

## **ПРИЛАГАНЕ НА РЕПРЕЗЕНТАТИВНИЯ МЕТОД ПРИ РАЗРАБОТКАТА НА МАТЕРИАЛИТЕ НА ГЛАВНА ДИРЕКЦИЯ НА СТАТИСТИКАТА**

### **Доклад до Върховния статистически съвет**

Доколкото можем да съдим, въз основа на сведенията, които имаме, разработката на резултатите от преброяването през 1926 г. и особено на картите “Ж” и “З” твърде закъснява. Тя едва ли ще бъде завършена към края на 1930 г., т.е. към времето, когато би следвало да се прави вече новото преброяване. Подобно закъснение за сегашното време се явява твърде нежелателно, във връзка с широките планове на Министерството на земеделието, а също, защото поставя под въпрос и самото преброяване за 1930 г. и специално стопанската част на последното. Без да се знаят резултатите на твърде смелия в много отношения опит от 1926 г., същият да бъде повторен е, разбира се, невъзможно.

Ето защо възможността за ускорено получаване, макар и на приблизителни резултати, от преброяването за 1926 г. би представлявала голям практически интерес – разбира се, при условие, че тази приблизителност не ще компрометира окончателните резултати от разработката и не ще предизвика нови разходи. Нашето мнение е, че подобна възможност има, ако се приложи тъй наречения “репрезентативен метод”, при което, може да се надяваме, дори би се постигнала твърде съществена икономия от парични средства, отпуснати за разработката на преброяването.

Що е “репрезентативният метод”? Изразено накратко, това е сурогатът, Ersatz'a на общото статистическо наблюдение (или в други случаи, на пълната обработка на вече събрания статистически материал), построен примерно на същия принцип, на който лежи и определянето качеството на голяма пратка от стока въз основа на анализиране на проби от същата. При приемане на каменни въглища, зърнени храни, брашна и пр. – от разни места на стоката, въз основа на определени и от дълга практика изпитани правила се вземат “пробите”, т.е. малки количества от стоката се подлагат на щателно изследване. Ако получените “проби” са още доста неудобни за анализ, то от тях, въз основа на аналогични правила, се извличат нови “проби” и т.н.. По този начин наместо общо изследване на цялата пратка от дадена пратка от дадена стока всъщност се изследва и изпитва само една десета или стотна част от нея и с това следователно се икономисват 10 или 100 пъти приблизително пари и време. Търговията с успех ползва този начин, понеже при нормални условия тя еднакво се нуждае както от абсолютна точност на анализа, тъй и от абсолютна точност при определяне теглото на стоката: при продажбата или покупката на хиляди тонове въглища нас свършено не ни

интересуват грамове или милиграмовете от теглото на въглицата и прилагането тук например на точни химически везни би било нещо повече от лудост. А приблизителна точност в анализа на стоката на основание на нейните “проби” се гарантира от изпитани в течение на дълги години от практиката правила за тази процедура. Методът на “пробите” вследствие на това е не само евтин и бърз, но и надежден, а често дори единствено възможен (например калорийността на въглицата се определя с изгарянето на част от тях в известен апарат, защото иначе трябва да изгорим цялото количество!).

В областта на практическото прилагане данните на статистиката ние в много случаи срещаме напълно аналогично положение на нещата.

*Първо*, огромен е броят на случаите, когато в действителност на нас свършено не ни е нужна абсолютна точност на ония статистически цифри, с които оперираме. Например практически ни е свършено безразлично съставлявало ли е населението на България сутринта на 1 януари 1927 г. точно 5,483,125 човека, или 5,480,000, или може би 5,500,000; не е важно дали в България в 1924 г. са се раждали 39.8, или 39.7, или 39.9 души на хиляда; съставлявала ли е общата сума на вноса за 1927 г. точно 6,133,042 лева, или може би 6,150,000 лв., имало ли е в Айтоска околия в 12 часа през нощта на 31 декември 1926 г. точно 5,311 свини, или може би 5,400, или пък наопаки 5,200 и т.н. Трудно би било дори да си въобразим такъв случай, когато за работата на държавника, учения или практика подобни дребни разлики биха имали каквото и да е реално значение.

*Второ*, само в много редки случаи резултатите от общите статистически наблюдения биват свършено точни: обикновено тяхната видима точност е чиста фикция. Както сочи например проф. Вестергорд,<sup>1</sup> почти всяко общо преброяване в голяма или малка степен е “репрезентативно”, тъй като всъщност то почти никога не обхваща всички обекти, които би трябвало да обхване – дали поради небрежността на статистическите органи, или благодарение злата воля на обекта на наблюдението, или най-сетне, предвид неяснотата в самото определяне признаците на обекта и неточностите в измерванията му.

Така например “Земеделската статистика на България за 1897 г.” е обхващала само 74,137,489.5 декара от 96,345,500.0 декара, които представляват истинската площ на България според геодезичните пресмятания, т.е. само 76.9%, и даже “Земеделската собственост през 1908 г.” е изпуснала пак 17.2% от площта на България. Главна дирекция на статистиката много находчиво, но едва ли с основание, е нарекла тази разлика (т.е. 22,208,010.5 дка в 1897 г. и 16,520,990.2 дка в 1908 г.) “неексплоатирана земя”, когато в действителност това е просто земя, която

---

<sup>1</sup> H. Westergaard und H. C. Nybble. Grundzuge der Theorie der Statistik, 2. Auflage, Jena 1928, S. 89.

по една или друга причина се е изплъзнала от статистическото наблюдение. В болшинството от статистическите публикации фигурира особена графа: “неизвестно”, “непоказано”, “неразпределено” и т.н., и всеки посветен знае, че ако статистическите учреждения не биха имали особено отвращение към тази графа, нейното съдържание би било много по-голямо: количество на нередовните, непълните, неясните и противоречиви статистически формуляри в действителност е много по-голямо, отколкото това е показано в статистическите публикации. Особено големи са, както се знае, пропуските и неопределеностите в областта на наблюдението над стопанските признаци, които могат да послужат за определяне на истинската доходност на предприятията, без да говорим вече за вторична статистика, която обработва материали, събрани от данъчните органи (например “данъчни декларации”). Така например I. S. Stamp,<sup>2</sup> като прави съпоставяне на статистиката на размерите на националния доход в главните държави на света, приема възможната грешка за едни до 10% (Англия, Германия), за други – до 20% (Франция), за трети до 30% (Япония), а за четвърти даже до 40% и повече (Канада, Италия, Испания, Австро-Унгария). И не без основание нашата Главна дирекция на статистиката публикува за подходящия данък само сведения за броя на данъкоплатците във всяка една категория на дохода и за сумата, платена от тях, но мълчи за общата сума, получена от всичките данъкоплатци годишен доход: това число би се получило смешно малко, едва ли 50% от истинския национален доход на България.

И по-нататък, *трето*, даже ако предположим, че произведеното статистическо наблюдение е абсолютно точно, все пак получените от нас цифри са точни само за момента, когато е произведено наблюдението. Ако допуснем, че на 12 часа през нощта на 31.XII. 1926 г. населението на България се равняваше на 5,484,125 души, то понастоящем, т.е. през юни 1928 г. това население е сигурно по-голямо. Главната дирекция на статистиката го оценява на 5,596,800 души към началото на 1928 г.; но точна ли е тази преценка? Нали същата Главна дирекция на статистиката преценяваше към средата на 1926 г. населението на България на 5,105,800 души (гл. “Месечни статистически известия”, дек. 1926 г.), а след преброяването, в следващия брой на същото списание (януари 1927 г.) е покачила тази цифра до 5,427,300, т.е. е признала направената грешка почти от 6%. Действителните количества на наличния добитък, домашните птици и пр. могат още по-значително да се отклоняват от скоро преди това установени посредством най-точните преброявания цифри, във връзка с определените сезони на клането, с изискванията на пиацата и със стойността на фуража. Така например критическият момент за свинете е Коледа и членовете на В.С.С., разбира се, добре знаят, че констатираното от преброяванията 1910 и 1920 г. огромно нарастване на броя на свинете за

---

<sup>2</sup> The Wealth and Income of chief Powers. - Journal of the Royal Statistical Society, 1919, July.

десетилетието в значителна степен се обяснява с това, че 1 януари по стар стил (1910 г.) се пада след, а 1 януари по нов стил (1920 г.) – преди Коледа.

Най-сетне, *четвърто*, възможни са случаи, когато стремежът към фиктивната точност на “общото” наблюдение, над всички обекти, може да стане причина на най-големи грешки и следователно само да влоши окончателните резултати. Такъв е например случаят с бюджетните изследвания, където “общо” и подробно изучаване на бюджетите на цялото население е просто технически невъзможно. Аз даже отчасти се опасявам, че и картата “Ж” се оказа не съвсем по силите на масовия тип на селския агент-преброител, с които разполагахме при извършване на преброяването. Аз смятам, че и досега е открит въпросът дали не щеше да е по-добре с карта “Ж” да се изследва само част от населението, но по-задълбочено и с помощта на по-квалифицирани сътрудници (например контролорите).

И така, статистическите числа, с които ние оперираме в повечето случаи, по самата си същност са *неточни* и са съпроводени от повече или по-малко *грешки*. Ако истинското число е А, то ни остава неизвестно, а фактически ние оперираме с величини А+а. Последният израз може да бъде

представен във вид  $A(1 + \frac{a}{A})$ , където  $\frac{a}{A}$  означава вече не абсолютната, а

относителната грешка на нашата цифра. Тази относителна грешка в зависимост от обстоятелствата може да се равнява и на 5%, и на 10%, понякога и на 50% и повече. Понякога ние не знаем дали тя е положителна или отрицателна; в други случаи ние можем със сигурност да определим нейния знак, т.е. да кажем в каква посока сме сгрешили. Така например броят на лицата на възраст са винаги кръгло число години (40, 50, 60 и т.н.), установен при преброяването, е винаги по-голям от истинския (грешка положителна); наопаки, броят на лицата, упражняващи така наречените срамни професии (кражба, проституция, просия), се оказва винаги по-малък от истинския (грешка отрицателна). Също така една отрицателна грешка ние можем да предположим в повечето случаи на стопанската статистика (количество на земята, на добитъка, размерите на дохода, особено на страничния, средното земеделско производство и пр.).

Естествено е да се постави въпросът, ако статистическите величини, с които оперираме на практика, са придружени винаги от по-голяма или по-малка грешка и ако при това в болшинството от случаите нам абсолютно точни цифри са съвсем ненужни, то следва ли винаги да се стремим към пълна точност и “изчерпателност” при събирането на първичния материал, определяйки, така да се каже, теглото на тоновете въглища с помощта на химически везни, с точност до един милиграм и не може ли обратното, в известни случаи в самото начало да се ограничим с по-сумарни, но затова пък и по-ефективни и бързо действащи “проби”? Необходимо е само да се определят пределите на възможната грешка и в действителност да сме

уверени, че избраният метод на приближения наистина ще ни даде резултата, който ще се отклони от истинския не в повече от някакъв предварително посочен процент (например 2, 5 или 10%). Тук именно лежат и ония трудности, които до днес смуцават статистиците-нематематици и спъват до известна степен по-широкото разпространение на репрезентативния метод в Западна Европа.

Ако изследователят в качеството на “проба” избира най-типичните или “средните” екземпляри от обектите на статистическото наблюдение, той може при достатъчен опит, учение и научна институция да достигне до изумително точни резултати. Но ако той няма тези качества, ако неговото “субективно средно” се отличава много от действителното, или, което е още по-лошо, той теоретически или практически е заинтересован в определен резултат от изследването си (и вследствие на това не избира “пробите” си безпристрастно) – цялата работа на репрезентативния метод ще бъде напълно компрометирана. Ето защо, поставен при условия, зависещи много от субективните качества на изследователя, репрезентативният метод не ще бъде никога признат за напълно убедителен и равноценен с общия.

Възможните предели на грешката могат да се определят само тогава, когато и “пробите” са избрани въз основа на известни обективни правила, независещи от волята и усмотрението на заинтересовани лица, т.е. когато това стане например точно по начина, по който се теглят печалбите от лотарията, както пада известен резултат на рулетката, числото на точките при заровете и т.н., с други думи, когато това стане “случайно”. Някоя отделно взета от нас “проба” при това може би ще се окаже по индивидуалните си признаци доста по-далече от средното ниво, отколкото например ако се избираха “типични случаи”, но средното от цялата “извадка” (така ще наричаме по-нататък репрезентативната маса, на руски “выборка”) благодарение на закона за “големите числа” ще бъде много по-близо до това ниво и във всеки случай теорията за вероятностите ще ни даде пълна възможност да определим за всяка “извадка”, с вероятност, колкото желаем близка до единица (т.е. достоверността), ония предели на грешката, които в никакъв случай няма да бъдат надминати. Теоретико-вероятностната страна на проблема днес за днес, след трудовете на професора от Лондонския университет А. Боули и тия на моя покоен учител А. А. Чупров-младши, трябва да се счита като окончателно изяснена: всички по-главни формули са изведени, всички предварителни условия са установени. И ако репрезентативният метод, така разработен от тия автори, не се ползва до днес още от всеобщо признание, например в Германия, причините на това явление следва да бъдат търсени не в недоразработеността на метода, а в субективни особености на болшинството представители на германската университетска статистическа наука от старото поколение и в частност на тяхното отвращение към всякаква математика (и особено към теорията на

вероятностите), тъй странно съчетавано с любовта им към цифрата.<sup>3</sup> Ние твърдим, че научно коректната постановка на репрезентативния метод, за да доведе до обективно правилни резултати (в пределите на една допустима и предварително процентно установена грешка), е възможна само тогава, когато при вземане на пробите е отстранен напълно личният произвол и когато следователно цялата “извадка” се извърши напълно, тъй както се теглят топките или билетчетата от затворена урна.

Възможно ли в подобно отъждествяване? Нали явленията на социалния живот са безкрайно разнообразни и сложни. Мислимо ли е да се отъждествяват описанието на една единица на наблюдение (семество, домакинство и пр.) с просто извличане на една топка от урната и именно да се отъждествява, а не само да се сравнява по аналогия? Това е възражението, което обикновено се прави от противниците на Боули-Чупаровската постановка на репрезентативния метод. За нас това възражение не се явява като основателно, и то ето защо. Съгласно определението на проф. Žižek статистиката е “мислене на групи” (Denken in Gruppen) или, ако се изразим с думите на руския статистик Журавски – “категорическо изчисляване” – в такъв смисъл, че всичките сами по себе си сложни единици на наблюдението се обединяват от нас, статистиците, в “групи” или “категории” въз основа на малкото, предварително точно установени (и зарегистрирани) *признаци* (Gruppenmerkmale): мъж, жена, жител на София, на Пловдивски окръг, на Варна, женен, неженен и пр. След

<sup>3</sup> Идеята за репрезентативен метод е пропагандирана отдавна от норвежкия статистик А. N. Kiaeg и от 1895 г. Международния статистически институт неведнъж се е занимавал с нея, но под влиянието предимно на германските членове на института последният до днес не е изнесъл окончателното си решение; впрочем това не пречело на практиката да прилага репрезентативния метод в твърде широки размери, и при това в болшинството от случаите напълно успешно. Киер е прокарал редица репрезентативни изследвания в Норвегия; Вестергорд в Дания; в Русия са известни репрезентативните изследвания, предприети от различни земства, например в Пенза (В. Громан); изследванията на селските стопанства в басейна на р. Сир-Дария, предприети от Министерство на земеделието в 1913-1917 г. във връзка с проектите за изкуствено напояване на Туркестан. Репрезентативният метод е бивал доста широко прилаган (С. А. Коген и пишещия тези редове отчасти). Най-последно, репрезентативният метод е бил доста широко приложен в девет грандиозни селскостопански преброявания през 1916 и 1917 г.

Теоретичният предговор към сборника на материалите от 1916 г., написан от С. С. Кон (по наст. прив. доц. в Прага), един от учениците на А. А. Чупров, се явява като най-добро от всички мен досега известни изложения за същността и техниката на репрезентативния метод. За жалост революцията попречи появяването на бял свят на този сборник и затова отделни отпечатъци от труда на С. С. Кон в настояще време представляват библиографическа рядкост. В Германия репрезентативният метод намира своето приложение в извършваната сега грандиозна общоимперска стопанска анкета.

Най-последно в англо-саксонските държави и особено в Англия, репрезентативният метод под името “Sampling” се използва много широко и се прилага например винаги в статистиката за работническите надници, за жилищата, за доходите и пр., и пр. За популяризацията му там твърде много е способствал споменатият по-горе проф. Боули, опрян от своя страна на огромния авторитет на проф. Пирсон и неговата школа.

Прилагане на репрезентативния метод при разработката на материалите...

това всички единици, влизайщи в дадена група, се събират и полученото число се признава за единствена характеристика на групата. Тъй например в групата на "български поданици женени мъже" ще влязат всички жители на България, колкото и да са различни по индивидуалните си качества един от друг, стига само да отговарят на по-горе споменатото определение за групата. Ако поискаме да съставим статистическа таблица, която да изобразява разпределението на българското население, например по поданство, пол и семейно положение – ние ще трябва относително ония жители, които се явяват като български поданици, да различаваме само следните 8 възможности:

1)	български поданик	мъж	женен
2)	"	"	неженен
3)	"	"	вдовец
4)	"	"	разведен
5)	"	жена	омъжена
6)	"	"	неомъжена
7)	"	"	вдовица
8)	"	"	разведена

и същите 8 възможности за всички не български поданици (заедно, в една група или отделно, по главни категории чужди поданици). Само тия 8, 16, 24, 32 или повече групи българско население ще съществуват за нас при съставянето на таблиците и всички 5.6 милиона индивидуално безкрайно своеобразни същества, населяващи днес България, ще бъдат разпределени от нас в само тия не много групи и *преброени* в тях, като че ли това са били просто 5.6 милиона еднакви топки, боядисани в 8, 16, 24 и т.н. различни цветове, или 5.6 милиона билети с написани върху тях 8, 16, 24 и т.н. различни номера. И ако от тия 5.6 милиона човешки същества са извади "случайно" "проба" в размер на 1/10 от общото им количество и се разпредели тази проба по своите 8, 16, 24 и т.н. рубрики – това ще бъде абсолютно същото, както ако от урна бяха изтеглени 560,000 топки и бяха преброени, колко от тях ще се окажат представители на всеки от дадените им видове цветове.

Ето защо могат със спокойна съвест и без колебание да бъдат приложени към Боули-Чупаровската постановка на репрезентативен метод всички ония теореми на теорията на вероятностите (начело със закона "за големите числа"), които се прилагат в различни комбинации на "случайното" изтегляне на топки или билети от една или няколко "урни". Тия теореми тук са приложими със същите модификации<sup>4</sup> и същите *ограничения*.

<sup>4</sup> Необходимо е да се различава например случай на теглене на билети с връщане на всеки изтеглен билет обратно в урната от случай без такова връщане. Схемата на последния е по-износна за нас, тъй като за същото количество тегления тя дава по-малка възможни пределни на грешката.

Тъкмо над тия ограничения, установени от теорията на вероятностите, които обуславят възможността за прилагането на репрезентативния метод, ние ще трябва да се спрем. Най-главните от тях са следните:

1) Прилагането на репрезентативния метод на дава непосредствено абсолютни числа, които характеризират цялата изследвана маса, но само относителни или средни числа. С негова помощ не е възможно например непосредствено да се определи общото население на България, но може да се намери % на мъжете от цялото население. Не може да се определи общата площ на културите в България, но може да се намери % на всяка една култура от общата площ, или пък процентното увеличаване на площта на културите сравнително с предишните “общини” преброявания за тази площ.

2) С помощта на репрезентативния метод може да се определя или *груповото разпределение* на масата по известни признаци (например разпределението на стопанствата според размерите на владенията: % стопанства с от 0 до 10 декара, с 10 до 20 декара и т.н.), или пък *средните величини*, характеризиращи масата (например средният размер на земевладенията изобщо, средният размер на владенията от 0 до 10 декара и т.н.)

За всеки един от тези 2 случая съществува особена формула, която определя възможните граници на грешката. Така например, ако общият брой единици в цялата маса е  $\sigma$ , броят на единиците, взети за репрезентативното изследване е  $s$ , а броят на единиците от интересувания ни вид, взет за репрезентативно изследване, е равен на  $m$  (например  $m$  – стопанства от 0 до 10 декара между  $s$  стопанства, взети за репрезентативно наблюдение от общо число  $\sigma$  стопанства в България), тогава най-малко в 995 случая от 1000 грешката, която ще направим при определяне на това каква част от масата представлява интересувания ни вид (в дадения случай каква част от всичките стопанства имат размер от 0 до 10 дека), т.е. относителната грешка на величината  $\frac{m}{s}$  няма да надмине

границите:

$$\pm 2 \sqrt{\frac{2 \frac{m}{s} \left(1 - \frac{m}{s}\right)}{s \cdot \frac{\sigma}{\sigma - s}}},$$

а в останалите 5 случая от 1000 тя ще ги надмине малко.

Формулата за грешката при определяне на *средните* чрез репрезентативния метод е вече по-сложна, понеже в нея вече влиза величина  $\mu$  “стандартно отклонение”, или “средна квадратична грешка” (което е едно и също) на репрезентативната маса и за нейното определяне трябва да се направят известни пресмятания над конкретния материал на самата репрезентативна маса; когато в първия случай грешката зависи само от



числеността на репрезентативната маса и от разпространеността на интересуващия ни вид (от величината  $\frac{m}{s}$ ).

3) Когато е възможно да употребим репрезентативно изследване, не можем предварително, само въз основа на априорните съображения, да решим каква именно част от цялата изследвана маса трябва да се отдели като репрезентативна такава: дали една десета, една двадесета, една петдесета и т.н. Както при анализа на някоя стока чрез “пробите” броят на тези проби зависи от степента на еднородността на партидата (колкото по-еднородна е партидата, толкова с по-малко проби можем да се ограничим), тъй също и при репрезентативното определяне на *средните величини* размерът на “извадката” пряко зависи от устойчивостта (или изменчивостта) на изследваната маса, а за тази устойчивост ние съдим по отношението между споменатото по-горе “стандартно отклонение” и средното аритметическо на “извадката”. Затова, преди да се установи окончателният план на репрезентативната обработка на събрания материал от общото наблюдение, или на едно ново чисто репрезентативно изследване, трябва да се предприеме, макар и не голямо, *пробно проучване* на наличния материал, например неговата разработка по няколко населени места или даже за цяла околия. При това, подобно проучване трябва да докосне всички преброявани от нас основни признаци, а не само няколко от тях. След като намерим нужните за нас конкретни средни величини, които приблизително характеризират изследваната статистическа маса, и като знаем предела на точността, който трябва да се постигне (например точност до 1%, до 5%, до 10% и т.н.), ще е вече лесно да се определи размерът на “извадката”, която ще бъде достатъчна, за да получим исканата точност.

Гореизложеното се отнася до определяне размера на “извадката” при намирането на *средните*. Колкото се касае до размера на “извадката” при определяне на *груповото разпределение на масата*, то за съответните пресмятания достатъчно е да се знаят само величините, които влизат в горната формула (т.е.  $m$ ,  $s$ ,  $\sigma$  и допустимата за величината  $\frac{m}{s}$  относителна грешка, която ние ще означим тук със символа  $\pm\delta$ ). Тогава, за да имаме в 995 случая от 1000 грешка, по-малка от  $\pm\delta$  (а в останалите 5 случая тя да надмине само малко тази величина), ний трябва да “извадим” от общата маса една част, големината на която се определя по формулата:

$$\frac{s}{\sigma} = \frac{1}{\sigma \cdot \frac{m}{s} \cdot \delta^2} \cdot \frac{1}{1 + \frac{s}{8 \left(1 - \frac{m}{s}\right)}}$$

Нека например  $\sigma$  е общият брой на селските стопанства в Царството и е равно на 600,000; нека  $\frac{m}{s}$  е разпространеността на стопанствата от 5 до 10

декара между всичките стопанства и се равнява на 1/10 или 10% и нека най-после,  $\pm\delta$ , допустимият предел на грешката при това определение, е също  $\pm 1/10$  или  $\pm 10\%$  (с други думи, нека практически за нас е безразлично дали процентът на стопанствата с 5 до 10 дка е точно на 10% или е някъде между 9 и 11%). Като поставим тези числа в нашата формула, ние в края на краищата ще получим:

$\frac{S}{\sigma} = 0.012$ , или 1.2% от пълната маса, т.е. от 600,000 стопанства да се изберат

“случайно“ само 7,200. Ако ли поискаме сега да определим същото отношение със същата точност не за цяла България, а само за определен окръг и ако общият брой на стопанствата в окръга е 40,000, тогава нашата формула ще ни даде

$\frac{S}{\sigma} = 0.153$ , т.е. 15.3%. В “извадката” ще трябва да се отделят 15.3% от всички

случаи, т.е. 6,120 стопанства от 40,000. Ако преминаваме към околия с 7,500 стопанства, то при същите предположения ще трябва да извадим най-малко 49% от всичките стопанства. Ако, най-сетне, ние имаме работа с още по-малка територия, обемаща например само 1,000 стопанства, то същата формула ще ни доведе до заключение, че ще трябва “да извадим” 878 стопанства от 1,000, в който случай, разбира се, прилагането на репрезентативния метод си е изгубило

смисъла. Положението ще бъде по-добро, ако  $\frac{m}{s}$  е 50%, а не 10%. Тогава би

било достатъчно да се вземат 444 стопанства от 1,000, което представлява вече една значителна икономия. По този начин нашите пресмятания ни позволяват да установим още едно съществено ограничение.

4) Репрезентативният метод е за предпочитане само когато броят на “извадените” единици е достатъчно голям. Колкото по-голяма е териториалната единица, за която се прави сводката, толкова, при други равни условия, по-голяма е икономията от прилагането на метода. Ако сводката се прави по твърде дребни териториални деления, прилагането на репрезентативния метод може да загуби всякакъв смисъл.<sup>5</sup>

И най-сетне, 5) за успеха на репрезентативното изследване необходимо е грижливо да се избегне фалшификацията на “пробите”, т.е.

<sup>5</sup> За да имаме възможност да прилагаме репрезентативния метод към по-дребните териториални деления, необходимо е или да увеличим наблюдаемата група (вместо групи, обгръщащи 1/10 от всичките случаи, да образуваме групи от не по-малко от 1/5 или 1/3 от всичките случаи), или пък да се съгласим на по-големи грешки. Наопаки, колкото повече единици на наблюдението се намират в района на репрезентативното изследване, толкова разделението на репрезентативната маса може да бъде по-детайлно. Нещо подобно се наблюдава впрочем и в комбинационните таблици.

Прилагане на репрезентативния метод при разработката на материалите...

несполучливият тенденциозен избор на единиците за репрезентативната маса. Както е показал руският опит, агентите, които работят на мястото при така наречения "експедиционен способ", са склонни да преброяват предимно онези селяни, които им се виждат по-интелигентни и от които може да се очаква следователно по-бърз и по-разбран отговор. Няма нужда да се казва, че такива лица са често по-заможни, вследствие на което целият събран материал може да получи едностранчива тенденция. Същото значение ще има отделянето в "извадката" предимно на формуляри, написани по-четливо и по-добре попълнени.

\*

Конкретно, предложението ми се свежда към следното:

"Да се образува една сравнително не многобройна подготвителна комисия от представители на науката и на статистическата техника, на която да се възложи да провери (ако е желателно) твърденията на настоящата записка, и в случай на съгласие с нея да се състави към идната сесия на В.С.С. *план* за репрезентативна разработка на резултатите от преброяването от 1926 г. (главно на карта "Ж"), (или на някоя друга закъсняла разработка на Г.Д.С.). Този план трябва да бъде нареден така, че да не се забавя и не пречи на "общата" разработка на целия материал, ако такава разработка все пак ще трябва да се извърши. Комисията ще трябва да получи за ръководство формите на проектираната таблична разработка на материала. На нея ще трябва да се съобщи още най-дребното териториално деление, което трябва да влезе в таблицата (окръг, околия и т.н.), а също и пределът на допустимата грешка при прилагането на репрезентативния метод.

По-нататък, на комисията трябва да се даде правото, да се запознае със състоянието на подлежащите на разработка попълнени формуляри и да се възползва от резултатите на завършената вече досега разработка. В случай на нужда Г.Д.С. ще трябва да извърши една пробна разработка на известна малка част от материала според указанията на комисията, но във всеки случай не повече от една околия."

Рискът от приемането на предложението ми е голям, понеже работата на комисията няма да изисква никакви значителни средства, а като резултат на нейната работа може да се получи едно голямо намаление на разходите за извършването на отделните статистически работи в Г.Д.С. и главно, може би, макар частично, Г.Д.С. ще сполучи да се отърве от основното бедствие на официалната статистика – от значителното закъснение на повечето от нейните публикации!

Може би, в случай на успех на предложението ми В.С.С. ще се съгласи да включи репрезентативното стопанско изследване в плана на следващото стопанско преброяване.

гр. Варна, 1 юни 1928 г.