

ОЦЕНКА НА РЕЗУЛТАТИ ОТ ИНТЕРНЕТ ИЗСЛЕДВАНИЯ

ВЕРА ВЕЛЕВА^x

При проектирането на съвременните емпирични социологически изследвания съвместно се използват знанията от социологическата теория, методите на статистиката и възможностите на съвременните информационни технологии.

Развитието на съвременните комуникации са привлекателни за организирането на различни социологически изследвания, включително и чрез непредставителни извадки. Техните предимства са големи поради минималните разходи на времеви, финансови и човешки ресурси в сравнение с традиционните изследвания. В тях могат да участват голям (практически неограничен) брой респонденти, без това да натоварва ресурсно изследването, а така също и изследователския екип.

В статията е представена идея за използване на разработените модели за калибриране на данни от извадки. Целта е подобряване качеството на информацията от извадки, за които не са приложени известните класически методи при тяхното формиране. Става дума за формиране на извадки, които се формират в резултат на доброволното участие на единици от определена генерална съвкупност („метод на отзовалите се“) при попълването на предложен в интернет инструментариум на изследване.

Приложението на калибрацията и резултатите от нея е илюстрирано с конкретен пример върху данни от проведено изследване по интернет, като е използван продукта g-CALIB, базиран в SPSS.

Крайната констатация е, че ако не може да се потвърди, че изследването след калибрацията има предствителен характер, то може да се каже, че след калибрацията обективността на данните в много по-голяма степен е гарантирана. Обработената с тази процедура информация носи всички добри качества на данни,

получени от извадки, за излъчването на които е приложен квотен подбор, съчетаван с определена степен на случайност в подбора на единиците за извадката.

***Ключови думи:** представителни и непредставителни извадки; претегляне на данни от извадки; постстратификация; калибрация на данни.*

Изучаването на съвременните обществените процеси, основано на постиженията в теорията и практиката на емпиричните изследвания, все повече изисква разработването и използването на нови научни подходи и методи за повишаване качеството на получаваната информация по отношение нейната обективност, достоверност, валидност и точност.

Изискванията за качество на информационния продукт от емпиричните социологически изследвания се налагат от важната им роля в обогатяването на знанията за обществото и неговото научно управление. Сложността и многообразието на взаимовръзките между елементите на обществената система и вероятностният характер в поведението на изследваните обекти на емпиричните социологически изследвания поставят в изключително трудна ситуация изследователите в стремежа им да създадат адекватни модели на изследваните процеси и тяхната прецизна реализация.

За преодоляването на трудностите по реализацията и получаването на данни с научна аргументация е необходимо при проектирането на емпиричните социологически изследвания да се прилага съчетаното използване на знанията от социологическата теория, методите на статистиката и възможностите на съвременните информационни технологии и комуникации.

Сред съвременните средства и методи за подобряването на качествените характеристики на данните от извадкови социологически изследвания са моделите за калибриране на данните. Те са един нов подход в обработката на данните от извадкови изследвания и научно обоснован метод за неутрализиране влиянието на различни смущаващи фактори върху изследователския процес. Става въпрос за фактори, които водят до разрушаване на представителността и точността на получаваните данни от извадки, използвани при провеждане на емпиричните социологически изследвания.

Социологически изследвания и извадков подход за тяхната реализация

При провеждането на емпирични социологически изследвания могат да се използват, както е добре известно, два основни подхода при определяне на обхвата на респондентите, от които се получават първичните данни:

Изчерпателен подход, чрез който в процеса на социологическото изследване се включват всички единици на изучаваната генерална съвкупност. Пример за такива изчерпателни изследвания са преброяванията на населението, както и изследвания върху малки организационни единици. За социологически изследвания върху малки организационни единици – предприятия, публични институции и пр. подобни изследвания са с най-голяма представителност и с най-високо качество на получените резултати. При тях не се получават стохастични грешки, които са характерни за извадкови изследвания.

Поради различни ограничения (ресурсни, времеви, достъпност и др.) и невъзможност за организиране на изчерпателни изследвания върху огромни масиви от населението или от цялото население в практиката на емпиричните социологически изследвания се прилагат предимно извадкови изследвания. Освен това, ако те са организирани и проведени в съответствие с принципите на социологическата теория е възможно да се получи дори по-точна и по-широка в съдържателно отношение информация за изучаваните съвкупности.

„Спецификата на обществените явления не е пречка за изучаването им чрез случайни извадки, тъй като повечето обществени явления не зависят от случайни фактори и са обект на съзнателна човешка дейност. Дори когато генералната съвкупност е била напълно обусловена от човешка намеса и нито един случай на съвкупността не е настъпил случайно не съществува пречка изучаването на общите свойства на съвкупността да стане непреднамерено, т. е. съзнателно избиране на случаите на извадката”. (Цонев,1970:74)

Приложението на този подход се „заключва във възможността да се изучи дадено явление или процес чрез включване в наблюдението само на част от единиците на интересуващата ни генерална съвкупност. Тези изследвания са научно обосновани само тогава, когато получената информация от включените в наблюдението единици, чрез формираната извадка, дава възможност получените от извадката характеристики да се отнасят (важат) за цялата генерална съвкупност. Това теоретично е допустимо и

възможно само когато формираната извадка е репрезентативна (представителна) на изучаваната генерална съвкупност, т.е. приложени са правилата на статистическия метод на изучаване”.(Калчев, 2005:70)

За да бъдат резултатите от изследвана извадкова съвкупност достоверни е необходимо не само да се определи обема на извадката, но и да се осигури случайност на избора на единиците в нея. Това условие е важно, тъй като от него в много голяма степен зависи точността на получените резултати.

За осигуряването на тези условия в практиката на емпиричните изследвания се използват различни организационни техники и методи за формиране на извадките. На практика не се прилага само един метод, а комбинация от тях. Това изисква познания на изследователя на различните схеми, технологии за формиране на извадките. Налага се тъй като в социалната действителност не всякога може да се приложи простия случаен подбор в чист вид. При това в определени случаи отстъплението от строгите правила на простия случаен (вероятностен) подбор дори подобрява резултатите, но това става само при определени правила. Става дума за изпълнението на процедури, които не само оптимизират разходите за провеждане на изследването (финансови и времеви), но и постигане на по-добри статистически резултати, или реализиране на т.нар. „дизайн ефект”.

В практиката на социалните изследвания освен представителни (репрезентативни) извадки се използват и непредставителни извадки. Известно за тях е, че получените изводи и направените заключения не могат да се генерализират за съвкупностите, от които те са излъчени. Такива са извадките при формирането на които е приложен типологичен или квотен подбор. Друга разновидност на този вид извадки са тези формиращи на „принципа на достъпността” или на „принципа на отзовалите се”.

Този вид непредставителни извадки имат своето място в провеждането на социологически изследвания независимо, че няма все още достатъчни научни доказателства за отнасяне на получените от тях резултати върху цялата изследвана съвкупност. Предстои да бъдат обсъждани техните възможности за приложението им, особено в случаите когато не може или не е целесъобразно да бъде проведено изчерпателно или представително извадково изследване.

При сегашните условия на все по-трудно събиране на данните от респондентите, вече са разработени и се прилагат нови математико-статистически методи (модели) за

редактиране и запълване на липсващи данни, импутация на данни, калибрация на данни от извадкови изследвания. Тези методи ще бъдат извиквани от потребностите на извадкови изследвания не само в случаите на директните, класически организационни начини (пряко интервю) за събиране на данните, но и поради все по-честото използване на новите технологии – телефонно интервю (САТІ), уеб-изследванията по интернет и др. електронни форми.

За необходимостта от претегляне на данни от извадкови изследвания

Приложението на статистическите извадкови методи в социалните изследвания най-често довежда до получаването на „непълнен масив“ от данни за разлика от тяхното приложение в естествените науки. Това предизвиква развитието през втората половина на миналия век на математико-статистически методи в посока към оптимизиране на ефектите от липсващи данни. Някои методи, свързани с осигуряването на ефективни оценки от непълнен масив с данни, са дублиране на данни, претегляне на данни, алгоритми за пряка оценка на загубените данни, импутация на данни чрез многомерно приписване на значения за липсващи данни, многомерно претегляне (калибриране - Calibration) на данните в непълните масиви.

Претеглянето на данните, получени от извадка е известна и често прилагана процедура за постигане на критериите за качество на информацията. Тя е особено необходима когато резултатите от проведеното емпирично изследване се разпростират (обобщават) за цялата генерална съвкупност. Претеглянето на данните се използва, за да се намали систематичното отклонение при оценките от извадково изследване (нестохастичната грешка).

В статистиката се предлагат различни известни решения за претегляне на данни от извадки, обединени под наименованието „методи за претегляне на извадкови данни“. Става дума за процедури, чрез които се цели да се постигне приближение или изравняване на извадково разпределение до разпределението на генералната съвкупност. По-конкретно с прилагането на традиционните методи за претегляне на данните от изследванията се преодоляват различията допуснати в подбора на единиците (ако има нарушение условия на случайния подбор), разликите в нивата на

отговаряне в отделните подгрупи и различията между оценките (статистическите характеристики) от извадката и параметрите на генералната съвкупност.

Съвременните нововъведения в развитието на математическата статистика и създаването на стандартни софтуерни продукти вече налагат нова генерация гъвкави процедури (модели), основаващи се на солидна теоретична база.

Една често срещана стандартна процедура за претегляне на данните в социалните изследвания (когато изследванията се провеждат в домакинствата) е изчисляване на постстратификационни тегла с отчитане на налични демографски и социални показатели.

Идеята за приложението на постстратификацията се основава на наличието на информация за основни демографски и социални признаци (пол, възраст, местоживееене, образование, икономическа активност, административни единици) за цялото население. Източник на такива данни е националната статистика или национални регистри на населението. Въз основа на такава информация по разновидностите на определен признак се формират определени страти сред населението.

Когато са налични няколко контролни променливи могат да се използват и по-сложни методи като се изследват линейни зависимости, например регресионен анализ. В този смисъл постстратификацията може да се разглежда като един „несложен праволинеен метод за калибрацията на данни”¹

Калибрация на данни от извадкови изследвания. Нов системен подход за претегляне на извадкови данни.

Съществен и открит стои въпросът, когато по различни причини част от първично избраните единици в извадката не се включват в изследването- не предоставят отговори. След публикуването на поредица от статии на специалисти в областта на теорията и практиката на статистическите извадкови изучавания в края на миналия век, идеята за калибрацията на данни от извадкови изследвания получава голям напредък. Според Девил и Серндал (Deville, J.C. and Särndal, 1992), калибрацията осигурява системен възглед за изследванията дори и в присъствието на различни грешки несвързани с проектирането на извадката.

В теоретичната основа на груповото и постстратификационното претегляне стоят инверсиите на вероятностите за включването на единиците в извадката. Калибрацията също използва тези идеи, но ги специфицира и доразвива.

Калибрацията (calibration) като метод за селектиране на тегла в извадкови изследвания е по-разпространена в статистическите служби на развитите страни. В последно време Евростат все повече препоръчва тази практика да бъде използвана от страните членки на ЕС при обработката на данни от извадкови изследвания.

Направеното проучване на различни литературни източници за използваните процедури за калибрацията на данните показва, че съществува много голям брой публикации по тази тематика. Една голяма част от тази литература се оказва твърде трудна за възприемане, тъй като в нея на много високо теоретично и аналитично (математическо) ниво се разглежда същността на калибрацията на данни. Поради това ние ще се спрем на нейната същност предимно от позицията на нейните функции и практическо приложение. Това ограничение се оправдава от по-голямата важност на практическите ефекти от прилагането на калибрацията и от факта, че всички процедури за калибриране на данни, както вече сеспомена, са налични и разпространени в софтуерните продукти – SAS, SPSS и др.

Най-общо в литературата „калибрацията” се определя като метод за повторно претегляне и се използва, когато има достъп до няколко променливи - качествени и количествени, на които трябва да се извършва съвместна корекция (Ardilly, 2006). Калибрирането се представя като набор от тегла за единиците в извадката, които отговарят за калибриране на известни крайни обобщаващи за съвкупността величини и такива, които оценителят довежда до съответствие (т.нар. проектиране в съответствие) със съвкупността, или, че дизайнът при меки условия, има асимптотично незначителен принос към средната квадратна грешка на оценителя. (Kott, 2006:133-142) С други думи, целта е постигане на безпристрастно проектиране и изпълнение на извадката. „Калибрацията чрез известни (*налични*) демографски показатели се използва за намаляване на изместеността на оценките, свързано с обхвата, частична компенсация на изместването, свързано с неотговарянето, което не може да се отстрани чрез претегляне и съгласуване на оценките с известни показатели. Тя позволява да се съкрати грешката на извадката в оценяването, като добре се корелира с налични контролни показатели.”²

Калибрирането на данните се разглежда като нов системен подход, който взима под внимание спомагателна информация за съвкупността и в стандартни условия осигурява включването на спомагателната информация в оценката. „Калибрацията е процес на целенасочено изменение на параметрите на извадковия план, като извадковите тегла, а също и на непараметричните модификации в състава на извадката, за намаляване на извадковите грешки и повишаване на точността и устойчивостта на извадковите оценки на статистическите показатели”(Степанов, 2009: 1) или калибрирането е специфичен начин да се отчете налична за изследваната съвкупност помощна информация и чрез различни стандартни настройки тази информация да бъде включена в получаването на крайните оценки. Трябва да се има предвид още, че същността на извършваното претегляне чрез процедурите за калибрация се състои в това, че теглата за всяка наблюдавана единица се изчисляват и присвояват по всички калибриращи признаци едновременно.

Приведените дефиниции за калибрацията относно нейната същност показват, че тя осигурява системно разглеждане на изследванията относно параметрите на извадката и изследваната съвкупност „даже в присъствието на различни грешки, несвързани с планираната извадка.”(Степанов, 2009: 6)

Въз основа на посочените определения за същността на калибрацията може да се заключи, че калибрацията представлява статистическа процедура, чрез която се преодоляват проблемите, свързани с липсващите данни за някоя променлива и проблемите, породени от неизпълнение или свръх набиране на единици за отделни групи от съвкупността. В резултат на балансиране на цялата извадкова съвкупност, в съответствие с нейни обобщаващи характеристики („съгласувани оценки”) се осигурява непремереността на подбора и се намалява/елиминира изместеността на получаваните оценки. Така в крайна сметка се постига по-голяма точност в стойностите на обобщаващите параметри на генералната съвкупност.

За процеса на калибрацията се използват, както наличните данни от дадено извадково изследване, така и допълнителна информация за характеристики на генералната съвкупност, като комбинирането им се извършва посредством калибрационни уравнения (функции). Използването на такава процедура предполага и е възможна при наличието на информация от допълнителни източници, която да характеризира в различни аспекти изследваната съвкупност – това са различните

административни регистри, статистически преброявания или др. изчерпателни записи за единиците в съвкупността.

Все още е много малък опитът в нашата страна в областта на калибрирането на данни при провеждане на извадкови изследвания и специално в областта на социологическите изследвания. Все пак първи опити в това отношение има при обработка на данните в някои представителни изследвания на Националния статистически институт (НСИ), които са включени и в европейската статистическа система (Евростат).

Калибрация на данни от интернет изследвания

Разгледаната накратко същност на процедурите за калибрацията на данни показва, че те са разработени и насочени за прилагане относно данни от представителни извадкови изследвания. Разбира се, както бе посочено, тези процедури оказват добри ефекти и върху грешки, които не са свързани само с планираната извадка - стохастичните грешки.

Това ни дава основание да представим калибриране на данни от извадки, за които не са приложени конвенционалните методи и критерии за формиране на извадка. Става дума за извадки от единици на една генерална съвкупност, в която всички имат достъп до предложения инструментариум, но участниците в изследването се включват по желание, т.е. за формиране на извадката действа известният „принцип на отзовалите се”.

Описаната ситуация може да се наблюдава в провеждането на изследвания като се използват интернет мрежите. Развитието на съвременните комуникации са привлекателни за организирането на различни социологически изследвания. Техните предимства са много големи, поради минималните разходи на времеви, финансови и човешки ресурси в сравнение с традиционните изследвания. В тях могат да участват голям (практически неограничен) брой респонденти, без това да натоварва ресурсно изследването, а така също и изследователския екип. Освен това тази технология за провеждане на изследвания преодолява и редица правни и организационни норми за достъп до единиците на изследване, тъй като участието им зависи единствено само от тяхното решение (воля) за участие.

Единствен основен недостатък на тези изследвания е, че получените резултати нямат теоретичната (методологична) база за тяхното разпространяване върху цялата генерална съвкупност, която е обект на изследването. Основанията са, че в тези случаи не всякога има информация за характеристиките, описващи разпределението на генералната съвкупност и не са спазени условията за равновероятностен подбор на единиците в извадката.

Понятието случайно се използва интуитивно, разбирано като подбор, свободен от систематични грешки в измерванията „За това ние обсъждаме проблема за случайността само във връзка с необходимостта да се избегнат различни видове волни и неволни систематични грешки. ... Следва да се отбележи, че при използване на случайни величини обикновените критерии за значимост стават пригодни в по-голяма или по-малка степен даже и в тези случаи, когато не е изпълнено предположението за нормално разпределение” (Bailey, 1959:170-172). И докато първото условие е преодолимо за практиката на социологическите изследвания при условието, че обработката на данните от тях се основава най-често на χ^2 (хи-квадрат) разпределението, то обсъждането на проблемите, свързани с обобщаването на данните от тези проучвания за изследваните генерални съвкупности е перспективно и очаква своята изследователска работа.

Няма съмнения, че интернет-технологиите в бъдеще ще се използват все повече за събиране на статистически данни и провеждане както на „представителни” чрез използването на регистри на единиците, така и „непредставителни” изследвания. За това може да се съди по международния интерес на провежданите работни срещи по тези проблеми. Последната (пета по ред) работна среща на тема „Интернет изследвания и методология на изследванията” се състоя в гр. Даеџеон в Република Корея на 11-12.09.2013 година. В работните доклади на тези срещи все по-голямо внимание се отделя и на методологическите проблеми, които трябва да се решават по отношение на качеството на данните от тези изследвания.

У нас като конкретен пример за използване на интернет мрежата за събиране на данни за населението може да се посочи проведеното през м. февруари 2011г. всеобщо преброяване на населението. По данни на НСИ близо 3 млн. души са се отзовали и са предоставили данни чрез предоставения електронен инструментариум за преброяване на населението.

Друг пример е проведеното през 2013 г. изследване сред населението, което има достъп до интернет по поръчка на Министерството на физическото възпитание и спорта (МФВС). Основна цел на изследването е да се проучи общественото мнение във връзка с актуализиране на нормативната уредба, планиране на политиките и програмите на държавните и местни власти и организациите по отношение на спорта в страната. Предмет на изследването е отношението на населението към спорта и спортната култура, както и към развитието на спорта и физическото възпитание в България.

Изследването обхваща българските граждани, използващи интернет. То е проведено през периода от 11 април до 2 май 2013 година.

Данните са събрани чрез онлайн анкета, като за посочения период са се отзовали и включили в изследването 9 945 души.

С цел да се осигури по-добра информативност и представителност на данните е приложена процедура за калибриране на резултатите. Чрез калибрацията е постигнато по-голямо съответствие на основните социално-демографски структури на извадката с тези на генералната съвкупност – българските интернет потребители. Калибрирането е извършено по пол, възраст и степен на завършено образование.

Тук ще приведем някои таблични резултати и изпълнените стъпки при калибрирането на данните от получената извадка.

Признаците за калибриране от посоченото изследване са пол, възраст, образование. Избрани са поради това, че най-често те се използват в социалните изследвания като стратифициращи при формирането на извадките и в голяма степен влияят върху мненията на анкетиранияте.

1. Изчисляване на абсолютни стойности и относителни дялове на тези признаци в генералната съвкупност. За получаването на показателите за абсолютното разпределение и структурата на интернет потребителите в страната по избраните признаци са използвани данни от проведеното изчерпателно преброяване на населението и жилищния фонд през 2011 година. В програмата на преброяването е включен въпрос, чрез който е събрана информация за достъпа на домакинствата и техните членове до интернет услуги.

Разпределение и структура на генералната съвкупност - ползващи интернет в България

Признаци		Структура - %	Брой
Вашият пол	Мъже	49,9	1702367
	Жени	50,1	1712045
	Общо	100,0	3414412
Вашето образование :	Висше	37,1	1100932
	Средно	48,3	1432117
	Основно и по-ниско	14,6	433158
	Общо	100,0	2966206
Вашата възраст е:	15-24 г.	24,7	692944
	25-64 г.	73,6	2065436
	65+ г.	1,7	47539
	Общо	100,0	2805919

2. Изчисляване на същите характеристики по данните от извадката

Структура и разпределение на анкетиранията съвкупност по пол, образование и възраст (преди калибрацията)

Признаци		Структура - %	Брой
Вашият пол е:	Мъже	59,8	5947
	Жени	40,2	3998
	Общо	100,0	9945
Вашето образование е:	Висше	67,0	6667
	Средно	27,4	2722
	Основно и по-ниско	5,6	556
	Общо	100,0	9945
Вашата възраст е:	15-24 г.	30,9	3074
	25-64 г.	68,5	6816
	65+ г.	0,6	55
	Общо	100,0	9945

3. Сравняване на данните от извадката и генералната съвкупност. От направените сравнения се установяват съществени различия между изчислените относителни дялове за отделните групи (в %) по разновидностите на избраните признаци.

4. Вземане на решение за калибрацията. Анализът на получените сравнителни резултати, получени от събраните извадкови данни и от изчерпателното преброяването на населението показват, че отделни групи от извадката ще влияят по-силно върху формираното обществено мнение по данни от извадката. За да се нормализират тези влияния се приема, че с прилагането на процедурите за калибрацията на данните от извадката ще се постигне по-добро представяне на отделните категории сред интернет потребителите.

5. Подготовка на софтуера за калибрацията. Конкретно в това изследване също е използван продукта g-CALIB, тъй като е базиран в SPSS и го прави достъпен. Независимо от универсалността на продукта за неговото приложение се правят съответни модификации в зависимост от задачите на калибрирането.

6. Подготовка на файла с данни от извадково изследване за калибриране и въвеждане на данните за генералната съвкупност в софтуера за калибрацията. Извършени са информационни процедури, с които се приготвят файловете с данни от извадката и преброяването на населението съгласно изискването за вход на софтуера.

7. Калибриране на данните. В тази стъпка се осъществява същинската част от процедурата по калибрацията. Чрез нея се изчисляват калибрационни тегла, които се отнасят до всяка единица. По този начин се уравновесяват теглата, а с това и тежестта на мненията на отделните социално-демографски групи, респективно данни, в общото мнение по изследваната проблематика.

8. Проверка за качество (ефективността) на изчислените калибрационни тегла. И при този вид изследвания се извършват съответните анализи на получените резултати от калибрацията на данните. Защото ако резултатите от калибрацията са задоволителни, отговарят на критериите на изследователя, следва прекратяване на процедурата. Ако те не отговарят на определените критерии, се извършват промени в някои параметри и процедурата се повтаря, докато се получат качествени калибрационни тегла.

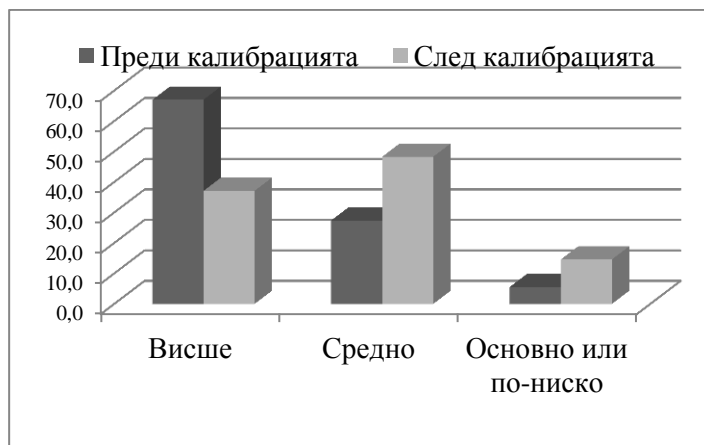
9. Прилагане на изчислените калибрационни тегла чрез претегляне на файла с данни от извадково изследване.

Структура (%) на анкетиранията съвкупност по пол, образование и възраст (след калибрацията)

Признаци		Структура - %	Брой
Вашият пол е:	Мъже	49,9	4958
	Жени	50,1	4987
	Общо	100,0	9945
Вашето образование	Висше	37,1	3691
	Средно	48,3	4802
	Основно и по-ниско	14,6	1452
	Общо	100,0	9945
Вашата възраст:	15-24 г.	24,7	2456
	25-64 г.	73,6	7321
	65+ г.	1,7	168
	Общо	100,0	9945

Получените калибрационни тегла се използват за претегляне на единиците от извадката. Така се постига и крайният ефект от калибрацията. Всички последващи обработки и анализи от извадковото изследване се извършват върху претеглени с калибрационните тегла единици, т.е. всички получавани в последствие аналитични таблици са претеглени и могат да бъдат подлагани на интерпретация без допълнително претегляне.

11. Установяване на разликите между изходните и калибрираните данни. За използваните калибрационни признаци се установява, че мъжете по-активно са участвали в интернет изследването - близо 60% от всички участници. По данни от преброяването на населението 2011 г. обаче, броят на мъжете и жените, в съвкупността на ползващите интернет е приблизително равен. От това следва, че мнението на мъжете в по-нататъшната обработка и анализ на данните ще определя и ще влияе в по-голяма степен при формирането на общественото мнение.



Структура (%) на изследваната съвкупност по степен на образование (преди и след калибрацията на данните)

При сравняването на данните от предходните две таблици, а също и от графиката могат да се направят аналогични заключения и по отношение на образователната структура на участващите в изследването лица. В интернет изследването много по-често са се отзовавали да участват хората с висше образование. Сред анкетиранията съвкупност лицата със средно и по-ниско образование са с по-малък относителен дял, отколкото е техният дял сред интернет потребителите в страната.

Подобни сравнения могат да се направят и по отношение на възрастовата структура на участниците в изследването - например, тук участниците в изследването на възраст 15-24 г. съставляват близо 31%, а по данни от преброяване на населението са близо 24,7%; за възрастовата група 25-64 г. участниците тези относителни дялове са съответно - 68,5% и 73,6%.

За да се преодолее влиянието на тези разлики в теглата на основните подсъвкупности (групи), поради неосигурена пропорционалност в представянето им в анкетиранията съвкупност, върху резултатите от изследването могат да се използват процедурите за калибрацията на данните. Тук в случая се допуска се, че данните от преброяването на населението дават достатъчно надеждна информация за размера и структурата на тази съвкупност (на интернет-потребителите) в страната по всички основни демографски и социални признаци.

След направената калибрация на данните достигането на критериите за приближаване на извадкови тегла към относителните дялове на съответните групи единици в генералната съвкупност (респ. вероятностите за включване в извадката)

може да се продължи с анализа на данните по включените в инструментариума на изследването променливи (индикатори).

За илюстрация как калибрацията (калибриращата информация) оказва влияние върху изследваните променливи, т.е. такива, които не са използвани при проектиране на дизайна на извадката (не са стратифициращи) ще представим разпределението (абсолютно и относително) на изследваните лица по техните отговори на въпрос относно заниманията им със спорт и честотата на тези занимания.

Разпределение и структура на изследваните лица относно заниманията им със спорт (преди калибрацията)

Занимание със спорт	Брой	Относителен дял-% (общо)	Относителен дял-% (отговорили)
1. Спортувам активно (вода системна тренировъчна и състезателна дейност)	2199	22,1	22,2
2. Спортувам редовно в свободното време	4018	40,4	40,5
3. Спортувам рядко	2382	24,0	24,0
4. Не спортувам	1315	13,2	13,3
Общо	9914	99,7	100,0
Не отговорили	31	0,3	
Общо анкетирани	9945	100,0	

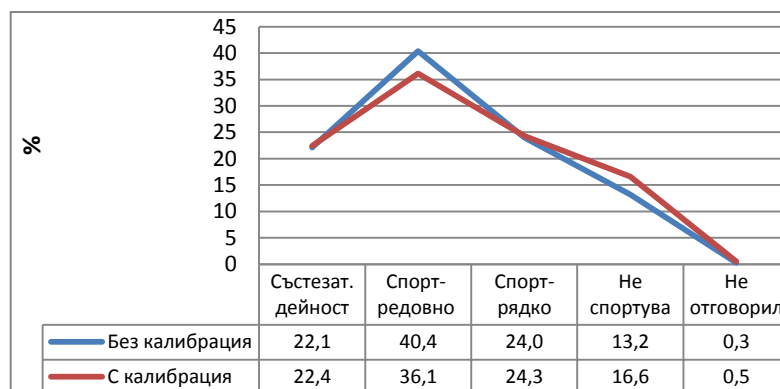
Сравняването на абсолютните и относителни показатели в двете таблици показва, че в резултат на извършената калибрация на данните по пол, възраст, образование (присвоените нови тегла на всяка участваща в изследването единица) са получени определени разлики в разпределението на анкетираните лица по разновидностите на изследвания въпрос.

В резултат на калибрацията на данните се установява условно нарастване на броя на лицата, които спортуват редовно в свободното си време - тяхното увеличение е с 424 души. Това нарастване на броя на единиците в тази подсъвкупност довежда до преразпределение на броя на случаите в подсъвкупността на тези, които участват в изследването и не се занимават активно със спорт. Техният брой намалява с 340. Намалява условно и броя на тези, които водят системна тренировъчна дейност - с 30 единици и тези, които спортуват рядко - с 36 единици.

Разпределение и структура на изследваните лица относно заниманията им със спорт (след калибрацията)

Занимание със спорт	Брой	Относителен дял-% (общо)	Относителен дял-% (отговорили)
1. Спортувам активно (вода системна тренировъчна и състезателна дейност)	2229	22,4	22,5
2. Спортувам редовно в свободното време	3594	36,1	36,3
3. Спортувам рядко	2418	24,3	24,4
4. Не спортувам	1655	16,6	16,7
Общо	9896	99,5	100,0
Не отговорили	49	0,5	
Общо анкетирани	9945	100,0	

Естествено абсолютното преразпределение води и до промени в структурата на изследваните лица по разновидностите на зададения въпрос за техните занимания със спорт. Най-голямо е различието между относителните дялове на лицата, които редовно се занимават със спорт, но не водят системна тренировъчна дейност. преди да бъдат изпълнени процедурите за калибрацията на данните техният относителен дял е 40,4%. След тяхното изпълнение той намалява до 36,1% или с над 4%.



Структура (в %) на изследваните лица по занимание със спорт преди и след калибрацията

Сред останалите групи изследвани лица с близо 2,5%, респ. теглата, се е увеличил дялът на тези, които не се занимават със спортна дейност.

От този кратък пример може да се направи заключението, че в изследването, което е реализирано на принципа на отзовалите се (на желаещите да участват), в първоначалния вариант на резултатите от формираната се извадка са надценени (почесто са участвали) лицата, които се занимават редовно със спорт и са подценени тези, които въобще не спортуват.

Въз основа на проведеното изследване са направени съответните заключения за развитието на професионалния и любителски спорт, мястото на физкултурата и спорта в съвременното общество, направленията в развитието на професионалния спорт, състоянието на спортната база и други.

От представения подход за получаване на оценки от интернет изследването възниква един изключително важен методологичен и практически проблем. Става въпрос доколко получените резултати от близо 10 хил. респонденти отговарят на условията да представят общественото мнение по тези проблеми. В момента няма теоретична и емпирична основа, която да даде ясен и обоснован отговор. Все пак това е едно от първите по-значими целенасочени социологически изследвания по интернет у нас, реализирано чрез непредставителна извадка и е направена калибрация на данните. Крайната констатация е, че ако не може да се потвърди, че изследването след калибрацията има представителен характер, то със сигурно след калибрацията обективността на данните в много голяма степен е гарантирана.

Може да се каже, че получената по тази технология информация носи всички добри качества на данни, получени от извадки, за излъчването на които е приложен квотен подбор, съчетаван с определена степен на случайност в подбора на единиците за извадката. В този ред на мисли може да се постави и въпросът доколко са представителни извадки, в които има голям брой неотговорили, или се прави подмяна на отказалите единици да участват в изследването.

БЕЛЕЖКИ

1. виж Руководство по подготовке статистических данных об использовании времени для оценки оплачиваемого и неоплачиваемого труда, Организация Объединенных Наций, Статистический отдел, Нью-Йорк, 2007 стр 160

2. пак там- стр.160

ИЗПОЛЗВАНА ЛИТЕРАТУРА

- Калчев, Й. (2005). Социологически траектории, София: УИ „Св. Кл. Охридски”
- Степанов, С.(2009). Калибровка выборки, Москва
- Цонев, В. (1970). Основи на репрезентативното изучаване, София
- Ardilly, P. (2006). Les technique de sondage, Raris
- Bailey, N. (1959). Statistical Methods in Biology, EUP
- Deville, J.C. and *Särndal*, C.E., (1992). Calibration estimators in survey sampling. Journal of the American Statistical Association.
- Ekholm, A. and Laaksonen, S. (1991). Weighting via response modeling in the Finnish Household Budget Survey. Journal of Official Statistics
- Kott, P (2006). Using calibration weighting to adjust for nonresponse and errors, Survey Methodology
- Lundstrom, S. and *Särndal*, C.E., (1999). Calibration as a standard method for treatment of nonresponse. Journal of Official Statistics

^x Вера Велева е PhD – главен асистент в катедра Социология към
Философски факултет на Югозападен университет „Неофит Рилски“