

МАТЕМАТИКАТА С ПОГЛЕД КЪМ СТАТИСТИКАТА

*Богдан Богданов**



В архивите на сп. „Статистика“ се намира едно забележително есе на известния наш математик дмн. проф. Николай Янев¹. Това есе е написано по повод Международната година на статистиката през 2013 година. Считам, че тази година по повод 141 години от създаването на българската държавна статистика полезно ще бъде отново да публикуваме това забележително есе. Причините и обясненията за това са накратко следните:

1. Да си припомним **съществуващата взаимозависимост** между математиката и статистиката, присъстващи в живота на стотици поколения от древността до наши дни.
2. Отново да акцентираме върху **фундамента на научни дисциплини**, важни за успешното развитие на всяка държава.
3. Да си припомним, че **успешното управление се основава на знанието** в основата, на което по принцип, са точните науки. В този смисъл математиката и статистиката са част от тях.
4. Да **проектираме бъдещето** като щрих на хоризонт, до който се достига с усвояване на инструментариума на точните науки.

* Д-р, заместник-председател на Националния статистически институт; e-mail: BBogdanov@nsi.bg.

¹ Дмн, професор, Институт по информатика и математика към Българската академия на науките; e-mail: yanev@math.bas.bg.

5. Да отворим широко **прозорците на познанието** за идващите след нас.

Нямам съмнение, че младото поколение от представители на науката ги очакват предизвикателствата на резултатите от бурното развитие на информационните технологии. Тези от тях, които поемат по трънливия път на изследователи на социалните явления и процеси, ги очакват вълнуващи моменти. Тези моменти ще бъдат свързани с усвояването на нови знания, както и със стремежа им да оставят своя следа в света на науката, където информационните иновации ще бъдат неотменима част от живота на обществото. Най-добрите ще станат и бъдат уникални в един нов дигитален свят, където мечтите се сбъдват и стават реалност. Те ще бъдат и съществена част от новите лидери, които ще построят един по-добър и справедлив свят.

Богдан Богданов

МЕЖДУНАРОДНА ГОДИНА НА СТАТИСТИКАТА - 2013²

Николай Янев

Обявяването на 2013 година от ООН за Международна година на статистиката (МГС) ни дава повод отново да се върнем към обсъждане на въпросите, отнасящи се до същността на Статистиката, нейния предмет, основни цели и задачи и обхвата на нейната приложимост. Това обяснява желанието на автора да стъпи на една по-широка платформа, като се постарее с известна доза атрактивност да привлече вниманието към проблемите и значението на Статистиката. Така че на всичко това би могло да се даде и подзаглавие „МГС през погледа на един математик“. Надявам се, че направените разсъждения ще бъдат интересни и полезни и за читателите на списание „Статистика“.

Във връзка с това да припомним, че 2000 година бе обявена за Международна година на Математиката. Това, че 21 век и въобще третото хилядолетие бяха посветени на Математиката, е не само признание за „царицата на науките“, но и изразява надеждата на човечеството за едно по-добро и справедливо бъдеще, основано на науката, в която Математиката е най-светлият символ на истинност, честност, справедливост и универсалност.

Последваха години, посветени и на други клонове на науката, дали своя принос в еволюцията на човешката цивилизация, за да стигнем до датата 20.10.2010 г., когато тържествено бе отбелязан Световният ден на Статистиката. Двадесет и първи век беше наречен век на информационното общество, а Статистиката играе много важна роля в него. Видя се, че един ден е малко, за да се разгледа всестранно тази роля и да се отдаде нужното внимание на Статистиката като част от Математиката и на нейните многобройни приложения, станали основа на много различни научни направления и всекидневни важни изводи, свързани с управленски прогнози и решения.

² Статията е подготвена въз основа на пленарен доклад, изнесен пред 42-рата пролетна конференция на Съюза на математиците в България. Бел. на редактора: В текста съзнателно е допуснато изписването с главни букви на наименованията на научните дисциплини.

Така естествено се стигна до определянето на 2013 година като Международна година на Статистиката (International Year of Statistics). Може би защото според Статистиката 13 не е фатално число?! Във всеки случай, ще имаме на разположение цяла година, през която се организират многобройни и разнообразни прояви за осмисляне на значението и ролята на Статистиката, на нейните постижения и приложения, за популяризиране на интереса към нейното по-задълбочено изучаване и по-всеобхватно използване, за нейното по-нататъшно развитие, свързано с развиващото се информационно общество. Разбира се, Съюзът на математиците в България (СМБ) не може да остане настрана от световните тенденции и този доклад е една от набелязаните прояви. И все пак, какво е това Статистика? Какъв е нейният предмет на дейност, нейните основни цели и задачи, обхватът на нейната приложимост? Кога, как и какво да се изучава? С други думи, как да се интегрира Статистиката в рамките на съществуващите образователни системи? Какво е положението в тази област у нас и по света? Как трябва да се развива Статистиката като наука? А кой и как трябва да се занимава с нейните приложения? За да не стигаме до популярния парадокс, с който най-често се осмива неправилното (понякога преднамерено) използване на статистиката: „Има три вида лъжа - обикновена лъжа, нагла лъжа и ... статистика“. А какви ли не епитети се срещат още: придворна статистика, подвеждаща статистика, поръчкова статистика, обслужваща статистика, погрешна статистика, „изсмукана от пръстите“ статистика и т.н.

Разбира се, да се отговори на тези въпроси в рамките на едно кратко есе не е лесно (колкото по-кратко, толкова по-трудно). И все пак, какво е това Статистика? За да отговорим на този въпрос, първо трябва да отбележим, че Статистиката има няколко лица, т.е. този термин има различни значения. За математиците Статистиката е математическа дисциплина, която в известен смисъл решава обратната задача на Теорията на вероятностите (ТВ). За да подчертаят това, те обикновено я наричат Математическа статистика (МС). Често тяхното обединение се нарича Стохастика, т.е. $ТВ + МС = Стохастика$. Това ще се опитаме да прецизираме по-нататък. За обикновения човек Статистика означава най-често някакви числови данни, от които се правят някакви изводи. И тук е „тънкият момент“, т.е. има две различни дейности - събиране на информация за даден обект или явление (статично или динамично) и изводи от събраната информация. Това също ще уточняваме по-нататък. Тук обаче трябва да отбележим, че терминът „статистика“ е добил гражданственост, особено на Запад, като една търговска марка, като събирателно понятие за всички тези математически и недотам

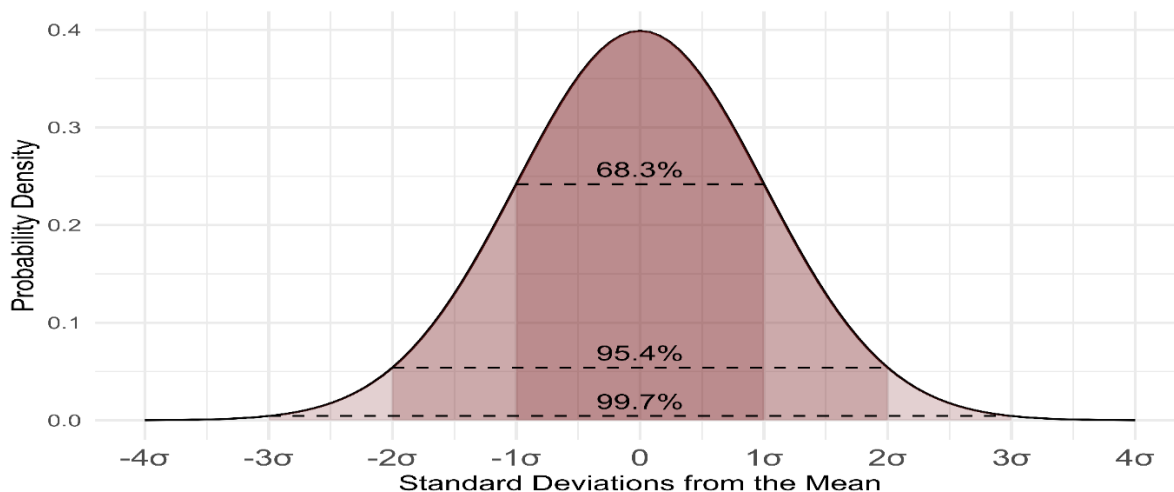
математически дейности. Така че като честваме Международната година на статистиката, ние трябва да разбираме Статистиката в цялата нейна многостранност.

Във всеки приличен университет днес има задължително катедра (или департамент) по Математика (където понякога се занимават и с ТВ и МС), а отделно има катедра (или департамент) по Статистика (където задължително се занимават с ТВ и МС). Освен това в по-големите университети, които имат медицински центрове, са създадени големи департаменти по Биостатистика. Към икономическите факултети (или центрове) има департаменти по Икономическа статистика и т.н. Така че в общественото пространство фигурират съчетания като медицинска статистика, инженерна статистика, икономическа статистика, държавна статистика, селскостопанска статистика, промишлена статистика, военна статистика, финансова статистика, борсова статистика, банкова статистика, търговска статистика (външна или вътрешна), макроикономическа статистика, спортна статистика и т.н. А всеки ден от медиите ни заливат с какви ли не потресаващи статистики (убийства, катастрофи, пожари, наводнения, бедствия, епидемии, болести, смъртност, корупция, кражби, безработица и т.н.). Има, разбира се, и „приятни“ статистики (раждания, бракосъчетания, ръст на БВП, спестявания, покупателна способност, ръст на заплати и пенсии, пътувания в чужбина и т.н.). Но като че ли вторите се губят в потока на първите (добрата вест не е новина за съвременната журналистика). Както отбелязахме, живеем в информационно общество. И както казват някои млади хора, няма защо да се учи: каквото ни трябва, намираме го в интернет. Решихме да проверим и написахме думата „Статистика“ в една от търсачките. Ето какво намерихме:

„Статистика

от Уикипедия, свободната енциклопедия

Диаграма на **нормално разпределение**, често използвано в статистиката



Статистиката е математическа дисциплина, която изучава добиването на информация чрез анализ и интерпретация на емпирични данни, използвайки **теорията на вероятностите**. Статистическата дейност включва също планирането и организирането на събирането на данни чрез проучвания и експерименти. Статистиката възниква във връзка с нуждите на емпиричната наука и се отличава от повечето клонове на математиката по своята приложна насоченост. Статистически методи се прилагат в широк кръг области като природните и обществените науки, държавното управление и бизнеса. Една от основните подобласти на статистиката е **описателната статистика**, която се занимава с обобщаването на систематизирани данни. То е от особена важност при емпиричните изследвания и описването на резултати от експерименти. Чрез методите на статистиката данните могат да бъдат анализирани, като се отчита случайността и несигурността на наблюденията, и въз основа на това да се правят изводи за съдържащи се в тях закономерности“.

Не зная кой е авторът на това описание, но то е доста неточно и непълно. И докато непълнотата може да бъде оправдана с желание и необходимост за краткост, то грешките могат да подведат непредубедения читател. Още първото изречение съдържа невярна информация, като определя Статистиката като математическа дисциплина. Както беше отбелязано, само една част от Статистиката, а именно Математическата статистика, е безспорно част от Математиката. Защото, когато срещаме например изрази като „статистика на кражбите“ или „борсова статистика“, те нямат нищо общо с Математиката. Те са просто съвкупности от данни, които също е прието да се наричат „статистика“ (при това най-често с някакво прилагателно). Всъщност етимологията на думата статистика е свързана с думите „state“ (държава, състояние) и „status“ (състояние,

положение). В този смисъл зараждането на статистиката може да бъде отнесено към Древния Китай, където още преди 5 000 години са започнали да събират сведения за състоянието на държавата. Ясно изразен образец на държавна статистика можем да намерим в Римската империя. Между другото, това е един от основните признаци на държавността. Днес всяка държава има специална институция с най-различни названия (министерство, агенция, институт, комитет, бюро, комисия), която се занимава с т.нар. държавна (или официална) статистика.

У нас държавната статистика започва да се развива от 1880 г. в Княжество България с учредяването на Статистическото отделение към Министерството на правосъдието. Оттогава, независимо от организационните промени на статистическата институция, тя събира, обработва и съхранява важна статистическа информация за демографските и икономическите процеси в страната. Сега статистическата дейност на държавата се осъществява главно от Националния статистически институт (НСИ), който е основан със Закона за статистиката, приет през 1991 г. от Народното събрание. По-подробно с историята, визията, целите, устройството, организацията и въобще с всички дейности на НСИ любознателният читател може да се запознае на интернет страницата на НСИ: <http://www.nsi.bg/index.php>. И ако не сте посещавали тази страница, ще останете приятно изненадани да разберете колко интересна информация има и каква огромна дейност се развива в областта на държавната статистика. Ето някои извадки: „Главната цел на развитието на НСС (Националната статистическа система) през периода 2008 - 2012 г. е да се усъвършенства разработването, производството и разпространението на статистическа информация за всички групи потребители чрез подобряване на институционалния капацитет на системата и ускорено внедряване на съвременни информационни и комуникационни технологии в условията на пълноправно членство на Република България в ЕС.

Официалната статистическа информация - продукт на НСС, е предназначена за анализи и научни изследвания, за планиране, прогнозиране и вземане на управленски решения на микро- и макроравнище. Статистическите информационни продукти и услуги са насочени към: органите на държавно управление и местната власт; бизнеса и неправителствените организации; научната и академичната общност; медиите; ЕС, ООН, МВФ и други международни организации; отделни граждани. Развитието на НСС през периода 2008 - 2012 г. предполага осъществяването на следните **основни цели**:

- Развитие на методологията и обогатяване на съдържанието на съществуващите статистически изследвания и внедряване на нови изследвания и показатели в

съответствие с нуждите на потребителите и приоритетите на Статистическата програма на ЕС 2008 - 2012 година.

- Усъвършенстване на производството, разпространението и съхранението на статистическа информация.
- Подобряване на средата, информационната инфраструктура и ресурсната осигуреност на НСС“.

Събирането на достоверна статистическа информация е важна и отговорна задача, но само по себе си това би било безполезно, ако получените статистически данни не се анализират, т.е. от тях да се получат съответните достоверни изводи и прогнози. А това вече не е възможно без използване на методите на Математическата статистика.

Както беше отбелязано, МС решава обратната задача на ТВ. Образно казано, те са като двете страни на една монета. Да се опитаме да изясним тази връзка, защото тя е много важна. Най-общо казано, предмет на ТВ и МС е математическият анализ на понятието „случайност“. ТВ и МС, или с други думи, Стохастиката, е тази част от Математиката, която създава и изследва математически модели на случайни явления, величини и процеси. Тук е моментът да припомним, че едно от големите постижения на ХХ век беше аксиоматизацията на ТВ, предложена от А. Н. Колмогоров през 1933 г. (VI проблем на Хилберт). Последвалото бурно развитие на ТВ превърна тази част от Математиката в естествен фундамент на цялото естествознание, защото случайността лежи в основата на самата природа и колкото повече се приближаваме до микросвета, толкова това става по-ясно. Съгласно Колмогоровската аксиоматика основните математически обекти в ТВ (случайни величини и случайни процеси) се разглеждат в т.нар. вероятностни пространства, където вероятността е определена като една неотрицателна, нормирана и адитивна мярка. По този начин в услуга на ТВ е впрегнат практически целият съществуващ математически арсенал, което естествено води до бурно развитие на самата ТВ. От друга страна, построяването на конкретно вероятностно пространство означава създаването на определен математически модел за даден стохастичен феномен. Това стимулира изключително многобройните и разнообразни приложения на ТВ. И тук естествено стигаме до МС. Образно казано, ТВ е изкуство за пресмятане на едни вероятности посредством други, т.е. при зададени начални разпределения на вероятностите се описват вероятностните разпределения на последващи по-сложни събития и процеси. А МС решава обратна в известен смисъл задача, т.е. по наблюденията над събитията, случайните величини или случайните процеси да оцени тези първоначални разпределения, чрез които се описва

математическият модел. Чрез МС се проверяват и хипотезите за адекватност, т.е. доколко математическият модел е близък до реалната действителност. При това основните методи на доказателство и изследване се основават на ТВ, като се използва, разбира се, и целият останал математически апарат.

Например, важна роля играят редица оптимизационни и апроксимационни методи. Може би е добре да илюстрираме тези идеи с един несложен реален пример.

Задача 1. При статистически качествен контрол се прави случайна извадка с обем n от партида с еднородни изделия. Известно е, че вероятността всяко отделно взето изделие да бъде дефектно, е p . Каква е вероятността в извадката да има точно k дефектни изделия? Това е една типична задача от ТВ. В този случай не е трудно да се покаже, че случайната величина X , равна на броя на дефектните изделия в извадката, има биномно разпределение $Bi(n, p)$, т.е. вероятностите $P(k) = P\{X = k\}$ са т.нар. биномни вероятности, $k = 0, 1, 2, \dots, n$. Ето как би изглеждала съответната ситуация от гледна точка на МС.

Задача 2. При статистически качествен контрол се прави случайна извадка с обем n от партида с еднородни изделия. При прегледа се оказва, че сред тях има точно k дефектни изделия.

а) Как да определим неизвестната вероятност p всяко едно отделно взето изделие да бъде дефектно?

б) Ако p^* е оценка за тази вероятност, как да определим точността на тази оценка?

в) Как да проверим хипотезата, че бракът е под 1%?

г) Как да различим две хипотези, при които бракът е съответно 2% или 5%?

Не е трудно да се предложи $p^* = k/n$ (често наричана честота) като една естествена оценка за p . Отговорът на останалите въпроси обаче не е толкова очевиден и изисква прилагането на определени статистически методи.

Съвременното развитие на ТВ и МС се обуславя от два мощни фактора: вътрешното развитие на теорията и непрекъснатия приток на нови и все по-сложни проблеми от заобикалящата ни действителност. Така са създадени редица нови направления както в самата математика, така и в колаборация с други науки като например Стохастичен анализ, Стохастични диференциални уравнения, Стохастична геометрия, Стохастични матрици, Финансова математика, Теория на риска и надеждността, Статистическа механика, Статистическа физика, Биостатистика, Икономическа статистика, Статистически контрол на качеството, Статистическа демография и т.н.

Заобикалящият ни свят и общество са стохастични по своята същност. Поведението на индивида или обществото не могат да бъдат описани с детерминистични функции, те имат ярко изразен стохастичен характер³. И тук стигаме до един от големите парадокси на човечеството: нашето възпитание по своя характер е строго детерминистично и това води до конфликти със заобикалящия ни стохастичен свят. На нас от малки ни втъпяват, че всичко е свързано в причинно-следствени връзки, пренебрегвайки изобщо категорията **случайност**.

По-възрастните сигурно си спомнят, а по-младите може би са чували за периода, когато под влияние на идеите на академик Лисенко в СССР и т.нар. социалистически лагер бяха разгромени редица научни направления, обявени като „буржоазни“ (и на първо място генетиката). Тогава беше издигнат и лозунгът: „Науката е враг на случайността!“ Под угроза за забрана са не само ТВ и МС, но и всички клонове на науката, където те се прилагат. Математиците Хинчин и Колмогоров обаче остроумно и своевременно издигат контратезата: „Да, науката е враг на случайността, но врагът трябва да се изучава. А това прави Теорията на вероятностите“. Това се харесало на диктатора Сталин, създател на тезата за „враговете на народа“, които били изпращани в концлагери. Казват, че всмуквайки от знаменитата си лула, Сталин заявил: „Правилно! Врагът трябва да се изучава!“ И това спасило ТВ и МС от пълен разгром, заедно с редица други области на науката.

Сега може да ни се струва нелепо, но тогава много учени са заплатили с живота си само за това, че са били истински учени. Този пример е ярко доказателство за вредата от фундаменталния детерминизъм, който се проявява в редица религиозни норми и ненаучни идеологии, донесли толкова беди на човечеството. И тук естествено стигаме до въпроса за стохастичното образование и възпитание, защото стохастичната култура е важна част от общата култура на индивида и нацията. Днес практически в почти всички университетски специалности се изучават в различна степен ТВ и МС. В развитите страни елементи на стохастиката са станали вече неизменна част от средното образование по математика, като в някои страни то започва от началните класове или дори от детските градини. В това отношение ние значително изоставаме от водещите световни тенденции.

³ Вж. Б. Димитров и Н. Янев. Вероятности и статистика. Издателство на СУ - първо изд. 1989, второ изд. 1996; Софтекс - трето изд. 2007, София.

В процеса на еволюцията у човечеството постепенно се развива стохастична интуиция вследствие на дългите наблюдения и опит, свързани с лова, земеделието, риболова, предсказването на времето и т.н. Човекът е свикнал да играе хазарт с природата и неслучайно хазартните игри са се появили от най-дълбока древност. Всъщност за начало на ТВ се счита 1654 г., когато започва кореспонденцията между Ферма и Паскал по повод на някои въпроси, поставени от кавалера де Мере и свързани с хазартните игри. По-нататъшното развитие на стохастиката е свързано с имената на Галилей, Хюйгенс, Моавр, Лаплас, Гаус, Поасон, Коши, Лобачевски, Чебишев, Марков, Ляпунов, Борел, Бернщайн, Леви, Хинчин, Винер, Фишер, Нейман, Пирсън, Гнеденко, Фелър, сред които отбелязахме вече фундаменталния принос на Колмогоров (1903 - 1987 г.), наричан често „баща на съвременната стохастика“.

Съвременната МС е силно свързана с развитието на информатиката. Сега всеки потребител на статистическа информация разполага с мощни пакети от статистически процедури, с които лесно и бързо могат да бъдат решени редица стандартни статистически задачи. Това доведе до развитието на нов дял от МС, наричан често компютърна статистика.

Като се връщам към спомените си за Световния статистически конгрес през далечната вече 1994 г., организиран от ISI (International Statistical Institute) в Чепълхил, Северна Каролина, не мога да не отбележа визитата в кампуса на компанията SAS, най-големият производител на статистически софтуер. По това време, както ни беше казано, там са работели над 3 000 души, които заедно със семействата си представляват едно малко модерно градче, разположено сред вековни гори и наброяващо над 10 000 жители. В демонстрационната зала на централния офис имаше над 150 монитора, на всеки от които се представяха последните постижения на фирмата в различни области на статистиката. Беше истинско удоволствие да се видят високопрофесионалните дискусии между участниците в конгреса (предимно статистици теоретици) и демонстраторите (специалисти по компютърна статистика). От някои участници в конгреса бяха изказани пожелания за разширяване на възможностите на съществуващия софтуер или предложения за нови статистически процедури. Това доведе моментално до предложения за сключване на договори за развитие на тези идеи. Много участници в конгреса получиха покани (сред тях и авторът) да изпратят своите доклади и други научни статии с цел евентуална разработка на нови статистически програми (авторът трябва да признае, че не се е възползвал от тази възможност). Това беше и ярка

демонстрация на сътрудничество между наука и бизнес. Всъщност SAS беше един от главните спонсори на конгреса.

Възможностите на съвременните компютри за бързо симулиране на случайни величини и процеси доведе до развитието на много нови дялове от статистиката и статистическото моделиране. По този начин практически всяка задача може да намери едно сравнително добро приближено решение. Особено силно се развиха т.нар. методи Монте Карло за пресмятане на интеграли, диференциални и интегрални уравнения. С подробности за организацията на честванията на Международната година на статистиката по цял свят любознателният читател може да се запознае на интернет сайта <http://www.statistics2013.org/>. На него са регистрирани над 2 000 научни организации, свързани със статистиката, много университети, научноизследователски институти, обществено-професионални статистически организации и други. България е добре представена в различните раздели, като трябва да отбележим преди всичко участието на ИМИ-БАН, НСИ, ФМИ-СУ, БСД (Българско статистическо дружество) и други.

Изумително е огромното количество хора и организации в цял свят, заети със статистическата информация в цялата гама от аспекти. Така например според някои оценки само в САЩ с ТВ и МС се занимават между 60 и 70% от професионалните математици. И въобще в САЩ бизнесът започва със статистика, развива се със статистика и прави финални заключения пак с помощта на статистиката. Всичко това естествено рефлектира върху сериозното количество издавана специализирана литература по Стохастика и съответните периодични списания с фундаментален или приложен характер. Интересно, че в сериозните списания по експериментални науки вече не се приемат статии, в които няма съответни статистически модели и изводи. За съжаление, стохастичната колегия у нас е твърде малка дори за мащабите на нашата страна, особено в областта на приложната статистика. Очертава се сериозна ниша, в която, да се надяваме, ще се появят млади и способни кадри. Един доказан начин за това е въвеждането на по-ранно и задълбочено запознаване с Вероятности и Статистика в средните училища. Да се говори и пише за Статистиката (особено пред професионалисти) не е нито лесно, нито просто, особено ако се търси по-нестандартен ракурс. От друга страна, всеки от нас има вече някакви изградени представи, а предложената тема е толкова широка, че спокойно може да обхване съдържанието на една дебела книга, където всички тези неща могат да бъдат разгледани по-подробно, в исторически аспект и с многобройни примери.

Целта на това кратко есе е по-скоро да представи някои въпроси и тезиси за размисъл и дискусии. Липсата на цитирана литература се обяснява просто с факта, че авторът не е ползвал такава (с изключение на това, което е останало в съзнанието от някъде-някога прочетеното). Останалото са собствени мисли, нахвърляни „експромпт“, не без известно удоволствие, защото да се говори и пише за Статистиката е наистина удоволствие. Разбира се, Международната година на статистиката е добър повод за сравнения, анализи и изводи, които ще бъдат последвани от дискусии и други мероприятия.