Osnovy teorii vyboročnogo metoda

A.G. Kovalevskij

Julkaistu sarjassa:
Učenie zapiski Gosudartsvennego Saratovskogo imeni N.G. Černyševskogo universiteta; tom2, vyp. 4: Fakul'tet hozâjstva i prava, sivut 60-138

Saratov: Saratovskij gosudarstvennyj universitet, 1924

УЧЕНЫЕ ЗАПИСКИ

ГОСУДАРСТВЕННОГО САРАТОВСКОГО имени Н. Г. ЧЕРНЫШЕВСКОГО УНИВЕРСИТЕТА.

ТОМ II

(по продолж. ТОМ XI „ИЗВЕСТИЙ САР. УН.“).

ВЫП. 4. ФАКУЛЬТЕТ ХОЗЯЙСТВА И ПРАВА.

Под редакцией проф. *И. А. Чуевского.*

Издание Правления Саратовского Государственного Университета.

САРАТОВ.
1924 г.

Основы теории выборочного метода.

А. Г. Ковалевский.

ПРЕДИСЛОВИЕ.

Выборочный метод является одной из наиболее важных и интересных проблем статистической методологии. Практические приложения этого метода в различных областях социальной хозяйственной статистики делаются все более и более систематичными и приобретают, в особенности в России, весьма видное значение. Однако, основания метода, получившие сравнительно недавно некоторое теоретическое освещение, не отличаются надлежащей ясностью и законченностью обоснования основных положений как раз применительно к явлениям социального порядка. Поэтому, быть может, не лишним является опубликование этой работы, имеющей целью систематическое, по возможности, строгое обоснование основ теории выборочного метода.

В числе факторов, позволивших мне довести эту работу до конца, очень существенным сказалось содружество математика Б. К. Ризенкампфа, предоставившего мне свои выводы для использования в настоящей работе. Из них я воспользовался затруднявшими меня важными доказательствами и положениями выборочного метода, изложенными в §§ 20, 21, 21, 26 и 27. Необходимо отметить, что эти результаты были получены Б. К. Ризенкампфом независимо от моих. За предоставление мне этих своих выводов и соображений, а также за ряд ценных указаний в затруднявших меня случаях, считаю приятным долгом выразить Б. К. Ризенкампфу свою глубокую признательность. Также приношу благодарность Президиуму Ф. О. Н. Саратовск. Госуд. им. Н. Г. Чернышевского Университета, предоставившему мне возможность напечатать эту работу в Ученых Записках Университета.

ГЛАВА I.

§ 1. О сплошных наблюдениях.

Современная статистическая методология для исследования массовых явлений вообще и социальных в особенности различает два способа наблюдения. Один из них—сплошное наблюдение, при котором по известной программе подвергаются описанию все единицы или экземпляры данной изучаемой совокупности; другой—частичное не сплошное наблюдение, при котором по известной программе подвергаются описанию не все экземпляры данной совокупности, а только некоторая часть.

Методы сплошных наблюдений, обслуживая последующие стадии статистического анализа и обобщений, дают прежде всего исчерпывающий счет, абсолютные числа как отдельных экземпляров совокупности, так и элементарных признаков их в пределах принятой программы.

Установленные из этих абсолютных чисел производные величины и сводные признаки, внутренние конструкции, взаимоотношения отдельных частей и групп, закономерности и причинно-следственные зависимости приобретают наиболее авторитетное значение, раз имеется уверенность, что все оттенки и групповые разновидности изучаемой совокупности получили надлежащее место, и исследователь не допустил грубых ошибок. Суб'ективное усмотрение в смысле одностороннего подбора экземпляров массы здесь отсутствует, так как нормально организованное сплошное наблюдение включает в описание все экземпляры и исчерпывает элементы совокупности с наибольшей практически возможной точностью. Поэтому, из всех методов статистического наблюдения принято считать сплошное исследование наиболее точным и это еще потому, что при его применении, если отсутствуют преднамеренные искажения, достигается наибольшее соответствие полученных результатов с действительностью, а этим в конечном счете и определяется достоинство того или другого метода наблюдения.

Однако, сплошные наблюдения в виде современных переписей, несмотря на заслуженную репутацию, имеют в практическом отношении целый ряд неудобств. Прежде всего переписи это грандиозные статистические операции, требующие большой затраты средств и труда многих тысяч лиц. Помимо большой стоимости организации переписей, также очень сложна и дальнейшая разработка полученного в результате их статистического материала. Обычно, считается нормальным сроком сплошной разработки переписного материала, судя по практике европейских переписей три, три с половиной года. Вследствие этого сплошные переписи производятся сравнительно редко—один раз в пять или десять лет. Когда же они производятся чаще, как у нас в России в последнее время, то случается, что предыдущие переписи остаются в значительной части не разработанными. Так, например, разработка сельско-хозяйственных переписей 1916 и 1917 г.г. не закончена; также не закончена пока разработка демографической и сольско-хозяйственной переписей 1220 г., когда мы уже накануне больших статистических исследований 1925 г.

В течение лет, когда разрабатывается материал переписи, добытые сведения становятся устаревшими, и результаты работы, не доходя еще до конца, становятся достоянием истории. В значительной мере это относится и к другим видам сплошного наблюдения в форме сплошных текущих записей.

§ 2. О частичных наблюдениях.

Стремление избежать эти неудобства сплошных наблюдений повели статистическую мысль по пути искания новых, менее громоздких приемов исследования над массовыми явлениями, которые при этом давали бы достаточное приближение к действительности при наименьшей затрате средств, труда и времени.

И мы видим, что на ряду с разработкой методологии сплошных наблюдений идут интересные попытки искания новых путей—замены сплошных исследований частичными—заманчиво обещающих статистику значительное упрощение его работы.

При осуществлении этих попыток организации и обоснования приемов частичного наблюдения всегда сталкивались с весьма трудно разрешимой задачей, как организовать исследование, чтобы добываемые результаты отличались бы надлежащей практически допустимой точностью и были бы достаточно, как говорят, репрезентативны.

Необходимость разрешения этой задачи возникает при всяком частичном наблюдении, так как, встав на путь частичного исследования и оторвавшись от привычного представления, в особенности при изучении социальных явлений о необходимости исчерпывающего наблюгения над всеми единицами совокупности, мы должны быть более осторожными и вправе требовать гарантии законности применения частичных приемов наблюдения.

Прежде всего этот вопрос возникает при плохо организованном сплошном наблюдении вообще и в частности тогда, когда частичность описания единиц массы произошла не по желанию исследователя, не по намеченному плану, а, так сказать, случайно, как это, например, часто бывает при не экспедиционных способах собирания первичного материала. Здесь исследователь в большинстве случаев не держит в руках всех нитей надзора и инструкционного влияния на ход работы. Благодаря организационным дефектам сплошное по мысли исследование, при таких приемах собирания об'екте данных, превращается, во-первых, в частичное наблюдение, а, во-вторых, эта частичность становится случайной как по количесву входящих в нее единиц, так и по характеру их. В таком частичном материале редка восможность соотносительного представительства различных оттенков и групп всей изучаемой совокупности и, наоборот, обычна односторонность. В заключение исследователь, не имея уверенности в об'ективности собранного материала, к тому же, вообще говоря, [1]) не может дать оценку точности результатов окончательных выводов. Следовательно, такое неорганизованное частичное наблюдение, ли-

---

[1]) В некоторых случаях такая оценка может быть дана по методу Bowley, если частичность выборки носила характер случайный в смысле теории вероятностей.

шенное необходимых теоретических основ, не может претендовать на научность.

Тот же вопрос о гарантиях точности и репрезентативности возникает и тогда, когда частичность наблюдения над отдельными единицами изучаемой совокупности происходит планомерно, как следствие преднамеренных действий исследователя, который по известной системе производит из всех единиц совокупности отбор подлежащих изучению единиц с целью, на основании изучения их, полученные сводные признаки и выводы распространить на все экземпляры изучаемой совокупности.

Такого рода преднамеренные частичные наблюдения известны в статистической практике под названием монографического и выборочного методов.

§ 3. Монографический метод.

При монографическом исследовании гарантией точности и соответствия с действительностью выводов, полученных на основе (монографического) изучения немногих единиц совокупности, считается, как известно, типичность отобранных единиц (хозяйств, сел и т. д.). Число их не играет роли, так как раз выбранные единицы типичны, то большое накопление их не имеет смысла. Да это при монографическом исследовании и не ставится целью, так как монографический метод не является мотодом массового наблюдения и принципиально ему противоположен. Монографический метод имеет дело „с применением естественно-научного единичного наблюдения, аналогичным тому, как оно прилагается к исследованию растений и менералов, следовательно, с допущением предположения, что наблюдение, произведенное над одним экземпляром, способно дать полное знание всех других экземпляров того же вида“ [1]). Однако, возможность подобных обобщений, в настоящее время оспаривается даже в природоведении, так как между отдельными экземплярами одного и того же вида существуют весьма значительные различия, даже и в этой области. Концепция типичности явлений природы, в противоположность явлениям общественным, постепенно изживается; например, по словам Гейнке, „современная зоология в корень порвала с шаблонной догмой, учившей, будто характер группы может быть познан путем исследования отдельных типических представителей“ [1]). Если значение „типичности“ колеблется в естественно-научной области, то спрашивается, можем ли мы обосновать на ней общеобязательные научные обобщения в области изменчивых общественных явлений, а в особенности, тех крайне индивидуальных признаков, которые входят в программу монографических исследований?

Типичность—понятие, которое допускает в своей интерпретации суб'ективный произвол исследователя, и, как говорит Г. Майр, „что предполагаемый тип действительно является таковым—в этом нет никакой гарантии“ [2]).

Типичность, как критерий и гарантия совпадения с действительностью полученных при монографическом исследовании результатов, является теоретически спорной и недостаточной. Ценность получен-

---

[1]). Г. Майр. Статистика и обществоведение т. I, изд. Ц. С. У. 1922 г. стр. 22.

[1]) А. А. Чупров. Очерки по теории статистики 2 изд., стр. 98, 99

[2]) См. по этому поводу аналогичные саображения: А. А. Кауфмана.—Теория и методы статистики 4 изд., стр. 437 и сл., А. Чаянова.—Опыт разработки бюджетов... т. I, стр. 65—66, С. Первушина.—О бюджетных исследованиях. Труды В. Э. О-ва. № 1—2. 1912 г., стр. 20 и сл.

ных результатов зависит от удачи или неудачи каждого конкретного случая исследования. Оценка достоинства этих результатов теорией метода не дается, и разными лицами может быть произведена различно. Поэтому, научное значение монографического метода, с точки зрения возможности оценки соответствия с действительностью достигаемых при помощи его результатов, очень сомнительно.

§ 4 Выборочный метод.

Второй прием частичных наблюдений, выборочный метод, старается разрешить вопрос о гарантиях точности получаемых результатов на основании более солидной предпосылки, именно—закона больших чисел.

Выборочный метод, как удачно определяет его А. А. Кауфман, есть „такой прием исследования массовых и в частности общественных явлений, при котором статистическому перечету, с одной стороны, подвергается только известная часть случаев изучаемого явления, и полученные на основании такого частичного учета выводы распространяются на все данное явление, и при котором, с другой стороны, подвергаемая перечету часть случаев настолько велика, что придает иследованию хотя уже не сплошной, но все еще массовый характер и открывает достаточный простор действию закона больших чисел“ [1]).

К этому следует добавить, что отбор части экземпляров исследуемого явления, или, как мы в дальнейшем будем называть, отбор единиц выборочной группы, на основании изучения которой полученные выводы распространяются на всю совокупность, должен гарантировать возможность полного устранения тенденциозности и суб'ективного усмотрения как наблюдателя, так и об'екта наблюдения. Иначе говоря, отбор единиц выборочной группы должен производиться так, чтобы устранялись постоянные или преднамеренные погрешности статистического наблюдения. При соблюдении этого положения достаточно большое число экземпляров выборочной группы будет нейтрализовать действие случайных отклонений, и полученные при помощи выборочного исследования результаты будут, сколь угодно, близки к таким же результатам сплошного наблюдения.

Если при этом отобранные единицы выборочной группы будут распределяться равномерно среди единиц всей совокупности, то тогда выборочная группа будет представлять всю массу единиц исследуемого явления или, как мы в дальнейшем будем называть, генеральную совокупность.

При соблюдении условия, что выборочная группа будет достаточно велика, чтоб исследование носило массовый характер, и при рационально-об'ективной системе отбора единиц ее, возникает принципиальная возможность замены во многих случаях сплошного исследования выборочным.

Эта возможность имеет безоговорочное значение при изучении явлений природы, к которым применяется статистический метод наблюдения, так как здесь естествоиспытатель, как правило, не испытывает нужды иметь дело с исчерпывающими сплошными наблюдениями. Возможность же замены сплошных исследований выборочными, применительно к изучению массовых социальных явлений, имеет некоторые ограничения. Так, при исследовании социальных масс выборочный метод не может дать исчерпывающего счета и описания всех

---

[1]) А. А. Кауфман. Сборник статей. „К вопросу о выборочном исследовании“. Москва 1915 г., стр. 353.

единиц состава массы; в этом случае сплошное наблюдение сохраняет свое значение. Что же касается получения средних и относительных величин, установления взаимоотношений отдельных частей массы, возможных закономерностей, сопостановлений и т. д., то здесь с успехом сплошное наблюдение может быть заменено выборочным.

Выборочное исследование может дать также, как мы это увидим ниже, с достаточной степенью точности и абсолютные цифры отдельных признаков единиц массы (наприм., число лошадей, десятин посева и т. д,), но вычисление их по данным выборочной группы, без предварительного знания общего числа единиц генеральной совокупности (наприм., числа дворов, хозяйств, сел и т. д.), невозможно [1]. Этим, в кратких чертах, и определяются функции выборочного метода, если у исследователя нет необходимости в исчерпывающем материале о всех единицах генеральной совокупности или очень дробных частях ее.

Но эта принципиальная возможность замены сплошных наблюдений выборочными становится научно-обоснованным фактом лишь тогда, когда мы можем измерять и оценивать достоинство результатов выборочного исследования, иначе говоря, когда мы имеем об'ективные априорные критерии оценки совпадения сводных признаков и обобщенных величин полученных в результате выборочного исследования, с такими же в генеральной совокупности. Без этой оценки выводы, полученные с помощью выборочного исследования, теряют общеобязательное значение.

§ 5. Выборочное исследование в практике земской статистики.

При соблюдении изложенных предпосылок выборочный метод является весьма ценным орудием статистического изучения различных сторон общественной жизни. На ряду с громоздскими сплошными исследованиями он представляет существенные преимущества и значительные удобства. Поэтому и неудивительно, что в статистической практике, впервые выдвинувшей идею выборочного исследования, мы находим разностороннее применение этой идеи и разработку основных положений метода.

В этом отношении русская земская статистическая практика, занимая очень видное место как по оригинальности применения нового метода, так и по существу методологических достижений, представляет значительный интерес в качестве иллюстраций приложения идеи выборочного метода к изучению крестьянского хозяйства.

В момент первоначального развития земских статистических исследований, когда русская статистическая нива была совершенно не вспахана, все стремления статистиков, конечно, были направлены на исчерпывающее сплошное статистическое изучение народо-хозяйственной действительности России.

В связи с этим мы видим вполне законное и естественно-необходимое увлечение методами сплошного наблюдения. На основе Московского и Черниговского образцов сплошных статистических исследований была изучена вся тогдашняя земская Россия. В это время мы видим кульминационный расцвет и всестороннее развитие мотодологии сплошных исследований в русской статистической практике. Но, не-

[1]) См. речь А. А. Чупрова.—„Выборочное исследование". Труды подсекции статистики—XII с'езда русских естествоиспыт. и врачей в Москве 1910 г. Чернигов 1912 г., стр. 160, 161.

смотря на подавляющее представление об исключительном значении сплошных подворных переписей, как единственного метода, гарантирующего необходимое качество статистического материала, желание глубже проникнуть в некоторые стороны народного хозяйства, особенно его динамики, заставило статистиков взяться за разработку частичных исследований, в частности—выборочного метода.

Здесь, за счет сокращения количества описаний, подлежащих изучению единиц генеральной совокупности, представлялось возможным увеличить глубину и об'ем программы статистического исследования рационально отобранной выборочной группы. Постепенно, вопреки господствующему взгляду о преимущественном значении сплошных исследований, методы сокращенных наблюдений стали все чаще и чаще применяться и медленно завоевывать симпатии статистиков.

Господствующий взгляд отразился в том характерном отношении, что частичные наблюдения, в форме-ли монографий, или собственно выборочных исследований, выступали, в большинстве случаев, не как самостоятельные методы, а как дополнительные, всегда (за исключением Вятского и Воронежского исследований) в сопровождении и на фоне сплошных исследований. Только недавно, главным образом, в после революционный период русской статистики, выборочное исследование получило все права гражданства и применяется не только как метод первичного наблюдения, но и как весьма удобный, дающий отличные результаты прием комбинационной разработки материала сплошных наблюдений. До этого времени неуверенность в выборочных наблюдениях была вполне понятна. С одной стороны твердо заученные, практически проверенные достоинства сплошных наблюдений, с другой стороны—новизна и неразработанность приемов, почти полное отсутствие теоретического освещения и обоснования метода не позволяли, конечно, ему конкурировать с сплошными исследованиями. И это тем более, что, за отсутствием априорных критериев точности результатов выборочных исследований, эта оценка давалась единственным эмпирическим приемом—путем сличения коэффициентов сплошного исчисления с коэффициентами, полученными на основании выборочного исследования.

Теперь же последняя задача может быть разрешена иначе, с соблюдением более строгих научных требований. Современная теория выборочного исследования, примыкая к авторитету теории вероятностей, знает более точные и об'ективные способы разрешения этой задачи, чем сличение сводных признаков сплошных и выборочных исследований.

При осуществлении идеи частичных исследований земские статистики прежде всего столкнулись с задачей, как организовать исследование, чтобы добытые результаты отличались надлежащей степенью точности, т. е. столкнулись с задачей найти удовлетворительные приемы отбора единиц выборочной совокупности и оценки, полученных на основании изучения ее, выводов. Решение этой задачи шло в двух направлениях. Представители первого направления, ранее всего возникшего в земской статистической практике, считали, что хорошие результаты могут быть получены, если отбирать в выборочную группу типичные для всей генеральной совокупности единицы. Представители второго направления находили, что хорошие результаты можно получить, лишь опираясь на закон больших чисел, т. е. при достаточно большом числе наблюдений, причем отбор единиц выборочной группы не должен быть связан никакой тенденцией.

Первое направление разделилось на два вида, один из которых вылился в самостоятельное ответвление, именно, в развитую систему подробных монографических описаний крестьянских бюджетов; второй, сохраняя в качестве основного принципа отбора типичность, постепенно усвоил необходимость привлечения массового наблюдения, и этим примкнул к выборочному исследованию. Это та разновидность частичных исследований, которая на Западе стала впервые известна благодаря трудам норвежского статистика Киаэр'а [1]).

В России первое статистическое обоснование этого вида выборочных исследований было дано А. И. Чупровым в докладе подсекции статистики IX с'езда естествоиспытателей и врачей в 1894 г. Проф. А. И. Чупров считал возможным производить исследование типических единиц (сел) не в качестве дополнительного приема наблюдения, а в качестве вполне самостоятельного метода, полезного даже и тогда, когда ему не предшествовало сплошного наблюдения, при условии возможности отбора типичных экземпляров массы [2])

§ 6. Выборочное исследование Вятской губ. 1900—1902 г.г.

Наиболее удачным осуществлением этого направления выборочных земских исследований считается повторное (на основе описания типических селений) выборочное исследование Вятской губ. в 1900—1902 г.г.

Сплошная сельско-хозяйственная перепись была произведена здесь в 80-х годах.

Выборочное исследование 1900—1902 гг. охватило около $^1/_5$ части всех селений губернии по 954 районам, установленным раньше независимо от выборочного исследования по сходству естественно-исторических факторов и экономических условий. Вот как характеризует это исследование и примененные при нем принципы отбора селений, в которых затем была произведена сплошная подворная перепись, один из главных участников вятского выборочного исследования А. Гурьев: „Частичное исследование, построенное на них (т. е. на известных основаниях отбора), имело характер, во-первых, массового исследования, так как оно охватило не менее 20% числа об'ектов, подлежащих изучению; во-вторых, равномерность распределения описываемых селений по территории обеспечивалась выбором их по однородным по естественно историческим и экономическим условиям района, и, в третьих, выбор селений для описания производился так, чтобы каждое выбранное селение являлось представителем особой группы селений, образованной по совокупности экономических признаков" [3]). Проверка типичности выбранных сел была сделана путем сопоставления средних и относительных величин, полученных в результате выборочного исследования с данными первой сплошной переписи 80-х годов, при чем, как говорит А. Кауфман: „получилось, в громадном большинстве случаев, весьма близкое совпадение, притом, естественно, возрастающее по мере перехода от волости к району и уезду, и это совпадение неопровержимо доказало, что отобранные 20% за-

---

[1]) А. Kiaer.—Die repräsentative Untersuchungsmethode, Allgemeines statistisches Archiv B. V. Tübingen. 1898

[2]) Труды подсекции статистики IX с'езда русских естествоиспытателей и врачей. А. Гурьев.—Происхождение выборочного исследования и первые его опыты в России. Вестник Статистики 1921 г. № 1—4, стр. 19.

[3]) А. Гурьев. Там же, стр. 39.

конно могут быть рассматриваемы, как типичные для всего населения, и что выводы из данных по отобранным селениям всецело могут быть распространяемы на всю губернию" [1]).

Действительно, те отрицательные стороны отбора по признаку типичности, о которых мы говорили выше, т. е. произвольность понятия типичности, допускающее поэтому различный суб'ективный подход исследователя, в вятском выборочном исследовании сведены в узкие пределы.

В сущности говоря, эта пресловутая типичность здесь и не причем.

В самом деле, исследователи имели от первого сплошного наблюдения разделение губернии на 954 более или менее однородных в хозяйственном отношении района.

Этот факт, как мы увидим ниже, имеет существенное значение для успешности выборочного исследования.

При наличии расчленения генеральной совокупности на ряд частных совокупностей с более или менее однородным составом единиц, отдельные районные варианты (величины исследуемых признаков) очень мало колеблятся около своих средних, и небольшая группа их будет репрезентативна всем экземплярам района.

Мало того, число отобранных селений из 954-х районов так велико ($^{1}/_{5}$), что выборочная группа по отношению к генеральной совокупности выразилась в 20%. Это очень большое число для выборочного исследования, при котором возможные случайные ошибки при выборе селений могли сгладиться в силу закона больших чисел.

При таком большом числе наблюдений, да еще при наличии районирования, всякий отбор по жребию или другому способу, лишь-бы он не был тенденциозным, дал бы такие же хорошие результаты. Кроме того, отбор селений в силу того же районирования был произведен равномерно по территории губернии. Типичность селений при вятском выборочном исследовании совершенно не определила положения дела.

Определяющими факторами были: наличие районирования, большое число единиц выборочной группы и равномерность их распределения по пространству генеральной совокупности.

Эти факторы при всяком другом не тенденциозном отборе создают необходимые, вполне достаточные условия для успешности выборочного исследования.

Выборочный метод не следует засорять сомнительными положениями монографии, так как понятие типичности не вытекает из идеи выборочного исследования, основывающегося на законе больших чисел, и даже последнему противоречит.

В этом отношении Н. С. Четвериков принципиально совершенно прав утверждая, что „внесение в выборочное исследование принципа монографии знаменует собою вместе с тем и внесение личного произвола и усмотрения со стороны исследователя, а тем самым уничтожает об'ективность результатов и устраняет возможность их научной оценки" [2]).

Возражающему против этого утверждения в отношении Вятского исследования А. Гурьеву следовало бы не ограничиваться вышепри-

---

[1]) А. А. Кауфман.—К вопросу о выборочном исследовании. Сборник статей 1915 г., стр. 367.

[2]) Н. С. Четвериков.—О выборочном исследовании. Вестник Статистики 1919 г. № 8—12, стр. 179.

См. также А. А. Чупров.—Выборочное исследование. Труды подсекции статистики XII с'езда естествоисп. и врачей стр. 161.

веденной оценкой А. А. Кауфмана, так как несколькими страницами ниже А. А. Кауфман стоит на той же точке зрения: „При отборе селений суб'ективный элемент неустраним“ и далее, после некоторых смягчающих для Вятского выборочного исследования обстоятельств, он продолжает: „Тем не менее, наилучшим всетаки приходится признать такие приемы, где отбор подлежащих описанию единиц или случаев исследуемого массового явления всецело предоставлен случаю—конечно, не в том грубожитейском смысле („что попадется под руку“), с каким мы встретились, говоря о лишенных планомерности частичных исследованиях, а в том, как этот термин понимается в теории вероятностей; иначе сказать—где отбор производится по какому нибудь механическому принципу, не стоящему ни в какой связи, ни прямой, ни косвенной, с целями и задачами данного исследования и устраняющему всякое влияние суб'ективного усмотрения“[1]).

§ 7. Выборочные исследования в России с механическим отбором.

К этим условиям отбора приближается второе направление земских выборочных исследований: крестьянских бюджетов в Калужской губ. (в Козельском у. в 1896 г., Лихвинском, Перемышльском и Калужском уездах в 1897 г.), проведенных А. В. Пешехоновым, известное выборочное исследование Пензенской губ. 1911—1913 г.г., организованное В. Г. Громаном, и близко примыкающее к этому исследованию выборочное исследование Донской области, произведенное в 1914—1916 г.г. В. Г. Швецовым. Завершением этого направления выборочных земских исследований был всероссийский опыт выборочного исследования крестьянских хозяйств при сельско-хозяйственной переписи 1916 г., произведенной при ближайшем участии земских статистических сил и под заметным влиянием идей В. Г. Громана.

Эти выбпрочные исследования характерны тем, что все они сопровождают краткую сплошную перепись, при чем отбор производится не типических селений, а хозяйств во всех селениях механическим способом. Так в предисловии описания Козельского уезда мы читаем: „исследователям было предложено подвергать подробному описанию один из каждых 10 дворов, при чем в выбор хозяйств не вносить ни какого суб'ективизма. Для этого предложено было описывать хозяйства, которые будут внесены в ведомость под номерами 1, 11, 21, 31 и т. д. Механичносгь выбора при отсутствии какого-либо предварительного исследования считалась при этом лучшим условием, которое могло при значительном числе описанных хозяйств обеспечить типичность их“[2]).

Интересно отметить, что А. В. Пешехонов применил механический отбор не потому, что у него к тому были серьезные теоретические соображения в преимуществе механического отбора перед всяким другим, а потому, что при постановке исследования у него не имелось данных для сознательного отбора единиц выборочной группы по признаку типичности. Последнее обстоятельство и явилось одной из существенных причин возникновения в земской статистической практике первого выборочного исследования с механическим отбором единиц выборочной группы.

---

[1]) А. А, Кауфман.—А вопросу о выборочном исследовании. Сборник статей стр. 373.

[2]) См. Статистическое описание Калужской губ. т I. Козельской уезд. Калуга 1898 г.

§ 8. Сопоставление выборочных исследований с типическим и механическим отбором.

Интересно теперь сопоставить по основным достижениям и теоретическому значению выборочные исследования Козельского у. и Вятской губ., как представителелей двух самостоятельных направлений земских выборочных исследований, которые в новейшей русской статистической практике оба довольно широко осуществляются.

Выборочное исследование Козельского уезда наглядно иллюстрировало возможность применения выборочного метода при достаточной численности выборочной группы и тогда, когда не было возможности найти, привычные до этого статистикам, типические об'екты наблюдения. Это был шаг вперед, расширявший возможное применение метода.

Самое большое в то время Вятское выборочное исслодование наглядно показало возможность применения выборочного метода не как дополнительного при сплошном наблюдении, а как в большой степени заменяющего функции последнего.

И в том и другом исследовании налицо основные предпосылки для успешного приложения метода, именно, в обоих осуществлено требование достаточно большого числа единиц выборочной группы и равномерность их распределения по пространству генеральной совокупности, при чем, в первом эта равномерность была обеспечена механическим отбором, а во втором наличием большого числа районов, на которые была разбита территория губернии.

Дальнейшее сопоставление оказывается уже не в пользу Вятского выборочного исследования, или вернее, не в пользу выборочного исследования, при котором отбор единиц наблюдения производится по принципу типичности. В самом деле, выборочное исследование с типическим отбором единиц выборочной совокупности, при условии достижения удовлетворительных результатов, безусловно, предполагает предварительное и притом дробное районирование генеральной совокупности, так как иначе нельзя наметить типичные об'екты наблюдения. Но очень часто исследователь не имеет достаточно хороших данных для осуществления этой ответственной и кропотливой работы; тогда выборочное исследование с типическим отбором или невозможно, или отбор выборочной группы будет настолько произвольным, что результаты выборочного исследования будут весьма сомнительными.

Районирование при всякой статистической работе всегда целесообразно, т. к. прежде всего оно позволяет сложные социальные массы разбить на относительно однородные части. В этом смысле оно полезно и для выборочного исследования с механическим отбором, которое, однако, может дать вполне удовлетворительные результаты и без предварительного районирования, так как в процессе наблюдения выборочное исследование с механическим отбором не зависит от районирования, ибо механический отбор одинаково распространяется на все районы и каждое, например, 11 хозяйство должно быть опрошено во всех районах.

В то же время преимущества районирования могут быть использованы во второй стадии работы, т. е. при разработке результатов выборочного наблюдения. Эта разработка может быть произведена по районам, материал для определения которых может дать само выборочное исследование. Здесь технический процесс районирования сокращается во столько раз, во сколько выборочное исследование сокращает работу по сравнению с сплошным наблюдением.

Если же предварительное, удовлетворительно произведенное районирование имело место, то им конечно следует воспозьзоваться и при выборочном исследовании с механическим отбором. Но и здесь механический отбор представляет некоторые преимущества, так как при нем совершенно устраняется необходимость технической работы по определенню типичных об'ектов внутри районов, при чем эти районы в данном случае вовсе не должны быть такими дробными, как при выборочном исследовании с отбором типичных единиц в выборочную группу.

Особенно выпукло выступают преимущества выборочного исследования с механическим отбором, когда возникает вопрос о теоретическом обосновании выборочного метода и оценке точности его результатов. Для того, чтобы выборочное исследование допускало об'ективную оценку точности своих результатов, необходимо, чтобы численность выборочной группы была бы достаточно велика, и чтобы отбор единиц ее исключал всякий произвол и тенденциозность. Этому требованию удовлетворяет случайный и механический отбор достаточно большого числа единиц выборочной группы. Но во всяком случае этому требованию не удовлетворяет отбор даже и большой части экземпляров массы по принципу типичности.

Выборное исследование с случайным отбором непосредственно опирается на авторитет теории вероятностей; как мы увидим ниже, косвенно возможно подвести под авторитет теории вероятностей также выборочное исследование с механическим отбором. Благодаря этому, как при механическом, так и случайном отборе мы получаем возможность обосновать целый ряд важнейших принципиальных предпосылок выборочного метода и ответить на такие существенные вопросы: что значит достаточно большое число единиц выборочной группы, и каково оно должно быть в каждом конкретном случае, как получить об'ективные теоретические критерии точности результатов выборочного исследования, т. е. как выяснить достоинства его относительных, средних и обсолютных величин и т. д. Все это заставляет нас придти к большей предпочтительности выборочного иследования с случайным или механическим отбором перед таковым же отбором единиц по признаку типичности. Последнее не может выйти из сферы эмпирических гаданий, иногда удачных, иногда нет, и не может претендовать, вообще говоря, на теоритически обоснованную общеобязательность своих результатов и обобщений.

§ 9. Выборочное исследование в построениях В. Г. Громана.

Счастливая идея—применить механический отбор, позволившая организатором выборочного исследования Козельского у. обойтись без отбора по признаку типичности, открыла новый путь для дальнейшего развития земских выборочных исследований, весьма интересным этапам которого явились получившие вполне заслуженную известность оригинальные работы В. Г. Громана.

Оригинальность построений Громана заключается прежде всего в комбинированном применении и разграничении методов сплошного, выборочного и монографического исследований.

В начале своих рассуждений В. Г. Громан выдвигает положение, что „выборочный и монографический методы, взятые сами по себе, не могут быть названы статистическими, так как не позволяют учесть с полной гарантией достоверности число явлений, характеризуемых теми или иными чертами"...

..."Комбинирование же методов выборочного исследования, монографического описания и сплошного учета при условии точного разграничения круга явлений, подлежащих исследованию каждым из этих методов, дает полную возможность как уловить закономерности общественной жизни, так и учесть все явления, отличающиеся друг от друга интересующими исследователя признаками" [1]).

Далее, исходя из факта закономерности и взаимной обусловленности общественных явлений, Громан классифицирует факторы, подлежащие статистическому изучению, на три группы: на факторы основные, или признаки причины,—производные и признаки следствия. "Логически ясно", говорит он, "что если мы хотим изучить признаки следствия, то должны разгруппировать явления по признакам причинам и если мы удачно выбрали причины, то тем меньше нужно показаний, чтоб изучить признаки следствия"...

..."Таким образом, данных о признаках следствиях может быть собрано меньше, чем об основных факторах. О последних же, дающих возможность произвести самую группировку явлений и учесть число явлений различных групп, должны быть собраны исчерпывающие данные" [2].)

Таким образом, предлагаемый В. Г. Громаном метод превращается в следующее трехстепенное исследование. Первая степень сплошное исследование фактов первого порядка. Вторая степень—выборочное изучение фактов второго порядка по значительной части случаев. Третья степень—монографическое изучение явлений—следствий по небольшой части случаев [3]).

Кроме соображений о синтезе методов статистического наблюдения и порядка их разграничения, мы находим в докладе Громана указания об относительной численности выборки и об основаниях отбора единиц ея [4]).

Таковы в главных чертах соображения Громана, как он их высказал в докладе на XII с'езде естествоиспытателей и врачей в 1910 г.

Не возражая по существу против возможности группировки хозяйственных факторов на производящие, производные и следствия, отметим, однако, что утверждение о том, что выборочный метод, взятый сам по себе, не является статистическим, так как не позволяет учесть с полной гарантией достоверности числа явлений, не вполне правильно. По отношению к монографическому методу, основывающемуся на типическом отборе, это утверждение может иметь место, но по отношении к выборочному нет. Всякий метод для того, чтобы быть статистическим, должен опираться на достаточно большое число наблюдений, логически это не означает что число наблюдений непременно должно быть исчерпывающим. Выборочный метод непосредственно вытекает из закона больших чисел и, опираясь на него, он является одним из методов массового статистического наблюдения. Полученные при помощи его средние и относительные велины теоретически равноценны таковым же, полученным на основе

---

[1]) Труды подсекции статистики XII с'езда естеств. и врачей. Доклад В. Г. Громана—о сочетании краткого сплошного и подробного исследований... стр. 208, 209.

[2]) Подворная перепись крестьянских хозяйств. Часть II. Пенза. 1913 г. Предисловие стр. III.

[3]) Труды подсекции статистики XII с'езда естеств. и врачей стр. 210.

[4]) Там же стр. 213.

сплошного наблюдения [1]). В сущности говоря, ни один метод статистического наблюдения не может дать с абсолютной гарантией достоверности числа явлений. Результат статистического опроса зависит не только от благих намерений исследователя, но и от об'екта наблюдения. Стоит, например, только вспомнить условия производства сельско-хозяйственной переписи 1920 г. и отношение к ней крестьян, чтобы усумниться в достоверности многих чисел полученных в результате этой переписи.

Несовсем ясно, о каких числах говорится в докладе В. Г. Громана. Если говорится о числе единиц генеральной совокупности, (наприм. хозяйств) то это верно, так как при выборочном исследовании их вообще точно получить нельзя, и они должны быть известными заранее. Получение данных о числе единиц массы есть неот'емлемая сфера сплошных наблюдений. Но все другие абсолютные числа об отдельных признаках единиц генеральной совокупности могут быть получены с достаточно большим приближением к таким же величинам, полученным на основании сплошного наблюдения. Ошибки средних, относительных и обсолютных величин, получаемых в результате выборочного исследования, всегда можно оценить, т. е. знать каждый раз в каких гриницах заключается истинное значение измеряемого признака. А это и есть те гарантии, которые считаются достаточными во всех опытных науках, широко применяющих практику приближенных вычислений.

Так, примерно, и смотрит на это и В. Громан, спустя год после доклада. В предисловии к подворной переписи крестьянских хозяйств Пензенской губ. находим: „если будет описано значительное число входящих в группу явлений, то соотношение между признаками в совокупности всех описанных явлений будет тоже самое, что и во всей группе. Если даже нас интересует абсолютное количество признаков во всех явлениях данной группы, то оно может быть получено простой счетной операцией, заключающейся в перемножении средних и относительных величин, полученных из совокупности описанных явлений, на все число явлений в группе“.

А если это так, то тогда группировка факторов на признаки причины, производные и признаки следствия принимает для разграничения сфер выборочного и сплошного исследований уже условное значение, так как, вообще говоря, большинство признаков причин, рассматриваемых В. Громаном (размер семьи, землевладение, скотовладение и друг.), могут быть с успехом учтены и выборочным методом; на долю сплошного наблюдения остается тогда только учет единиц статистического наблюдения. Но и эту точку зрения нельзя выдержать до конца, так как все зависит от ставимых нами целей исследования. Если нас интересуют статистические признаки совокупности без отношения их к отдельным единицам ее, то, зная общее число единиц этой совокупности, мы можем обойтись, конечно, и без сплошного наблюдения. Если же, помимо данных характеризующих генеральную совокупность в целом, нам надо иметь сведения о все более и более мелких частях ее (о селах, обществах и даже отдельных хозяйствах), т. е. когда на сцену выступают идеографические моменты, то мы вынуждены будем в конце концов произвести сплошное, исчерпывающее описание, как об этом и говорит В. Г. Громан на III стр. указанного предисловия.

[1]) См. аналогичные соображения Р. М. Орженцкого. Труды подсекции статистики XII с'езда естеств. и врачей в прениях по докладу В. Громана стр. 36.

Впрочем, как мы увидим ниже, это положение прямо вытекает из смысла выборочного исследования, именно: при одной и той же величине практически допускаемой ошибки относительная численность [1]) выборочной группы тем меньше, чем больше численность генеральной совокупности, и при небольшой численности генеральной совокупности относительная численность выборочной группы может стать настолько большой, что выборочное исследование фактически должно перейти в сплошное (см. вторую таб. в II главе).

На основе изложенных построений Громана было произведено им исследование Пензенской губ. в 1911—1913 г.г. Это исследование было пятистепенным, при чем, в зависимости от степени полноты программ были произведены: сокращенное, краткое, подробное, специальное и бюджетные, описания. Каждая более полная программа включала все вопросы менее полной. Порядок описания хозяйств был следующий: краткому описанию подвергалось каждое третье хозяйство; подробные описания составлялись на каждое третье, кратко описываемое; специальный бланк составлялся на каждое третье, подробно описываемое хозяйство; для бюджетного описания должны были быть выбраны по 25 хозяйств на каждый уезд, на все же остальные хозяйства составлялись сокращенные описания [2]). Выборочное исследование, следовательно, здесь трехстепенное; отбор установлен для первой степени каждого 3-го хозяйства, для второй—9-го и для третьей—каждого 27-го хозяйства.

Нельзя не отметить, что вся эта схема в общем сложна, а разделение выборочного исследования на три степени в значительной мере произвольно. Это—во-первых; а во-вторых, почему в отбор попали 3, 9 и 27 хозяйства, почему не 4, 12 и 36, или какая нибудь другая комбинация. Необходимость произведенного отбора ни откуда не вытекает.

Впрочем, этот вопрос следует задать не только этому исследованию, но и всем другим выборочным исследованиям с механическим отбором единиц выборочной группы, произведенным до Пензенского и после него.

Пропорция отбора выборочной группы обычно определяется на практике на основании опыта, эмпирически, с помощью, как говорит А. А. Кауфман, статистического чутья и чувства цифры. Конечно, подсознательные, интуитивные переживания играют видную роль в процессе творчества, но однако, ими ограничиваться нельзя.

Рационализация вопроса о числе и способах отбора имеет для выборочного метода такое же важное значение, как и рационализация оценки точности получаемых результатов.

Теоретическое разрешение этих вопросов освобождает выборочный метод от последних остатков элемента произвольности и делает его научно обоснованным методом.

§ 10. Новейшие выборочные исследования.

По более простой схеме (чем Пензенская) была произведена Всероссийская сельско-хозяйственная перепись 1916 г. Из общего числа крестьянских хозяйств 95% было переписано по краткой программе (сплошная перепись), 3% по карточке № 2 и 2% по карточ-

---

[1]) Под относительной численностью мы понимаем отношение численности выборочной группы к численности генеральной совокупности.

[2]) Подворная перепись крестьянских хозяйств. Предисловие стр. I. Пенза 1913 г

ке № 3 (выборочные исследования). Отбор хозяйств для описания по этим формам производился механически.

К глубокому сожалению, эти выборочные исследования 1916 г. остались в общем не разработанными, и результаты первого Всероссийского опыта выборочного метода до сих пор неизвестны.

Второй опыт применения во Всероссийском масштабе выборочного метода был произведен уже в после революционный период русской статистики Центральным Статистическим Управлением и его местными статистическими организациями в 1919 году. С этого времени в том или другом виде, в особенности в текущей сельско-хозяйственной статистике, выборочные исследования получили в русской статистической практике небывалый размах применения.

Выборочные исследования, проведенные Ц. С. У, опираются на выработанные земской статистикой методологические положения, причем, видимо, одинаково признаны оба направления, как с отбором по принципу типичности, так и с механическим отбором единиц выборочной группы. К первому примыкает выборочная перепись 1919 г. и последующие выборочные гнездовые исследования динамики крестьянских хозяйств, к второму, с некоторыми отклонениями,—выборочные исследования, произведенные в практике текущей сельско хозяйственной статистики.

По тщательности своего выполнения последние выборочные исследования значительно уступают ранее произведенным, что, впрочем, принимая во внимание работу статистических органов в тяжелых революционных условиях, вполне естественно и неизбежно. К тому же результаты этих выборочных исследований не оценены.

Возможность применения выработанного земской статистической практикой эмпирического приема проверки путем сопоставления результатов выборочного исследования с сплошным наблюдением, строго говоря, сомнительна, так как с одной стороны сплошное наблюдение (с. х. перепись 1920 г.) сама нуждается в подкреплении и проверке, а с другой, необычайная в последнее время изменчивость сельско-хозяйственных элементов создает к этому большие препятствия. Нельзя быть уверенным, чему обязано возможное различие в коэффициентах дефектам-ли выборочного исследования, или же фактически происшедшим изменением.

Априорные приемы оценки пока еще в статистической практике не применяются. Да об этом трудно и говорить, раз еще не достигнута достаточно удовлетворительная налаженность организации выборочного наблюдения, а следовательно нет гарантий отсутствия преднамеренных погрешностей.

Поэтому ценность добытых современными выборочными исследованиями данных не известна и возбуждает сомнения; это особенно относится к распространенным (для всей генеральной совокупности) итогам, оценки которых также не имеется.

При современных выборочных исследованиях, когда сомнительна эмпирическая проверка путем сопоставления коэффициентов выборочного исследования с сплошным, которая, вообще говоря, довольно произвольна, особенно чувствуется необходимость независимых от сплошного наблюдения теоретически обоснованных приемов оценки точности результатов выборочного исследования,

В последнее время в Русской статистической практике появилось еще весьма важное применение идеи выборочного метода в самой ценной и интересной области статистического анализа, именно, при комбинационной разработке материала, получаемого в результате сплошного наблюдения. Особенность этого применения принципа выборочного исследования заключается в том, что здесь выборочный метод выступает не как метод первичного наблюдения, а пользуется уже готовым первичным материалом, при чем отбор единиц выборочной группы производится не в натуре, а по карточкам сплошного наблюдения, преимущественно механически.

Впервые выборочная комбинационная разработка была применена Г. И. Баскиным при Статистическом Бюро Самарского земства в 1911—1913 г.г.

В 1916 г. Отдел переписи М-ва земледелия, столкнувшись с необходимостью разработать 18 миллионов подворных карт, полученных в результате сельско-хозяйственной переписи 1916 г., встал на путь обоснования выборочной комбинационной разработки.

Под руководством С. С. Кона были поставлены интересные, многообещавшие исследования по дисперсии элементов крестьянского хозяйства. Эти исследования в общем, кажется, были закончены, но к настоящему времени опубликована только, к сажелению, одна глава [1]) (методологическое введение), которая представляет большой интерес по существу затронутых в ней основоположений выборочного метода и в частности по критическим замечаниям относительно неудовлетворительности эмпирических приемов проверки достаточности установленного процента выборки и точности результатов выборочной комбинационной разработки.

Выборочная комбинационная разработка была применена к материалу сельско хозяйственной переписи 1917 г., но осталась не законченной. В настоящее время такая разработка производится по материалу сельско-хозяйственной переписи 1920 года.

Особо следует отметить очень интересную, в методологическом отношении работу, предпринятую Отд. Тек. Сельско-хоз. Статистики Ц. С. У. в 1923 г., по сопоставлении площадей посева и численности скота по одним и тем же хозяйствам переписи 1920 г. и выборочного (тридцатидворного) исследования 1921 г., включившего, кроме сведений за 21 г., аналогичные данные (по воспоминаниям) за 1920 г.

Здесь выборочное исследование принимает на себя новую обязанность: проверить совпадение или расхождение показаний, относящихся к одному и тому же предмету, времени и даваемых одними и теми же лицами.

В данном случае выборочное исследование выступает уже с контрольными целями по отношению к вызывающему некоторые сомнения сплошному исследованию.

К аналогичной цели, для определения точности материала переписи 1920 г. и для проверки данных 30-тидворного обследования, привлекалось и выборочное динамическое исследование.

Последнее сопоставление, не лишенное известного практического интереса, теоретически будет законно лишь тогда, когда имеются гарантии достаточной точности результатов выборочного исследования. Как же можно установить эту точность? Оставаясь на эмпирической почве, это можно сделать единственным приемом — сопоставления коэффициентов выборочного и сплошного исследований, но тут мы, очевидно, вращаемся в кругу, так как сплошное исследование само проверяется данными выборочного наблюдения.

---

[1]) С. С. Кон.—К вопросу о применении выборочного метода при разработке сельско-хозяйственных переписей 1917 г.

Первый прием, т. е. сопоставление одних и тех же карточек переписи 20 г. с воспоминаниями, относящимися тоже к 20-му году по выборочному исследованию 1921 г., не вызывает принципиальных сомнений, так как здесь делается сопоставление не коэффициентов выборочной группы с коэффициентами сплошного наблюдения, а установление идентичности показаний одних и тех же хозяйств. Здесь как бы исскуственно создается репрезентативность выборочной группы с генеральной совокупностью, которая ограничивается тем же числом единиц и их полным соответствием.

Факт расхождения или совпадения показаний может быть установлен вполне законно; но если полученные коэффициенты расхождения полагают распространить на всю генеральную совокупность, то для законности этого должна существовать уверенность в репрезентативности выборочной группы всей массе, т. е. должна быть предварительная оценка достаточности отбора и результатов выборочного исследования.

Второй прием, вызывая вообще принципиальные возражения, в данном случае из порочного круга выйти не может без произвольного допущения, что или перепись совершенно верна или выборочное исследование вполне удовлетворительно. Этой неприятности можно избежать только тогда, когда будут применяться способы оценки репрезентативности выборочного исследования и точности его относительных и средних величин независимо от сопоставления его результатов с результатами сплошного наблюдения. Эти способы и критерии точности, позволяющие выборочному методу получить самодовлеющее значение, дает теория вероятностей.

§ 11. Метод Bowley.

Все вышеизложенное приводит нас к необходимости теоретической постановки и разрешения поднятых основных задач выборочного исследования. Первым, реально вступившим на путь научного обоснования выборочного метода, был известный английский ученый А. Bowley, который указал на основе применения теории вероятностей, или вернее теории ошибок, более строгие, чем эмпирические приемы, априорные критерии точности результатов выборочного исследования.

Теория А. Bowley [1]) дает возможность при известных условиях найти, во первых, в зависимости от данной численности выборочной группы, те теоретически допустимые пределы, за которые не выйдут отклонения результатов выборочного исследования от наивероятнейшего значения измеряемых признаков с вероятностью, как угодно, близкой к единице (т. е. достоверности) и, во вторых, дает возможность, отправляясь от допустимой величины этих отклонений, определить достаточное число единиц выборочной группы.

Для решения этих основных задач выборочного метода предпосылки теории А. Bowley требуют соблюдения следующих условий: во первых, предполагаемые к описанию единицы выборочной группы выхватываются из единиц генеральной совокупности при ничем не ограничиваемой свободной игре случая, а, во вторых, подвергнутые регистрации единицы должны возвращаться после каждого испытания назад в общую массу, так что возможно, что некоторые единицы генеральной совокупности могут подвергнуться перечету несколько раз. Словом, здесь требуется соблюдение условий неизменной вероятности, иначе говоря, схемы возвращенных шаров.

[1]) см. Elements of statistics. London. 1902. стр. 308—313.

Теория погрешностей устанавливает, что, при соблюдении этих условий, погрешность выводов будет тем меньше, чем больше будет число наблюдений (т. е. единиц выборочной группы), а именно, величина погрешности будет обратно пропорциональна квадратному корню из числа наблюдений. При этих условиях численность генеральной совокупности не играет никакой роли; все дело здесь в абсолютном числе наблюдений.

Логически вытекает, что по теории Bowley, для получения идеально точных результатов следовало бы подвергнуть описанию безконечное число единиц.

В статистической практике исследуемые совокупности, обыкновенно, конечны (за исключением случаев, встречающихся в ботанике, зоологии, растениеводстве и в некоторых других естественных науках, где совокупность практически может быть бесконечной).

Таким образом получается на первый взгляд несколько неудобный вывод, что даже при сплошном исследовании поголовно всех единиц данной совокупности результаты всетаки не будут вполне верными. Но надо иметь ввиду. чти построения A. Bowley опираются на предположение, что регистрируемые единицы возвращаются обратно в общую массу и следовательно единицы могут подвергнуться описанию по нескольку раз; а при таком предположении число исследуемых единиц может быть доведено до сколь угодно большого числа.

Как бы то ни было, но в теории Bowley статистика мало утешает необходимость повторной регистрации единиц, однажды уже попавших в описание. Даже, если бы статистику понадобились совершенно точные результаты, то ясно, что он не стал бы, опираясь на теорию Bolwey, регистрировать бесчисленное множество экземпляров, да при том еще некоторые из них по несколько раз, а предпочел бы попросту произвести сплошное исследование.

С другой стороны, как было упомянуто выше, в основе построений Bowley лежит еще другое требование, чтобы единицы отбирались случайно (по жребию). Соблюдение этого требования на практике, в случае не гнездовой выборки, встречает большие неудобства, так как такой способ отбора предполагает наличие подробного списка всех единиц, входящих в генеральную совокупность, чтобы сделать жребиевку тех единиц, которые после этого должны быть подвергнуты описанию. Это связано с значительной работой, которая излишне осложняет применение выборочного метода.

На практике для достижения об‘ективности отбора, а также для равномерного распределения единиц выборочной группы по всему пространству генеральной совокупности применяется известный нам механический отбор. Но при этом способе отбора оказывается нарушенным не только первое требование A. Bowley (случайность отбора), но и второе требование, т. е. принцип возвращаемости единиц после каждого наблюдения, так как при механическом отборе ни одна единица не может быть зарегистрирована более одного раза, если не прибегать к ничем неоправдываемому произволу.

Таким образом, в сущности, невозможно найти способ отбора, достаточно удобно осуществимый на практике, при котором соблюдались бы в значительной мере оба требования A. Bowley. Выборочный метод в построениях A. Bowley недостаточно гибок и, пожалуй, слишком формален, чтобы учитывать требования практического удобства.

§ 12. Задачи настоящего исследования

Все эти соображения заставляют искать других, более удобных схем обоснования выборочного метода, применительно к капризным условиям социальной статистики, т. е. побуждают заняться разработкой более общего случая выборочного метода, когда подвергнутые однажды описанию единицы вторично уже не регистрируются, т. е. когда налицо условия переменной вероятности, соответствующие схеме невозвращенных шаров.

Если исходить из принципа невозвращаемости единиц, то оказывается, что для достижения данной точности существенную роль играет не только абсолютное число отобранных единиц, но и пропорция их по отношению ко всей массе генеральной совокупности[1]).

Далее оказывается, что в этом случае формулы погрешностей, выведенные в предположении случайного отбора, легко могут быть приспособлены к случаю механического отбора. При этом здесь обнаруживается, что наиболее точные выводы получаются как раз при механическом отборе, а не при случайном. Поэтому в дальнейшем разработка выборочного метода дается преимущественно в предположении механического отбора, применительно не только к относительно однородным совокупностям, но и к неоднородным, когда для достижения хороших результатов полезно изучаемую совокупность разбить на более однородные частные совокупности.

Так как применение выборочного метода для получения относительных и средних величин имеет свои особенности, то в дальнейшем, несмотря на то, что выборочный метод для относительных величин можно рассматривать как частный случай более общего случая выборочного метода для средних величин, им обоим посвящены специальные главы.

В последней главе трактуется применение теоретических положений выборочного метода на практике при организации выборочного исследования и при разработке результатов наблюдения; здесь рассматриваются вопросы: определение численности выборочной группы, оценка точности распространенных средних и относительных величин, получение распространенных итогов и их оценка.

Чтение II и III глав предполагает знакомство с элементами высшей математики и особенно с теорией вероятностей, включая математические ожидания (преимущественно теорем о математическом ожидании суммы и произведения[2]).

Следует отметить, что в дальнейшем изложении в качестве меры погрешностей мы будем пользоваться, ради удобства и практического значения, средне-квадратической погрешностью, под которой мы понимаем квадратный корень из математического ожидания квадрата погрешности[3]). Для сокращения речи мы будем во всем дальнейшем изложении называть эту погрешность просто средней погрешностью.

---

[1]) Этот вопрос в русской статистической литературе вызвал оживленное обсуждение. См.:

А. А. Чупров.—Выборочное исследование.
Труды подсекции статистики на XII с‘езде естеств. и врач. стр. 163.

А. А. Кауфман.—К вопросу о выбор. исслед. Сборник статей, стр. 359—361.
его-же —Введение в теоретическую статистику, стр. 342—347.

Е. Е. Слуцкий.—Статистика и математика. Статистический Вестник 1915—16 г. кн. 3—4, стр. 116, 117.

С. С. Кон.—К вопросу о применении выборочного метода при разработке сельско-хоз. переписей стр. 13 (примечание).

[2]) см. А. А. Марков.—Исчисление вероятностей. 3-е изд. стр. 36—58.

[3]) там-же стр. 229.

Так как в русской литературе да отчасти и в иностранной, мерой случайных погрешностей служит часто модуль, то считаем не лишним напомнить, что между модулем M и средней ошибкой σ существует соотношение

$$M = \sqrt{2}\,\sigma.$$

Пользуясь этим соотношением, легко вычислить по таблице интеграла[1])

$$\frac{2}{\sqrt{\pi}}\int_0^x e^{-t^2}dt$$

вероятность того, что ошибка не превзойдет $t\sigma$ (т. е. t—кратную среднюю ошибку).

Для этого, выражая σ через M, получаем

$$t\sigma = \frac{t}{\sqrt{2}}M$$

и ищем значение интеграла для $x = \frac{t}{\sqrt{2}}$.

Чтобы избавить читателя от этих вычислений, приводим ниже краткую табличку значений этих вероятностей P для различных значений t.

t	Вероятность	t	Вероятность	t	Вероятность
0,0	0,000	1,0	0,683	2,0	0,954
0,1	0,080	1,1	0,729	2,2	0,972
0,2	0,159	1,2	0,770	2,4	0,984
0,3	0,236	1,3	0,806	2,6	0,991
0,4	0,311	1,4	0,838	2,8	0,995
0,5	0.383	1,5	0,866	3,0	0,9973
0,6	0,451	1,6	0,890	3,5	0,99953
0,7	0,516	1,7	0,911	4,0	0,999937
0,8	0,576	1,8	0,928	5,0	0,9999994
0,9	0,632	1,9	0,943	—	—

Пример. Найти вероятность того, что ошибка не превосходит трех средне-квадратических отклонений, т. е. 3σ.

В данном случае $t = 3$, следовательно, искомая вероятность по нашей таблице будет 0,9973.

Если расчитать, в скольких случаях фактическая ошибка выйдет за пределы 1σ, 2σ, 3σ и 4σ, то окажется, что за пределы одной σ ошибка может выйти в среднем в $^1/_3$ случаев, за пределы двух σ ошибка выйдет лишь в $^1/_{22}$ части случаев, ошибка более трех σ может быть только в $^1/_{370}$ части случаев, более четырех—в $^1/_{16000}$ части случаев и т. д. Из последнего следует, что ошибка более трех средних отклонений—3σ почти невероятна; поэтому 3σ практически можно принять за высший предел фактических погрешностей.

[1]) см. А. А. Чупров.—Очерки по теории Статистики стр. 424, А. А. Марков—Исчисление вероятностей стр. 375.

ГЛАВА II.

Об относительных величинах.

I. Средняя ошибка в схеме невозвращенных шаров.

§ 13. Предварительные замечания.

При научной разработке выборочного метода применительно к относительным величинам мы будем оперировать, как это обычно принято, с урнами и шарами. Необходимо здесь же сказать, что введение урн и шаров нисколько не умаляет ценности и общности наших выводов, так как ясно, что при определении пропорции единиц генеральной совокупности, обладающих данным признаком по такой же пропорции в выборочной группе, природа единиц генеральной совокупности не играет никакой роли.

Степень точности, полученных при выборочном исследовании выводов существенно зависит от способа отбора единиц выборочной группы. Для возможности об'ективной оценки этой степени точности необходимо, чтобы отбор единиц выборочной группы совершался по способу, гарантирующему отсутствие тенденции и произвола. Такими способами отбора являются, например, механический отбор и отбор единиц по жребию, т. е. случайный отбор. При механическом отборе отсутствует элемент случайности и потому, при применении его, оценка точности выводов не может быть—без некоторых допущений—произведена на основе теории вероятностей. Иначе обстоит дело при случайном отборе. При нем все зависит, так сказать, от игры случая и следовательно à priori можно заключить, что здесь такая оценка возможна.

Прежде чем произвести эту оценку, остановимся вкратце на некоторых положениях теории вероятностей. Теория вероятностей учитывает, как известно, лишь случайные погрешности. Поэтому, производя на основе теории вероятностей оценку какого либо приближенного равенства, нам необходимо прежде всего убедиться, что наше равенство свободно от постоянной погрешности. Для этого достаточно, как известно, показать, что математическое ожидание возможной погрешности нашего равенства равно нулю. При выполнении последнего условия случайная погрешность нашего равенства может быть учтена при помощи средней погрешности, которая равна, как известно, корню квадратному из математического ожидания квадрата случайной погрешности. После этих кратких замечаний перейдем к основной задаче настоящей главы.

§ 14. Вывод основных формул.

Представим себе урну, содержащую N шаров, из которых M белых. Число M предполагается неизвестным. С целью определить пропорцию $\frac{M}{N}$ белых шаров в урне мы вынимаем из нее на удачу n шаров ($n < N$), и пусть среди них оказалось m белых.

Разсматривая пропорцию белых шаров в выборочной группе, как приближенное выражение той же пропорции в генеральной совокупности, будем иметь следующее приближенное равенство

$$\frac{M}{N} \underset{\sim}{=} \frac{m}{n} \ ^{1)} \qquad (1)$$

Это равенство, как мы сейчас покажем, свободно от постоянной погрешности. В самом деле погрешность нашего равенства будет

$$\frac{m}{n} - \frac{M}{N}.$$

Вычисляя математическое ожидание этой погрешности, будем иметь

$$\text{м. о.}\left(\frac{m}{n} - \frac{M}{N}\right) = \text{м. о.}\frac{m}{n} - \frac{M}{N} = \frac{1}{n}\,\text{м. о.}\, m - \frac{M}{N}. \qquad (2)$$

Далее будем искать м. о. m. Из теории вероятностей известно, что вероятность появления m белых шаров среди вынутых n шаров, когда в урне среди общего числа N шаров содержится M белых, определяется выражением

$$\frac{C_M^m \cdot C_{N-M}^{n-m}}{C_N^n},$$

где знаками C_M^m и т. д. мы пользуемся для обозначения числа сочетаний из M элементов по m и т. д. [2]).

Давая m все значения от 0 до N включительно составим сумму произведений различных значений числа m на их вероятности; тогда получим

$$\text{м. о. } m = \sum_{m=o}^{n} \frac{C_M^m\, C_{N-M}^{n-m} \cdot m}{C_N^n}.$$

Так как для всех слагаемых знаменатель C_N^n одинаков, то предыдущее равенство можно представить так:

$$\text{м. о. } m = \frac{\sum\limits_{m=o}^{n} C_M^m\, C_{N-M}^{n-m} \cdot m}{C_N^n}.$$

Ставя себе целью упростить числителя этого выражения покажем, что

$$\sum_{m=o}^{n} C_M^m\, C_{N-M}^{n-m} \cdot m = M \cdot C_{N-1}^{n-1}. \qquad (3)$$

---

1) Знаком $\sim$ мы хотим подчеркнуть, что раренство приближенное.

2). См. А. А. Марков. Исчисление вероятностей. Изд. III. 1913 г. стр. 113. § 21. Задача I.

Имеем

$$(1+t)^M = 1 + C_M^1 t + C_M^2 t^2 + \dots + C_M^m t^m + \dots + C_M^M t^M.$$

Дифференцируя это тождество по t, получим

$$M(1+t)^{M-1} = C_M^1 + 2\,C_M^2 t + \dots + m\,C_M^m t^{m-1} + \dots + M\,C_M^M t^{M-1}. \tag{4}$$

Далее имеем

$$(1+t)^{N-M} = 1 + C_{N-M}^1 t + C_{N-M}^2 t^2 + \dots + C_{N-M}^{n-m} t^{n-m} + \dots +$$

$$+ C_{N-M}^{N-M} t^{N-M}.$$

Перемножая последние два равенства почленно, и выставляя на вид коэффициент при t^{n-1}, получим

$$M(1+t)^{N-1} = C_M^1 + \left(C_M^1 C_{N-M}^1 + C_M^2\, 1.2.\right) t + \dots +$$

$$+ \left(\sum_{m=0}^{n} C_M^m C_{N-M}^{n-m} \cdot m\right) t^{n-1} + \dots$$

С другой стороны

$$M(1+t)^{N-1} = M + M\,C_{N-1}^1 t + \dots + M\,C_{N-1}^{n-1} t^{n-1} + \dots$$

Сравнивая коэффициенты при t^{n-1} последних двух выражений мы и получим искомое равенство (3).

На основании этого равенства будем иметь

$$\text{м. о. } m = \frac{M\,C_{N-1}^{n-1}}{C_N^n} \tag{5}$$

или

$$\text{м. о. } m = \frac{M\,\dfrac{(N-1)!}{(n-1)!\,(N-n)!}}{\dfrac{N!}{n!\,(N-n)!}}$$

и наконец

$$\text{м. о. } m = n\,\frac{M}{N}. \tag{6}$$

Подставляя полученный результат в выражение (2) получим окончательно

$$\text{м. о.}\left(\frac{m}{n}-\frac{M}{N}\right)=\frac{1}{n}\cdot n\,\frac{M}{N}-\frac{M}{N}=0\ .$$

Таким образом мы доказали, что равенство (1) не содержит постоянной погрешности.

Теперь перейдем к оценке случайной погрешности равенства (1). Называя среднюю ошибку через σ будем иметь

$$\sigma^2=\text{м. о.}\left(\frac{m}{n}-\frac{M}{N}\right)^2.$$

Но

$$\text{м. о.}\left(\frac{m}{n}-\frac{M}{N}\right)^2=\text{м. о.}\left(\frac{m^2}{n^2}-2\,\frac{M}{N}\cdot\frac{m}{n}+\frac{M^2}{N^2}\right)=$$

$$=\text{м. о.}\,\frac{m^2}{n^2}-2\,\frac{M}{N}\,\text{м. о.}\,\frac{m}{n}+\frac{M^2}{N^2}=$$

$$=\frac{1}{n^2}\,\text{м. о.}\,m^2-2\,\frac{M}{N}\cdot\frac{1}{n}\,\text{м. о.}\,m+\frac{M^2}{N^2}\,.$$

Далее на основании выражения (6) получим

$$\text{м. о.}\left(\frac{m}{n}-\frac{M}{N}\right)^2=\frac{1}{n^2}\,\text{м. о.}\,m^2-\frac{M^2}{N^2}\,. \qquad (7)$$

Таким образом для вычесления математического ожидания квадрата погрешности нам остается лишь вычислить м. о. m^2. Как и раньше, когда мы вычислили м. о. m, очевидно, будем иметь:

$$\text{м. о.}\ m^2=\sum_{m=o}^{n}\frac{C_M^m\,C_{N-M}^{n-m}}{C_N^n}\cdot m^2=\frac{\sum\limits_{m=o}^{n}C_M^m\,C_{N-M}^{n-m}\cdot m^2}{C_N^n}$$

Покажем теперь, что

$$\sum_{N=o}^{n}C_M^m\,C_{N-M}^{n-m}\,m^2=M\left[C_{N-1}^{n-1}+(M-1)\,C_{N-2}^{n-2}\right]. \qquad (8)$$

В самом деле, умножая тождество (4) на t, получим

$$M\,(1+t)^{M-1}\,t=C_M^1\,t+2\,C_M^2\,t^2+\ldots+m\,C_M^m\,t^m+\ldots,$$

откуда дифференцируя по t, найдем

$$M\left[(1+t)^{M-1}+(M-1)\,(1+t)^{M-2}\,t\right]=C_M^1+2^2\,C_M^2\,t+\ldots+$$

$$+ m^2 C_M^m t^{m-1} + \dots$$

Кроме того имеем

$$(1+t)^{N-M} = 1 + C_{N-M}^1 t + \dots + C_{N-M}^{n-m} t^{n-m} + \dots$$

Перемножая последние два равенства почленно, получим

$$M\left[(1+t)^{N-1} + (M-1)(1+t)^{N-2} t\right] = C_M^1 + \left(C_M^1 C_{N-M}^1 + 2^2 C_M^2\right)t + \dots$$

$$\dots + \left(\sum_{m=0}^{n} m^2 C_M^m C_{N-M}^{n-m}\right) t^{n-1} + \dots \qquad (9)$$

С другой стороны

$$M(1+t)^{N-1} = M + M C_{N-1}^1 t + \dots + M C_{N-1}^{n-1} t^{n-1} + \dots$$

и

$$M(M-1)(1+t)^{N-2} t = M(M-1) t + M(M-1) C_{N-2}^1 t^2 + \dots +$$

$$\dots + M(M-1) C_{N-2}^{n-2} t^{n-1} + \dots$$

Складывая эти два выражения почленно, найдем

$$M\left[(1+t)^{N-1} + (M-1)(1+t)^{N-2} t\right] = M + M\left[C_{N-1}^1 + (M-1)\right]t + \dots$$

$$\dots + M\left[C_{N-1}^{n-1} + (M-1) C_{N-2}^{n-2}\right] t^{n-1} + \dots$$

Сравнивая коэффициенты при t^{n-1} в выражении (9) и в последнем мы и получим искомое выражение (8).

На основании его будем иметь

$$\text{м. о. } m^2 = \frac{M C_{N-1}^{n-1} + M(M-1) C_{N-2}^{n-2}}{C_N^n} =$$

$$= M \frac{C_{N-1}^{n-1}}{C_N^n} + M(M-1) \frac{C_{N-2}^{n-2}}{C_N^n}$$

Здесь первое слагаемое можно заменить на основании равенств (5 и (6) через n. $\frac{M}{N}$, второе же легко упрощается следующим образом:

$$M(M-1)\frac{C_{N-2}^{n-2}}{C_N^n} = M(M-1)\frac{(N-2)!\,n!\,(N-n)!}{(n-2)!\,(N-n)!\,N!} = (n-1)\,n\,\frac{M(M-1)}{N(N-1)}.$$

Окончательно будем иметь

$$\text{м. о. } m^2 = n\frac{M}{N} + (n-1)\,n\,\frac{M(M-1)}{N(N-1)}$$

Подставляя этот результат в выражение (7), получим

$$\text{м. о.}\left(\frac{m}{n} - \frac{M}{N}\right)^2 = \frac{1}{n}\frac{M}{N} + \frac{n-1}{n}\frac{M(M-1)}{N(N-1)} - \frac{M^2}{N^2} =$$

$$= \frac{M}{N}\left[\frac{1}{n} + \left(1-\frac{1}{n}\right)\frac{M-1}{N-1} - \frac{M}{N}\right] =$$

$$= \frac{M}{N}\left[\frac{1}{n}\left(1-\frac{M-1}{N-1}\right) + \frac{M-1}{N-1} - \frac{M}{N}\right] =$$

$$= \frac{M}{N}\left[\frac{1}{n}\cdot\frac{N-M}{N-1} - \frac{N-M}{N(N-1)}\right] =$$

$$= \frac{M}{N}\,\frac{N-M}{N-1}\left(\frac{1}{n} - \frac{1}{N}\right) = \frac{M}{N}\,\frac{N-M}{N-1}\cdot\frac{N-n}{n.\,N} =$$

$$= \frac{\frac{M}{N}\,\frac{N-M}{N}}{n}\cdot\frac{N-n}{N-1}.$$

Здесь $\frac{M}{N}$ есть вероятность вынуть белый шар при одном испытании. Называя эту вероятность через p и полагая $\frac{N-M}{N} = 1-p=q$, получим

$$\text{м. о.}\left(\frac{m}{n} - \frac{M}{N}\right)^2 = \frac{pq}{n}\,\frac{N-n}{N-1}.$$

Извлекая отсюда квадратный корень мы и получим искомую среднюю ошибку равенства (1)

$$\sigma = \sqrt{\frac{pq}{n}\,\frac{N-n}{N-1}}\,[1]. \tag{10}$$

---

[1]) ср. Czúber Wahrscheinlichkeitsrehnung, B. I, S. 163—164. Б. Ястремский. О дисперсии в схеме невозвращенных шаров. Вестник Статистики 1923 г., книга XV стр. 294.

Так как на практике N число большое, то пренебрегая в знаменателе единицей рядом с N будем иметь

$$\sigma = \sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)} \quad . \qquad (11)$$

В частном случае, если N весьма велико по сравнению с n, то дробью $\frac{n}{N}$ (или точнее — $\frac{n-1}{N-1}$ можно пренебречь и выражение ошибки примет вид

$$\sigma = \sqrt{\frac{pq}{n}} \quad . \qquad (12)$$

В такой же форме представляется ошибка в „схеме возвращенных шаров".

Зная ошибку σ равенства (1) можно представить его, как это часто принято, в виде

$$\frac{M}{N} \approx \frac{m}{n} \pm \sigma \quad . \qquad (13)$$

Умножая обе части его на N, получим

$$M \approx N\frac{m}{n} \pm N\sigma \quad .$$

Здесь M представляет, как известно, абсолютное число белых шаров в урне, $N\frac{m}{n}$ — его приближенное значение, а $N\sigma$ — среднюю ошибку этого приближенного значения.

Если бы ошибки равенства (1) распределялись по закону Гаусса, то вероятность того, что действительная ошибка его не превосходит $t\,\sigma$, выражается, как мы уже говорили на стр. 80, интегралом

$$\frac{2}{\sqrt{\pi}} \int_{0}^{\frac{t}{\sqrt{2}}} e^{-t^2}\, dt$$

Однако, ошибки равенства (1), строго говоря, не подчиняются закону Гаусса; но если величины n и N—n не очень малы, что в практике выборочного исследования обычно и бывает, то отклонение от закона Гаусса будет столь ничтожным, что им можно пренебречь и считать этот закон выполняющимся. Правильность этого положения непосредственно усматривается из вывода средней ошибки в схеме невозвращенных шаров, который помещен в книге Czuber'a „Wahrscheinlichkeitsrechnung und ihre Anwendüngen..." (B. I. S. 163—164).

Впрочем, о вероятности, что ошибка не превосходит $t\sigma$ можно составить себе некоторое представление и независимо от закона Гаусса. Путь к этому открывает теорема Чебышева, по которой эта вероятность больше величины

$$1 - \frac{1}{t^2}\ ^{1)}$$

§ 15. Сопоставление формул в схеме возвращенных и невозвращенных шаров.

Сравнивая среднюю ошибку

$$\sigma = \sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right)}$$

в схеме невозвращенных шаров со средней ошибкой

$$\sigma = \sqrt{\frac{pq}{n}}$$

в схеме возвращенных шаров, мы замечаем, что первая ошибка зависит не только от численности выборочной группы n, но и от ея относительной численности $\frac{n}{N}$. Множитель $\sqrt{1 - \frac{n}{N}}$, которым первая формула отличается от второй, очевидно, всегда меньше единицы. Отсюда следует, что при равной численности выборочной группы n, средняя ошибка в схеме невозвращенных шаров всегда меньше средней ошибки в схеме возвращенных шаров. Если n составляет ничтожную долю N, то множитель $\sqrt{1 - \frac{n}{N}}$ близок к единице и следовательно, обе ошибки почти равны. Если же n приближается к N, т. е. когда выборочное исследование приближается к сплошному, то множитель $\sqrt{1 - \frac{n}{N}}$, а вместе с ним и соответствующая ошибка приближается к 0, как и должно быть. Иной результат получается из второй формулы, которая при $n = N$ дает ошибку равную $\sqrt{\frac{pq}{N}}$, что и понятно, так как, вторая формула выведена в предположении, что шары возвращаются в урну.

$\frac{n}{N}$	$\sqrt{1-\frac{n}{N}}$
0,0	1,00
0,1	0,95
0,2	0,89
0,3	0,84
0,4	0,77
0,5	0,71
0,6	0,63
0,7	0,55
0,8	0,45
0,9	0,32
1,0	0

Для более наглядного сравнения этих формул приводим таблицу значений множителя $\sqrt{1 - \frac{n}{N}}$ при различных значениях отношения $\frac{n}{N}$.

Если бы мы пожелали по заданной ошибке вычислить численность выборочной группы, то при разных численностях генеральной совокупности вторая формула давала бы для n одинаковые значения, первая же разные. В нижеприводимой таблице даны численности выборочной группы, при численностях генеральной совокупности 5.000, 50.000, 500,000, 5.000.000 и заданных средних ошибках $1/50$, $1/100$, $1/1000$,

[1]) См. А. А. Марков. Исчисление вероятностей изд. 3. стр. 63.

когда вероятности p и q равны $\frac{1}{2}$. Рядом с численностями выборочной группы приводятся и относительные ея численности в %.

N	при σ_2 =						при σ_1 =					
	$^1/_{50}$.		$^1/_{100}$.		$^1/_{1000}$.		$^1/_{50}$.		$^1/_{100}$.		$^1/_{1000}$.	
	n.	%	n.	%	n.	%	n.	%	n.	%	n.	%
5000	625	12,5	2500	50,0	250000	5000	555	11,1	1666	33,3	4902	98,0
50000	„	1,2	„	5,0	„	500	617	1,2	2381	4,8	41666	83,3
500000	„	0,1	„	0,5	„	50	624	0,1	2487	0,5	166666	30
5000000	„	0,01	„	0,05	„	5	625	0,01	2499	0,05	238095	4,8

При рассмотрении этой таблицы подтверждается, между прочим, факт отмеченный нами на стр. 74, что относительная численность выборочной группы тем меньше, чем больше численность генеральной совокупности.

§ 16. Случай нескольких урн или районов.

Предшествующие формулы средних ошибок выведены в предположении отбора единиц выборочной группы на удачу. Однако, случайный отбор, как мы видели при разборе метода Bowley, практически почти не осуществим и это обстоятельство заставляет заменить случайный отбор механическим, как наиболее удобным. Но тут возникает вопрос при каких условиях механический отбор равносилен случайному. Чтобы ответить на него, введем некоторую классификацию тех совокупностей, с которыми нам придется иметь дело.

Если единицы с интересующим нас признаком расположены среди единиц генеральной совокупности более или менее случайно (не тенденциозно), то такого рода совокупности мы будем называть относительно однородными, а самое расположение единиц с интересующим нас признаком среди остальных единиц генеральной совокупности будем называть случайным. Примером такой совокупности может служить состав урны с тщательно перемешанными шарами.

Если единицы с интересующим нас признаком расположены не случайно, а как нибудь иначе, то здесь мы можем встретиться с двумя случаями. Может быть, что в одном случае единицы с интересующим нас признаком распределены среди единиц генеральной совокупности более равномерно, чем при случайном их распределении. Такого рода совокупности мы назовем однородными, а распределение единиц с интересующим нас признаком среди всех единиц—равномерным. Примером такого распределения могут служить белые и черные квадраты на шахматной доске. Такие совокупности представляются практически маловероятными, но теории приходится их учитывать. Во втором случае может быть, что единицы с интересующим нас признаком в общей массе расположены менее равномерно чем при случайном их распределении. Такого рода совокупности, почти исключительно встречающиеся на практике назовем не однородными, а распределение единиц с интересующим нас признаком среди всех единиц назовем не равномерным.

Если генеральная совокупность относительно однородна, в вышеизложенном смысле, то ясно, что механический и случайный отбор будут эквивалентны и следовательно, для таких совокупностей, найденные нами формулы средних ошибок применимы и при механическом отборе единиц.

Если отбросить пока тот маловероятный случай, когда материал генеральной совокупности более однороден, чем при случайном распределении единиц, то остается лишь рассмотреть неоднородные совокупности, имеющие наибольшее практическое значение. Так как для таких совокупностей, при осуществлении механического отбора, применение найденных формул было бы не законно, то естественно попытаться расчленить такую совокупность на ряд частных совокупностей или, как мы будем говорить ниже, районов, которые можно было бы считать относительно однородными. Применяя к каждому из таких районов механический отбор, мы можем вычислить для них ошибки относительных величин по вышеприведенным формулам, соединив потом эти относительные величины и ошибки по известным правилам мы получим относительную величину и ошибку для всей генеральной совокупности.

Таким образом, возвращаясь к схемам теории вероятностей, мы приходим к необходимости решить следующую задачу. Пусть перед нами S урн. Назовем через

$$N_1 \;,\; N_2 \;,\; N_3 \;,\; \ldots\ldots \;,\; N_s$$

числа шаров в каждой из них и через

$$M_1 \;,\; M_2 \;,\; M_3 \;,\; \ldots \;,\; M_s$$

соответственно числа белых шаров в них.

Из этих урн извлекается n шаров, причем вынутый шар не возвращается в ту урну, из которой он извлечен. Пусть при этом из числа n извлекаемых шаров n_1 приходится на первую урну, n_2 — на вторую и т. д., n_s — на S—ную урну. Если среди n_1 шаров оказалось m_1 белых, среди n_2 шаров оказалось m_2 белых и т. д., то для какой либо i—той урны (где под i можно разуметь любое из чисел 1, 2, . . . S) имеем на основании предшествующих выводов:

$$\frac{M_i}{N_i} \approx \frac{m_i}{n_i} \pm \sqrt{\frac{p_i\, q_i}{n_i}\left(1 - \frac{n_i}{N_i}\right)} \tag{14}$$

и

$$M_i \approx N_i\,\frac{m_i}{n_i} \pm N_i \sqrt{\frac{p_i\, q_i}{n_i}\left(1 - \frac{n_i}{N_i}\right)} \tag{15}$$

Называя через M общее число белых шаров во всех урнах и вспомнив, что ошибка суммы равна корню квадратному из суммы квадратов ошибок слагаемых будем иметь

$$M = \sum_{i=1}^{s} M_i \cong \sum_{i=1}^{s} N_i \frac{m_i}{n_i} \pm \sqrt{\sum_{i=1}^{s} N_i^2 \frac{p_i\, q_i}{n_i} \left(1 - \frac{n_i}{N_i}\right)} \quad (16)$$

откуда, обозначив $N = \sum_{i=1}^{s} N_i$, где N число шаров во всех урнах, получим:

$$\frac{M}{N} \cong \frac{\sum_{i=1}^{s} N_i \frac{m_i}{n_i}}{\sum_{i=1}^{s} N_i} \pm \frac{\sqrt{\sum_{i=1}^{s} N_i^2 \frac{p_i\, q_i}{n_i} \left(1 - \frac{n_i}{N_i}\right)}}{\sum_{i=1}^{s} N_i} \quad (17)$$

§ 17. Наивыгоднейшая пропорция отбора.

В связи с предыдущей задачей возникает весьма интересный вопрос:

Нельзя ли подобрать пропорции $\frac{n_i}{N_i}$ $(i=1, 2, \ldots S)$ вынимаемых шаров так, чтобы погрешности равенств

$$M \cong \sum_{i=1}^{s} M_i \quad \text{и} \quad \frac{M}{N} \cong \frac{\sum_{i=1}^{s} M_i}{\sum_{i=1}^{s} N_i}$$

были бы наименьшим?

Очевидно, достаточно исследовать одно из этих равенств, т. к. погрешность второго получается из погрешности первого делением на N_i . Кроме того вместо самой погрешности можно исследовать квадрат погрешности. Итак, ставя себе целью найти минимум погрешности формулы (16) будем искать minimúm выражения

$$\sum_{i=1}^{s} N_i^2 \frac{p_i\, q_i}{n_i} \left(1 - \frac{n_i}{N_i}\right) = \sum_{i=1}^{s} N_i^2\, p_i\, q_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) \quad (18)$$

Очевидно, что погрешность эта будет наименьшей и даже равной 0, когда $n_i = N_i$ $(i=1, 2, 3, \ldots, s)$, но этот случай нам не интересен. Мы будем искать условный минимум, а именно, будем искать minimúm выражения (18) при условии, что общее количество вынутых шаров есть величина постоянная. т. е, $n = \sum n_i = Const.$

Поступая по правилам отыскания условного maxima и minima нам надо будет найти minimúm выражения

$$\sum_{i=1}^{s} N_i^2\, p_i\, q_i \left(\frac{1}{n_i} - \frac{1}{N_i}\right) + \lambda \sum_{i=1}^{s} n_i$$

где λ неопределенный множитель.

Дифференцируя это выражение по каждому из переменных n_i ($i = 1, 2, 3,\ldots, S$) и приравнивая результаты нулю получим систему уравнений вида

$$-N_i^2\, p_i\, q_i\, \frac{1}{n_i^2} + \lambda = 0 \qquad (i = 1, 2, 3,\ldots,s)$$

или

$$N_i^2\, p_i\, q_i\, \frac{1}{n_i^2} = \lambda,$$

что можно представить также в виде

$$\frac{p_i\, q_i}{\left(\frac{n_i}{N_i}\right)^2} = \lambda.$$

Отсюда, взяв обратную величину и извлекая квадратный корень, получим

$$\frac{\frac{n_i}{N_i}}{\sqrt{p_i\, q_i}} = \frac{1}{\sqrt{\lambda}} \qquad i = (1, 2, 3,\ldots, s)$$

Подставляя в левую часть этого равенства вместо i последовательно 1, 2, 3,...., s получим систему уравнений, правые части которых равны одному и тому же числу $\frac{1}{\sqrt{\lambda}}$. Поэтому будем иметь

$$\frac{\frac{n_1}{N_1}}{\sqrt{p_1\, q_1}} = \frac{\frac{n_2}{N_2}}{\sqrt{p_2\, q_2}} = \ldots = \frac{\frac{n_s}{N_s}}{\sqrt{p_s q_s}} \tag{19}$$

Итак, мы видим, что погрешности числа M и пропорции $\frac{M}{N}$ будут наименьшими если пропорции вынимаемых шаров $\frac{n_i}{N_i}$ будут пропорциональны соответствующим средне-квадратическим отклонениям $\sqrt{p_i\, q_i}$ единичных испытаний.

Чтобы вычислить числа $n_1, n_2, \ldots, n_s$ представим равенства (19) в таком виде

$$\frac{n_1}{N_1\sqrt{p_1\, q_1}} = \frac{n_2}{N_2\sqrt{p_2\, q_2}} = \ldots = \frac{n_s}{N_s\sqrt{p_s\, q_s}} \tag{20}$$

Суммируя числителей и знаменателей, легко получим общую величину входящих в это равенство отношений:

$$\frac{\sum\limits_{i=1}^{s} n_i}{\sum\limits_{i=1}^{s} N_i\sqrt{p_i\, q_i}} = \frac{n}{\sum\limits_{i=1}^{s} N_i\sqrt{p_i\, q_i}}.$$

Приравнивая каждое из отношений только что найденной величине получим ряд равенств вида:

$$\frac{n_i}{N_i\sqrt{p_i\, q_i}} = \frac{n}{\sum\limits_{i=1}^{s} N_i\sqrt{p_i\, q_i}} \qquad (i = 1, 2, \ldots, s)$$

Откуда легко находим пропорции отбора

$$\frac{n_i}{N_i} = \frac{n\sqrt{p_i\, q_i}}{\sum\limits_{i=1}^{s} N_i\sqrt{p\, q_i}} \qquad (i = 1, 2, \ldots, s) \tag{21}$$

и численности районных выборочных групп

$$n_i = \frac{n\, N_i\sqrt{p_i\, q_i}}{\sum\limits_{i=1}^{s} N_i\sqrt{p_i\, q_i}} \qquad (i = 1, 2, \ldots, s) \tag{22}$$

Этими формулами следует, конечно, пользоваться тогда когда по старым статистическим материалам известны, существовавшие до выборочного исследования, пропорции $p_i = \frac{M_i}{N_i}$, а следовательно и q_i. Если же о пропорции p_i, q_i никаких сведений не имеется или же эти сведения не надежны, то естественно не отдавать предпочтения ни одному району и считать все p_i равными некоторой постоянной величине p. В этом случае формулы (21) и (22) обращаются в

$$\frac{n_i}{N_i} = \frac{n}{N}$$

и

$$n_i = \frac{n\,N_i}{N} \qquad (i = 1, 2, \ldots, s) \tag{23}$$

Отсюда следует, что если пропорции во всех районах равны, то погрешности результатов будут наименьшими когда пропорции отбора во всех районах будут одинаковы.

§ 18. Целесообразность районирования.

Покажем теперь, что в случае районирования погрешность будет несколько меньше чем при отсутствии его. Для упрощения задачи предположим, что пропорции отбора во всех районах одинаковы, т. е $\frac{n_i}{N_i} = \frac{n}{N}$. Здесь, как и в предыдущем §, безразлично исследовать ли погрешность равенства (16) или (17). Обратимся к рассмотрению погрешности равенства (17). Так как, по нашим предположениям, $\frac{n_i}{N_i} = \frac{n}{N}$, то эта погрешность представится ввиде:

$$\sigma = \frac{\sqrt{\Sigma \frac{N}{n} N_i\, p_i\, q_i \left(1 - \frac{n}{N}\right)}}{N} = \frac{\sqrt{\frac{N}{n}\left(1 - \frac{n}{N}\right) \Sigma N_i\, p_i\, q_i}}{N} =$$

$$= \sqrt{\frac{1}{Nn}\left(1 - \frac{n}{N}\right) \Sigma N_i\, p_i\, q_i} \tag{24}$$

Заменяя в сумме, стоящей под радикалом q_i через $1-p_i$ получим:

$$\Sigma N_i\, p_i\, q_i = \Sigma N_i\, p_i\,(1-p_i) = \Sigma N_i\, p_i - \Sigma N_i\, p_i^2 \tag{25}$$

Далее, замечая что

$$p = \frac{M}{N} = \frac{\Sigma M_i}{\Sigma N_i}; \quad p_i = \frac{M_i}{N_i}; \quad M_i = N_i\, p_i$$

получим

$$p = \frac{\Sigma N_i\, p_i}{\Sigma N_i},$$

откуда

$$\Sigma\, N_i\, p = \Sigma\, N_i\, p_i \quad \ldots \tag{26}$$

и следовательно

$$\Sigma\, N_i\, (p - p_i) = 0\,. \tag{27}$$

Эти равенства нам ниже понадобятся.

Возвращаясь к (25) можно на основании равенства (26)

$\Sigma\, N_i\, p_i$ заменить через $\Sigma\, N_i\, p$,

тогда получим

$$\Sigma N_i\, p_i\, q_i = \Sigma N_i\, p - \Sigma N_i\, p_i^2 = p\, \Sigma N_i - \Sigma N_i \left(p_i - p + p\right)^2 =$$

$$= N\, p - \Sigma \left\{ N_i \left[(p_i - p)^2 + 2\,(p_i - p)\, p + p^2 \right] \right\} =$$

$$= N\, p - \Sigma\, N_i\, (p_i - p)^2 - 2\, p\, \Sigma\, N_i\, (p_i - p) - N\, p^2\,.$$

Откуда ни основании равенства (27) получим

$$\Sigma\, N_i\, p_i\, q_i = N\, p\, q - \Sigma\, N_i\, (p_i - p)^2\,.$$

Подставляя найденный результат в (24) будем иметь

$$\sigma \sqrt{\frac{1}{N\, n}\left(1 - \frac{n}{N}\right)\left[N\, p\, q - \Sigma\, N_i \left(p_i - p\right)^2 \right]}\,. \tag{28}$$

Откуда окончательно

$$\sigma \sqrt{\frac{pq}{n}\left(1 - \frac{n}{N}\right) - \frac{\Sigma\, N_i\, (p_i - p)^2}{N\, n}\left(1 - \frac{n}{N}\right)}\,. \tag{29}$$

Замечая в этой формуле, что $\frac{pq}{n}\left(1 - \frac{n}{N}\right)$ есть квадрат погрешности отношения $\frac{M}{N}$, когда содержимое всех урн соединено в одну урну и шары вынимаются на удачу, приходим к выводу, что погрешность пропорции $\frac{M}{N}$, в случае нескольких урн с различными веротностями вынуть белый шар, будет несколько меньше чем погрешность той же пропорции $\frac{M}{N}$, когда все шары соединены в одну урну. Отсюда и вытекает целесообразность расчления генеральной совокупности на частные или районирования при постановке выборочного исследования.

В частном случае, когда исследуемая совокупность настолько однородна, что пропорции $p_i = \frac{M_i}{N_i}$ отдельных районов мало отличаются от пропорции $p = \frac{M}{N} = \frac{\Sigma M_i}{\Sigma N_i}$ всей совокупности, то член

$$\frac{\Sigma N_i (p_i - p)^2}{N n}\left(1 - \frac{n}{N}\right)$$

становится весьма малым по сравнению с предшествующим членом (см. форм. 29) и им можно пренебречь. В этом случае районирование, конечно, бесполезно.

§ 19. Целесообразность равномерного распределения регистрируемых единиц в общей массе.

Выводы предыдущего §-а обнаруживают, что чем равномернее распределены единицы выборочной группы в генеральной совокупности, тем меньше будут погрешности окончательных результатов. В самом деле, предположим, что генеральная совокупность разбита на ряд S равночисленных районов. Мы можем исследовать эту совокупность следующими двумя способами. Один раз выбираем единицы выборочной группы n из единиц генеральной совокупности N по жребию, не обращая внимания на районирование. Другой раз, принимая во внимание районирование, будем брать единицы также по жребию, но с условием, чтобы из каждого района было взято по $\frac{n}{s}$ единиц выборочной группы. Очевидно, что во втором случае единицы выборочной группы будут выхвачены из общей массы более равномерно. В первом случае возможно, что жребием будут захвачены единицы генеральной совокупности в одном районе чаще, в другом реже или даже совсем не будут захвачены. Во втором случае этого не может быть, так как мы ставим обязательным условием, чтобы из каждого района было взято $\frac{n}{s}$ единиц. Но из формулы (29) следует, что во втором случае погрешность будет меньше.

Этим собственно и доказывается, что чем равномернее распределяются единицы выборочной группы по пространству генеральной совокупности, тем точнее будут окончательные выводы.

Отсюда следует, что наиболее точные выводы для всей совокупности получатся при идеально равномерном распределении единиц выборочной группы в пределах генеральной совокупности, иначе говоря, при механическом отборе.

Эти разсуждения позволяют, до некоторой степени, произвести оценку выводов, полученных при механическом отборе. В самом деле, хотя мы и не в состоянии вычислить ошибку, получающуюся при механическом отборе, непосредственно, но зато мы можем утверждать, что эта ошибка не превзойдет той, которая сопряжена со случайным отбором, что практически можно считать достаточным. В следующей главе мы рассмотрим этот вопрос детальнее.

ГЛАВА III.

О средних величинах.

§ 20. Вывод основных формул.

Предположим, что генеральная совокупность состоит из N единиц, имеющих общий им всем измеримый признак. Назовем величину этого признака (варианту) через x и пусть значения ея для отдельных единиц совокупности будут

$$x_1, \quad x_2, \; . \; . \; . \; . \;, \quad x_N \; . \tag{30}$$

С целью определить среднюю величину признака

$$x_{ср.} = \frac{x_1 + x_2 + \; . \; . \; . + x_N}{N} \tag{31}$$

возьмем из общего числа единиц генеральной совокупности наудачу (по жребию) n единиц, составляющих выборочную группу, и пусть значения признака x для этих же единиц будут

$$x_{\alpha_1}, \quad x_{\alpha_2}, \; . \; . \; . \; . \;, \quad x_{\alpha_n}$$

Ряд указателей $\alpha_1, \alpha_2, \; . \; . \; . \; \alpha_n$ представляет, очевидно, одно из сочетаний N элементов 1, 2, . . . , N по n. Рассматривая среднее значение признака в выборочной группе, как приближенное значение искомой средней $x_{ср.}$ будем иметь

$$x_{ср.} = \frac{x_{\alpha_1} + x_{\alpha_2} + \; . \; . \; . + x_{\alpha_n}}{n} \; . \tag{32}$$

Вводя для обозначения суммы n каких либо вариант ряда (30) знак $\overset{n}{\Sigma}x$, можно последнее равенство кратко переписать так:

$$x_{ср.} \approx \frac{\overset{n}{\Sigma}x}{n} \; . \tag{33}$$

Покажем, что это приближенное равенство свободно от постоянной погрешности. В самом деле, погрешность этого равенства будет:

$$\frac{\sum\limits^{n} x}{n} - x_{ср.}$$

Взяв математическое ожидание этой погрешности (см. гл. II § 13) будем иметь

$$м. о. \left(\frac{\sum\limits^{n} x}{n} - x_{ср.} \right) = м. о. \frac{\sum\limits^{n} x - n\, x_{ср.}}{n} = \frac{1}{n} м. о. \left(\sum^{n} x - \sum^{n} x_{ср.} \right) =$$

$$= \frac{1}{n} м. о. \sum^{n} (x - x_{ср.}) = \frac{1}{n} \sum^{n} м. о. (x - x_{ср.}) =$$

$$= \frac{1}{n} \sum^{n} (м. о. x. - x_{ср.}) .$$

Так как под x следует разуметь любое из N значений x_1, x_2, . . . , x_N и так как все эти значения равновероятны, то вероятность каждого из них будет $\frac{1}{N}$; поэтому

$$м. о. x = \frac{x_1 + x_2 + \ldots + x_N}{N} = x_{ср.}$$

и следовательно

$$м. о. \left(\frac{\sum\limits^{n} x}{n} - x_{ср.} \right) = \frac{1}{n} \sum^{n} (м. о. x - x_{ср.}) =$$

$$= \frac{1}{n} \sum^{n} (x_{ср.} - x_{ср.}) = 0$$

Оценим теперь случайную погрешность равенства (33). Для этого надо, как известно, вычислить среднюю ошибку его σ, которая равна математическому ожиданию квадрата погрешности равенства (33).

Имеем:

$$\sigma^2 = м. о. \left(\frac{\sum\limits^{n} x}{n} - x_{ср.} \right)^2 = м. о. \left(\frac{\sum\limits^{n} x - n\, x_{ср.}}{n} \right)^2 = \frac{1}{n^2} м. о. \left(\sum^{n} x - \sum^{n} x_{ср.} \right)^2 =$$

$$= \frac{1}{n^2} м. о. \left\{ \sum^{n} \left(x - x_{ср.} \right)^2 + 2 \sum^{C_n^2} \left(x' - x_{ср} \right) \left(x'' - x_{ср} \right) \right\} ,$$

где во второй сумме знак C_n^2 указывает число слагаемых, а значки $'$ и $''$ означают, что варианты x' и x'' различны.

Далее,

$$\text{м. о.}\left|\sum^{n}\left(x-x_{ср}\right)^{2}+2\sum^{C_n^2}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)\right|=$$

$$=\text{м. о.}\sum^{n}\left(x-x_{ср}\right)^{2}+2\ \text{м. о.}\sum^{C_n^2}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)=$$

$$=\sum^{n}\text{м. о.}\left(x-x_{ср}\right)^{2}+2\sum^{C_n^2}\text{м. о.}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)=$$

$$=n.\ \text{м. о.}\left(x-x_{ср}\right)^{2}+2\frac{n.\,n-1}{1.2}\ \text{м. о.}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right). \qquad (34)$$

Теперь, так как вероятность любого из квадратов разности $x-x_{ср}$ такая же как и вероятность соответствующего значения x, то

$$\text{м. о.}\left(x-x_{ср}\right)^{2}=\frac{\sum^{N}\left(x-x_{ср}\right)^{2}}{N} \qquad (35)$$

Далее, очевидно, что различных произведений вида $\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)$ будет столько, сколько возможно составить сочетаний из N элементов x_1, x_2, ... , x_N по два; поэтому вероятность любого из таких произведений будет $\frac{1}{C_N^2}$ и следовательно

$$\text{м. о.}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)=\frac{\sum^{C_N^2}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)}{C_N^2}.$$

Сумму $\sum^{C_N^2}\left(x'-x_{ср}\right)\left(x''-x_{ср}\right)$ легко вычислить. В самом деле,

$$\frac{\sum^{N}}{N}-x_{ср}=0$$

или

$$\sum^{N}x-N\,x_{ср}=0$$

$$\sum^{N}\left(x - x_{cp}\right) = 0 .$$

Возвышая последнее выражение в квадрат, получим

$$\sum^{N}\left(x - x_{cp}\right)^2 + 2\sum^{C_N^2}\left(x' - x_{cp}\right)\left(x'' - x_{cp}\right) = 0 .$$

Откуда

$$\sum^{C_N^2}\left(x' - x_{cp}\right)\left(x'' - x_{cp}\right) = -\frac{\sum^{N}\left(x - x_{cp}\right)^2}{2} .$$

Следовательно,

$$\text{м. о.}\left(x' - x_{cp}\right)\left(x'' - x_{cp.}\right) = -\frac{\sum^{N}\left(x - x_{cp}\right)^2}{2\,C_N^2} = -\frac{\sum^{N}\left(x - x_{cp}\right)^2}{\frac{2.N(N-1)}{1\,.\,2.}} =$$

$$= -\frac{1}{N-1}\,\frac{\sum^{N}\left(x - x_{cp}\right)^2}{N} . \qquad (36)$$

Принимая во внимание найденные результаты, будем иметь

$$\sigma^2 = \frac{1}{n}\,\frac{\sum^{N}\left(x - x_{cp}\right)^2}{N} - \frac{n-1}{n\,(N-1)}\,\frac{\sum^{N}\left(x - x_{cp}\right)^2}{N}$$

Обозначая для краткости $\frac{\sum^{N}\left(x - x_{cp}\right)^2}{N}$ через δ^2,

будем иметь окончательно

$$\sigma = \sqrt{\frac{\delta^2}{n}\cdot\frac{N-n}{N-1}} . \qquad (37)$$

Здесь $\delta^2 = \frac{\sum^{N}(x - x_{cp})^2}{N}$ представляет, очевидно, квадрат средне-квадратического отклонения единичного наблюдения от x_{cp}. Формула эта была найдена впервые А. А. Чупровым в 1917 г. [1]). Вывод А. А. Чупро-

[1]) См. С. С. Кон. К вопросу о применении выборочного метода к разработке сельско-хозяйственных переписей. Петроград 1917 г. стр 22, 23.

ва не был опубликован. Независимо от А. А. Чупрова ту же формулу нашел Б. К. Ризенкампф, ему же принадлежит и вышеприведенный вывод ея.

Так как N число обычно большое, то принебрегая 1 вычитаемой из N можно предшествующую формолу переписать так:

$$\sigma = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)} \tag{38}$$

Возвращаясь теперь к равенству (33) и выставляя на вид его ошибку будем имет

$$x_{ср.} = \frac{\sum\limits^{n} x}{n} \pm \sigma \tag{39}$$

Умножая обе части этого равенства на N и имея ввиду, что

$$x_{ср.} = \frac{\sum\limits^{N} x}{N}$$

получим

$$\sum^{N} x = N\, x_{ср.} = N \frac{\sum\limits^{n} x}{n} \pm N\sigma \tag{40}$$

Здесь $\sum\limits^{N} x$ есть сумма вариант генеральной совокупности, т. е., так называемое, абсолютное число, $\frac{\sum\limits^{n} x}{n}$ — его приближенное значение и $N\sigma$ средняя погрешность абсолютного числа.

Что касается закона распределения ошибок равенства (39), то здесь следует сказать об этом тоже, что было сказано в § 14 стр. 88, т. е., что при n и N—n не очень малых ошибки будут следовать закону Гаусса. Правильноть этого положения вытекает до известной степени из теоремы о пределе вероятности [1]).

Посмотрим, во что обратится формула (37), когда $N = \infty$ (практически, когда численность генеральной совокупности весьма велика, сравнительно с численностью выборочной группы n).

Имеем:

$$\sigma = \sqrt{\frac{\delta^2}{n}\frac{N-n}{N-1}} = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n-1}{N-1}\right)}.$$

При $N = \infty$ член $\frac{n-1}{N-1}$ пропадет и следовательно получим

[1]) А. А. Марков. Исчисление вероятностей. III изд. стр. 88, § 17.

$$\sigma = \sqrt{\frac{\delta^2}{n}}\ .$$

Аналогичный результат получился бы если единицы выборочной группы после каждого испытания возвращались бы назад в генеральную совокупность.

Сравнивая выражения ошибок

$$\sigma = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)} \quad \text{и} \quad \sigma = \frac{\delta}{\sqrt{n}}$$

можно относительно них сказать все то, что уже говорилось в § 15.

§ 21. О вычислении δ на основании эмпирических данных.

Уместно здесь несколько остановиться на вычислении

$$\delta^2 = \frac{\sum\limits^{N} (x - x_{ср.})^2}{N}$$

Очевидно, что если известны не все N значений варианты x, то точное значение δ^2 определить нельзя. Но δ^2 представляет собою средний квадрат отклонений отдельных вариант от их среднего значения. Приравнивая этот средний квадрат отклонений тому среднему квадрату отклонений, который может быть найден на основании вариант выборочной группы будем иметь приближенно:

$$\delta^2 \cong \frac{\sum\limits^{n} \left(x - x_{ср}\right)^2}{n}$$

где вместо $x_{ср}$ придется взять, конечно, его приближенное значение

$$x_{ср} = \frac{\sum\limits^{n} x}{n}\ .$$

Можно, однако, вычислить δ^2 по вариантам выборочной группы, исходя из более строгих соображений. Положим для краткости $\frac{\sum\limits^{n} x}{n} = \xi$, тогда будет иметь

$$\text{м. о.} \sum^{n} \left(x - \xi\right)^2 = \text{м. о.} \sum^{n} \left[\left(x - x_{ср}\right) - \left(\xi - x_{ср}\right)\right]^2 =$$

$$= \text{м. о.} \sum^{n} \left[\left(x - x_{ср}\right)^2 - 2\left(\xi - x_{ср}\right)\left(x - x_{ср}\right) + \left(\xi - x_{ср}\right)^2\right] =$$

$$= \text{м. о.} \left| \sum^{n} \left(x - x_{cp}\right)^2 - 2\left(\xi - x_{cp}\right) \sum^{n} \left(x - x_{cp}\right) + n\left(\xi - x_{cp}\right)^2 \right| .$$

Но

$$\sum^{n}\left(x - x_{cp}\right) = \sum^{n} x - \sum^{n} x_{cp} = n\frac{\sum^{n} x}{n} - n\ x_{cp} = n\ \left(\xi - x_{cp}\right) .$$

Поэтому делая подстановку, получим

$$\text{м. о.} \sum^{n} \left(x - \xi\right)^2 = \text{м. о.} \left| \sum^{n} \left(x - x_{cp}\right)^2 - n\left(\xi - x_{cp}\right)^2 \right| =$$

$$= \sum^{n} \text{м. о.} \left(x - x_{cp}\right)^2 - n\ \text{м. о.} \left(\xi - x_{cp}\right)^2 ,$$

но выше мы видели, что

$$\text{м. о.} \left(x - x_{cp}\right)^2 = \frac{\sum^{N}\left(x - x_{cp}\right)^2}{N} = \delta^2$$

и кроме того

$$\text{м. о.} \left(\xi - x_{cp}\right)^2 = \text{м. о.} \left(\frac{\sum^{n} x}{n} - x_{cp}\right)^2 = \sigma^2 = \frac{\delta^2}{n}\ \frac{N - n}{N - 1} ,$$

поэтому

$$\text{м. о.} \sum^{n} \left(x - \xi\right)^2 = n\ \delta^2 - \frac{n\ \delta^2}{n} \frac{N - n}{N - 1} = \delta^2 \left(n - 1\right) \frac{N}{N - 1}$$

откуда

$$\delta^2 = \frac{\text{м. о.} \sum^{n} \left(x - \xi\right)^2}{\left(n - 1\right) \frac{N}{N - 1}} .$$

Заменяя в этом выражении м. о. $\sum^{n}\left(x - \xi\right)^2$ тем частным значением $\sum^{n}\left(x - \xi\right)^2$, которое получается по вариантам выборочной группы, получаем окончательно

$$\delta^2 \cong \frac{\sum^{n}\left(x - \xi\right)^2}{\left(n - 1\right) \frac{N}{N - 1}} \qquad (41)$$

где $\xi = \frac{\sum\limits^{n} x}{n}$. При n, N сколько нибудь больших это значение δ^2 мало отличается от

$$\delta^2 \cong \frac{\sum\limits^{n} \left(x - \xi\right)^2}{n}.$$

§ 22. Распространение полученных выводов на случай относительных величин.

Покажем, что найденные нами формулы могут быть распространены и на случай относительных величин. В самом деле, представим себе, что из урны, сдержащей N шаров, среди которых M белых, вынимается на удачу n шаров и пусть среди последних оказаловь m белых. Сопоставляя с каждым шаром некоторую варианту, которая в случае белого шара пусть равна 1, а в случае черного—0, будем иметь

$$\sum^{N} x = M, \quad \sum^{N} x^2 = M, \quad \sum^{n} x = m$$

Далее пологая $\frac{M}{N} = p$, получим

$$x_{cp} = \frac{\sum\limits^{N} x}{N} = \frac{M}{N} = p,$$

$$\frac{\sum\limits^{n} x}{N} = \frac{m}{n} \cong p,$$

$$\delta^2 = \frac{\sum\limits^{N} \left(x - x_{cp}\right)^2}{N} = \frac{\sum\limits^{N} x^2}{N} - x_{cp}^2 = \frac{M}{N} - p^2 = p - p^2 = pq$$

где $q = 1 - p$.

На основании всего этого формулы (39) и (40) примут вид:

$$\frac{M}{N} \cong \frac{m}{n} \pm \sqrt{\frac{pq}{n} \frac{N - n}{N - 1}}$$

$$M \cong N \frac{m}{n} \pm N \sqrt{\frac{pq}{n} \frac{N - n}{N - 1}},$$

что в точности совпадает с найденным во II-й главе.

Посмотрим во что обратится в случае относительных величин формула (41). Что касается левой части этой формулы, то она обратится в pq. Далее будем иметь

$$\sum^{n} (x - \xi)^2 = \sum^{n} x^2 - n\,\xi^2 ,$$

но

$$\sum^{n} x^2 = m , \quad \xi = \frac{\sum^{n} x}{n} = \frac{m}{n}$$

поэтому

$$\sum^{n} \left(x - \xi\right)^2 = m - \frac{m^2}{n} = \frac{m\left(n - m\right)}{n}$$

и окончательно

$$pq \underset{s}{=} \frac{m\left(n - m\right)}{n\left(n - 1\right)\frac{N}{N - 1}} ,$$

что практически при сколько нибудь больших n и N мало отличается от обычно принимаемого

$$pq \underset{s}{=} \frac{m\left(n - m\right)}{n^2} .$$

§ 23. Средняя ошибка в случае нескольких районов.

Все предшествующие выводы построены на предположении, что при выборке n единиц выборочной группы из N единиц генеральной совокупности не было тенденциозного подбора, который может возникнуть не только вследствии произвольности подбора единиц, но даже и при механическом их выборе. Иначе говоря, предполагалось, что избранные n единиц взяты на удачу (т. е. по жребию или другим способом, гарантирующим полную случайность отбора).

Очевидно, что если единицы (варианты) расположены в общей массе случайно, то механический отбор эквивалентен случайному. В этом случае случайный отбор можно заменить механическим, вычисляя погрешности по формулам, выведенным для случайного отбора. Однако, на практике материал редко бывает настолько относительно однородным чтобы его можно было считать равноценным тщательно перемешанному материалу (как в урне с шарами). В этом случае, механический отбор вариант не равносилен случайному отбору и потому применение формул, выведенных в предположении случайного отбора, незаконно, если не доказаио, что ошибка при механическом отборе не превышает ошибки при случайном отборе.

Но в таком случае, естественно, разбить материал на ряд частей, в пределах которых можно считать материал относительно однородным. Тогда возможно будет к каждой из таких частных совокупностей применить механический отбор и вычислить по известным правилам среднюю величину признака и его среднюю ошибку для всей генеральной совокупности.

Таким образом, мы приходим к решению следующей задачи:

Предположим, что исследуемая совокупность разбита на S районов. Назовем через

$$N_1,\ N_2,\ \ldots\ldots,\ N_S$$

общие числа единиц в каждом районе. Дальше положим, что

$$N_1 + N_2 + N_3 + \ldots\ldots + N_S = N\ ,$$

кроме того предположим, что из этих районов берется на удачу (если в районах материал относительно однороден, то такой отбор может быть заменен механическим) соответственно

$$n_1,\ n_2,\ n_3,\ \ldots\ldots n_S$$

единиц, причем положим, что

$$n_1 + n_2 + n_3 + \ldots\ldots + n_S = n\ .$$

Тогда на основании предшествующих выводов будет иметь для какого либо i—того района (на основании 40 и 37)

$$\overset{N_i}{\Sigma} x \underset{S}{=} N_i \frac{\overset{n_i}{\Sigma} x}{n_i} + N_i \sqrt{\frac{\delta_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1}} \qquad (42)$$

где в каждую из сумм $\overset{N_i}{\Sigma} x$ и $\overset{n_i}{\Sigma} x$ входят лишь варианты i—того района, а δ_i^2 есть средняя ошибка единичного наблюдения для i—того района.

Складывая подобные выражения для всех районов, соединяя при этом средние ошибки по известным правилам, получим

$$\overset{N}{\Sigma} x = \sum_{i=1}^{S} N_i \frac{\overset{n_i}{\Sigma} x}{n_i} \pm \sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1}} \qquad (43)$$

Откуда, разделив на N, получим

$$x_{ср.} \cong \frac{\sum\limits_{i=1}^{S} N_i \dfrac{\sum\limits^{n_i} x}{n_i}}{N} \pm \frac{1}{N}\sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i} \cdot \frac{N_i - n_i}{N_i - 1}} \qquad (44)$$

Формулами (43) и (44) исчерпывается решение нашей задачи.

Если считать число N_i достаточно большим, то единицей, вычитаемой из него можно пренебречь и формулы (43) и (44) примут вид:

$$\sum^{N} x \cong \sum_{i=1}^{S} N_i \frac{\sum\limits^{n_i} x}{n_i} \pm \sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)} \qquad (45)$$

$$x_{ср.} = \frac{\sum\limits^{N} x}{N} \cong \frac{\sum\limits_{i=1}^{S} N_i \dfrac{\sum\limits^{n_i} x}{n_i}}{N} \pm \frac{1}{N}\sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)} \qquad (46)$$

§ 24. Наивыгоднейшие пропорции отбора.

Найдем теперь какими должны быть числа n_i, чтобы ошибка

$$\sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)}$$

была бы наименьшая. При этом поставим условием, чтобы было $n = \sum n_i =$ Const. Поступая по правилам отыскания условного maxima и minima, как это мы делали в § 17, получим

$$\frac{\dfrac{n_1}{N_1}}{\delta_1} = \frac{\dfrac{n_2}{N_2}}{\delta_2} = \ldots\ldots = \frac{\dfrac{n_S}{N_S}}{\delta_S} \qquad (47)$$

Отсюда следует, что ошибка будет наименьшая когда пропорции вынимаемых единиц в районах будут пропорциональны среднеквадратическим отклонениям δ в отдельных районах.

Из (47) получаем по правилам алгебры

$$\frac{n_1}{N_1 \delta_1} = \frac{n_2}{N_2 \delta_2} = \ldots\ldots = \frac{n_S}{N_S \delta_S} = \frac{\sum\limits_{i=1}^{S} n_i}{\sum\limits_{i=1}^{S} N_i \delta_i} = \frac{n}{\sum\limits_{i=1}^{S} N_i \delta_i}$$

или

$$\frac{n_i}{N_i\ \delta_i} = \frac{n}{\Sigma N_i\ \delta_i}\text{, откуда } n_i = \frac{n\ N_i\ \delta_i}{\sum\limits_{i=1}^{S} N_i\ \delta_i} \tag{48}$$

и

$$\frac{n_i}{N_i} = \frac{n\delta_i}{\Sigma N_i\ \delta_i}\ . \tag{49}$$

По этим формулам можно вычислить для отдельных районов числа отбираемых единиц n_i или пропорции $\frac{n_i}{N_i}$, если известны по имеющемуся материалу δ_i для отдельных районов. Когда же о δ_i ничего не известно, то естественно, не отдавая предпочтения ни одному δ_i, считать их все равными некоторой постоянной величине δ. Подставляя это δ в (48) и (49), найдем

$$n_i = \frac{n\ N_i}{\sum\limits_{i=1}^{S} N_i} = \frac{nN_i}{N}\ ;\ \frac{n_i}{N_i} = \frac{n}{N}\ . \tag{50}$$

Последний результат показывает, что если все δ одинаковы, то ошибка средней будет наименьшая, когда пропорции извлекаемых единиц будут во всех районах одинаковы.

§ 25. Целесообразность районирования.

Полагая теперь в формуле (46) пропорции одинаковыми и равными $\frac{n}{N}$ получим для ея ошибки

$$\sigma = \frac{1}{N}\sqrt{\sum_{i=1}^{s} N_i^2\,\frac{\delta_i^2}{n_i}\left(1-\frac{n_i}{N_i}\right)} = \frac{1}{N}\sqrt{\frac{N}{n}\left(1-\frac{n}{N}\right)\sum_{i=1}^{s} N_i\ \delta_i^2} \tag{51}$$

Сумму $\sum\limits_{i=1}^{s} N_i\ \delta_i^2$ легко представить в виде

$$N\ \delta^2 - \sum_{i=1}^{s} N_i\ (\xi_i - x_{ср.})^2$$

где N — численность генеральной совокупности, $\delta^2 = \frac{\sum\limits^{N}(x - x_{ср.})^2}{N}$

среднее отклонение единичного наблюдения для нея и $\xi_i = \frac{\sum^{N_i} x}{N_i}$ среднее для i — того района.

В самом деле, имеем

$$N\,\delta^2 = \sum^{N}\left(x - x_{ср.}\right)^2 = \sum_{i=1}^{s}\sum^{N_i}\left(x - x_{ср.}\right)^2 = \sum_{i=1}^{s}\sum^{N_i}\left(x - \xi_i + \xi_i - x_{ср}\right)^2 =$$

$$= \sum^{s}\sum^{N_i}\left[\left(x-\xi_i\right)^2 + 2\left(\xi_i - x_{ср.}\right)\left(x-\xi_i\right) + \left(\xi_i - x_{ср.}\right)^2\right] =$$

$$= \sum_{i=1}^{s}\left[\sum^{N_i}\left(x-\xi_i\right)^2 + 2\left(\xi_i - x_{ср.}\right)\sum^{N_i}\left(x-\xi_i\right) + N_i\left(\xi_i - x_{ср}\right)^2\right];$$

но $\sum^{N_i}\left(x-\xi_i\right) = \sum^{N_i} x - N\xi_i = o$, поэтому

$$N\,\delta^2 = \sum_{i=1}^{s}\left[\sum^{N_i}\left(x-\xi_i\right)^2 + N_i\left(\xi_i - x_{ср.}\right)^2\right].$$

Имея теперь в виду, что $\sum^{N_i}\left(x - \xi_i\right)^2 = N_i\,\delta_i^2$, получим

$$N\,\delta^2 = \sum_{i=1}^{s} N_i\,\delta_i^2 + \sum_{i-1}^{s} N_i\left(\xi_i - x_{ср.}\right)^2$$

откуда окончательно

$$\sum_{i=1}^{s} N_i\,\delta_i^2 = N\,\delta^2 - \sum_{i=1}^{s} N_i\left(\xi_i - x_{ср.}\right)^2.$$

Подставляя этот результат в (51), будем иметь

$$\sigma = \frac{1}{N}\sqrt{\frac{N}{n}\left(1 - \frac{n}{N}\right)\left[N\,\delta^2 - \sum_{i=1}^{s} N_i\left(\xi_i - x_{ср.}\right)^2\right]}$$

или

$$\sigma = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right) - \frac{\sum\limits_{i=1}^{s} N_i\left(\xi_i - x_{ср.}\right)^2}{n.\, N}\left(1 - \frac{n}{N}\right)} \qquad (52)$$

Замечая, что $\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)$ есть квадрат погрешности $x_{ср.} = \frac{\Sigma x}{N}$, когда районирование не имеет места, видим, что при районировании ошибка получается несколько меньшая чем при отсутствии его.

§ 26. Более точное решение предыдущего вопроса.

Для более детального исследования вопроса о целесообразности районирования обратимся к более точным формулам, а именно, будем исходить из формулы (44). Для простоты положим, что все n_i и N_i равны между собой (т. е. численности районов одинаковы); тогда

$$n_i = \frac{n}{S} \text{ и } N_i = \frac{N}{S}$$

Подставляя эти значения в (44), получим

$$\sigma = \frac{1}{N}\sqrt{\sum_{i=1}^{s} N_i^2 \frac{\delta_i^2}{n_i} \frac{N_i - n_i}{N_i - 1}} = \frac{1}{N}\sqrt{\sum_{i=1}^{s} \frac{N^2}{S^2} \frac{\delta_i^2}{\frac{n}{S}} \frac{\frac{N}{S} - \frac{n}{s}}{\frac{N}{S} - 1}} =$$

$$= \frac{1}{N}\sqrt{\sum_{i=1}^{s} \frac{N^2}{S^2} \frac{s}{n} \delta_i^2 \frac{N - n}{N - S}} = \frac{1}{N}\sqrt{\frac{N}{n} \frac{N - n}{N - S} \sum_{i=1}^{s} \frac{N}{S} \delta_i^2}.$$

Но мы нашли, что $\sum\limits_{i=1}^{s} N_i \delta_i^2 = N\delta^2 - \sum\limits_{i=1}^{s} N_i \left(\xi_i - x_{ср.}\right)^2$,

а так как в нашем случае $N_i = \frac{N}{S}$, то имеем $\sum\limits_{i=1}^{s} \frac{N}{S} \delta_i^2 = N\delta^2 -$

$$- \frac{N}{S} \sum_{i=1}^{s} \left(\xi_i - x_{ср.}\right)^2$$

и следовательно

$$\sigma = \frac{1}{N}\sqrt{\frac{N}{n} \frac{N-n}{N-s}\left[N\delta^2 - \frac{N}{s}\sum_{i=1}^{s}\left(\xi_i - x_{ср}\right)^2\right]}$$

или

$$\sigma = \sqrt{\left[\delta^2 - \frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s} \right] \frac{1}{n} \frac{N-n}{N-s}} . \tag{53}$$

Так, следовательно, выражается ошибка среднего когда генеральная совокупность разбита на s равных частей, причем, пропорции выборки в этих частях также одинаковы. В этой формуле $\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s}$ есть средне-квадратическое колебание районных средних около общей средней генеральной совокупности.

Предположим теперь, что материал всей массы однороден в той же степени, в какой однороден всякий тщательно перемешанный материал, т. е. варианты расположены так, как будто бы они расположены совершенно случайно, иначе говоря, в полном беспорядке. В таком случае отклонения районных средних ξ от общей средней x_{cp} будут совершенно случайны. Для вычисления $\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s}$ следовало бы знать кроме общей средней также и районные средние. Но так как даже при случайных колебаниях районных средних величина

$$\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s}$$

может иметь множество значений, в зависимости от установившихся при районировании районных средних, то чтобы иметь представление о величине $\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s}$ при случайном распределении материала найдем среднее значение величины

$$\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s}$$

при всех возможных комбинациях величины ξ_i $(i = 1, 2, \ldots, s)$. Иначе говоря, найдем математическое ожидание

$$\frac{\sum\limits_{i=1}^{s} (\xi_i - x_{cp})^2}{s} .$$

Имеем,

$$\text{м. о.}\frac{\sum\limits_{i=1}^{s}\left(\xi_i-x_{cp}\right)^2}{S}=\frac{1}{S}\text{ м. о. }\sum_{i=1}^{s}\left(\xi_i-x_{cp}\right)^2=\frac{1}{S}\sum_{i=1}^{s}\text{ м. о. }\left(\xi_i-x_{cp}\right)$$

Таким образом, дело сводится к нахождению м. о. $\left(\xi_i-x_{cp}\right)^2$, т. е. к нахождению математического ожидания квадрата отклонения районной средней ξ_i от общей средней x_{cp}; но ξ_i есть среднее из $\frac{N}{s}$ вариант и математическое ожидание квадрата отклонения такой средней от общей средней может быть найдено по формуле (37);

Для этого достаточно заменить в этой формуле n через $\frac{N}{s}$. Таким образом получим

$$\text{м. о.}\left(\xi_i-x_{cp}\right)^2=\frac{\delta^2}{\frac{N}{s}}\;\frac{N-\frac{N}{s}}{N-1}.$$

Умножая числителя и знаменателя на s, получим

$$\text{м. о.}\left(\xi_i-x_{cp}\right)^2=\frac{\delta^2}{N}\;\frac{s\,N-N}{N-1}=\delta^2\frac{s-1}{N-1}. \tag{54}$$

Далее

$$\sum_{i=1}^{s}\text{ м. о. }\left(\xi_i-x_{cp}\right)^2=s\,\delta^2\,\frac{s-1}{N-1}$$

и окончательно

$$\text{м. о.}\frac{\sum\limits_{i=1}^{s}\left(\xi_i-x_{cp}\right)^2}{S}=\frac{1}{S}\sum_{i=1}^{s}\text{ м. о. }\left(\xi_i-x_{cp}\right)^2=\delta^2\frac{s-1}{N-1}.$$

Подставляя этот результат в (53) вместо $\dfrac{\sum\limits_{i=1}^{s}\left(\xi_i-x_{cp}\right)^2}{S}$, найдем

$$\sigma=\sqrt{\left(\delta^2-\delta^2\frac{s-1}{N-1}\right)\frac{1}{n}\;\frac{N-n}{N-s}}=\sqrt{\frac{\delta^2}{n}\left(1-\frac{s-1}{N-1}\right)\frac{N-n}{N-s}},$$

или

$$\sigma=\sqrt{\frac{\delta^2}{n}\;\frac{N-n}{N-1}}.$$

Отсюда мы заключаем, что если варианты в исследуемой совокупности расположены в генеральной совокупности в полном безпорядке, совершенно случайно, т. е. исследуемая совокупность относительно однородна, то районирование, вообще говоря, бесполезно.

Предположим теперь, что средне-квадратическое колебание районных средних около общей средней больше того, которого следует ожидать при случайном распределении вариант, т. е., что

$$\frac{\sum\limits_{i=1}^{S}(\xi_i - x_{ср.})^2}{S} > \delta^2\,\frac{S-1}{N-1}$$

тогда по формуле (53) будем иметь

$$\sqrt{\left[\delta^2 - \frac{\sum\limits_{i=1}^{S}(\xi_i - x_{ср.})^2}{S}\right]\frac{1}{n}\,\frac{N-1}{N-S}} < \sqrt{\frac{\delta^2}{n}\,\frac{N-n}{N-1}}\,,$$

т. е. ошибка общей средней будет в этом случае меньше той, которую следует ожидать при случайном распределении вариант.

Итак, когда средне-квадратическое колебание районных средних больше того, которое ожидается при случайном распределении вариант,

$$\left(\text{т. е. } \frac{\sum\limits_{i=1}^{S}(\xi_i - x_{ср.})^2}{S} > \delta^2\,\frac{s-1}{N-1}\right),$$

то ошибка средней при районировании будет меньше ошибки средней когда районирование не имеет место.

Предположим, наконец, что при тех же самых предположениях, средне-квадратическое колебание районных средних меньше того, которое ожидается при случайном распределении вариант, т. е.,

$$\frac{\sum\limits_{i=1}^{s}(\xi_i - x_{ср.})^2}{S} < \delta^2\,\frac{s-1}{N-1}\,;$$

Тогда, очевидно, что ошибка средней при районировании будет больше ошибки средней когда районирование не имеет место. В этом случае районирование, очевидно вредно [1]).

Любопытно отметить, что рассмотренные три случая дисперсии районных средних вполне аналогичны трем видам дисперсии рядов по схеме Lexis'a.

---

[1]) Впрочем, такой случай на практике вряд ли возможен.

§ 27. Сопоставление механического отбра со случайным.

Выше мы не раз видели, что если материал генеральной совокупности относительно однороден, то механический и случайный отбор равносильны. Далее в § 19 мы, до известной степени, показали, что в случае неоднородного материла механический отбор дает меньшую ошибку чем случайный. Оставался лишь невыясненным тот же вопрос в применении к материалу, который мы называем однородным, т. е., когда разные варианты располагаются в генеральной совокупности более равномерно, чем при случайном их распределении. Так как предшествующие соображения относительно механического отбора не были вполне строго обоснованы, то постараемся разобрать этот вопрос глубже.

Механический отбор можно истолковать, как деление генеральной совокупности на весьма большое число малых районов, причем все районы равны и из каждого района берется по одной варианте, предполагая, что в пределах малых районов варианты расположены случайно (а это практически весьма близко к истине), очевидно, безразлично откуда брать варианту из начала ли, середины или конца этого маленького района. Но в таком случае мы можем найти ошибку средней при механическом отборе, если в формуле (53) положим число районов S равным n. Делая указанную подстановку получим, что ошибка общей средней $x_{cp.}$ при механическом районе будет равна

$$\sigma = \sqrt{\left[\delta^2 - \frac{\sum\limits_{i=1}^{s}\left(\xi_i - x_{cp}\right)^2}{n}\right] \cdot \frac{1}{n}} \qquad (55)$$

Правда для практических целей эта формула не годится, т. к. нельзя определить районных средних по одной варианте в каждом районе, но для теоретических рассуждений она вполне достаточна.

Если материал однороден, как тщательно перемешанные шары в урне, то член

$$\frac{\sum\limits_{i=1}^{s}\left(\xi_i - x_{cp}\right)^2}{n},$$

по вышедоказанному, может быть заменен выражением $\delta^2 \frac{S-1}{N-1}$, причем вместо S необходимо, конечно, подставить n, тогда формула (55) даст нам

$$\sigma = \sqrt{\left[\delta^2 - \frac{\sum\limits_{i=1}^{s}\left(\xi_i - x_{cp}\right)^2}{n}\right] \frac{1}{n}} = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n-1}{N-1}\right)}.$$

Это же есть выражение ошибки средней, когда единицы выхватываются из генеральной совокупности случайно.

Таким образом, в случае относительно—однородного материала механический отбор единиц и случайный отбор равноценны.

Если

$$\frac{\sum_{i=1}^{n}\left(\xi_i - x_{ср}\right)^2}{n} > \delta^2 \frac{n-1}{N-1},$$

то, очевидно, материал наш нельзя назвать однородным, так как дисперсия средних для малых районов, из которых берется по одной варианте, около общей средней $x_{ср.}$, будет больше чем при случайном распределении вариант. Формула (55) показывает, что в этом случае

$$\sigma = \sqrt{\left[\delta^2 - \frac{\sum_{i=1}^{n}\left(\xi_i - x_{ср}\right)^2}{n}\right]\frac{1}{n}} < \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n-1}{N-1}\right)},$$

т. е., если материал не однороден, то ошибка средней при механическом отборе меньше чем при случайном отборе.

Пусть теперь средне-квадратическое колебание средних для малых районов меньше того, которое ожидается при случайном распределении вариант, т. е.

$$\frac{\sum_{i=1}^{n}\left(\xi_i - x_{ср}\right)^2}{n} < \delta^2 \frac{n-1}{N-1}.$$

В этом случае материал, очевидно, более однороден, чем всякий тщательно перемешанный материал. Легко видеть, что в этом случае

$$\sigma = \sqrt{\left[\delta^2 - \frac{\sum_{i=1}^{n}\left(\xi_i - x_{ср}\right)^2}{n}\right]\frac{1}{n}} > \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n-1}{N-1}\right)}.$$

Таким образом, если материал более однороден, чем тщательно перемешанный материал, то ошибка средней при механическом отборе больше чем ошибка при случайном отборе, т. е. механический отбор вреден.

Для пояснения последнего случая рассмотрим пример из области относительных величин, причем ограничимся крайним случаем, когда вероятности p_i для всех малых районов равны между собой и равны следовательно вероятности p для всей совокупности.

Если не обращать внимания на наши районы, то при случайном отборе, соответствующим случаю схемы невозвращенных шаров, ошибка частости будет

$$\sigma = \sqrt{\frac{pq}{n} \cdot \frac{N-n}{N-1}}.$$

В случае схемы возвращенных шаров ошибка частости будет

$$\sigma = \sqrt{\frac{pq}{n}} ,$$

т. е. больше чем в предыдущем случае. Наконец, в случае механического отбора ошибку частости получим если в формуле (55) положим $\xi_i = p_i = p$, $x_{ср.} = p$ и $\delta^2 = pq$.

Таким образом будем иметь:

$$\sigma = \sqrt{\frac{pq}{n}} ,$$

т. е. ошибка получилась такая же, как в случае схемы возвращенных шаров. Это и понятно, в самом деле, мы предположили, что вероятности для всех районов одинаковы и из каждого района вынимается по одному шару (варианте), но это, очевидно, равносильно ряду последовательных выниманий шара из одного какого нибудь района, так как для других районов вероятности такие же. Таким образом, рассматриваемый случай равносилен схеме возвращаемых шаров. Этот пример поясняет также, до известной степени, почему при районировании ошибка может получиться иной раз меньше чем при отсутствии такового.

§ 28. Исследование вопроса о целесообразности механического отбора с другой точки зрения.

Предположим, что варианты в генеральной совокупности колеблятся около некоторого изменчивого от района к району среднего уровня. Поясним это примером.

Предположим, что мы ставим себе целью определить среднюю величину ряда вариант расположенных на прямой. Любую варианту x можно считать состоящей из трех частей: из $x_{ср.}$ плюс отклонение α переменного уровня от $x_{ср.}$ и плюс случайное отклонение β варианты от переменного уровня [1]; т. е. варианта

$$x = x_{ср.} + \alpha + \beta \qquad (56).$$

Разобьем весь ряд вариант на ряд равных частей или районов. В пределах района α можно считать грубо величиной постоянной, так как $x_{ср.} + \alpha$ остается почти постоянным, а следовательно и α остается постоянным. Тогда из (56) имеем

$$x - x_{ср.} = \alpha + \beta$$

Найдем математическое ожидание квадрада этой величины для всей совокупности, тогда будем иметь

[1]) См. Б. Ястремский.—Косвенный метод определения устойчивости или изменяемости статистических рядов. Вестник статистики, 1919 г. № 4—7 стр. 3—6.

$$м.\ о.\ (x - x_{ср.})^2 = м.\ о.\ (\alpha + \beta)^2 = \tag{57}$$
$$= м.\ о.\ (\alpha^2 + 2\,\alpha\beta + \beta^2) = м.\ о.\ \alpha^2 + 2\ м.\ о.\ \alpha\beta + м.\ о.\ \beta^2,$$

но мы знаем, что м. о. $(x - x_{ср.})^2 = \delta^2$. Дальше, в пределах уже всей генеральной совокупности α нельзя считать постоянной. Называя математическое ожидание квадрата отклонения переменного уровня от постоянного (т. е. от $x_{ср.}$) через $\triangle^2$ будем иметь м. о. $\alpha^2 = \triangle^2$.

Считая теперь, что случайное отклонение от переменного уровня β не зависит от α будем иметь

$$м.\ о\ \alpha\beta = м.\ о\ \alpha\ .\ м.\ о.\ \beta.$$

Так как всякий средний уровень подбирается так, чтобы отклонения от него в одну сторону уравновешивались бы отклонениями в другую, то м. о. $\alpha = о$ и м. о $\beta = о$. Математическое ожидание β^2 есть, очевидно, квадрат средне-квадратического отклонения вариант x от среднего переменного уровня, называя это отклонение через δ' будем иметь м. о. $\beta^2 = \delta'^2$.

Следовательно, подставляя все эти значения в (57), получим

$$\delta^2 = \triangle^2 + \delta'^2 \tag{58}$$

Рассмотрим теперь разность $x - \alpha$. Эта разность представляет собой значение варианты, освобожденной от отклонения переменного уровня от постоянного среднего уровня. Найдем среднее значение величин $x - \alpha$ для i—того района. Имеем

$$\frac{\sum\limits^{\frac{N}{n}} (x_i - \alpha)}{\frac{N}{n}} = \frac{\sum\limits^{\frac{N}{n}} x_i - \sum\limits^{\frac{N}{n}} \alpha}{\frac{N}{n}} = \frac{\sum\limits^{\frac{N}{n}} x_i}{\frac{N}{n}} - \frac{\sum\limits^{\frac{N}{n}} \alpha}{\frac{N}{n}} \tag{59}$$

но $\dfrac{\sum\limits^{\frac{N}{n}} x_i}{\frac{N}{n}}$, очевидно, есть районная средняя, которую мы обозначили через ξ_i. Дальше в пределах района α мы считаем постоянной, а поэтому

$$\frac{\sum\limits^{\frac{N}{n}} \alpha}{\frac{N}{n}} = \alpha\,.$$

Называя среднее значение вариант i—того района, освобожденных от колебаний α—переменного уровня через $\bar{\xi}_i$ и замечая, что

$$\bar{\xi}_i = \frac{\sum\limits^{\frac{N}{n}} (x_i - \alpha)}{\frac{N}{n}}$$

будем иметь на основании предшествующих обозначений, после подстановки в формулу (59)

$$\bar{\xi}_i = \bar{\xi}_i - \alpha \cdot \tag{60}$$

Найдем теперь математическое ожидание величины

$$\sum_{i=1}^{n} (\xi_i - x_{cp})^2 \cdot$$

Имеем

$$\text{м. о} \sum_{i=1}^{n} (\xi_i - x_{cp})^2 \cdot = \sum_{i=1}^{n} \text{м. о} (\xi_i - x_{cp})^2 =$$

$$= \sum_{i=1}^{n} \text{м. о.} (\xi_i - \alpha - x_{cp} + \alpha)^2 =$$

$$\sum_{i=1}^{n} \text{м. о.} \left\{ \left[(\xi_i - \alpha) - x_{cp} \right] = \alpha \right\}^2 .$$

Подставляя вместо $\xi_i - \alpha$ его значение из (60) получим

$$\text{м. о.} \sum_{i=1}^{n} \left(\xi_i - x_{cp} \right)^2 = \text{м. о.} \left[\left(\bar{\xi}_i - x_{cp} \right) + \alpha \right]^2 =$$

$$= \sum_{i=1}^{n} \text{м. о} \left[\left(\bar{\xi}_i - x_{cp} \right)^2 + 2\alpha \left(\bar{\xi}_i - x_{cp} \right) + \alpha^2 \right] =$$

$$= \sum_{i=1}^{n} \left[\text{м о.} \left(\bar{\xi}_i - x_{cp} \right)^2 + 2\alpha \text{ м. о.} \left(\bar{\xi}_i - x_{cp} \right) + \alpha^2 \right] \tag{61}$$

Так как $\bar{\xi}_i$ есть среднее значение вариант i—того района, освобожденных от колебаний переменного уровня, и т. к. при освобождении вариант от колебаний переменного уровня материал становится однородным, то можно к выражению м. о. $\left(\overline{\xi_i} - x_{cp}\right)^2$ применить формулу (54), выведенной как раз в том предположении, что материал однороден, с той лишь разницей, что δ придется заменить через δ', так как при освобождении вариант от колебаний α переменного уровня, остаются лишь случайные колебания около постоянного среднего уровня x_{cp} и средне-квадратическое этих колебаний обозначено было нами через δ'. Кроме того вместо S надо поставить, очевидно, n, так как у нас $S = n$. Сделав указанные постановки, получим

$$\text{м. о.} \left(\overline{\xi_i} - x_{cp}\right)^2 = \delta'^2 \frac{n-1}{N-1} .$$

Перейдем теперь к вычислению м. о. $\left(\overline{\xi_i} - x_{cp}\right)$. Имеем

$$\text{м. о.} \left(\overline{\xi_i} - x_{cp}\right) = \text{м. о.}\ \overline{\xi_i} - x_{cp} .$$

Но м. о. $\overline{\xi_i}$, являясь средней из всех возможных значений $\overline{\xi_i}$, очевидно, равно x_{cp}. Поэтому

$$\text{м. о.} \left(\overline{\xi_i} - x_{cp}\right) = 0 .$$

Подставляя найденные результаты в формулу (61) будем иметь

$$\text{м. о.} \sum_{i=1}^{n} \left(\overline{\xi_i} - x_{cp}\right)^2 = \sum_{i=1}^{n} \left[\delta'^2 \frac{n-1}{N-1} + \alpha^2\right] =$$

$$n\, \delta'^2 \frac{n-1}{N-1} + \sum_{i=1}^{n} \alpha^2 .$$

Разделим обе части равенства на n, тогда

$$\text{м. о.} \frac{\sum\limits_{i=1}^{n} \left(\overline{\xi_i} - x_{cp}\right)^2}{n} = \delta'^2 \frac{n-1}{N-1} + \frac{\sum\limits_{i=1}^{n} \alpha^2}{n} .$$

Приняв теперь во внимание, что последнее слагаемое есть квадрат среднеквадратического отклонения переменного уровня от постоянного, который обозначен нами в формуле (58) через $\triangle^2$, найдем окончательно

$$\text{м. о.} \frac{\sum\limits_{i=1}^{n} \left(\overline{\xi_i} - x_{cp}\right)^2}{n} = \delta'^2 \frac{n-1}{N-1} + \triangle^2 .$$

Подставляя в формулу (55) для ошибки при механическом отборе вместо

$$\frac{\sum\limits_{i=1}^{n}\left(\xi_i - x_{cp}\right)^2}{n}$$

среднее из всех возможных значений этого выражения, т. е. его математическое ожидание[1]) получим, принимая во внимание формулу (58)

$$\sigma = \sqrt{\left(\triangle^2 + \delta'^2 - \delta'^2\frac{n-1}{N-1} - \triangle^2\right)\frac{1}{n}}$$

или

$$\sigma = \sqrt{\frac{\delta'^2}{n}\;\frac{N-n}{N-1}} \tag{62}$$

Это и есть ошибка средней при механическом отборе.

Сравним полученное выражение ошибки с тем, которое получается при случайном отборе n единиц из общей массы N единиц.

Так как для всей совокупности средне-квадратическое отклонение одного наблюдения есть $\delta^2 = \delta'^2 + \triangle^2$, то ошибка средней здесь будет (по 37)

$$\sigma = \sqrt{\frac{\delta'^2 + \triangle^2}{n}\;\frac{N-n}{N-1}} = \sqrt{\frac{\delta'^2}{n}\;\frac{N-n}{N-1} + \frac{\triangle^2}{n}\;\frac{N-n}{N-1}} \tag{63}$$

Первое слагаемое, как видно из (62) представляет ошибку средней при механическом отборе. Эта ошибка, как мы видели, зависит только от случайных колебаний вариант около переменного среднего уровня. Второе же слагаемое зависит только от колебаний самого переменного среднего уровня около постоянного среднего уровня. Так как в (62) этого слагаемого нет, то ясно, что при механическом отборе совершенно устраняется влияние колебаний переменного уровня на величину средней.

При случайном же отборе ошибка средней слагается из двух частей. Одна из них зависит исключитлльно от случайных колебаний вариант около переменного среднего уровня, другая же часть зависит только от колебаний переменного среднего уровня около постоянного среднего уровня.

Иначе предшествующие положения можно выразить так: при случайном отборе всегда сказывается (в большой или меньшей степени) влияние на среднюю неоднородность материала, при механическом отборе вариант влияние неоднородности материала на среднюю устраняется полностью, что дает полное обоснование применения механического отбора и его преимуществах перед случайным отбором при постановке выборочного исследования.

---

[1]) Делаем мы это потому что истинное значение этой суммы нам неизвестно и при данном среднем переменном и постоянном уровнях, средне-квадратическом отклонении δ' случайных колебаний около переменного уровня, мы можем только определить математическое ожидание этой суммы.

Впрочем, необходимо заметить, что предшествующий вывод основан на предположении, что дисперсия δ' случайных колебаний остается во всей совокупности постоянной. Такие совокупности по терминологии K. Pearson'a называются равноизменчивыми. На практике же такие совокупности весьма редки и в большинстве случаев приходится иметь дело с, так называемыми, разноизменчивыми[1]) совокупностями, которые характеризуются тем, что мера случайных колебаний δ' около переменного уровня не остается постоянной во всей генеральной совокупности, меняясь от места к месту. Для таких совокупностей предшествующие выводы ослабляются. Но ведь всякую разноизменчивую совокупность мы можем всегда разбить на некоторое число частных совокупностей, в пределах которых δ' может меняться так мало, что их можно считать равноизменчивыми. При таком условии предшествующие положения остаются в силе и для разноизменчивых совокупностей.

Предыдушие выводы также вполне справедливы и в случае относительных величин, так как мы показали в § 22, что случай относительных величин есть частный случай выборочного исследовлния для средних величин.

Выяснив, что лучшим способом отбора единиц выборочной совокупности является механический отбор, будем строить наши предположения, что в дальнейшем осуществляется этот отбор.

---

[1]) см. Е. Е. Слуцкий. Теория корреляции стр. 83

ГЛАВА IV.

Применение выводов предшествующих глав на практике.

I. Постановка или организация выборочного исследования.

§ 29. Об определении численности выборочной группы при отсутствии районирования.

При организации всякого выборочного исследования прежде всего возникает вопрос, какова должна быть численность выборочной группы, чтобы окончательные выводы отличались бы надлежащей степенью точности.

Мы знаем, что между численностью выборочной группы и точностью окончательных результатов существуют некоторые соотношения, которые имеют различный вид в зависимости от того—производится ли случайный отбор или механический и имеет ли место районирование или нет. Самым простым случаем является тот когда районирование не имеет место и единицы выборочной группы отбираются на удачу. В этом случае ошибка средней (или относительной величины) определяется равенством

$$\sigma = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)}\ .$$

Ошибка же общего итога получится умножением этой ошибки на численность генеральной совокупности. Таким образом чем меньше будет ошибка средней, тем меньше будет и ошибка общего итога и наоборот. Ввиду такого соотношения между этими ошибками, в дальнейшем можно будет ограничиться исследованием лишь ошибки средней.

Если мы хотим, чтобы эта ошибка была меньше наперед заданной величины ε, то n надо подобрать так, чтобы удовлетверялось неравенство

$$\sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)} \leq \varepsilon.$$

Решая это неравенство относительно n, получим

$$\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right) < \varepsilon^2\ ;\ \frac{\delta^2}{\varepsilon^2}\left(1 - \frac{n}{N}\right) < n;\ \frac{\delta^2}{\varepsilon^2} \leq n + \frac{\delta^2}{\varepsilon^2 N}\ n\ ;$$

$$\frac{\delta^2}{\varepsilon^2} \leq n\left(1 + \frac{\delta^2}{\varepsilon^2 N}\right); \text{ откуда } n \geq \frac{\frac{\delta^2}{\varepsilon^2}}{1 + \frac{\delta^2}{\varepsilon^2 N}}$$

или разделив числителя и знаменателя на $\frac{\delta^2}{\varepsilon^2 N}$

$$n \geq \frac{N}{1 + \frac{\delta^2}{\varepsilon^2 N}}.$$

Так как мы стремимся получить возможно малое значение n, то следует взять

$$n = \frac{N}{1 + \frac{\delta^2}{\varepsilon^2 N}}. \tag{64}$$

В этом случае ошибка σ в точности будет равна ε. Из (64) легко получим и пропорцию отбора

$$\frac{n}{N} = \frac{1}{1 + \frac{\delta^2}{\varepsilon^2 N}}. \tag{65}$$

Выражеиия (64) и (65) показывают, что для вычисления численности выборочной группы n и пропорции отбора $\frac{n}{N}$ необходимо знать численность генеральной совокупности N и дисперсию вариант δ около среднего значения их. Если δ нам не известна, то для получения представления о ней можно в порядке подготовительных работ или разработать некоторое число карточек предшествующего сплошного наблюдения или, если и это невозможно, то в натуре произвести исследование небольшого числа вариант и, по этим данным приближенно вычислить δ[1]).

Формулы (64) и (65) выведены нами в предположении случайного отбора. Но зная, что погрешность при механическом отборе меньше чем при случайном мы можем быть уверены, что при вычисленных по формулам (64) и (65) значениях n и $\frac{n}{N}$ ошибки не превзойдут ε, что нам и нужно.

Впрочем, необходимо заметить, что если материал можно считать относительно однородным, то, как следует из выводов § 27, ошибки среднего при механическом и случайном отборе, вообще го-

---

[1]) Этому вопросу посвящена неопубликованная часть работы С. С. Кона.

воря, совпадут. Различие между ошибками при том или другом способе отбора может быть лишь тогда, когда материал не однороден, но и в этом случае можно довольно близко подойти к действительной ошибке при механическом отборе если расчленить генеральную совокупность на относительно однородные частные совокупностуи; речь об этом будет в следующем §. Такое расчленение генеральной совокупности на частные трудно выполнимо если не имеется достаточных данных для суждения о характере неоднородности материала; но в таком случае для вычисления n лучше всего, конечно, воспользоваться формулой

$$n = \frac{N}{1 + \frac{\delta^2}{\varepsilon^2 N}},$$

которая может быть и даст для n значение несколько преувеличенное, но зато будет иметься гарантия, что ошибка среднего не превзойдет наперед заданной величины ε.

§ 30. Сопоставление со схемой Bowley.

Если бы мы для вычисления n воспользовались формулой Bowley

$$\sigma = \sqrt{\frac{\delta^2}{n}},$$

то как и раньше имели бы

$$\sqrt{\frac{\delta^2}{n}} \leq \varepsilon ; \quad \frac{\delta^2}{n} \leq \varepsilon^2 ; \quad \frac{\delta^2}{\varepsilon^2} \leq n \text{ или } n \geq \frac{\delta^2}{\varepsilon^2}$$

и

$$\frac{n}{N} \geq \frac{\delta^2}{\varepsilon^2 N} ; \quad \frac{N}{n} \leq \frac{\varepsilon^2 N}{\delta^2} .$$

Или, так как мы стремимся получить наименьшее значение n, то следует положить

$$n = \frac{\delta^2}{\varepsilon^2} \text{ и } \frac{n}{N} = \frac{\delta^2}{\varepsilon^2 N} \qquad (66) .$$

Посмотрим теперь какое соотношение будет существовать между n вычисленным по формуле $n = \frac{\delta^2}{\varepsilon^2}$ и тем n, которое получается по формуле (64), если в той и другой формуле наперед заданная ошибка ε будет иметь одно и тоже значение.

Из формулы (66) получим

$$\frac{N}{n} = \frac{\varepsilon^2 N}{\delta^2}$$

из формулы же (64) получим

$$\frac{N}{n} = \frac{\varepsilon^2 N}{\delta^2} + 1\,,$$

откуда, сравнивая эти две формулы и обозначая n вычисленное по формуле (66) через n' будем иметь

$$\frac{N}{n} = \frac{N}{n'} + 1 \qquad (67)$$

Это и будет искомое соотношение между n в схеме невозвращенных и возвращенных шаров.

Замечая, что отношение $\frac{N}{n}$ показывает насколько единиц генеральной совокупности приходится одна единица выборочной группы, можно, на основании формулы (67), сказать, что если в схеме возвращенных шаров одна единица выборочной группы приходится на $\frac{N}{n'}$ единиц генеральной совокупности, то в схеме невозвращенных шаров одна единица выборочной группы придется на $\frac{N}{n'} + 1$ единиц генеральной сопокупности, т. е. на одну единицу больше. Например, если по формуле Bowley нужно брать каждый 5, 10, 20 и т. д. двор, то для получения той же степени точности по нашим формулам следовало бы брать 6, 11, 21 и т. д. двор.

Называя пропорцию отбора в схеме невозвращенных шаров через P, а в схеме возвращенных шаров через P' и замечая, что

$$P = \frac{n}{N} \text{ и } P' = \frac{n'}{N}\,,$$

будем иметь на основании формулы (67)

$$\frac{1}{P} = \frac{1}{P'} + 1$$

откуда

$$P' = \frac{P}{1 - P}\,,$$

т. е. если пропорция отбора в схеме невозвращенных шаров есть P, то для достижения той же степени точности, пропорцию отбора в схеме возвращенных шаров должно быть

$$\frac{P}{1 - P} .$$

Пользуясь этим соотношением найдем, что если % отбора в схеме невозвращенных шаров будет

0, 10, 20, 30, 40, 50, 60, 70, 80, 90, 100,

то соответствующий % отбора в схеме возвращенных шаров для достижения той же степени точности должен быть

0, 11, 25, 43, 67, 100, 150, 233, 400, 900, ∞ .

Предшествующие рассуждения имели целью определить n по заданной ошибке среднего. Посмотрим какие изменения нужно внести, если мы вычисляем n по заданной ошибке распространенного итога. Пусть мы хотим, чтобы ошибка распространенного итога σ не превосходила бы ε'. Как известно, ошибка распространенного итога связана с ошибкой среднего соотношением:

$$\sigma_{\text{ср.}} = N\sigma ,$$

т. е. ошибка распространенного итога в N раз больше ошибки среднего или ошибка среднего в N раз меньше ошибки распространенного итога. Поэтому, если мы хотим, чтобы ошибка распространенного итога была бы меньше ε' нужно, чтобы ошибка среднего была меньше $\varepsilon = \frac{\varepsilon'}{N}$; таким образом вопрос сводится к предыдущему.

§ 31. Вычисление численности выборочной группы в случае нескольких районов.

Перейдем теперь к тому случаю, когда материал неоднороден. В этом случае, как мы знаем, целесообразно провести районирование. Может оказаться, что для разных районов δ будут одинаковы или различны. Разберем сперва тот случай когда δ одинаковы. В этом случае, нам известно, что следует установить для всех районов один и тот же % отбора, т. е.

$$\frac{n_i}{N_i} = \frac{n}{N} .$$

Ошибка распространенного итога и среднего выражается при районировании формулой (45) и (46). Возьмем выражение ошибки для средних:

$$\sigma_{\text{ср.}} = \frac{1}{N} \sqrt{\sum_{i=1}^{S} N_i^2 \frac{\delta_i^2}{n_i} \left(1 - \frac{n_i}{N_i}\right)} .$$

Его можно переписать так:

$$\sigma_{ср.} = \frac{1}{N} \sqrt{\sum_{i=1}^{S} N_i \, \delta_i^2 \, \frac{N_i}{n_i}\left(1 - \frac{n_i}{N_i}\right)},$$

так как по нашим предположениям $\delta_i = \delta$, а все $\frac{n_i}{N_i} = \frac{n}{N}$, то имеем

$$\sigma_{ср.} = \frac{1}{N} \sqrt{\sum_{i=1}^{S} N_i \, \delta^2 \, \frac{N}{n}\left(1 - \frac{n}{N}\right)}$$

Вынося $\delta^2 . \frac{N}{n}\left(1 - \frac{n}{H}\right)$ общим множителем за знак суммы и замечая, что $\Sigma N_i = N$, получим

$$\sigma_{ср.} = \frac{1}{N} \sqrt{\delta^2 \frac{N}{n}\left(1 - \frac{n}{N}\right) N}$$

или

$$\sigma_{ср.} = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)}.$$

Получилось такое же выражение ошибки какое получается при отсутствии районировании. Это обстоятельство позволяет вычислитить n в данном случае по тем же формулам, по которым мы вычисляли его не имея представления о характере материала, т. е. по формулам (64) и (65).

Пусть теперь в разных районах δ будут различны и предположим, что они известны из старых источников. В этом случае, как мы знаем, целесообразно выбирать численности районных выборочных групп n_i по формулам (48) или (49):

$$n_i = \frac{n \, N_i \, \delta_i}{\sum\limits_{i=1}^{S} N_i \, \delta_i} \quad \text{и} \quad \frac{n_i}{N_i} = \frac{n \, \delta_i}{\sum\limits_{i=1}^{S} N_i \, \delta_i} \qquad (68)$$

где $n = \Sigma \, n_i$ есть численность выборочной группы для всей генеральной совокупности.

Однако, желание упростить выборочное исследование может побудить нас выбирать n_i так, чтобы пропорции $\frac{n_i}{N_i}$ были бы во всех районах одинаковы и равны $\frac{n}{N}$. В этом случае формула (46) переходит, как мы знаем в (51), которую можно переписать так:

$$\sigma = \sqrt{\frac{\Sigma N_i \delta_i^2}{N.\, n}\left(1 - \frac{n}{N}\right)}$$

Полагая в этой формуле для сокращения

$$\frac{\Sigma N_i \delta^2}{N} = \overline{\delta}^2,$$

получим

$$\sigma_{ср.} = \sqrt{\frac{\overline{\delta}^2}{n}\left(1 - \frac{n}{N}\right)}.$$

В этом случае $\overline{\delta}^2$ представляет взвешенную среднюю из отдельных районных δ_i^2. Поэтому, если вместо отдельных δ_i^2 ввести их взвешенное среднее $\overline{\delta}^2$, то ошибка средней (и распространенного итога) выражается прежней формулой с заменой δ^2 на

$$\overline{\delta}^2 = \frac{\Sigma N_i \delta_i^2}{N}.$$

Следовательно, и в этом случае, для вычисления n можно пользоваться формулами (64) или (65), заменив только δ^2 на $\overline{\delta}^2$.

§ 32.

Если же мы хотим использовать все средства для уменьшения n, то, конечно, надо будет брать разные пропорции по формулам (68). В этом случае удобнее исходить из ошибки распространенного итога (45):

$$\sigma = \sqrt{\Sigma N_i^2 \frac{\delta_i^2}{n_i}\left(1 - \frac{n_i}{N_i}\right)} = \sqrt{\Sigma N_i^2 \delta_i^2 \left(\frac{1}{n_i} - \frac{1}{N_i}\right)}$$

подставляя сюда вместо n_i его значение из (65), получим

$$\sigma = \sqrt{\Sigma N_i^2 \delta_i^2 \left(\frac{\Sigma N_i \delta_i}{n N_i \delta_i} - \frac{1}{N_i}\right)} = \sqrt{\Sigma \frac{N_i \delta_i \Sigma N_i \delta_i}{n} - \Sigma N_i \delta_i^2};$$

вынося в первой сумме за Σ множитель

$$\frac{\Sigma N_i \delta_i}{n},$$

получим

$$\sigma = \sqrt{\frac{\Sigma N_i \delta_i}{n} \cdot \Sigma N_i \delta_i - \Sigma N_i \delta_i^2}$$

или короче

$$\sigma = \sqrt{\frac{(\Sigma N_i \delta_i)^2}{n} - \Sigma N_i \delta_i^2}\ .$$

Если хотим , чтобы ошибка распространенного итога не превосходила бы ε', то надо подобрать n так, чтобы было

$$\sqrt{\frac{(\Sigma N_i \delta_i)^2}{n} - \Sigma N_i \delta_i^2} \geq \varepsilon'$$

или

$$\frac{(\Sigma N_i \delta_i)^2}{n} - \Sigma N_i \delta_i^2 \leq \varepsilon'^2 ; \frac{(\Sigma N_i \delta_i)^2}{n} < \varepsilon'^2 + \Sigma N_i \delta_i^2$$

откуда

$$\frac{n}{(\Sigma N_i \delta_i)^2} \geq \frac{1}{\varepsilon'^2 + \Sigma N_i \delta_i^2}$$

и окончательно

$$n \geq \frac{(\Sigma N_i \delta_i)^2}{\varepsilon'^2 + \Sigma N_i \delta_i^2}\ .$$

Вот каким должно быть n, чтобы ошибка распространенного итога не превосходила бы ε'. Если хотим, чтобы ошибка среднего не превосходила ε, то помня, что $\varepsilon' = N\varepsilon$ получим для этого случая

$$n \geq \frac{(\Sigma N_i \delta_i)^2}{N^2\varepsilon^2 + \Sigma N_i \delta_i^2}$$

n у нас, как и прежде, численность всей выборочной группы. Выборочные же группы n_i для отдельных районов и пропорции выборки в них определятся по формулам (65), т. е.

$$n_i = \frac{n N_i \delta_i}{\Sigma N_i \delta_i} \quad \text{и} \quad \frac{n_i}{N_i} = \frac{n\delta_i}{\Sigma N_i \delta_i}$$

Изложенное распространяется и на случай относительных величин с заменой δ_i^2 через $p_i\, q_i$.

II. Разработка результатов выборочного исследования.

Задача настоящего отдела заключается в вычислении распространенных средних или относительных величин, распространенных итогов и оценке точности найденных результатов. Начнем с простейшего.

§ 33. Случай одного района.

В данном случае нам нет надобности предполагать, что материал однороден, так как погрешности мы будем оценивать по формулам выведенным для случайного отбора, а заменяя случайный отбор механическим мы рискуем лишь тем, что действительные ошибки, как было показано раньше, будут меньше вычисленных по нашим формулам. Переходя к вычислению среднего или относительного значений мы примем без дальнейших рассуждений, что распространенные на всю генеральную совокупность средние и относительные равны с известным приближением таким же средним и относительным найденным нами на основании выборочной группы.

Таким образом, если генеральная совокупность N, выборочная группа—n; итог для выборочной группы $\sum\limits^{n} x$ (для относительных чисел это будет m) и итог для генеральной совокупности $\sum\limits^{N} x$ (для относительных это дадет M), то будем полагать

$$\frac{\sum\limits^{N} x}{N} \approx \frac{\sum\limits^{n} x}{n} \pm \sigma \quad \text{и} \quad \frac{M}{N} \approx \frac{m}{n} \pm \sigma$$

где σ означает соответствующую ошибку. От этих выражений умножая на N переходим к распространенным итогам:

$$\sum\limits^{H} x \approx N\frac{\sum\limits^{n} x}{n} \pm N\sigma \quad \text{и} \quad M \approx N\frac{m}{n} \pm N\sigma \;.$$

Последние слагаемые этих выражений будут также соответствующими ошибками. Для вычисления ошибок придется, конечно, на основании соображений § 29, воспользоваться формулой

$$\sigma = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)}\,,$$

которую для практических целей лучше всего переписать так:

$$\sigma = \sqrt{\frac{\delta^2}{n} \; \frac{N-n}{N}} \; .$$

В случае относительных величин следует положить

$$\delta^2 = p\,q \cong \frac{m\,(n-m)}{n^2}$$

и формула ошибки для относительных величин примет вид:

$$\sigma = \sqrt{\frac{m\,(n-m)}{n^2} \; \frac{N-n}{N}} \; .$$

Величину δ удобнее всего вычислять пользуясь способом моментов [1]). Этим способом можно также воспользоватся и для вычисления средних. Называя через ν_1 и ν_2 нецентральные моменты, а через μ_1 и μ_2 центральные, будем как известно, иметь

$$\frac{\sum^{n} x}{n} = \mu_1 \quad \text{и} \quad \delta^2 = \mu_2 = \nu_2 - \nu_1^2 \; .$$

§ 34. Случай нескольких райнов.

В этом случае для какого нибудь i—того района будем иметь:

$$\frac{\sum^{N_i} x}{N_i} \cong \frac{\sum^{n_i} x}{n_i} \pm \sigma_i \quad \text{и} \quad \frac{M_i}{N_i} \cong \frac{m_i}{n_i} \pm \sigma_i \; .$$

$$\sum^{N_i} x \cong N_i \frac{\sum^{n_i} x}{n_i} \pm N_i\,\sigma_i \quad \text{и} \quad M_i \cong N_i \frac{m_i}{n_i} \pm N_i\,\sigma_i \; .$$

Сумируя, получаем для всей генеральной совокупности

$$\sum^{N} x = \sum_{i=1}^{S} \sum^{N_i} x \pm \sqrt{\sum_{i=1}^{S} \left(N_i\,\sigma_i \right)^2} \; ;$$

---

1) И. Yule. ·Yntroduction to the theory of statistics. 1912 г. 108—112, 134—142.
Е. Е. Слуцкий—Теория корреляции—стр. 10. 19 и след.
В. В. Голубев—Введение в математическую статистику—стр. 57, 58.
Р. Орженцкий—Сводные признаки стр. 192 и след.
Его же — Учебник математической статистики—стр. 54, 57.

$$M = \Sigma M_i \pm \sqrt{\Sigma (N_i \sigma_i)^2}.$$

Разделив же на N найдем распространенные средние

$$\frac{\sum\limits^{N} x}{N} = \frac{\sum\limits_{i=1}^{S}\sum\limits^{N} x}{N} \pm \frac{1}{N}\sqrt{\sum\limits_{i=1}^{S}\left(N_i \sigma_i\right)^2}.$$

Аналогично для относительных величин:

$$\frac{M}{N} = \frac{\Sigma M_i}{N} \pm \frac{1}{N}\sqrt{\Sigma\left(N_i \sigma_i\right)^2}.$$

В Заключение скажем, что этими формулами можно воспользоваться и для групповой разработки материала, причем, в этом случае под n_i нужно будет понимать численности групп в выборочной совокупности, под N_i — численности групп в генеральной совокупности и под σ_i — ошибки средних или относительных величин в группах.

Численности N_i групп в генеральной совокупности находятся по правилам, изложенным в главе об относительных величинах в виде произведений $N\frac{n_i}{n}$

Ошибки этих численностей найдутся по формуле (11), которая в применении к рассматриваемому случаю даст для N_i ошибку.

$$\sigma_{N_i} = \sqrt{\frac{n_i\,(n - n_i)}{n^3}\left(1 - \frac{n}{И}\right)}.$$

Здесь важно заметить, что вычисляя для какой нибудь группы общий итог по формуле

$$\sum^{N_i} x \underset{S}{=} N_i \frac{\sum\limits^{n_i} x}{n_i} \qquad (*)$$

нельзя будет ошибку этого итога считать равной произведению

$$N_i\ \sigma_i$$

где σ_i ошибка среднего $\frac{\sum\limits^{n_i} x}{n_i}$, так как в данном случае значение N_i будут известны не точно, а приближенно с вышеуказанной погрешностью.

Однако, соединяя по известным правилам погрешности N_i и $\frac{\sum\limits^{n_i} x}{n_i}$, получим для погрешности равенств (*) значение

$$\sqrt{\left(N_i\,\sigma_i\right)^2+\left(\xi_i\,\sigma_{N_i}\right)^2}\ ^{1)}$$

где $\xi_i = \frac{\sum\limits^{n_i} x}{n_i}$. Имея групповые итоги $\sum\limits^{N_i} x$ можно получить по формулам (стр. 132) итог $\sum\limits^{N} x$ и среднюю $\frac{\sum\limits^{N} x}{N}$ для всей генеральной совокупности.

---

[1]) Это значение погрешности можно найти следующим образом: Полагая $\frac{\sum\limits^{n_i} x}{n_i} = \xi_i$ можно равенство (*) переписать так

$$\sum^{N_i} x = N_i\,\xi_i$$

Далее пусть $\triangle N_i$ и $\triangle\xi_i$ будут действительными ошибками величин N_i и ξ_i. Если считать эти ошибки величинами малыми, то действительная ошибка произведения $N_i\,\xi_i$ представится в виде

$$N_i \triangle \xi_i + \xi_i \triangle N_i\,.$$

Взяв квадрат этой ошибки и вычислив его математическое ожидание получим, как известно, ошибку произведения $N_i\,\xi_i$.

$$\text{Но м. о. } \left(N_i \triangle\xi_i + \xi_i \triangle N_i\right)^2 =$$

$$= N_i^2 \text{ м. о. } \left(\triangle\xi_i\right)^2 + 2N_i\,\xi_i \text{ м. о. } \left(\triangle\xi_i \triangle N_i\right) + \xi_i^2 \text{ м. о. } \left(\triangle N_i\right)^2$$

Математическое ожидание величины $\left(\triangle\xi_i\right)^2$ есть, очевидно, квадрат средней ошибки величины $\frac{\sum\limits^{n_i} x}{n_i}$, которая была обозначена нами через σ_i.

Аналогично, м. о. $\left(\triangle N_i\right)^2 = \sigma_{N_i}^2$. Считая, теперь, величины ξ_i и N_i независимыми будем иметь

$$\text{м. о.}\left(\triangle\xi_i \triangle N_i\right) = \text{м. о. } \triangle\xi_i \text{ м. о. } \triangle N_i\,.$$

Но математические ожидания величин $\triangle\xi_i$ и $\triangle N_i$, очевидно, равны 0, следовательно, При этом, однако, необходимо заметить, что ошибки этих выражений не могут быть найдены так просто, ка раньше, ибо групповые итоги $\Sigma_{x}^{N_i}$ не независимы между собою. [2])

§ 35.

Для иллюстрации применения формул и техники расчета рассмотрим следующий пример.

Предположим, что мы произвели выборочное исследование с целью определения среднего количества лошадей на хозяйство, пропорций безлошадных, однолошадных и т. д. хозяйств и общего количества лошадей с соответствующей оценкой погрешностей в N-ском уезде, который дзя достижения возможной однородности материала был разбит на четыре района.

Пусть результаты этого выборочного исследования даны таблицей:

№№ районов.	Общее число хозяйств N_i	Число обслед. хозяйств n_i	Группировка х-в по числу лошадей (по данным выбор. группы).					
			0	1	2	3	4	5
I	16817	1682	632	883	135	21	9	2
II	20635	1032	387	547	83	12	2	1
III	15421	1030	387	542	83	12	5	1
IV	10734	2150	808	1132	171	25	11	3

$$\text{м. о.}\left(N_i\,\Delta\xi_i+\xi_i\,\Delta N_i\right)^2=\left(N_i\,\sigma_i\right)^2+\left(\xi_i\,\sigma_{N_i}\right)^2.$$

Извлекая отсюда корень квадратный мы и получим, указанное нами, выражение ошибки произведения $N_i\,\dfrac{\Sigma_x^{n_i}}{n_i}$.

[2]) Можно было бы найти ошибки общего итога Σ_x^{N} и среднего $\dfrac{\Sigma_x^{N}}{N}$ введя в рассмотрение коэффициенты корреляции между групповыми итогами, но такой путь являясь очень сложным не имеет в то же время никакого практического значения, так как ошибки этих величин могут быть найдены ранее и независимо от групповой разработки материала.

1. Вычисление средней и ея ошибки:

Для I района

v	f	fv	fv2
0	632	0	0
1	883	883	883
2	135	270	540
3	21	63	189
4	9	36	144
5	2	10	50
	1682	1262	1806

$$\frac{\Sigma fv}{\Sigma f} = \frac{\overset{n_1}{\Sigma x}}{n_1} = \mu_1 = \frac{1262}{1682} = 0{,}75;\ \mu_2 = \frac{\Sigma fv^2}{\Sigma f} =$$

$$= \frac{1806}{1682} = 1{,}07;\ \delta^2 = \mu_2^1 = \mu_2 - \mu_1^2 = 1{,}07 - 0{,}56 =$$

$$0{,}5;\ \delta = 0{,}7071.\ \sigma_{\mu_1} = \sqrt{\frac{\delta^2}{n}\left(1 - \frac{n}{N}\right)} =$$

$$= \delta\sqrt{\frac{N-n}{nN}} = 0{,}7071\sqrt{\frac{16817-1682}{16817.1632}} = 0{,}0163.$$

$$\text{Поэтому } \frac{\overset{N_1}{\Sigma}}{N_1} = 0{,}75 + 0{,}0163.$$

Для II района.

v	f	fv	fv²
0	387	0	0
1	547	547	547
2	83	166	332
3	12	36	108
4	2	8	32
5	1	5	25
	1032	762	1044

$$\frac{\overset{n_2}{\Sigma x}}{n_2} = \frac{\Sigma fv}{\Sigma f} = \mu_1 = \frac{762}{1032} = 0{,}738;\ \mu_2 = \frac{1044}{1032} =$$

$$= 1{,}011;\ \delta^2 = 1{,}011 - 0{,}544 = 0{,}47;\ \delta = 0{,}6856.$$

$$\sigma = 0{,}6856\sqrt{\frac{19603}{21.295.320}} = 0{,}0204.$$

$$\frac{\overset{\Sigma_2}{\Sigma x}}{\Sigma_2} = 0{,}738 \mp 0{,}0204.$$

Для III района.

v	f	fv	fv²
0	387	0	0
1	542	542	542
2	83	166	332
3	12	36	108
4	5	20	40
5	1	5	25
	1030	769	1047

$$\mu_1 = \frac{769}{1030} = 0{,}746\ ;\ \mu_2 = \frac{1047}{1030} = 1{,}016$$

$$\delta^2 = 1{,}016 - 0{,}556 = 0{,}46;\ \delta = 0{,}6782$$

$$\sigma = 0{,}6782\sqrt{\frac{14.391}{15.883.630}} = 0{,}0203.$$

$$\frac{\overset{N_3}{\Sigma x}}{N_3} = 0{,}74 \pm 0{,}0203.$$

Для IV района.

v	f	fv	fv²
0	808	0	0
1	1132	1132	1132
2	171	342	684
3	25	75	225
4	11	44	176
5	3	15	75
	2150	1608	2292

$$\mu_1 = \frac{1608}{2150} = 0{,}747 \quad ; \quad \mu_2 = \frac{2292}{2150} = 1{,}066$$

$$\delta^2 = 1{,}066 - 0{,}558 = 0{,}5; \; \delta = 0{,}7071$$

$$\sigma = 0{,}7071 \sqrt{\frac{8584}{23.078.100}} = 0{,}0134.$$

$$\frac{\overset{N_4}{\Sigma x}}{N_4} = 0{,}75 \pm 0{,}0134.$$

2. Вычисление общего количества (или распространенного итога) лошадей в районах, в уезде и средней лошадности для уезда.

Распространенный итог для какого нибудь i—того райна $\overset{N_i}{\Sigma x}$, как нам известно, определяется по формуле

$$\overset{N_i}{\Sigma x} = N_i \, \frac{\overset{n_i}{\Sigma x}}{n_i} + N_i \, \sigma_i \; . \; .$$

Подставляя последовательно в нее наши значения будем иметь

для I района $\overset{N_1}{\Sigma x} = (0{,}75 \pm 0{,}0163). \; 16.817 = 12.613 \pm 274.$

„ II „ $\overset{N_2}{\Sigma x} = (0{,}738 \pm 0{,}0204). \; 20.635 = 15.228 \pm 421.$

„ III „ $\overset{N_3}{\Sigma x} = (0{,}746 \pm 0{,}0203). \; 15.421 = 11.504 \pm 313.$

„ IV „ $\overset{N_4}{\Sigma x} = (0{,}747 \pm 0{,}0134). \; 10.734 = 8.018 \pm 143.$

Распространенный итог для всего уезда или генеральной совокупности определится по формуле:

$$\overset{N}{\Sigma x} = \sum_{i=1}^{s} \overset{N_i}{\Sigma x} \pm \sqrt{\sum_{i=1}^{s} \left(N_i \; \sigma_i \right)^2} \, .$$

Подставляя наши значения получим:

$$\sum_{i=1}^{4}\sum^{N_i} x = 47.363; \sqrt{\sum_{i=1}^{4}\left(N_i \quad \sigma_i\right)^2} = \sqrt{274^2+421^2+313^2+143^2} =$$

$$= \sqrt{370735} = 608.$$

Следовательно,

$$\sum^{N} x = 47.363 \pm 608,$$

где 608 есть абсолютная величина ошибки, относительна же ее величина. равная $\frac{608.100}{47.363}$, будет равна 1,28%.

Среднюю для всего уезда получим по формуле

$$\frac{\sum^{N} x}{N} = \frac{\sum_{i=1}^{s}\sum^{N_i} x}{N} \pm \frac{\sqrt{\sum_{i=1}^{s}\left(N_i \quad \sigma\right)^2}}{N}.$$

Переставляя наши значения найдем

$$\frac{\sum^{N} x}{N} = 0{,}74 \pm 0{,}0095 \qquad (N = 63.607)$$

Относительная ошибка будет равна 1,28%.

3. Найдем теперь пропорции безлошадных хозяйств; для примера ограничимся лишь I-м районом нашего уезда.

В этом случае, как известно, нужно воспользоваться формулой

$$\frac{M_i}{N_i} = \frac{m_i}{n_i} \pm \sqrt{\frac{p_i \, q_i}{n_i} \quad \frac{N_i - n_i}{N_i}} = \frac{m_i}{n_i} \pm \sqrt{\frac{m_i(n_i - m_i)}{n_i^{\,3}} \quad \frac{N_i - n_i}{N_i}}$$

или подставляя числа будем иметь для I района

$$\frac{M_1}{N_1} = \frac{632}{1682} \pm \sqrt{\frac{632(1682-632)}{1682^3} \quad \frac{16.817-1682}{16.817}} =$$

$$= \frac{632}{1682} \pm \sqrt{0{,}000012} = 0{,}3757 \pm 0{,}0034.$$

Относительная ошибка равна 0,9049%.

Аналогично поступая для остальных районов мы можем получить и для них пропорцию безлошадных хозяйств. Применяя же формулу

$$M = N\frac{m}{n} \pm N\sigma$$

мы определим численность безлошадных хозяйств в районах с оценкой погрешностей, а отсюда легко получим такой же результат и для всего уезда, т. е. для генеральной совокупности.