

МИКРОДАНИИ. МЕТАДАНИИ

Боян Захариев*



Още отпреди присъединяването ни към ЕС нашето общество беше започнало да развива активен интерес към данните за други страни - основно европейски страни, с които бихме искали да се сравняваме. С превръщането на България в пълноценен участник в обмена на данни в рамките на Евростат и в богатия календар от статистически изследвания статистическите данни постепенно започнаха да предизвикват все по-голям интерес сред широк кръг потребители. Това е свързано и с „откриването“ от медиите на данни, които и преди са били достъпни, но са оставали незабелязани и некомментиранни. Същевременно разбирането на дефинициите, които стоят зад публикуваните данни, става все по-голямо предизвикателство. Много от дефинициите в социално-икономическата статистика, които се използват от Евростат и НСИ, не са тривиални и правилното им разбиране изисква немалко усилия. С нарастването на обема на събираните и предоставяните за ползване данни от официалните статистически служби, нараства и обемът на метаданните. Наред с това протича и процес на усложняване на дефинициите. Самото събиране и организиране на метаданните се превръща в сложна задача, която е предмет на стандартизиране¹. Самите метаданни образуват сложни логически структури и понякога разбирането на една дефиниция не може да стане без разбиране на други части от структурата. Организацията на метаданните ражда нови - още по-сложни и абстрактни структури - йерархии, мрежи, (логически) карти².

Това прави разбирането (истинското разбиране) на статистическата информация трудно не само за редовия потребител, но и за журналисти и дори за анализатори, чиято роля е да интерпретират и обясняват. Интересът на редовия потребител на медийна статистическа информация е да разбере къде се намира спрямо репери, които му помагат да се ориентира в света спрямо някакви референтни групи. Перспективата на този вид ползватели към света зависи от мястото, на което се намират - в обществото и на картата.

Националните медии често отразяват и коментират статистически данни, които сравняват България с други страни. Това е отговор на интересите на жителите на големите градове, които са и основен потребител на медиите и искат да знаят как живеят спрямо останалите европейци. Жителите на по-малките населени места също се интересуват от това, още повече че в днешно време много от тях имат близки в чужбина, на които разчитат за допълване на доходите си. Регионалните медии нерядко се фокусират върху сравнения

* Д-р по социология, Институт *Отворено общество* - София; e-mail: bzhariev@osi.bg.

¹ Scanu, M. S. M. and G. Brancato. 'Statistical Metadata: a Unified Approach to Management and Dissemination', *Journal of Official Statistics* (31:2), 2015, pp. 325 - 347.

² Hay, D. *Data Model Patterns. A Metadata Map*, Elsevier, 2006.

между тяхната административна област и останалите, особено с тези, които представляват за тях нещо като национален репер за качеството на живота. В България в по-малките градове и в селските райони медийната аудитория е любопитна да сравни живота си и с живота в по-големите и по-развити градски центрове и най-вече с живота в столицата. Нуждата от подобна информация се задоволява предимно от публикации на националната статистика. Тази информация за разлика от информацията на страницата на Евростат е директно достъпна на български език. По-малките медии рядко могат да си позволят да направят превод от английски или друг език на специализирани и трудни за разбиране наименования и цели текстове.

Затова не е учудващо, че в много случаи медиите предоставят информацията, без да я модифицират, например преписвайки от страницата на НСИ или от публикувани от НСИ анализи. Единственото оригинално нещо в подобни материали е заглавието. В останалите случаи публикациите почти винаги съдържат забавни недоразумения. Когато информацията е просто преписана от анализ на НСИ или прессъобщение, тя естествено не съдържа обичайните грешки и недоразумения, но се поражда въпросът до каква степен читателите я разбират. До каква степен журналистите в самата медия разбират какво се казва в публикуваната информация? В крайна сметка ролята на медиите освен да отразяват е и да анализират и коментират информацията, съобразявайки се с интересите и възможностите на своята аудитория.

Мястото на официалната статистика в потока от информация

Официалната статистика никога не е била единственият източник на количествена информация за икономиката и обществото. Но днес повече от всякога, сме заобиколени от необозрими потоци информация, огромната част от която не е съставена от числа, а от думи и образи. Никой не може да претендира, че може да предложи това, което някои теоретици наричат 360-градусово тримерно представяне на реалността³. Светът е станал прекалено сложен и става все по-сложен. Обзорно и цялостно представяне е невъзможно. В този контекст би било наивно да се твърди, че официалната статистика съдържа някакъв, дори частично пълен, обзор на социалната и икономическата действителност. Следователно от практически неограничените аспекти на реалността - било то в природата, икономиката или обществото - ние избираме да наблюдаваме само малка част. Но официалната статистика съдържа нещо по-важно - информация, която макар да е нищожна част от информационния поток, е проверена и надеждна. А това никак не е за пренебрегване. Обогащването на наличните статистически данни протича паралелно с масовото разпространение на повърхностна информация, която понякога съдържа и числа. Числата вдъхват почти рефлекторно доверие, което лесно може да бъде измамно, защото носят претенцията, че нещо е било измерено, а не само интуитивно оценено. При тези условия ползването на информацията неизбежно е фрагментарно. Това се отнася дори до ползването на масивите от числа и категории, които наричаме данни. Различни аудитории получават достъп до различни фрагменти от информацията.

Какво съдържа светът на официалната статистика и колко е голям? Как е подреден и какъв е неговият дял сред складирана и циркулираща в света информация. В компютърните науки се използва понятието „онтология“, което действително е заимствано от философията. В компютърните науки онтология означава често „набор от символи или обекти в определена среда“. Като дял на философията онтологията се занимава с всичко, което съществува в най-широк смисъл. Именно в този смисъл понятието „онтология“ е много подходящо за описанието на света на метаданните в официалната статистика. Действително

³ Stahl, R. and P. Staab. *Die Vermessung des Datenuniversums. Datenintegration mithilfe des Statistikstandards SDMX*, Springer Vieweg, 2017.

това, което е дефинирано, и това, което се измерва, е всичко, което съществува - всичко, за което има данни всичко, което е наблюдаемо. Отделният наблюдател или отделният ползвател потенциално може да се интересува от друг аспект на действителността и да отдели свой ресурс, за да събере данни за него. В днешно време огромни масиви от данни са частно притежание. Все пак официалната статистика представя една картина на цялата икономика и на цялото общество в мащаби, в които това не е достъпно от други източници.

Нека погледнем набързо какво съществува в онтологията на Евростат.

Евростат съдържа към 4 600 масиви с данни с общо над 1.2 милиарда единични статистически данни (стойности) - най-често числа⁴. Това вероятно се равнява на минимум 2.4 милиона мегабайта, или 2.4 хил. терабайта информация⁵. За сравнение, за ден в Туитър се изпращат около 1 милиард съобщения, които съдържат повече информация от цялата база данни на Евростат. В момента Евростат съдържа информация по девет статистически области, оформена в табличен вид. Зад тези таблици стоят големи изследвания, съдържащи микроданни за различни изследвани единици - сред тях хора, жилища, фирми, други организации.

Във Фейсбук всяка минута потребителите споделят няколко милиона единици съдържание. Активността в Туитър е приблизително от същия мащаб, но съобщенията са по-кратки - формално средният обем информация на едно съобщение е по-малък. За същото време в Гугъл се правят над 5 милиона запитвания. Това означава, че въпросите, зададени на Гугъл за 10 - 20 минути, измерени в битове информация, се равняват приблизително на целия масив от данни, акумулиран до момента в Евростат. Активността в Туитър произвежда такъв обем информация в рамките на малко повече от час.

Така че онтологията на официалната статистика заема само малка част от света на информацията. Разбира се, истинският мащаб на присъствие на официалната статистика в потока от информация е по-голям. Трябва да добавим копирането, препредаването и интерпретирането на тази информация в медиите, социалните мрежи и останалата част от глобалното информационно пространство. Нямам възможност да оценя дори приблизително мащабите на проникване на тази информация - например в социалните мрежи. Бих рискувал обаче да предположа, че такава оценка не би променила съществено извода за нейния реален дял.

Искушавам се да кажа - от световите на информацията, защото мисля, че глобалното информационно пространство е силно сегментирано не толкова поради езикови бариери, колкото поради това, че хората имат навика да навигират в морето от информация, избирайки маршрути, които най-много им харесват - с други думи, хората някак си „попадат“ най-често на информация, която потвърждава вече съществуващите им убеждения, а от целия масив от информация, на която попадат, разбират това, което могат да разберат, по начин, по който могат да го разберат. Краят на последното изречение не е тавтология и не е част от елитистка концепция за информацията и разбирането - просто светът е станал невъобразимо сложен и препълнен с необозрима информация.

Силата на микроданните - гориво за производство на нови данни

В данните, предоставяни от официалната статистика, може да има много повече възможности дори от всички, които се съдържат в многобройните променливи и техните възможни съчетания за произвеждане на нови променливи. В огромните масиви от данни,

⁴ <http://ec.europa.eu/eurostat/help/first-visit/database>

⁵ Информация за размера на някои от важните бази данни в Евростат се съдържа и в *LFS USER GUIDE VOLUME 9 (EUROSTAT)*.

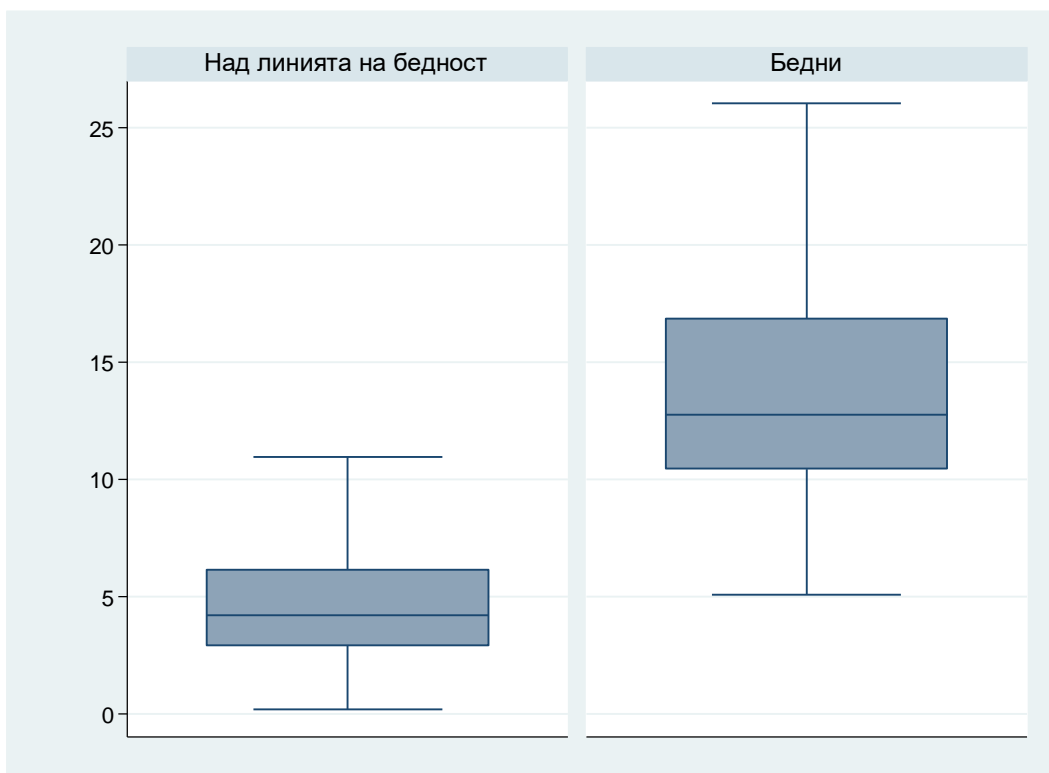
които се събират и публикуват от официалната статистика, се съдържа потенциално информация, която е с много порядъци по-голяма.

Дори ако вземем сами примерите, свързани с хората, и най-малките единици, в които те са организирани - домакинствата, наличните променливи са толкова много и покриват такъв широк спектър от теми, че не е по силите на никоя отделна организация да ги анализира. Затова предоставянето на микроданни на различни потребители за приложни анализи и научни изследвания има голям потенциал да информира и подрежда все по-хаотичния обществен дебат и да произвежда нови знания. Данните от всички големи изследвания се използват в продължение на години за цели и анализи, които надхвърлят целите, за които са били създадени. Дори когато данните изглеждат вече достатъчно стари, за да не представляват (на пръв поглед) интерес от гледна точка на непосредственото правене на политика, вземане на решения и създаване на позиции по определени обществени теми, те продължават да се използват за проследяване на дългосрочни обществени тенденции, за създаване на нови теории и за историята. В този си втори живот подобни данни понякога внезапно отново добиват политическа актуалност.

Към данните по-различни начини могат да се добавят нови възможности за анализ - например чрез свързването с други масиви от данни. Могат да се създават и съвсем нови данни. С това нямам предвид произвеждането на нови променливи чрез прости логически и математически операции върху съществуващите. Един от начините това да стане е чрез симулиране на данни. Домакинствата са много подходяща „подложка“ за всякакви симулации. В случая ще дам пример със симулация на потреблението на електроенергия, която позволи да се направи анализ на енергийната бедност.

Симулирането дава много богати възможности за използване на данните в неочаквани области. Симулирането е подходящо в сфери, в които се ползват нормативни описания или прогнози, базирани на известни процеси. Такъв е случаят с потреблението на електроенергия. Симулирането на потреблението на ниво домакинства като минимум изисква използването на подробен регистър от домакинства с информация за техния състав и възрастова структура. Допълнителни възможности в конкретния пример с потреблението на електроенергия дава информацията за размера на жилището, неговото качество и състояние, макар че в приведенния пример такава информация не е използвана, тъй като симулацията не включва използването на електроенергия за регулиране на температурата в жилището (в този случай размерът на жилището и неговото качество и състояние дава полезна допълнителна информация за симулацията), а само за готвене, съхранение на храната, осветление, пране и подгряване на вода за миене и къпане. Друг важен индикатор, който се среща в EU-SILC и има пряко отношение към енергийната бедност, е възможността да се покриват разходи за поддръжка на жилището. Наред с наема, разходите за покупка на жилище и покриване на ипотечен кредит тук попадат и текущите разходи за ток, вода и отопление. Отделно има и въпрос дали домакинството се ограничава при отоплението на жилището поради недостиг на средства - въпрос, който е формулиран по начин, който позволява да се направят и заключения за потребителското поведение на домакинството, а именно - дали то си налага ограничения, за да намалява разходите си.

Фиг. 1. Дял на разходите за електроенергия при хората, живеещи в бедни домакинства и в домакинства над линията на бедност



Фиг. 1 не може да бъде направена без данните както от EU-SILC, така и от допълнителната симулация.

На фиг. 1 е показано какъв дял от располагаемия си доход плащат ежемесечно за електроенергия хората, живеещи в домакинства над линията на бедност, и тези, чиито располагаем месечен доход на глава от домакинството е под линията на бедност. Вижда се, че почти всички, живеещи над линията на бедност отделят за електроенергия под 10% от располагаемия си месечен доход, докато при хората, живеещи в бедни домакинства, е точно обратното - почти всички отделят за електроенергия над 10% от располагаемия си месечен доход. Десетпроцентният критерий е един от основните при определяне на енергийната бедност, макар че съществува и разнообразие от различни други дефиниции, голяма част от които също може да бъде използвана върху данните от EU-SILC. В случая целта не е да навлезем в анализа в енергийната бедност, а да илюстрираме възможностите за анализ, които дават предоставените от НСИ микроданни. В този контекст интересното във фиг. 1 е, че в двете категории от ляво надясно виждаме класификацията на бедни и небедни, която е невъзможна без използването на цялата информация от сложния агрегат на доходите в EU-SILC и цялата непроста дефиниция на показателя „риск от бедност“. По ординатата виждаме относителния дял от дохода, отделян за електроенергия, за изчисляването на който е необходима информацията от симулирането на потреблението на електроенергия. Подобна симулация не възпроизвежда реалното потребление, а някакво нормативно определено (нормално) потребление. В случая с енергийната бедност това се оказва в някои отношения предимство, а не недостатък. При определянето на енергийната бедност по нормативен път за бедни се считат и домакинства, които при реалното си потребление биха се лишавали от важни домакински процеси, свързани с енергийното потребление, за да могат да се вместят в крайно ограничени си бюджет. Съответно домакинства, които се отнасят небрежно и неикономично към потреблението на електроенергия, не биха попаднали сред енергийно

бедните, въпреки че месечните им разходи за електроенергия биха надхвърлили някакъв нормативно дефиниран праг, например от 10%, както е в случая. Тук е важно да се отбележи едно съществено ограничение при ползването на извадкови данни за планиране на определена политика. То се състои именно в това, че става дума за извадка, колкото и голяма да е тя. Добре направените, достатъчно големи извадки дават статистически валидна информация за състоянието на обществото и икономиката, но те не ни казват как може да бъде реално запазена с информацията определена политика, която желае да използва критериите и дефинициите, използвани върху извадковите данни.

За конкретика нека погледнем нашия пример с енергийната бедност. На базата на информацията за домакинствата от EU-SILC и направената симулация откриваме, че около 530 хил. домакинства биха попаднали в категорията на енергийно бедните и биха имали нужда от някакво подпомагане. За да може обаче тази политика реално да бъде възможна, е необходимо в някакви изчерпателни официални регистри да се съдържа информация, която позволява дефиницията за енергийна бедност да бъде използвана, или ако такава информация към момента не се съдържа в никой регистър или комбинация от регистри, да е възможно да се проектира някакъв разумно скъп процес, с който определена държавна структура или друга организация би могла да събира и обработва съответната информация за **цялото население**. Една политика може да бъде замислена, оценена и дори остойностена на базата на извадкови данни, но изпълнението ѝ в общия случай не може да бъде запазено от такива данни. Формулирано по този начин това наблюдение звучи почти очевидно, но според мен на европейско и национално ниво няма да е трудно да се намерят примери, когато то не се отчита при планирането на някакви политики и поставянето на цели. С помощта на извадковите проучвания през някакъв период от време (често ежегодно, но понякога по-рядко и при всички случаи със закъснение във времето) се вижда какво се е случило, но самият механизъм на прилагане на политики остава черна кутия и невъзможно да се определи какво се случва текущо.

Метаданни, дефиниции и нови разбирания за обществото

Както видяхме „реалността“, създавана от статистиката, със сигурност е ограничена и твърде малка в сравнение с глобалната информационна реалност, но тя е сред най-подредените и съгласувани и сред най-малко фрагментирани реалности - поне докато не попадне в информационния поток и не бъде преработена от него. Тази статистическа онтология освен това има сериозни последствия за голямата политическа и икономическа реалност. Неслучайно индикаторите, за които се събират нужните данни и които след като преминават през предвидените статистически процедури и операции се публикуват, не са еднакво приемливи за всички държави членки. Политиците имат различни мотиви да виждат (или да искат да виждат) едни индикатори и да не виждат (да не искат да виждат) други. Първата причина е, че страните, в които управляват, могат да изглеждат зле през призмата на един или друг индикатор. Могат да нямат предимство в сравнение със страни, от чието население техните избиратели искат да живеят по-добре или са убедени, че живеят по-добре, и биха били много разочаровани, ако някой индикатор показва, че това не е така. В някои случаи определен индикатор наистина може да е по-подходящ да улови проблемите в едно общество отколкото в друго.

Участието в системата на Евростат въвежда нов тип „относителни“ индикатори, които съотнасят положението в България с нови референти, свързани с ЕС - сред тях граждани на държавите от ЕС, членове на различни домакинства и социални групи в отделните страни от ЕС и в ЕС като цяло.

Не е ясно, колко от хората в България, които живеят добре по българските стандарти, се чувстват удовлетворени от този факт или просто са сменили референтната си група и очакванията си, гледайки към жителите на други, по-развити държави в ЕС. Интересът на

национални медии към ситуацията в други страни от ЕС подсказва, че този интерес може да е споделен от голяма част от тяхната аудитория. В дилемата за използването на абсолютни или относителни индикатори за бедност и социално изключване присъства и възможността като референт в относителните индикатори да се използва ситуацията в други, по-богати или по-развити страни⁶. В подобни случаи е възможно хората, които не отговарят на определен стандарт, да се окажат (значително) мнозинство.

Ще цитирам предговора на една много влиятелна книга, написана преди разширяването на ЕС към Централна и Източна Европа, където значението на социалните индикатори, събирани от официалната статистика за всички държави членки, е описано по впечатляващ и вдъхновяващ начин. Предговорът е написан от белгийския социален министър, тъй като самата книга е създадена в контекста на Белгийското председателство на Съвета на ЕС:

„Утвърждаването от Европейския съюз на набор от общоприети и еднакво дефинирани социални индикатори ще представлява важна стъпка към постигането на социалните цели на Европа. Социалните индикатори ще осигурят рамка, в която държавите членки могат да развият социалната си политика и политиката за заетост по начин, който отразява националните условия и въплъщава целите за социална кохезия на ЕС.“⁷

По-нататък авторите казват, че тези индикатори ще са от значение и за националните правителства и за местните и регионалните общности, за социалните партньори, за неправителствените организации, за гражданското общество и за отделните европейски граждани и че тези индикатори ще изпратят сигнал към бъдещите държави членки (т.е. и към нас), че това е начинът да се измерват постиженията на Европейския социален модел - „един завършен модел, който е част от Европейската цивилизация“.

Може би днес, след новите разширявания на ЕС, глобалната икономическа криза и състоянието на процеса на икономическа и социална интеграция, тези констатации звучат малко преувеличено и нереалистично. Постепенното въвеждане на немонетарни индикатори за социално включване предшества с много десетилетия присъединяването на България към ЕС и се основава на по-ранни регионални инициативи за използване на общи индикатори, например в скандинавските страни⁸. Тези индикатори отразяват една сложна предистория на политически дебати и консенсус първо в отделни страни, после в групи от страни и в ЕС като цяло. В генезиса им са утаени разбирания за смисъла на социалното приобщаване и социалното изключване от няколко десетилетия и от различни места в Европа, но общо взето - от най-развитите в икономическо и социално отношение страни в ЕС, които най-често са пионери в подобни проучвания.

На този фон не е учудващо, че не всички индикатори са еднакво популярни или еднакво използвани в различните държави - членки на ЕС. Например в принос към стратегията „Европа 2020“ държавите са избрали да използват различни индикатори вместо риска от бедност или социално изключване. Много от страните са се спрели на конкретен компонент от този сложен индикатор. Други са избрали съвсем различни индикатори, които според тях по-добре отговарят на предизвикателствата пред социално включване. Както споменах, неявните мотиви за такъв избор може да са най-разнообразни. Дания например е избрала да използва индикатора „хора, живеещи в домакинства с нисък интензитет на трудовата активност“. Избор, който подсказва, че в Дания интензивната работа би трябвало

⁶ Atkinson, A. and E. Marlier. *Analyzing and Measuring Social Inclusion in Global Context*, United Nations, New York, 2010, p. 10.

⁷ Atkinson, A. and E. Marlier. *Analyzing and Measuring Social Inclusion in Global Context*, United Nations, New York, 2010.

⁸ Atkinson, T., B. Cantillon, E. Marlier and B. Nolan. *Social Indicators: The EU and Social Inclusion*, Oxford University Press, 2002.

да бъде от най-добрите средства за преодоляване на социалното изключване. Ирландия е избрала съвсем друг акцент, вземайки от индикатора за бедност и социално изключване именно другите два компонента - материалните лишения и риска от бедност, т.е. очевидно ирландското правителство е преценило, че интензивността на трудовото участие не е толкова решаваща или не е основен проблем в тяхната страна. Съответно Латвия е избрала друга комбинация от компонентите на сложния индикатор - риска от бедност и живота в домакинства с нисък интензитет на труда. Германия е предпочела изобщо да не ползва индикатора за бедност или социално изключване, а вместо това да намали броя на дългосрочно безработните, Швеция - да намали дела на жените и мъжете на възраст 20 - 64, които не участват на пазара на труда или са дългосрочно безработни, или са в дългосрочен отпуск по болест, Обединеното кралство използва собствени индикатори, които са свързани с детската бедност⁹.

В стратегията си до 2020 година България поставя акцент именно върху риска от бедност. По този индикатор страната ни е сред най-бедните в ЕС, като през различните години заема различни места в челото на тази класация наред с няколко други страни, сред които Румъния, Латвия и Гърция. Въпреки това дистанцията с основната група от страни в Европейския съюз и със средните показатели за ЕС не изглежда непреодолима. Тъй като рискът от бедност или социално изключване зависи повече от разпределението на доходите, отколкото от абсолютния им размер, подобряването на този индикатор е силно зависимо от социалните трансфери и данъчната политика, а в България те не са благоприятни по отношение на целите за намаляване на неравенствата (страната ни изпитва затруднения с постигането на поставените цели).

Първоначално при въвеждането на новия индикатор за бедност или социално включване в началото на второто десетилетие на 21-ви век дистанцията между България и Румъния и останалите държави от ЕС по този показател е много голяма, основно поради многокомпонентния индикатор за материални лишения. Може би това е била причината в началото този индикатор да се приема с резерви от политическия елит в двете страни. Впоследствие обаче се очертава тенденция, която в ретроспекция изглежда лесно предвидима (това може да е обичайната ретроспективна илюзия). В страни като България и Румъния, които наблягат на общия икономически растеж и слабо се интересуват от неравенството в обществото, се оказва по-лесно да бъдат подобрени фиксирани индикатори като тези за материални лишения, които реално измерват абсолютна бедност, макар и при много висок праг, калибриран спрямо представите за добър живот в богати страни от Западна Европа.

Бедността е една от най-дискутираните и най-влиятелните теми в социалната статистика. За „бедност“ има многобройни дефиниции, които са пряко свързани със съответните програми за намаляване на бедността и с цял набор от съпътстващи и подпомагащи политики. В този смисъл изборът на дефиниция на бедността има изключително важно значение.

Може да се каже, че българската дефиниция за бедност, използвана за политически цели през годините е еволюирала, както и в много други страни, от разбирането на бедността като недояждане към бедността като монетарен феномен. Използваните от Евростат дефиниции за бедност я разглеждат многостранно - като монетарен феномен, като ниска позиция в обществената йерархия или обществената подредба на доходите и възможностите, като отдалеченост от „центъра“ на обществото, като неравенство и като невъзможност да си позволиш неща, които се считат за норма на живота в дадено общество.

⁹ Примерите са от Darvas, Z. 'Why is it so hard to reach the EU's 'poverty' target?' *Policy contribution, Issue No. 1*, 2017.

Разбирането на обществото като структура, състояща се от център и периферия, е особено мощен инструмент на социалноориентираното политическо мислене. Ето един пример за подобна дефиниция с ясен европейски произход, заложен в един от най-важните стратегически документи на българското правителство - Националната стратегия за намаляване на бедността и насърчаване на социалното включване:

„Социалното изключване е процес, при който определени хора са изтласкани на ръба на обществото и са възпрепятствани от пълно участие в него поради своята бедност или поради липсата на базови компетенции и възможности за учене през целия живот, или в резултат от дискриминация. Това ги отдалечава от възможността за работа, доходи и образование, както и до социални и общностни мрежи и дейности. Те имат ограничен достъп до власт и до органите, вземащи решение, и така често се чувстват безвластни и неспособни да упражняват контрол върху решенията, които влияят върху техния ежедневен живот.“¹⁰

За измерване на относителната бедност се използва най-популярният и най-често цитиран индикатор, който днес носи наименованието „риск от бедност“. Дефинирането на пребиваването в тази група като вид риск създава интересно двусмислие. Ако допуснем, че не всички от попадащите под така дефинираната линия на бедност са бедни, а само че сред тази група има **риск** от бедност, трябва да можем да уточним каква част от тях действително стават бедни - за това ни трябва допълнителен критерий.

В този случай е съвсем уместен въпросът кои са рисковите фактори, които увеличават вероятността членовете на едно домакинство да се окажат в групата на бедните, т.е. какъв е рискът дадено домакинство да се окаже в риск от бедност. Може да забележим, че условната вероятност за попадане сред бедните, когато човек е безработен или работи малка част от времето, или когато надхвърли определена възраст и т.н. е по-голяма, т.е. хората, които отговарят на тези критерии или попадат в тези категории, са в по-голям риск да изпаднат в бедност. Тук понятието „риск“ действително означава риск. Всъщност от много публикации на Евростат и НСИ е ясно, че рискът от бедност е чисто и просто индикатор за бедност с малко по-особено наименование.

Накрая ще дам пример с дефиниция, свързана с регионалните различия в бедността. За България това е особено важно, тъй като разликите в развитието на различни части от територията на страната ни са важен проблем, еднакво припознат от обществото и политическия елит. НСИ редовно публикува информация за риска от бедност по географски области.

Анализът на НСИ сочи, че в България поне в първите години след присъединяването към ЕС, които съвпадат и с въвеждането на нови индикатори от Евростат, най-голям брой домакинства изпитват ограничения по отношение на почивката извън дома и възможностите да посрещнат неочаквани разходи (за неотложен ремонт на жилището или колата, подмяна на пералня или хладилник, внезапно заболяване и др.) със собствени средства, както и при отоплението на жилището. Не са малко и домакинствата, които не могат през ден да си позволят, ако желаят, консумацията на месо, пиле или риба или вегетарианския им еквивалент или имат проблеми с плащането навреме на разходите за жилището.

В изданията на НСИ се посочва, че при изчисляване на линията на бедност за всяка област сама за себе си е приложен същият метод както при линията на бедност на национално ниво - 60% от средния общ разполагаем нетен доход на домакинствата в областта¹¹. Този подход има важни последици за изводите по отношение на бедността. Линиите на бедност в отделните административни области на страната силно се различават една от друга, защото на областно ниво има значителни разлики в размера и разпределението

¹⁰ Национална стратегия за намаляване на бедността и насърчаване на социалното включване 2020.

¹¹ НСИ, *Индикатори за бедност и социално включване, основани на европейско изследване за доходите и условията на живот (EU-SILC) в контекста на отворения метод за координация*, 2011, с. 7.

на доходите. Медианата на доходите и 60% от нея варират силно, като през годините няколко от областите на страната, разположени около най-големите градове, имат линия на бедност над средната за страната, а останалите области имат линия на бедност под средната за страната. С най-високи доходи и най-висока линия на бедност е град София. В общия случай колкото по-висока линия на бедност използваме, толкова повече хора ще се окажат под нея, т.е. ще бъдат определени като бедни. Ако за всички области в страната се ползваше националната линия на бедност, регистрираните нива на бедността в повечето от тях (по-конкретно на 22 - 23 области, чиято линия на бедност се оказва под средната за страната) щеше да се окаже много по-висока. Бедността, измерена по този начин в столицата, наред с Варна, Пловдив и Бургас и две-три други области, щеше да изглежда по-ниска. Хората, живеещи в по-ниско урбанизирани и по-слабо икономически развити райони на страната, и местните политически лидери щяха да виждат по-големи разлики между тях и други части на страната, към които мигрира част от трудоспособното местно население. По принцип и двата подхода за изчисляване на линиите на бедност - ползване на национална за територията на цялата страна или ползване на различни регионални линии - са приемливи в зависимост от конкретната аналитична задача или политическия контекст. Европейската дефиниция за риск от бедност предполага използването на различни национални линии на бедност. Това очевидно е предопределено и от факта, че между държавите членки съществуват големи различия, а в социалната сфера, както и в много други, свързани с нея политики, степента на интеграция е много слаба. Що се отнася до бедността, въпреки общия пазар и мобилността, Европейският съюз продължава да бъде сбор от отделни общества и икономики, в които доходите, качеството на живота, условията на живот и възможностите за участие се различават значително. Няма обща данъчна политика, няма обща политика за социалните трансфери или за предоставянето на социални услуги. В рамките на отделна страна нещата не са чак толкова очевидни. Ясно е, че и тук важи правилото, че при изчисляването на относителната бедност (на риска от бедност) човек се счита за беден в рамките на обществото или общността, в която живее. Вероятно в по-бедните райони на България животът в много отношения е по-евтин. Това едва ли важи за много от важните стоки, без които едно домакинство не може, но би трябвало да важи за жилищата и наемите и за услугите, чиито цени зависят от обичайните доходи на съответното място. От друга страна, като част от една-единствена и при това не много голяма икономика е нормално хората да искат да се ориентират спрямо някаква обща норма. Стремещът към мобилност и пътуване са ключови за много икономически и социални процеси, а те предполагат всеки доход, придобит, където и да било в страната, да се разглежда еднакво (за да има обща картина). За да бъдем по-конкретни - линията на бедност в Монтана е много по-ниска от тази в област София, а линията на бедност на област София е по-ниска от тази в столицата. Между селата, разположени в област Монтана и близките до тях села от област София и между крайградските квартали на София и близките до тях села на област София няма никакви разлики нито в правилата за правене на бизнес, нито в трудовите отношения и социалното осигуряване, нито в пенсионната система и социалните трансфери. Много хора, които са бедни в гр. София според регионалната линия на бедност, ще могат да си позволят повече и да живеят по-добре от хора, живеещи в близките села, които не са бедни според линията на бедност на област София. Всъщност този аргумент дори не изисква районите да са съседни, само конструирах примера така, за да има пълна яснота за какво става въпрос.

При изчисляването на бедността в национален план за бедни навсякъде се считат тези, които са бедни според националния праг на бедност. Те не се припокриват, а само се застъпват с тези, които се оказват бедни според регионалните линии на бедност. За бедните според националната линия на бедност не се публикува регионална статистика, т.е. не се отбелязва в кои части на страната живеят, а те са концентрирани в много голяма степен в бедните части на страната. Освен това, когато разглеждаме съставния индикатор за бедност и социално изключване, веднага забелязваме, че другите му компоненти не подлежат на регионализация. Няма утвърден метод интензитетът на икономическата активност на

домакинствата да се изчислява спрямо някаква териториално определена норма (някаква обичайна за административната област интензивност на трудовата активност). Подобен метод не съществува и за материалните лишения - няма изискване жителите на столицата да имат повече средства за посрещане на неотложни нужди или повече дни почивка извън дома през годината. Критериите за материални лишения са еднакви за всички жители на страната.

Нека спомена все пак, че използването на регионални линии на бедност има предимството директно да насочва към един важен за икономиката теоретичен въпрос, който е от съществено значение за избора на една или друга икономическа политика. Това е въпросът дали икономически по-развитите райони успяват да намалят в по-голяма степен относителната бедност. Тук няма възможност да се впуснем в цялата дълбочина на този въпрос, само ще отбележа, че графиките, публикувани от НСИ, навеждат на мисълта, че такава явна връзка не съществува. И пак напомням, че в случая нямам за цел да дискутирам кои дефиниции са по-добри, а само да илюстрирам колко важен е този избор и как той може да насочва към определени публични възприятия на действителността, към определени изследователски въпроси и дори към определени важни решения и политики. Все пак отбелязвам, че ми се струва полезно НСИ да публикува и разпределението на бедните по административни области според националната линия на бедност, ако това е технически възможно, както и разпределението на хората, живеещи в бедност и социално изключване.

Краят на числата?

Този материал е посветен на годишнината от създаването на българската статистика. На рожден ден не е редно гостите да бъдат през цялото време сериозни. Затова за финал искам да поздравя НСИ с празника с искрено, може би малко себично пожелание и едно намигване.

Искреното ми пожелание като потребител на информация за изследователски цели е НСИ да остане все така отворен към споделяне на микроданни за научни и научно-приложни изследвания, дори да има наднационални (глобални) тенденции в обратната посока. Всъщност по-правилно е да кажа - въпреки че вече се наблюдават такива тенденции. Не бива развитието на социалните науки да бъде възпрепятствано от действията на непочтени хора, които злоупотребяват с лични данни. Все пак трябва да отбележа, че злоупотребите са свързани изцяло с данни, които не идват по линия на официалната статистика, с данните от огромните информационни океани, за които споменах, в сравнение, с които официалната статистика е една капка.

Накрая ми се ще да задам следния полусериозен въпрос **Какво ще стане със статистиката в епоха, когато числата могат да изчезнат?** Такъв процес се наблюдава отдавна във високите сфери на математиката и от известно време - в ниските сфери на ежедневната комуникация, която вече се слива с медийната. Съвременната математика, започвайки още от геометрията и топологията и стигайки до теорията на множествата и теорията на категориите, наред с абстрактната алгебра и като цяло много други новопоявяващи се раздели на висшата математика на практика не използват числа, дори не използват обекти, които са производни на числата¹². В ежедневието и в масовите медии, от друга страна, все повече доминират изображенията - или без всякакви числа (под формата на снимки, клипове, кратки съобщения и др.), или под формата на т.нар. инфографики, в които едно число е заобиколено от много други символи и изображения - да кажем примерно, че в тях битовите информация, отделени за числата, са под 0.01%, докато в една традиционна статистическа таблица почти всички битове информация са заделени за числата. Нямам

¹² Вж. например Spivak, D. *Category Theory for Scientists*, 2013.

точна представа каква част от битовите информация в социалните мрежи са свързани с числа. Предполагам, че тази част е относително малка и вероятно намаляваща (това признавам е открит въпрос). Предполагам, че там мисията на числата (били те относително малко или повече) е спомагателна. Освен това може да дискутираме какъв смисъл имат числата, изолирани от връзките си с други числа, които в потока от сегментирана информация не могат да бъдат нито възстановени, нито проверени.