МАТЕМАТИКА И МАТЕМАТИЧЕСКО ОБРАЗОВАНИЕ, 2019
MATHEMATICS AND EDUCATION IN MATHEMATICS, 2019
Proceedings of the Forty-eighth Spring Conference
of the Union of Bulgarian Mathematicians
Borovetz, April 1–5, 2019

НЯКОИ АКЦЕНТИ В ПРИЛОЖНИТЕ СТАТИСТИЧЕСКИ ИЗСЛЕДВАНИЯ

Велизар Павлов

Наличието на огромна по обем статистическа информация в публичното пространство и нуждите от компетентен статистически анализ във всички области на науката и живота изискват адекватно образование и знания по статистика за широк кръг ползватели. В настоящата статия са поставени някои акценти в приложните статистически изследвания: представителност; специфика на статистическите закономерности и зависимости; математически и статистически различия; сравнителен анализ, основан на експериментални и контролни групи; нормалност на емпиричните разпределения; линейна корелация – сила и значимост на коефициентите на корелация; регресионни модели – адекватност и значимост на регресионните коефициенти. Целта е да се помогне на всички заинтересовани читатели да повишат своите компетенции в областта на приложната статистика и да ги предпази от допускането на някои често срещани грешки в статистическите анализи.

1. Въведение. Думата *математика* произлиза от старогръцката дума $\mu\alpha\theta\eta\mu\alpha$ (*матема*), която означава „знание, изучаване“. До около 17-ти век в Европа под „математика“ повече се е разбирало това, което днес се нарича астрология, но с повишаването на научната ѝ приложимост тя започва да се разглежда самостоятелно, като Гаус я нарича „кралица на всички науки“. Понастоящем няма всеобщо приета дефиниция за математика. Разминават се мненията дори на професионални математици. По-скоро се правят опити за конкретизиране на определени области на математиката. Може да се каже, че чистата математика се занимава с абстрактни постановки и задачи, които нямат практическа насоченост. Апаратът обаче, който тя развива, има голям обхват на приложимост в различни дялове на науката и живота.

Този факт дава големи възможности за развитието на съвременната приложна математика, която най-общо се занимава с разработване на математически методи за решаване на задачи от практиката. Важна нейна част е емпиричната математика, която се основава на събиране или симулиране на данни, последващ анализ и/или тестване на научни предположения. В голяма степен това е предмет и на статистиката. Така се преплитат приложната математика и статистиката, която пък се основава на теорията на вероятностите. Някои учени определят статистиката като дял на математиката, а други като самостоятелна наука. Този спор едва ли ще намери някога решение. Сигурно е обаче, че съвременното математическо образование изисква и образование по статистика. Не напразно Хърбърт Уелс казва: „Статистическото мислене е толкова необходимо на цивилизования човек, колкото и умението му да чете и пише.“

В подкрепа на това съждение е огромната по обем статистическа информация в публичното пространство – обобщени данни, оценки, изводи и заключения, които ни поднасят печатните и електронни медии. Например данни за жизнен стандарт; икономически ръст; трудова заетост; безработица; цени на недвижимата собственост; валутни курсове; средна работна заплата; продажби на автомобили; рейтинги на институции (в това число на висшите училища в България), партии, политици; информация, свързана със спортни състезания – класирания, брой победи, загуби, отбелязани попадения; демография – смъртност, раждаемост, данни за населението по постоянен и/или настоящ адрес, по области, общини, населени места; образование – резултати от външното оценяване, брой ученици и/или студенти; данни за извършени престъпления, мигранти, пътно-транспортни произшествия, наложени наказания, платени глоби; данни за пациенти, брой болнични легла, здравни пътеки и много други.

По същество тази информация има първичен характер. Статистическият анализ изисква формиране на изследвана съвкупност, най-често извадка, и прилагане на съответни статистически методи за извличане на изводи и заключения. Днес практически няма област от научното познание и живота, в която да не се използват успешно статистическите методи.

В тази връзка в учебните планове и програми на почти всички университетски специалности са включени един или няколко курса по статистика. Например в икономически специалности обикновено се включват курсове по Статистика, Компютърна статистика, Иконометрия и курсов проект, който в чужбина завършва със студентска публикация.

В България се забелязва свиване и дори неглижиране на обучението по статистика, както в средното, така и във висшето образование. В същото време в различни статистически анализи в публичното пространство или такива, включени в изследователски разработки – публикации, дипломни работи на студенти или учители за повишаване на квалификацията, дисертационни трудове и др., често се правят грешки. От тази гледна точка е особено важно да не допускаме да ни подвежда „стъкмистиката“ на онези, които съзнателно или не, използвайки некоректно статистически данни, манипулират нашето съзнание. В тази връзка в настоящата работа са поставени някои акценти в приложните статистически изследвания, като се надяваме те да бъдат от полза на всички заинтересовани читатели и да повишат своите компетенции в областта на приложната статистика.

2. Разграничения и уточнения. Приложните статистически изследвания се основават на теорията на вероятностите и математическата статистика [2, 3]. Тази теория е трудно разбираема от масовия потребител и има консервативен характер за разлика от приложенията в практиката, които са динамични, разнообразни, многобройни и винаги съдържат в себе си риск за грешка. С оглед повишаване на степента на разбираемост на статистическата теория и нейната връзка с практиката, правим някои разграничения и уточнения.

В теорията се разглеждат т. нар. **теоретични (вероятностни) разпределе-**

ния, където важна роля играят **случайните величини**, които могат да приемат определени стойности със съответни вероятности. В зависимост от вида на теоретичното разпределение, тези вероятности се пресмятат по съответни теоретични формули.

В практиката се изследват т. нар. **емпирични статистически разпределения**, където понятието **признак** се явява аналог на понятието случайна величина. При това признаците приемат определени значения със съответни честоти на проявления, в резултат на проведено статистическо емпирично наблюдение. Важно е да отбележим, че огромната част от тези наблюдения са извадкови.

По силата на Закона за големите числа с нарастване на броя на опитите (наблюденията, обема на извадката) полученото емпирично разпределение се приближава към съответно теоретично разпределение (теоретичен модел) [13]. В случай на установена близост, при определен риск за грешка, цялата известна теория за дадено теоретично разпределение се прехвърля върху съответното емпирично разпределение. Това дава големи възможности за познавателния процес.

Подобна е практиката например в медицината, където при установено заболяване се прехвърля съществуващата теория за лечение от това заболяване и ако не е допусната грешка, положителните резултати са налице.

Тази наша интерпретация дава ясен отговор на един често срещан въпрос „Защо е нужно да се изучава теория?“ Задачата за свързване и прилагане на теорията в практиката е една от най-важните в човешкото познание. Тази задача може да бъде решавана успешно само от хора, които имат теоретична подготовка и умения да я прилагат в практиката, при което винаги обективно съществува риск за грешка. Причината за този риск е свързана с факта, че в практиката винаги влияят и случайни фактори, които дори не могат да бъдат предвидени. Ето защо собственият ни житейски опит сочи, че в живота не винаги

$$2 + 2 = 4.$$

3. Някои акценти в приложните статистически изследвания. В този раздел са поставени акценти върху някои важни специфични моменти в приложните статистически изследвания.

3.1. Извадкови статистически изследвания. Както вече отбелязахме, в практиката огромна част от статистическите изследвания са извадкови. В тази връзка следва да разграничаваме репрезентативните (представителни) изследвания от останалите. Първите дават възможност да се характеризира съответната генерална съвкупност, докато при вторите статистическият анализ е валиден само и единствено за изследваната извадка.

Изискванията за представителност на една извадка са тя да е случайна (основана на случаен подбор) и да е с достатъчно голям обем [10]. В практиката тези изисквания не винаги са изпълними. Например при изследване на съвкупности от хора изискването за случаен подбор се трансформира в изискване извадката да възпроизвежда основните характеристики и пропорции на съответната генерална съвкупност [8, 9]. За тази цел следва да бъдат проверени различията между относителните дялове в извадката и съответните дялове в генералната съвкупност по всички използвани основни признаци [16]. По отношение на второто изискване за представителност, на практика се търси баланс между размера на максимално допустимата грешка и обема на извадката [12]. Например тестваме системите за сигурност на няколко чисто нови току що произведени автомобили и правим извод за всички останали или анализираме резултатите от приложено ново лечение (подход, методика) върху определен брой пациенти и пренасяме този нов подход към всички останали. Важно е да отбележим, че този процес винаги съдържа риск за грешка, която не можем да избегнем поради факта, че не е възможно да тестваме всички произведени автомобили или пък всички пациенти, болни от дадена болест.

В по-голямата си част приложните статистически изследвания нямат представителен характер. Те са по-лесни за организация и провеждане, изискват значително по-малки ресурси, свързани с финанси, време или хора. Тези изследвания също могат да бъдат показателни, особено когато са серийни през малки периоди от време.

3.2. Статистически закономерности и зависимости. Статистиката най-общо изследва масовите явления, които се подчиняват на т. нар. **статистически закономерности**. Те се проявяват като съществени и устойчиви съотношения между отделните проявления на масовото явление в пространството и във времето. Например съжденията: „Цените на стоките и услугите в София са по-високи от цените на съответните стоки и услуги в Габрово.“ или „С нарастване на ръста при хората се увеличава и теглото.“ се отнасят за масови явления и отразяват статистически закономерности. Те не се отнасят за отделните проявления на масовото явление, т.е. за отделните случаи. Напълно възможно е да съществува стока(и) или услуга(и), чиято цена в Габрово да е по-висока от цената на същата стока или услуга в София или пък да съществуват хора високи и слаби или ниски и пълни, но това в никакъв случай не противоречи на общата статистическа закономерност или зависимост, която е валидна за преобладаващия брой единици от дадена съвкупност. Намерените изключения не променят тази валидност! В тази връзка често се пораждат излишни спорове.

Този основен принцип на статистиката съществено я различава от другите математически науки, където един закон (закономерност) е закон, ако е изпълнен при всяко проявление на дадено масово явление. Например Питагоровата теорема е изпълнена за всички правоъгълни триъгълници **без изключение**!

Важно различие между функционалните (математически) зависимости и статистическите е свързано с факта, че при функционалните зависимости на всяко определено значение на фактора x съответства винаги, на всяко място, по всяко време едно и също точно определено значение на резултата y. Например, ако разгледаме функционалната зависимост $y = 2x + 3$ за различни стойности на x, получаваме

X	0	1	2	3	...
Y	3	5	7	9	...

т.е. ако $x = 2$, то винаги $y = 7$. Това не е така при статистическите зависимости. Например, нека разгледаме дадена статистическа зависимост между ръста и теглото при хората

x – ръст (cm)	170	165	160	172	165	180	...
y – тегло (kg)	70	72	58	75	68	92	...

от която лесно се вижда, че е напълно възможно за една и съща стойност на фактора x – ръст да имаме различни стойности на резултата y – тегло. В случая за $x = 165$

при един индивид имаме тегло $y = 72$, а при друг $y = 68$, т.е. двама души имат един и същ ръст, но различно тегло.

3.3. Математически и статистически различия. Математическите различия между числовите характеристики или относителни дялове в изследвани извадки не са основание за съответни изводи [5, 9]. В този раздел е поставен акцент върху различията между средните стойности (аритметични) и относителните дялове в две извадки, които играят важна роля в практиката.

Средните стойности са важни числови характеристики за обосноваване на изводи в сравнителни анализи, основани на извадки. Най-често се сравняват средните стойности в две независими извадки (например контролна и експериментална група), в зависими извадки (експериментална група в началото и в края на даден период) или средната стойност в една извадка с наложен стандарт относно нейния размер.

Нека предположим, че съгласно наложен стандарт (БДС) към производител на електрически прекъсвачи средният брой включвания и изключвания, на които трябва да издържат прекъсвачите е 9000. По данни от случайна извадка с обем 50 изделия е установено, че средният брой на включванията и изключванията, на които издържат прекъсвачите е 8917, при което е известно още, че стандартното отклонение е $\sigma = 380$. Естественият въпрос е дали има проблем, свързан с качеството на произвежданите прекъсвачи? Основавайки се на математическите различия между средната стойност от наложения стандарт и установената средна стойност в извадката можем да направим извод, че е налице такъв проблем. Дали това е така?

Правилният подход е да се формулират две хипотези [16]. Първата (нулева) хипотеза, която гласи, че различията между средната стойност в извадката $\overline{X}_S$ и средната стойност от наложения стандарт са незначими

$$H_0 : \overline{X}_S = 9000,$$

и алтернативна хипотеза, която гласи, че тези различия са значими

$$H_1 : \overline{X}_S \neq 9000.$$

Изборът за приемане на една от двете хипотези и отхвърляне на другата се основава на критерия на Стюдънт, при което за емпиричната характеристика се получава

$$t_{emp} = \frac{|\overline{X} - \overline{X}_S|\sqrt{n}}{\sigma} = \frac{|9000 - 8917|\sqrt{50}}{380} = 1.54$$

Теоретичната характеристика $t_T = 2.01$ се определя от таблица със стойности на теоретичното t-разпределение [19], при естествен риск за грешка $\alpha = 0.05$, двустранна критична област $(P = 1| - \alpha/2 = 1 - 0.025 = 0.975)$ и степени на свобода $k = n - 1 = 50 - 1 = 49$. Лесно се вижда, че

$$t_{emp} = 1.54 < t_T = 2.01.$$

Следователно H_0 се приема, т.е. различията при средните стойности са незначими. От гледна точка на практиката това означава, че стандартът е спазен. Този извод се различава от първоначалния извод, основан на математическите различия между двете средни стойности.

Сравнителните анализи, основани на относителни дялове, също играят важна роля в статистическите изследвания. Най-често се сравняват относителни дялове в две независими извадки (например контролна и експериментална група), в зависими извадки (експериментална група в началото и в края на даден период) или относителен дял в една извадка с наложен стандарт относно неговия размер.

Нека предположим, че дадена компания има намерение да пусне нов продукт на пазара и желае да установи дали има значими различия в относителните дялове на хората, които го одобряват в две различни населени места с приблизително еднакво население. За тази цел са направени две независими случайни извадки с достатъчно голям обем по 170 души. След разясняване на същността на продукта 80 души от първата извадка са го одобрили, а от втората извадка това са направили 95 души.

Основавайки се на математическите различия в двата дяла на одобрение изводът е, че одобрението във второто населено място е по-високо от първото. Дали обаче това е така?

Правилният подход е да се формулират две хипотези [16]. Първата (нулева) хипотеза, която гласи че различията в относителните дялове на одобрение са незначими, т.е.

$$H_0 : \pi_1 = \pi_2$$

и алтернативна хипотеза, която гласи, че тези различия са значими

$$H_1 : \pi_1 \neq \pi_2.$$

Изборът за приемане на една от двете хипотези и отхвърляне на другата се основава на критерия на Пирсън, при което теоретичните честоти се пресмятат по формулите

$$f_{t11} = \frac{170 * 175}{340} = 87.5, \quad f_{t12} = \frac{170 * 160}{340} = 82.5,$$
$$f_{t21} = \frac{170 * 175}{340} = 87.5, \quad f_{t22} = \frac{170 * 160}{340} = 82.5,$$

които за удобство се поставят в скоби в таблицата, съдържаща съответните емпирични честоти

Населено място / отношение	Одобрили	Не одобрили	Общо
Първо населено място	80 (87.5)	90 (82.5)	170
Второ населено място	95 (87.5)	75 (82.5)	170
Общо	175	160	340

За емпиричната характеристика по формулата на Пирсън се получава

$$\chi^2_{emp} = \sum \frac{(f_e - f_t)^2}{f_t} = \frac{(80-87.5)^2}{87.5} + \frac{(95-87.5)^2}{87.5} + \frac{(90-82.5)^2}{82.5} + \frac{(75-82.5)^2}{82.5} = 2.64$$

Теоретичната характеристика $\chi^2_T = 5.02$ се определя от таблица със стойности на теоретичното χ^2-разпределение [19], при естествен риск за грешка $\alpha = 0.05$, двустранна критична област и степени на свобода $k = (r-1)(c-1) = (2-1)(2-1) = 1$, където r е броят на редовете (населени места) в таблицата с данни, а c е броят на колоните (одобрение/неодобрение) в същата таблица. Лесно се вижда, че

$$\chi^2_{emp} = 2.64 < \chi^2_T = 5.02,$$

което означава, че се приема нулевата хипотеза, т.е. различията в относителните дялове на хората, одобряващи новия продукт са статистически незначими. Този извод се различава от първоначалния извод, основан на математическите различия между наблюдаваните относителни дялове.

3.4. Експериментални и контролни групи. Приложните статистически изследвания често са свързани с експериментиране на нови средства и методи за въздействие, които да оказват благоприятен ефект [4]. Експерименталният дизайн обикновено включва формирането на две групи – експериментална, в която се прилага новото въздействие и контролна, в която се прилагат традиционни методи на въздействие. Състоянието на признаците, върху които се въздейства, се измерва най-малко два пъти – в началото, преди прилагане на експерименталното въздействие (първо изследване), и след определен период от време, който изследователят счита че е необходим, за да се доловят настъпилите промени (второ изследване).

За ефекта на новото въздействие се извършват **пет сравнения** [8, 9], най-често по отношение на средната стойност на наблюдаван количествен признак или относителен дял. **Първото** сравнение е свързано с проверка за статистическа значимост на измененията в експерименталната група (начало-край), а **второто** – за същото в и контролната група (начало-край). Важно е да отбележим, че сравнителният анализ между двете групи задължително трябва да се предхожда от проверка за незначими различия между двете групи („равен старт“) в началото на периода (**трето сравнение**). Ако липсват статистически значими различия в началото на прилагане на експерименталното въздействие, то сравнителният анализ е съдържателен. В противен случай той губи своя смисъл. **Четвъртото** сравнение е свързано с установяване на статистическата значимост на различията между двете групи в края на периода (при второто изследване). На тази база се формират изводи за различията в ефекта на двете въздействия (методики). **Последното сравнение** е най-съществено от гледна точка на различията в ефекта на експерименталното и контролното (традиционното) въздействие. То е свързано с установяване на статистическата значимост на различията между прирастите в двете групи в края на експерименталния период.

В практиката не винаги е възможно да бъде формирана контролна група. Тогава се работи само с експериментална група, при което основното сравнение е свързано с проверка за значимост на различията между началото (първо изследване) и края на периода (второ изследване).

Важно е да отбележим, че при работа с експериментални и контролни групи, те трябва да бъдат хомогенни (еднородни) по състав спрямо признаци, които оказват влияние върху очакваните резултати. Например спрямо признаците „пол“ и „възраст“ при оценяване на двигателните качества при хората. В този случай е недопустимо провеждането на какъвто и да е сравнителен анализ между групи, съдържащи мъже и жени, т.е. сравнителният анализ трябва да се провежда отделно за мъжете и отделно за жените.

3.5. Нормалност на емпиричните разпределения. Задачата за изследване на близостта между дадено емпирично разпределение и теоретичното нормално разпределение е една от най-важните в приложната статистика.

Проверката за близост до нормалното разпределение може да бъде направена с помощта на емпиричния критерий, основан на коефициентите на асиметрия и ексцес, за който е известно, че не е много прецизен [1]. За по-голяма сигурност могат да се използват критериите на Пирсън (при групирани данни) [21] или на Колмогоров-Смирнов (при негрупирани данни) [14, 21].

Установяването на близост до нормалното разпределение обосновава по-нататъшното приложение на параметрични критерии за проверка на хипотези, както използването на корелационния коефициент на Пирсън в евентуален корелаиционен, регресионен или факторен анализ. В противен случай следва да се използват непараметрични критерии и коефициент на Спирмън или други коефициенти в зависимост от спецификата на данните [22].

3.6. Линейна корелация. Сила и значимост. Статистическите зависимости още се наричат **корелационни** или просто **корелации**. В случая, когато те имат линеен характер и за определяне на тяхната степен на сила най-често се използват корелационните коефициенти на Пирсън и на Спирмън [22]. Коефициентът на Пирсън се използва, когато се изследва корелция между количествени признаци, всеки от които има разпределение, близко до нормалното. Коефициентът на Спирмън се използва в случаите на рангово скалирани признаци или при количествени признаци, за които не е известно, че имат разпределение, близко до нормалното. Видът на зависимостта (права или обратна) се определя от знака на коефициента на корелация r, а в зависимост от получената стойност могат да се определят следните степени на силата на корелацията

$$|r| = \begin{cases} 0 < |r| \leq 0.3 & \text{слаба корелация} \\ 0.3 < |r| \leq 0.7 & \text{умерена (средна) корелация} \\ 0.7 < |r| < 1 & \text{силна корелация} \end{cases}$$

Важно е да отбележим, че **сила** и **значимост** на корелацията са **съвършено различни** понятия. При извършване на корелационен анализ е задължително да се изследва значимостта на корелацията, чрез съответна проверка на хипотеза за значимост на корелационния коефициент [16]. Незначимите коефициенти не носят особено съдържание. В приложните статистически изследвания често се установяват слаби или умерени значими корелациии. Те са предпоставка за съставяне на съответни регресионни модели. Слабите и умерените корелации са естествен резултат, който показва, че върху наблюдавания резултативен признак влияят повече от един факторен признак. Този факт дава възможности за конструиране на многофакторни регресионни модели.

В случая на нелинейни корелации посочените корелационни коефициенти са неприложими. В този случай може да се използва подход на трансформация на изследваните признаци [15].

3.7. Регресионни модели. Адекватност на модела. Значимост на регресионните коефициенти. При съставяне на регресионни модели е задължително изследването за адекватност на модела и значимост на регресионните коефициенти [18], а в случаи на линейни многофакторни модели – евентуално наличие на колинеарност (мултиколинеарност) между използваните факторни променливи [20]. Модели, за които не е доказана адекватност и регресионни коефициенти, за които не е доказана значимост, не трябва да бъдат тълкувани.

В случаи на линейни многофакторни модели е задължителна проверката за наличие на колинеарност (мултиколинеарност) между използваните факторни променливи. Това явление може да се ограничи, чрез намаляване на броя на факторните променливи или след предварително проведен факторен анализ [20].

Заключение. Надяваме се настоящата работа, без да има претенции за изчерпателност, да помогне на широк кръг ползватели да правят разлика между компетентен и некомпетентен статистически анализ. Според Йордан Калчев – дългогодишен водещ експерт и изследовател в Националния статистически институт, в некомпетентния статистически анализ: обектът (съвкупността) и предметът на изследване са неясни; получените обобщаващи резултати не произтичат от първичните данни и се представят непълно; направените изводи не следват от получените резултати; формулираните препоръки не произтичат от направените изводи; илюстрациите са цветни и анализите винаги са луксозно подвързани; натрапва се убеждението, че анализите са разработени професионално; те могат да се използват само тенденциозно.

ЛИТЕРАТУРА

[1] К. Гатев. Въведение в статистиката. София, 1995.
[2] В. Гнеденко, А. Гешев. Теория на вероятностите и математическа статистика. Пловдив, 1994.
[3] Б. Димитров, Н. Янев. Вероятности и статистика. София, СОФТЕХ, 2007.
[4] К. Калинов. Статистически методи в поведенческите и социалните науки. София, НБУ, 2001.
[5] Т. Калоянов. Статистика. София, Тракия–М, 2004.
[6] В. Павлов. Приложна статистика. София, ПРЕПРЕС, 2013.
[7] В. Павлов, В. Михова. Приложна статистика със SPSS. Русе, Авангард принт, 2016.
[8] И. Съйкова, С. Тодорова. Статистическото изследване. София, 1994.
[9] И. Съйкова, А. Стойкова, С. Съйкова. Статистическото изследване на зависимости. София, УНСС, 2002.
[10] Л. Закс. Статистическое оценивание. Москва, Статистика, 1976.
[11] А. Кобзарь. Прикладная математическая статистика. Москва, ФИЗМАТЛИТ, 2012.
[12] А. Колмогоров. Основные понятия теории вероятностей. Москва, Наука, 1974.
[13] Г. Крамер. Математические методы статистики. Москва, Мир, 1975.
[14] Э. Леман. Проверка статистических гипотез. Москва, Наука, 1979.
[15] М. Мелник. Основы прикладной статистики. Москва, Енергоатомиздат, 1983.
[16] Дж. Себер. Линейный регресионный анализ. Москва, 1980.
[17] Н. Хастингс, Дж. Пикок. Справочник по статистическим распределениям. Москва, Статистика, 1990.
[18] D. Gujarati. Basic Econometrics. Military Academy, West Point, Boston, USA, 2003.
[19] D. L. Harnett, A. K. Soni. Statistical methods. Indiana, USA, Addison-Wesley Publishing Company, 2009.
[20] D. Moore, G. McCabe. Introduction to the Practice of Statistics. New York, W. H. Freeman & Co, 1989.

Велизар Павлов
Катедра Приложна математика и статистика
Русенски университет
ул. Студентска, 8
Русе
e-mail: vpavlov@uni-ruse.bg

SOME ACCENTS IN APPLIED STATISTICAL RESEARCH

Velizar Pavlov

The availability of vast amount of statistical information in the public domain, the need for competent statistical analysis in all areas of science and life requires adequate education and knowledge of statistics for a wide range of users. The paper presents some accents in applied statistical research: the representativeness of the sample surveys, the specificity of the statistical laws and dependencies, the difference between mathematical and statistical differences, the comparative analysis based on experimental and control groups, the importance of checking the normality of the distributions, the strength and significance of the correlation coefficients in the case of linear correlation, the adequacy of regression models and the significance of regression coefficients. The aim is to help all interested readers to increase their competencies in applied statistics and to prevent them from accepting some common mistakes in statistical analyzes.

Key words: Applied Statistics, Data Analysis, Applied Mathematics