки представления, соответствия. Они имеют место только в несплошном наблюдении, когда мы по обследованной части совокупности желаем судить о всей совокупности единиц. Ошибки репрезентативности характеризуют размер расхождения между показателями или характеристиками, полученными на основе несплошного (в частности, выборочного) наблюдения, и соответствующими показателями (характеристиками) по всей статистической совокупности. Причина этого вида ошибок заключается в том, что обследованию подвергаются не все единицы данной статистической совокупности, а только некоторая ее часть. Обследованная часть данной совокупности по своему составу, т. е. по структуре, может отличаться от состава всей совокупности. Например: на заводах треста 25 000 рабочих. Общая средняя месячная заработная плата рабочих по всем заводам составляет 898 руб. В порядке выборки обследовали 900 рабочих. На основе этого обследования установлена средняя месячная заработная плата рабочих в 900 руб. По нашему определению разность между средней заработной платой, полученной на основе несплошного наблюдения (900 руб.), и соответствующим показателем по всей совокупности (898 руб.) составляет ошибку репрезентативности, которую обозначим через Δ . Таким образом, для данного примера ошибка репрезентативности Δ составляет $(900 - 898) = 2$ руб.

В некоторых видах несплошного наблюдения ошибки репрезентативности, как правило, носят односторонний характер: характеристики, полученные на основе этих наблюдений, будут, как правило, уклоняться всегда в одну сторону. Например, если в анкетах, розданных покупателям обуви, поставить вопрос о про-

должительности носки обуви, то очевидно, что в большей части возвратятся анкеты от покупателей, у которых обувь пришла в негодность в короткий срок. Следовательно, средний срок носки обуви на основе этих анкет будет преуменьшен. В подобных случаях ошибки могут достигать больших размеров, и их нельзя уменьшить путем увеличения объема обследуемой части данной совокупности. Несмотря на увеличение количества обследованных единиц данной совокупности, ошибки репрезентативности попрежнему будут носить систематический, односторонний характер.

Иное положение для выборочного наблюдения, в котором единицы для обследования отбирают в случайном порядке, независимо от воли и желания исследователя. В этом случае ошибки репрезентативности носят случайный характер. Они одинаково возможны как со знаком плюс, так и со знаком минус. Они колеблются и по абсолютной величине: у одного исследователя ошибка репрезентативности может быть больше или меньше, чем у другого исследователя. В силу этого к выборочному наблюдению применимы теоремы теории вероятностей, закон больших чисел, чего нельзя сказать в отношении других видов сплошного наблюдения.

Сущность закона больших чисел заключается в том, что при достаточно большом объеме выборки обобщающие характеристики, полученные на основе выборочного наблюдения, будут сколь угодно мало отличаться от соответствующих характеристик по всей совокупности. Пользуясь законом больших чисел, мы можем, увеличивая объем выборки, регулировать пределы возможной ошибки репрезентативности, довести ее до минимальных (допустимых) размеров. Зная предел допустимой ошибки, мы можем на основе закона больших чисел определить необходимую численность выборки.

Закон больших чисел дает математическое обоснование выборочного метода в статистике. Иных взглядов держатся буржуазные статистики. Они утверждают, что закон больших чисел является основой выборочного метода и всей статистики как науки. Исходя из этого, они утверждают, что статистика в одинаковой мере применима к явлениям общественной жизни и к любым явлениям природы. Они считают главным условием для статистики наличие большого количества наблюдений без их обобщения на основе познания объективных законов общественного развития. Такая постановка вопроса является в корне неправильной. Методом любой науки, в том числе и статистики, является марксистский диалектический метод. В буржуазном обществе методом статистики, как и других общественных наук, являются метафизический метод и различные позитивистские направления.

В отличие от буржуазной статистики, где господствует голый эмпиризм, увлечение и незаконное применение математики и математической статистики, советские статистики учитывают, что основой статистического исследования должны быть исторический

материализм и марксистско-ленинская политическая экономия, что закон больших чисел в различной степени применяется к явлениям природы и явлениям общественной жизни. Чем сложнее явление, тем с большей осторожностью надо применять методы математической статистики.

4. Разработка теоретических основ выборочного метода в нашей стране

Приемы выборочных наблюдений были известны русским людям с древних пор. В отношении организации выборки, как одного из способов статистического наблюдения, им принадлежит большая заслуга.

До наших дней дошли документы, свидетельствующие о применении выборочного метода на Руси более 300 лет тому назад.

В «Актах» хозяйства боярина Морозова (1648 г.) даны указания по определению урожайности хлебов выборочным методом. Организация контрольных (т. е. выборочных) обмолотов описана в Уставе конюшенной канцелярии («Регламент конюшенный», изданный в 1733 г.). Эти выборочные наблюдения, правда, были еще весьма примитивными, но для того времени являлись новым методом статистического исследования.

Широкое применение выборочный метод получил в русской земской статистике (конец XIX — начало XX в.). Земские статистики, несмотря на ряд серьезных ошибок и недостатков в их работе, многое сделали в разработке теории и организации статистической работы и особенно в отношении собирания прекрасных массовых статистических данных о крестьянских хозяйствах дореволюционной России (подворные переписи крестьянских хозяйств). В своей работе земские статистики часто применяли выборочные наблюдения, разрабатывали и совершенствовали способы выборочных обследований. Впервые так называемая типическая выборка с предварительной разбивкой всей обследуемой совокупности на более или менее однородные группы была применена в 1900 г. в Вятской губернии. Эта выборка дает более точные результаты по сравнению с обычной, т. е. собственно случайной, выборкой. Механический отбор, имеющий организационные преимущества по сравнению с собственно случайной выборкой и обеспечивающий соблюдение принципа непреднамеренного отбора, был впервые применен калужскими земскими статистиками в 1896 г.

После победы Великой Октябрьской социалистической революции в нашей стране началась организация государственной статистики. В 1918 г. было организовано Центральное статистическое управление. Для собирания статистических данных ЦСУ стало проводить специальные статистические наблюдения (сплошные переписи) и выборочные обследования. Особенно много выборочных наблюдений было в сельскохозяйственной статистике: выборочные

весенние и осенние опросы крестьянских хозяйств, бюджетные обследования крестьян, контрольные обмолоты и т. п. Советские статистики разработали на научной основе выборочное измерение урожайности зерновых и пропашных культур путем наложения метровок на посевы и ряд других выборочных наблюдений (контрольные обходы после переписи скота, бюджетные обследования семей рабочих, служащих и колхозников и др.).

Как было отмечено выше, выборочный метод опирается на закон больших чисел. Основные и наиболее общие положения закона больших чисел разработаны нашими русскими учеными математиками-академиками П. Л. Чебышевым, А. М. Ляпуновым, А. А. Марковым. Дальнейшая разработка теории вероятностей произведена трудами крупнейших советских математиков — акад. С. Н. Бернштейна, акад. А. Н. Колмогорова и А. Я. Хинчина.

Теорему Чебышева применительно к выборке можно сформулировать так: с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом числе независимых наблюдений средняя величина изучаемого признака, полученная на основе выборки, будет сколь угодно мало отличаться от средней величины изучаемого признака по всей совокупности.

Напомним известные из курса математической статистики основные понятия вероятности, о которой идет речь в теореме Чебышева. Вероятность — мера объективной возможности наступления определенного события в определенных, могущих повторяться неограниченное число раз условиях. Вероятностью события A называется отношение числа удачных (благоприятных) исходов (M) к числу всех исходов (N). Это можно записать так:

$$P(A) = p = \frac{M}{N}.$$

Вероятность находится в пределах от нуля до единицы

$$0 \leq P(A) \leq 1.$$

Если вероятность события A мала (другими словами, близка к нулю), то это событие A наступает редко. Такое событие называется маловероятным. Практически его можно считать невозможным. Если вероятность A близка к единице, то это событие наступает часто (в большинстве случаев). При $P(A) = 1$ считают событие A достоверным. Можно рассчитывать наверняка на его наступление.

Заслуга П. Л. Чебышева заключается в том, что он, доказав в 1867 г. свою теорему, сформулировал закон больших чисел в общем виде. Правда, значительно раньше (в 1713 г.) швейцарским математиком Якобом Бернулли была опубликована подобная теорема, но только для одного частного случая, а именно в отношении доли изучаемого признака, характеризующая соотношение между так называемой частотой и вероятностью. Применительно к выборочному наблюдению теорему Якоба Бернулли можно

сформулировать так: с вероятностью, сколь угодно близкой к единице, можно утверждать, что при достаточно большом числе независимых наблюдений доля изучаемого признака, полученная на основе выборочного наблюдения, будет сколь угодно мало отличаться от доли данного признака во всей совокупности. Итак, в теореме Бернулли речь идет о соотношении доли в выборке (частость) и доли данного признака во всей совокупности (вероятность). Например, можно говорить о соотношении между долей (процентом) всхожести семян по выборочным наблюдениям и процентом всхожести семян по всему семенному фонду.

Крупнейшие математики мира (Пуассон, Лаплас, Гаусс и другие) длительно и упорно работали после Якоба Бернулли в области теории вероятностей, имели ряд достижений, но все обобщить теорему Бернулли им не удалось. Данная задача была блестящим образом разрешена русским математиком П. Л. Чебышевым. Он дал доказательство закона больших чисел в общем виде. Из теоремы Чебышева можно легко получить теорему Бернулли как частный случай.

Теорема Чебышева будет более подробно изложена несколько ниже, после ознакомления с основными показателями (характеристиками), получаемыми нами в результате статистического наблюдения.

5. Генеральная совокупность и выборка из нее

Переходя к вопросам организации выборочного наблюдения, мы должны установить некоторые основные понятия, некоторые показатели (характеристики), получаемые в результате статистического наблюдения. Если обследовать все без исключения единицы, то мы получим общую совокупность единиц, которую называют генеральной совокупностью. Часть генеральной совокупности, подлежащая выборочному обследованию, называется выборочной совокупностью. Численность (объем) генеральной совокупности будем обозначать через N , а численность выборочной совокупности (численность или объем выборки) обозначим через n . В статистическом обследовании можно поставить две задачи: 1) определение среднего размера изучаемого признака, например средней выработки или средней заработной платы рабочих, средней урожайности, средней удойности коров; 2) определение доли, удельного веса изучаемого признака в данной совокупности, например удельного веса породного скота, доли сортовых посевов, процента брака.

Средний размер изучаемого признака в генеральной совокупности будем называть средней генеральной. Обозначим ее через \bar{X} . Средняя генеральная находится по формуле средней арифметической невзвешенной, а именно:

$$\bar{X} = \frac{\sum X}{N}.$$

В случае группового распределения средняя генеральная находится по формуле средней арифметической взвешенной:

$$\bar{X} = \frac{\sum XM}{\sum M},$$

где M — частоты (веса), т. е. количество единиц в каждой группе.

Средний размер изучаемого признака в выборочной совокупности будем называть средней выборочной. Обозначим ее через \tilde{X} . Средняя выборочная тоже может быть простой (невзвешенной) $\tilde{X} = \frac{\sum X}{n}$; или взвешенной $\tilde{X} = \frac{\sum Xm}{\sum m}$. Ошибка репрезентативности (Δ) есть разность между выборочной средней и генеральной средней.

Пример. Общая численность рабочих на заводах треста 25 000 чел. ($N=25\,000$ чел.). Их средняя месячная заработная плата (средняя генеральная) составляет 898 руб. ($\bar{X}=898$ руб.). В порядке выборки обследовали 9000 чел. (численность выборки $n=9000$ чел.). Их средняя месячная заработная плата (выборочная средняя) составляет 900 руб. ($\tilde{X}=900$ руб.). Ошибка репрезентативности при определении средней заработной платы составляет $\Delta = \tilde{X} - \bar{X} = 900$ руб. — 898 руб. = 2 руб.

Предположим, что нужно выяснить долю, т. е. удельный вес изучаемого признака. Обозначим, как и прежде, общую численность, т. е. численность генеральной совокупности, через N . Количество единиц, обладающих данным признаком, обозначим через M . Тогда долю единиц, обладающих данным признаком в генеральной совокупности (p), можно определить так:

$$p = \frac{M}{N}.$$

Для выборки будем иметь аналогичные показатели:

n — численность выборки;

m — частота (число единиц, обладающих данным признаком в выборке);

W — частота (доля единиц, обладающих данным признаком в выборке).

Разность между частностью (W) и долей данного признака в генеральной совокупности (p) представляет собой ошибку репрезентативности (Δ). Поясним эти показатели на следующем примере: на тех же заводах (см. предыдущий пример) из общей численности рабочих в 25 000 чел. мужчины составляют 19 750 чел. Из 9000 рабочих, попавших в выборку, мужчины составляют 7200 чел.

Таким образом, можем записать:

численность генеральной совокупности (общее количество рабочих на заводах) $N=25\,000$ чел. Число единиц, обладающих

данным признаком в генеральной совокупности, $M=19\ 750$ чел. Доля данного признака в генеральной совокупности

$$p = \frac{M}{N} = \frac{19\ 750}{25\ 000} = 0,79 = 79\%.$$

Численность выборки (количество обследованных рабочих на заводах треста) $n=9000$ чел. Частота, т. е. число единиц, обладающих данным признаком в выборке (количество мужчин среди обследованных 9000 рабочих), $m=7200$ чел. Частость, т. е. доля данного признака в выборке (удельный вес мужчин среди обследованных 9000 рабочих),

$$W = \frac{m}{n} = \frac{7200}{9000} = 0,80 = 80\%.$$

Ошибка репрезентативности — разность между частостью и долей, т. е. разность между удельным весом мужчин в выборке и в генеральной совокупности, равна $0,80-0,79=0,01=1\%$.

Остановимся, далее, на показателях пестроты или колеблемости изучаемого признака, на показателях вариации.

Пример. Заработная плата рабочих в двух бригадах характеризуется следующими данными:

Первая бригада		Вторая бригада	
№ п/п	Заработная плата (в руб.)	№ п/п	Заработная плата (в руб.)
1	$X_1 = 300$	1	$Y_1 = 480$
2	$X_2 = 400$	2	$Y_2 = 490$
3	$X_3 = 500$	3	$Y_3 = 500$
4	$X_4 = 600$	4	$Y_4 = 510$
5	$X_5 = 700$	5	$Y_5 = 520$

Фонд заработной платы

$$\Sigma X = 2500 \text{ руб.}$$

$$\Sigma Y = 2500 \text{ руб.}$$

Вычислим среднюю заработную плату по каждой бригаде как среднюю арифметическую простую.

По первой бригаде будем иметь

$$\bar{X} = \frac{\Sigma X}{N} = \frac{2500}{5} = 500 \text{ руб.}$$

По второй бригаде будем иметь

$$\bar{Y} = \frac{\Sigma Y}{N} = \frac{2500}{5} = 500 \text{ руб.}$$